Heart Rate Measurement from Videos using Feed Forward Back Propagation Neural Network with Artificial Bee Colony Optimization

Gaganjot Kaur

A thesis submitted to
Auckland University of Technology
in fulfilment for the degree of
Doctor of Philosophy

2022

School of Engineering

Abstract

The overall aim of the research reported in this thesis was to build a new non-invasive method for Heart Rate monitoring using videos and investigate the algorithms for selecting a region of interest to enable the extractions of more useful features for the measurement of Heart Rate. Analysis of these image features will assist in the creation of a more effective database for monitoring Heart Rate.

Heart Rate is important for monitoring and diagnosing medical conditions of a individuals health. Several techniques are available to measure heart rate, these used sensors, which need to be in direct contact with the surface of the skin, that may cause discomfort and soreness to the patient, especially for sensitive skin patients. However, recent advances in computer vision have shown that several physiological factors linked to the heart may be evaluated without invasive methods. Heart rate monitoring can be done using video observation on the subject by analysing specific facial parts like eyes, forehead, cheeks etc. The literature has observed and shown that the facial colour changes, leading to redness from the normal face colour when the heart rate goes up. However, the changes are so minute that they cannot be observed from the naked eye, the frame responsible for detecting needs to be magnified.

This research work has utilised Euler video modification as it has been cited as one of the most efficient techniques for magnifying specified portions of the face. The research was carried out by recording 30 participants, and a pulse oximeter was used to measure the heart rate.

The research first extracts the frames from the input video, and then face detection is done using cascaded object detection. Then, evaluate the face's forehead using histogram equalization, magnifies it, and optimize the magnified part using Swarm Intelligence oriented Artificial Bee Colony (ABC). Modified grouping behaviour has been presented, and a novel fitness function based on the pixel distribution has been applied. Due to the shortage of samples for the training section as the global pandemic affected the entire world, the training has been done by conventional neural networks that work with Levenberg Back Propagation algorithm.

The result section has been designed based on frame analysis with true-positive rate, false-positive rate, Accuracy and Kappa coefficient. The results contain the analysis based on the variation of the total number of samples in the data repository. The results compared analysis of the algorithms used in this research, Convolutional Neural Network (CNN) and Feed Forward Back Propagation (FFBPNN) with optimization method ABC and without optimization. Overall, the results compared with the proposed algorithm FFBPNN-L with ABC compared to CNN with ABC that TPR has been improved by 8.24%, Kappa coefficient has been improved by 4.58%. The limited dataset shows better results for FFBPNN-L, such that ABC with FFBPNN-L resulted in 4.25% better accuracy than ABC with CNN. Thus, FFBPNN-L with the integration of ABC shows improved results than when ABC is integrated with CNN. CNN extracts the features and does not require any external feature extraction support. This outcome with better performance of FFBPNN over CNN is mainly due to the smaller dataset size available for the research.

The findings of this research, which demonstrate methodologies that can be used to monitor heart rate, also identify directions for future work in medical engineering.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ABC	Artificial Bee Colony
AF	Atrial Fibrillation
AI	Artificial Intelligence
ANN	Artificial Neural Network
BCG	Ballistocardiograph
BPA	Business Process Automation
bpm	Beats Per Minute
BSS	Blind source separation
CNN	Convolution Neural Networks
COD	Cascade Object Detector
CVD	Cardiovascular Diseases
DoG	Difference-of-Gaussians Pyramid
DRMF	Discriminative Response Map Fitting
E2VM	Enhanced Eulerian Video Magnification
ECG/EKG	Electrocardiogram
EMMS	Efficient Motion Magnification System
EVM	Eulerian Video Magnification
FBPNN	Feed Forward Back Propagation Neural Networks
FIR	Finite Impulse Response
FPR	False Positive Rate
HCI	Human To Computer Interaction
HOG	Histogram of Oriented Gradients
HR	Heart Rate

HRV	Heart Rate Variability
IBI	Inter Beat Interval
ICA	Independent Component Analysis
IIR	Infinite Impulse Response
iPPG	Imaging PPG
LED	Light Emitting Diode
LVM	Linear Video Magnification
MI	Myocardial Infarction
MIT	Massachusetts Institute of Technology's
ML	Machine Learning
MOMBAT	Monitoring Utilising Modelling and Bayesian Tracking
MRF	Markov Random Fields
ncPPG	Non-Contact PPG
NLMS	Normalized Least Mean Squares
PBM	Phase-Based Magnification
PCA	Principal Component Analysis
PD	Photodetector
PPG	Photoplethysmography
RMSE	Root Mean Square Error
ROI	Region Of Interest
rPPG	Remote Photoplethysmography
RSA	Respiratory Sinus Arrhythmia
SAAMD	Software As a Medical Device
SI	Swarm Intelligence
SIANN	Space Invariant Artificial Neural Networks

SIFT	Scale Invariant Feature Transform
SL	Supervised Learning
SpO_2	Oxygen Saturation
STFT	Short-Time Fourier Transform
STVEN	Spatio-Temporal Video Enhancement Network
SURF	Speeded Up Robust Features
SVM	Support Vector Machines
TPR	True Positive Rate
UBCM	Upper Body Classification Model

Attestation of Authorship

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person except where explicitly defined in the acknowledgements, nor material which to a substantial extent has been submitted for the award of any other degree or diploma of a university or other institution of higher learning.

Gaganjot Kaur

May 2022

Acknowledgements

I wish to express my gratitude to my supervisors, Dr Jeff Kilby and Dr Hakilo Sabit for their patience, support, guidance, and advice throughout this research. Their immense knowledge and plentiful experience have encouraged me in all the time of my academic research and daily life.

I would also like to express my gratitude to my parents, my husband, and my son. Without their tremendous understanding and encouragement in the past few years, it would be impossible for me to complete my study.

Due to the nature of research involving of human participants, ethics approval (Application Number 21/424) was sought and granted by the Auckland University of Technology Ethics Committee (AUTEC) Ethics on 09 December 2021.

Chapter 1
Introduction

1.1 Background: Overview of Clinical Heart Rate Measurement

Measuring a person's Heart Rate (HR) is critical for determining an individual's physiological and pathological status [1; Chapter 1, 2; Chapter 9]. The heart is one of the most vital organs in the human body. Heart Rate is the number of heartbeats per minute (bpm). The heart's primary purpose is to circulate blood throughout the body by beating at a resting rate of 60 to 100 bpm [3; Chapter 4]. The two methods for measuring HR in the clinical environment are the Electrocardiogram (ECG) and the Photoplethysmography (PPG).

ECG is one of the oldest diagnostic instruments still used in medicine today, with the earliest records going back to 1902 [4]. Recording the heart's electrical activity or producing ECG is called electrocardiography. ECG employs conductive electrodes that are connected to the patient's body in a predetermined and standardized manner to detect and record the change in electric potential between various electrodes caused by the electric activity of cardiac muscle fibres over time. ECG is also represented as EKG (elecktrokardiogram) based on the German spelling.

The clinical technique for evaluating cardiac properties of the heart is ECG signals, known for their excellent accuracy and simplicity of interpretation. Despite its drawbacks, the ECG is widely regarded as the best method for determining HR, Heart Rate Variability (HRV) or InterBeat Intervals (IBI). HR is calculated by measuring the beats per minute (bpm), and HRV is the time in milliseconds between successive heartbeats [3], as shown in Figure 1.1 (top), also referred to as RR interval [5, 6].

PPG measures heart pulses using a pulse oximeter that illuminates the skin and measures changes in light absorption and is usually placed on fingertips, toes, and earlobe [7]. The PP interval is the time interval between two successive pulses, shown in Figure 1.1 (bottom).

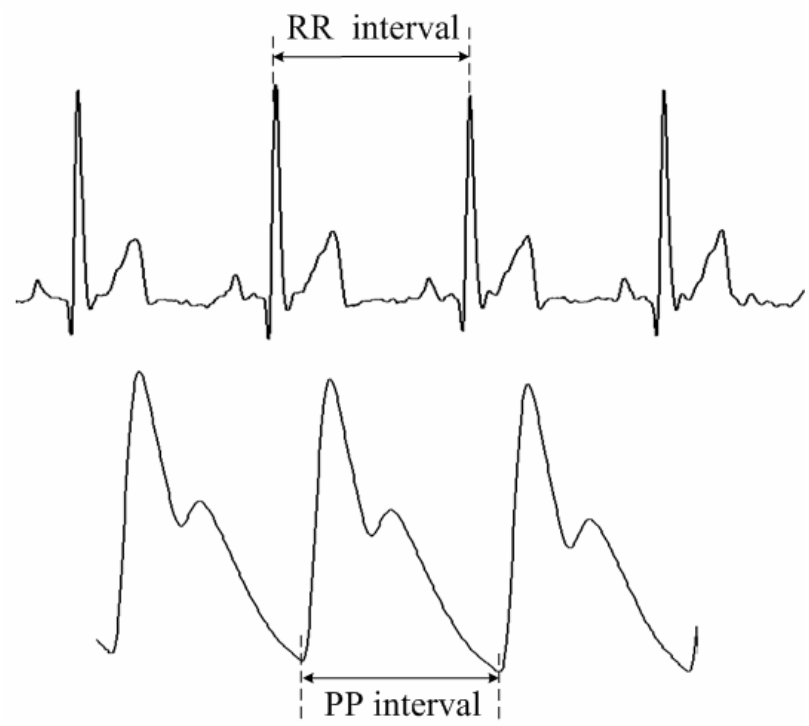


Figure 1.1 Electrocardiography signal showing RR interval (top) and Photoplethysmography signal indicating PP interval (bottom) [8]

1.2 History of Heart Rate and Electrocardiogram

John Floyer (1649–1734) devised the "physician pulse watch", a forerunner to the chronometer, utilised to reliably measure the patient's heart rate in the early eighteenth century [9; Chapter 10]. Stephen Hales (1677–1761) supplied an essential demonstration of heart rate variability when he published an article in 1733 [9]. Stephen stated that the space between heartbeats and blood pressure fluctuated over the respiratory cycle. Carl Ludwig (1816–1895) built on these findings by declaring unequivocally that the pulse elevated during inhalation and reduced during expiration, setting the groundwork for what would later be known as Respiratory Sinus Arrhythmia (RSA) [9; Chapter 1].

Willem Einthoven (1860–1927), who used galvanometers to record the first ECG in 1901, marked the change from past work done to modernity in the science of electrocardiography [10]. One of his first concepts, weighing around 270 kg and requiring five people to operate, is shown in Figure 1.2. Einthoven was the first to utilize the present terminology of (P, Q, R, S, T).

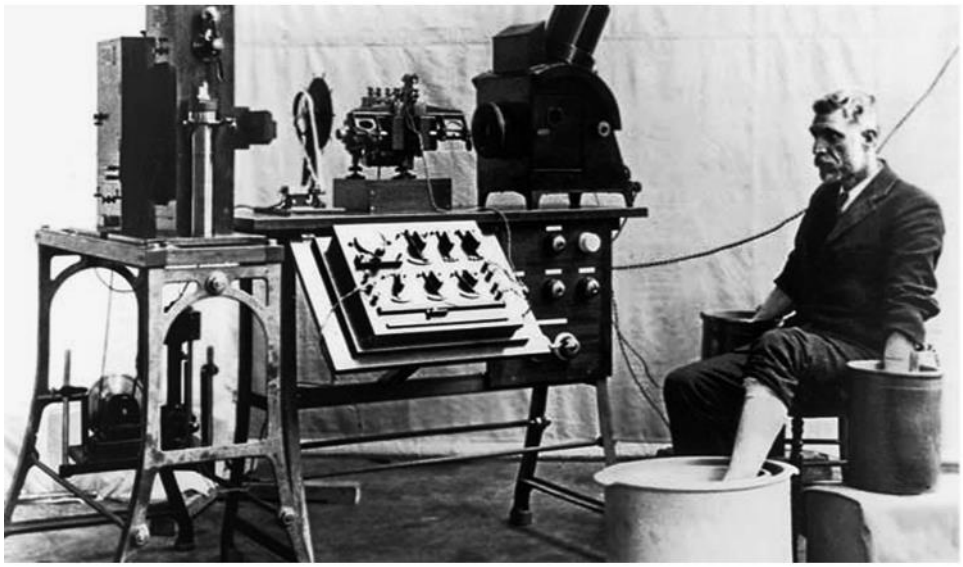


Figure 1.2 Galvanometer Electrocardiogram Generator [9]

Dr Willem Einthoven, a Dutch physiologist who was inspired by Waller's work, modified the capillary electrometer, even more, demonstrating five deflections that he termed ABCDE, as shown in Figure 1.3 [4].

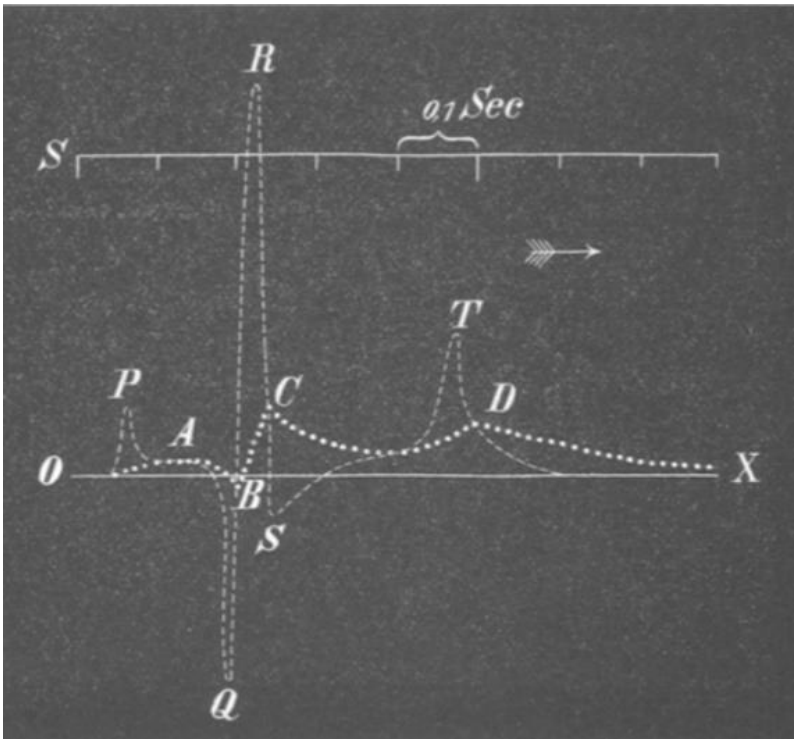


Figure 1.3 Labelling of deflections using letters ABCD and PQRST [9]

Dr Willem used a mathematical correction to account for inertia in the capillary system, resulting in the curves which are observed nowadays. He renamed deflections ABCDE using the new alphabet series PQRST, which followed the mathematical tradition which was established by Descartes [11]. Einthoven invented the name "electrocardiogram" to represent these wave shapes during the Dutch Medical Meeting of 1893.

ECG recording techniques improved throughout time, and in 1961, Norman Holter (1914–1983) invented a device (which still retains his name, Holter) that allowed long-term ECG data to be registered [12]. Figure 1.4 shows the historical milestones of EKG (or ECG) development from 1842 to 1954.

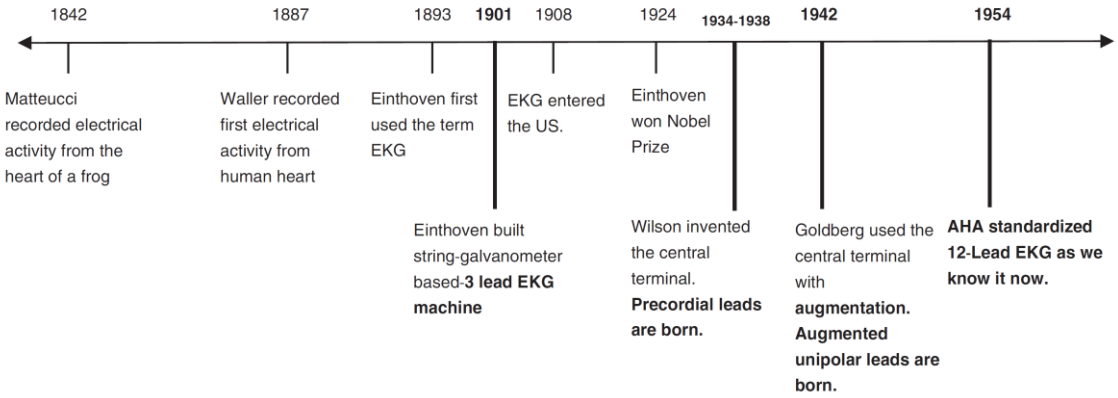


Figure 1.4 Timeline of clinical development of EKG (also known as ECG) [4]

Techniques for recording ECGs have improved over time. It took half a century for Einthoven's original ECG to evolve into the 12-lead electrocardiogram [4], which is also presently used in medicine. In recent years, research in the field of HR has continued apace. Different research has validated its clinical benefit, and new approaches have been presented. Experiments have also been done to explain its physiological basis further.

1.3 History of Heart Rate and Photoplethysmography

Photoplethysmography (PPG) is not a discovery; it is almost 150 years old, but it was revolutionised in the 20th century for clinical use [13]. Real-time optical blood flow

monitoring was first used in the late 1800s by having people hold their hands up to a candle in a dark room to see the vascular structure and blood flow.

Alrick Hertzman first described the PPG. in 1937 [14], which led to a series of published papers [15-17]. PPG takes advantage of the fact that the absorption of light by a transilluminated tissue varies with its blood content to detect vascular changes with the photoelectric cell, as shown in Figure 1.5. Hertzman named it the *'photoelectric plethysmograph'* based upon his belief and early observations that its creation was linked to blood volume changes. The advantage is taken from the fact that the absorption of light by a transilluminated tissue varies with its blood content, content, to detect vascular changes with the photoelectric cell.

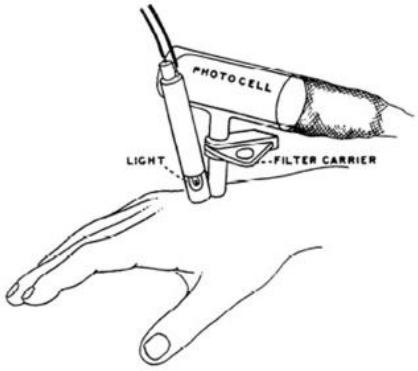


Figure 1.5 The photoelectric plethysmograph is in position over the skin of the hand [15]

It is important to note that Hertzman was working from the simplistic model of light that was prevalent at the time. Before, the quantum complexities and light scattering characteristics were generally appreciated [18]. Hertzman's theories were derived from Beer-Lambert's Law of Light [19]. The primary assumptions of the latter were that light absorption is directly proportional to the path length, the concentration of substances, and the light absorption by each of those substances.

In 1979 Challoner, et al. [20] worked on an optical measurement technique that can be used to detect blood flow and volume in the skin. It has been used for clinical applications, with the technology utilised in commercially available medical devices such as pulse oximeters. The basic form of PPG technology requires only a few optoelectronic components. A light source to illuminate the skin tissue and a

photodetector to measure the minor light intensity variations associated with perfusion changes in the catchment volume. PPG is most often employed non-invasively and operates at a red or a near-infrared wavelength. The most recognised waveform feature is the peripheral pulse, and it is synchronised to each heartbeat.

A significant advance in the clinical use of a PPG-based technology by Aoyagi, et al. [21] and Yoshiya, et al. [22] came with introducing a pulse oximeter, a non-invasive method for monitoring patients' arterial oxygen saturation. The desire for small, reliable, low-cost and simple-to-use non-invasive cardiovascular assessment techniques are key factors that have helped re-establish photoplethysmography. The developments in semiconductor technology, i.e., light-emitting diodes (LED), photodiodes, and phototransistors, have improved the PPG probe design's size, sensitivity, reliability, and reproducibility. Figure 1.6 shows a brief visual timeline for the history of PPG sensors since the 1800s.

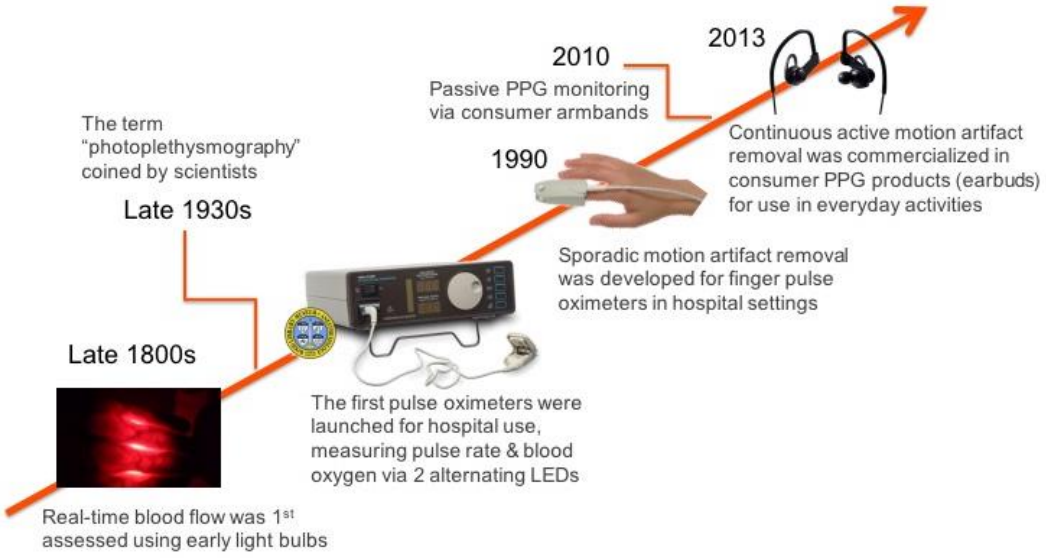


Figure 1.6 A brief visual timeline for the history of PPG sensors [23]

1.4 Non-Contact methods of Heart Rate Measurement

For the monitoring of HR, researchers used various algorithms and methods to calculate bpm from different data sets like videos, audio, and iris. These all focused on the non-invasive method for the measurement of heart rate.

1.4.1 Heart Rate Measurement using Video recordings

Several researchers [24-28] have worked on HR measurement using contactless devices or remote methods such as using a video camera. In 2005 Liu, et al. [29]began working with motion magnification, creating a technique that acted like a microscope for visual motion. To achieve motion magnification, there is a need to measure visible movements accurately and group the pixels to be modified. Then each pixel in a frame is assigned to one of the motion clusters, or layers, which are then stacked on top of each other in the outputting video to describe motion. The output video was built using the estimated movements to project each pixel back to a reference frame and ultimately retender the original video to see the form and characteristics of the magnified changes invisible to the human eye. In 2012 Wu, et al. [30] developed a non-contact method Eulerian Video Magnification (EVM) at the Massachusetts Institute of Technology, Computer Science and Artificial Intelligence Laboratory (MIT-CSAIL), which enhances the visual effect of time-varying blood flow for a region of interest (ROI) on the human body by amplifying red, green and blue colour channels of the video.

1.4.2 Heart Rate Measurement Using Audio

With the advancement in technology, researchers also used audio signals for HR measurement based on modelling vowel speech signals and processing several physiological parameters. Davletcharova, et al. and James, et al. [31, 32] used Short-Time Fourier Transform (STFT) to detect the maximum peaks of the formants in the process of detecting the heart rate.

Audio or speech signals of an individual are used to measure HR. In machine learning, it requires audio data and the HR for the training and testing of the model [33].

1.4.3 Heart Rate Measurement using Iris

Researchers used vision-based technology to measure physiological signals, in which pupillometry has been utilised extensively in medical and psychophysiology research. Also, the ocular pulse has been observed to see the changes in HR [34, 35]. For heart rate extraction, Parnandi, et al. [36] used pupillary diameter fluctuations. Heart rate

variability was calculated from the power spectrum of the time series of pupillary diameters. But the lighting conditions affect the results. When the subject blinks during that time, pupil area cannot be detected, and HR calculation is not possible.

The visual aspects have been the area of interest in the last couple of years. They have become more efficient by analysing the videos rather than putting voice capturing sensors for audio and observing ocular pulse or pupillometry fluctuations. Nowadays, the variation from the normal HR is very common and may symbolize heart-related disorders. Thus, effective daily monitoring of the HR is critical to recognize the potential heart disease at an early stage. Several studies have attempted to integrate a touchless video input to predict the HR of the subjects [37-39]. Inspired by such research, tries to design HR analysis using video monitoring in which video is transformed into several frames followed by integration of image processing techniques.

1.5 Pros and Cons of Clinical HR Measurement Methods

The cardiac data required for further HR analysis is often collected using one of the two techniques utilised in the clinical setting: ECG or PPG. Although founded on distinct concepts and assessing different phenomena, both approaches offer trustworthy findings when correctly conducted on the one hand but are constrained by several considerations, the most significant of which is the requirement for physical contact with the subject on the other. Because they must come into touch with the skin, they can be restricting. Contact-based techniques can be unpleasant or distracting, such as with newborns. These procedures are non-invasive, but they need touch with human skin, which may harm those with sensitive skin (e.g., neonates). It may also irritate (e.g. a fitness tracker for individuals) or distract (e.g., when worn in a professional environment). These limitations, combined with the growing demand for ubiquitous measuring, both within and outside the hospital environment of human physiological parameters and the option of use in commercial environments, have led researchers worldwide to seek a way to optimize and free them from their limitations.

1.6 Research Methodology

As discussed in section 1.4, limitations of current clinical HR measurement techniques are due to the ability to contact skin, special trained staff requirements, and non-reliable results in particular conditions, resulting in the need for a non-contact heart rate measuring method. Using image processing from videos, it's possible to extract valuable information for measuring HR. Data extraction from frames allows more meaningful information to be extracted by the researcher.

During a change in the heart rate that could also lead to a condition which could be uncomfortable for a user, all the vital signs for the detection could not be viewed from the naked eye. To resolve this issue, video magnification algorithms were presented earlier. The research area still has abilities and scope for expansion, which inspires this thesis's problem statement. This thesis refines the magnification problem to improved segmentation or Region of Interest selection, followed by the training and classification of the segmented portions to detect the heart rate $\text{hr} \in \{60 - 120\}$ beats per minute. The problem is described mathematically in equation (1.1).

$$f = argopt \sum_{i=1}^{t} f_{set} \ \forall \in R_i \tag{1.1}$$

where, $\forall$ is the universal quantification symbol represents all values of the samples that has been considered, t is total number of frames extracted, $\in$ is the element, f_{set} is the feature set that belongs to extracted ROI R_i [40, 41].

The function takes the input argument as the feature set and aims to optimize the existing feature set via selecting the current portion of the segmentation.

The optimal *f* is required to pass to the training algorithm to create training instances which will be further utilised to classify the ground truth values. Optimal *f* doesn't need to produce high classification accuracy due to the tuning architecture of propagation layers. Selection of the number of layers requires a statistical analysis through regression, and hence it becomes necessary to tune the training mechanism as per the best available data. The problem statement extends to train the system based on the ground truths of the dataset's extracted features/segmented frames.

The mathematical representation can be found using equation (1.2).

$$f' = argmax \sum_{i=1}^{t} CA \ \forall \ G_i \tag{1.2}$$

where, *CA* is the classification accuracy, *t* is total number of extracted frames, , G_i is the ground truth. The aim is to maximize the chances of the classified sample matching its ground truth values. In other words, the problem is to increase the classification accuracy of the test samples.

A monitored video is useless if it cannot conclude a statement. To analyse the video, the video must be processed, and the processed video must go through an inference engine to conclude a statement. Figure 1.7 shows the schematic diagram of stages undertaken in research.

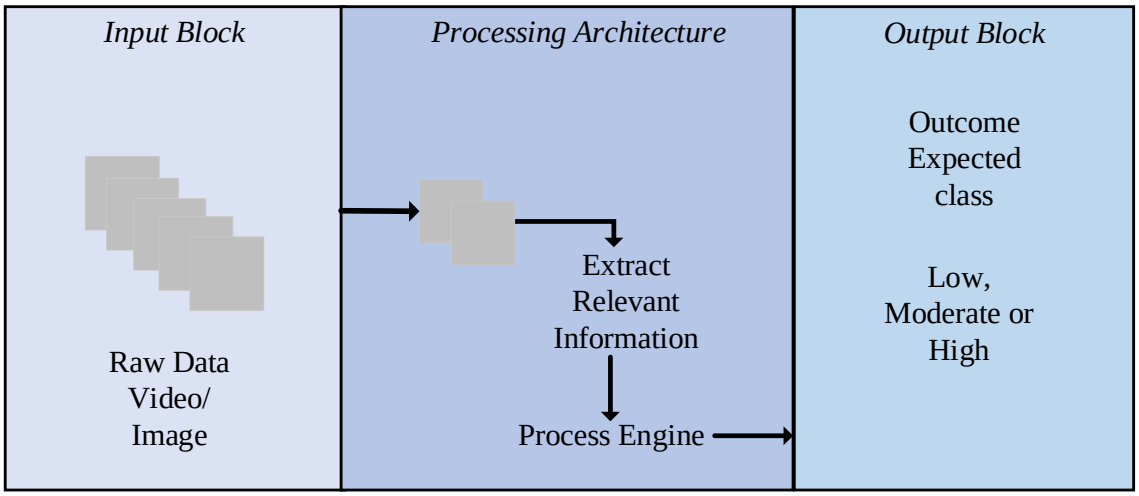


Figure 1.7 Schematic diagram of the various stages undertaken in this research

1.6.1 Input Block

The input set contains the objects and its description. The input might be a video, frames of a video, audio segments, images etc. The proposed work has utilised videos as the main input source, and hence from now onward till the last chapter, this thesis refers to data as video frames. When it comes to medical data analysis, video surveillance has gained popularity [42, 43]. The visual aspects of the video provide a

significant amount of information that becomes useful when it comes to training the system.

1.6.2 Processing Architecture

The processing architecture block takes the judgement based on the previous information under the hood of a stored repository. A stored repository is created using a training algorithm. For instance, Feed Forward Neural Network is a training algorithm that intakes two inputs before running the training engine.

(a) The attribute set: is a collection of features, represented in equation (1.3) as follows:

$$f = \sum_{i=1}^{t} R_d . Fv \ \forall \ G_i \tag{1.3}$$

where G_i is all the ground truth for which the feature vector containing 'n' number of features has been extracted from the raw data R_d.

(b) The ground truth values: until a machine does not learn what to call a set of attributes, it cannot predict the label of the attribute set and hence, each feature vector set $Fv \ \epsilon \ f$ is tagged with its ground truth value G_i.

The Levenberg algorithm, also known as the Levenberg-Marquardt algorithm, is a widely used method for solving nonlinear least-squares problems. The algorithm iteratively adjusts the parameter estimates to minimize the sum of squares of the residuals. The steps involved in the Levenberg algorithm are as follows:

1. Choose an initial estimate for the parameters,
2. Compute the Jacobian matrix
3. Compute the residual vector
4. Compute the approximated Hessian matrix
5. Compute the update rule
6. Update the parameter estimates by setting

The Levenberg-Marquardt algorithm solves nonlinear least squares problems by iteratively adjusting the parameter estimates to minimize the sum of squares of the

residuals. In order to ensure that the approximated Hessian matrix, J^TJ is invertible, the algorithm introduces another approximation to the Hessian matrix,

$$HJ \approx J^TJ + \mu I \tag{1.4}$$

where μ is always positive and called the combination coefficient, and I is the identity matrix.

This approximation ensures that the elements on the main diagonal of the approximated Hessian matrix are larger than zero, thereby guaranteeing that the matrix H is always invertible. The update rule of the Levenberg-Marquardt algorithm, which is expressed in equation (1.5) as follows:

$$w_k + 1 = w_k - \left(J_k^T J_k + \mu I\right)^{-1} J_k^T e_k \tag{1.5}$$

As a combination of the steepest descent and Gauss-Newton algorithms, the Levenberg-Marquardt algorithm switches between the two algorithms during the training process. The Levenberg algorithm iteratively adjusts the parameter estimates by combining the steepest descent and Gauss-Newton algorithms to minimize the sum of squares of the residuals. The algorithm switches between the two algorithms based on the value of μ, which is always positive and called the combination coefficient. The algorithm guarantees that the approximated Hessian matrix is invertible by introducing an approximation to the true Hessian matrix [44].

1.6.3 Output Block

The output block delivers the results based on the input fed to the previous block, where the training and classification are done in any determining architecture. The results provide the output in the form of one of the expected classes, namely normal, moderate, and high.

1.7 Research Objectives

Based on the literature review, the following research objectives are listed as follows:

(a) To study and analyze video magnification methods and algorithms shown in the state of art techniques.
(b) To enhance the magnified video in terms of ROI selection by applying various strategies.
(c) To perform training and classification through various multi-class classification techniques and to identify the best suitable classification algorithm for the utilized dataset.
(d) To evaluate and compare the quantitative parameters of this research.

1.8 Organization of the Thesis

The thesis is divided into eight chapters as follows:

Chapter 1 introduces heart rate measurement techniques, and it highlights the limitations of existing methods used for heart rate measurement.

Chapter 2 covers the heart structure, clinical methods for heart rate measurement and new emerging techniques.

Chapter 3 presents research articles in contrast to proposed work and reviewed the literature that are relevant to video magnification and the training and classification mechanism.

Chapter 4 covers detailed information how data was acquired, and which equipment and setup was considered during data acquisition of this research. To perform training and classification through various multi-class classification techniques and to identify the best suitable classification algorithm for the utilized dataset against Heart Rate.

Chapter 5 covers detailed information on the technical design of the selection of ROI from the face video and optimization.

Chapter 6 discusses the feature extraction, training, and classification.

Chapter 7 summarizes the results and outcomes of the proposed architecture. The details discussed in terms of True Positive Rate, False Positive Rate, Accuracy and Kappa are summarized concerning variation in the number of layered architectures used in the analysis.

Chapter 8 presents the overall findings, future work, limitations, and conclusion of this research.

1.9 Contribution of Thesis

The health of a country's population is one of the major concerns of any government around the world. With the help of the modern technology, it is now even more possible to save human life before it is lost. In the same context, heart attacks have been observed to be very common among the population. The health education sector along with academic research and industry views it as an opportunity to improve the prevention of heart attacks.

This research uses the identification of heart rate via changes over the forehead in terms of its colour which cannot be visualized with naked eyes. It becomes essential to magnify the video footage in order to identify the change in colour. A machine learning based training algorithm was required to predict the possibilities based on the acquired data that has been provided to it against different heart beats or rates. To train a system, most appropriate features are to be passed to a training engine in case of semi-supervised approach. Though there has been a systematic way of classifying the heart rates and making decisions against it, the proposed work contributes in the following manner.

(a) Design a Swarm Intelligence based improved algorithm that uses ABC with a novel fitness function.
(b) Validations over the proposed algorithm have been made based on quantitative parameter analysis.
(c) Comparison of proposed work has been made based on other algorithms.

Chapter 2
Heart Rate Measurement

2.1 Introduction

While Electrocardiogram (ECG) is concerned with the heart's electrical activity, which causes contractions of the heart muscle, for this research, it is necessary to describe the structure of the human heart in general and the blood flow in the body. Therefore, this chapter presents an overview of heart rate measurement to provide background and context for this research. Vital signs are physiological tests that examine a human body's basic functioning and are essential for monitoring human health. They are markers of a person's overall physical health, and they enable the diagnosis of diseases and the tracking of recovery progress. Body temperature, respiration rate, blood pressure, and pulse rate or heart rate are the four primary vital indicators routinely checked by a healthcare expert [45, 46].

2.2 Heart Structure

The heart is the main organ of the cardiovascular system, a network of blood vessels that pumps blood throughout the body. It also works with other physiological systems to control your heart rate and blood pressure. The heart's conduction system is like the electrical wiring as it controls the rhythm and pace of the heartbeat. The parts of the heart normally beat in order: Contraction of the atria (atrial systole) is followed by contraction of the ventricles (ventricular systole), and all four chambers are not contracting or relaxed; that phase is diastole. The heartbeat begins in a specific cardiac conduction system and travels throughout the myocardium via this system. The sinoatrial node[1], the internodal atrial pathways[2], the atrioventricular node[3], the bundle

[1] The sinoatrial node is a specialized myocardial structure that initiates the electrical impulses to stimulate contraction and is found in the atrial wall at the junction of superior cavel vein and the right atrium.
[2] Internodal pathways are the connecting pathways that form a direct connection between the sinoatrial node and the atrioventricular node in the right atrium of the heart.
[3] The atrioventricular node electrically connects the heart's atria and ventricles to coordinate beating in the top of the heart; it is part of the electrical conduction system of the heart.

of His[4] and its branches, and the Purkinje system[5] are the structures that make up the cardiac conduction system [47; Chapter 29], shown in Figure 2.1.

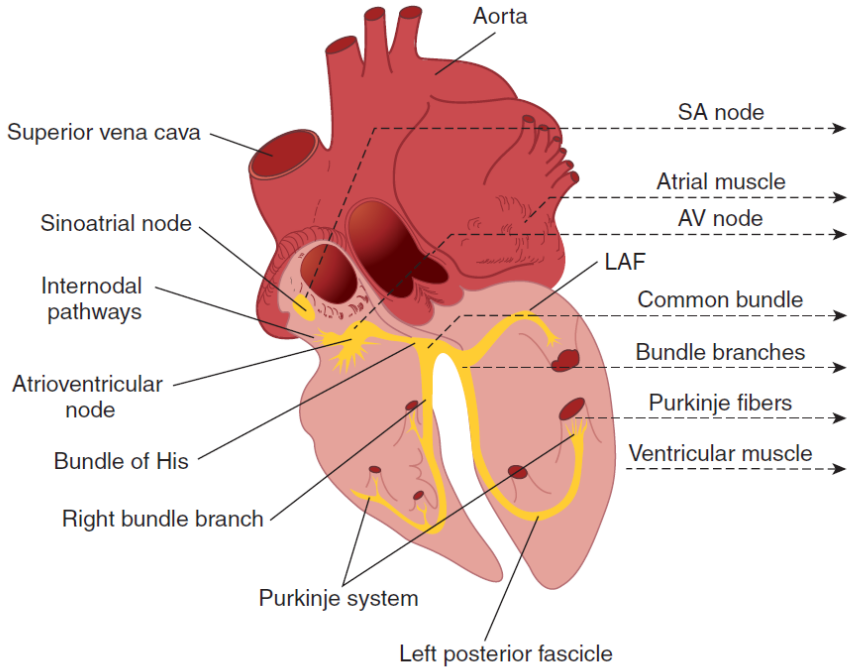


Figure 2.1 Cardiac Conduction System [47; Chapter 29]

The blood circulation throughout the body is not automatic; it is activated by electrical impulses that cause the heart muscle contractions. How the blood flows in the body is discussed in the next section.

2.2.1 Blood Flow through the Heart

When the heart beats, blood flows through the blood vessels because the heart pumps the blood, known as the blood circulatory system or cardiovascular system [48]. The vessels are flexible, muscular tubes that transport blood throughout the body. There are three main types of blood vessels:

[4] The bundle of His is a part of the electrical system of the heart. It is a collection of cells that carry electrical signals from the AV node to the bundle branches.

[5] The Purkinje system is a network of specialized muscle cells that carry cardiac impulses to the ventricles of the heart and cause them to contract.

- **Arteries:** transports oxygen-rich blood from the heart to all parts of the body. They branch out multiple times, getting smaller as they carry blood away from the heart and towards the organs.
- **Veins:** these are blood vessels that return blood to the heart; this blood contains less oxygen and is high in waste that must be expelled or eliminated. The veins become larger and larger as you get closer to the heart. The superior vena cava is a big vein that transports blood from the head and arms to the heart, whereas the inferior vena cava transports blood from the abdomen and legs.
- **Capillaries:** are tiny, thin blood vessels that connect the arteries and veins. Our organ's cells can pass oxygen, nutrients, carbon dioxide, and other waste products through their thin walls.

2.3 Types of Heart Rate Measurement

There are three types of observations used to measure heart rate. Different activities in the body are observed like electrical activity of the heart, blood flow and skin colour change.

2.3.1 Clinical Methods for Heart Rate Measurement

The two standard methods used for clinical heart rate measurement are Electrocardiogram (ECG) and Photoplethysmography (PPG).

- ***The heart's electrical signals:*** are recorded using ECG, a standard method used to monitor the heart's health. It is done in clinics or hospitals by doctors, trained nurses, or technicians [49]. The ECG technology has become one of the most widely used diagnostic methods in the clinical cardiovascular domain for detecting heart diseases and tracking the heart's electrical activity. The ECG is primarily used to identify a few cardiovascular diseases (CVD), including myocardial infarction (MI), heart failure, arrhythmia, atrial fibrillation (AF), and ventricular hypertrophy [50]. A noiseless ECG signal is required for effective assessment by the clinician. Filtering methods such as Infinite Impulse Response (IIR) and Finite Impulse Response (FIR) aid in removing noise from

the ECG signal. The repolarisation and depolarisation of various muscle tissues [51] from each part of the ECG waveform when electrical pulses are delivered. Every pulse has feature waves represented with P, Q, R, S, and T. ECG signal contains a P wave, a PR interval, the QRS complex, and an ST-segment. The P wave shows the sequential activation of both left and right atria, created by depolarization or contraction of atriums. QRS complex shows the ventricular depolarization, whereas the ST-segment and T wave reflects ventricular repolarization and are produced by relaxation of the ventricles. The PR interval is between the P wave's first deflection and the QRS complex's first deflection. One R peak represents one beat while reading the ECG signal, and the number of R peaks in the signal over one-minute equals the HR, shown in Figure 2.2. RR interval is the duration between two consecutive R-waves of the QRS signal on an ECG due to inherent sinus node features and autonomic factors and is measured in milliseconds.

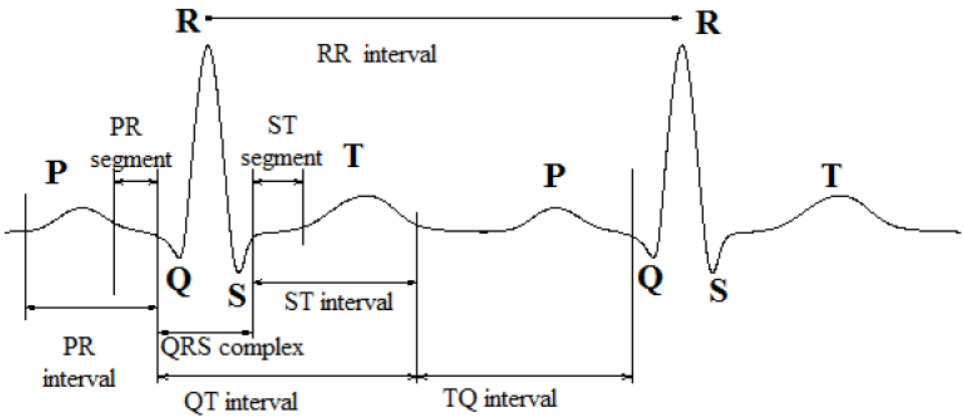


Figure 2.2 QRS and RR interval [50]

- ***Blood flow*** can be observed by using Photoplethysmography (PPG), a simple and low-cost optical technique that can be used to detect blood volume changes in the microvascular bed of tissue and is easy to use as no special training is needed [52]. PPG technique has been used in a variety of commonly available medical devices to measure oxygen level, blood pressure, and cardiac output. As a photoelectric detection method, PPG aims to detect the changes in blood volume in the arterioles and capillaries by illuminating the living body tissues using optoelectronic equipment and then measuring the intensity of the reflected light to obtain the blood volume pulse signal. PPG is popular as an alternative

HR monitoring technique because it is easy to use as there is no need for clinical set-up, can wear comfortably, and is low cost as it uses light-emitting diode (LED) as light source and photodetector for receiving the volumetric blood changes as shown in Figure 2.3 [53, 54].

PPG works in two modes; the first is the transmission mode, which can produce a reasonably good signal, and the measuring site may be constrained. In the transmission mode measurement, the infrared light-emitting diode and photodetector (PD) are placed on opposite sides of the measured human tissue, as shown in Figure 2.3. The PD detects residual light from the source after its absorption by the tissue. For the sensor to be functional, it must be placed on the body where transmitted light may be detected easily, such as on the fingertip or earlobe. The second mode is reflectance mode, in which the IF/RD LEDs and the PD measuring the intensity of the reflected light are placed side by side on the same body surface [55].

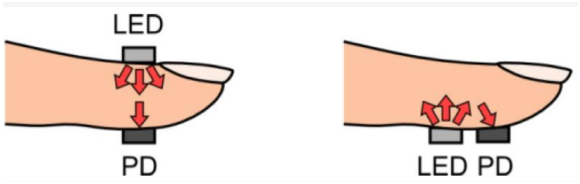


Figure 2.3 Light-emitting diode (LED) and photodetector (PD) transmission and reflectance-mode photoplethysmography [54]

PPG's contact requirements limit its applicability, despite the large variety of applications. First, because movement affects data, heart rate can only be detected if the patient or person stays stable or does not move the PPG device. Because of the requirement to remain stable limits the PPG for use during sports and other moving activities as it shows inaccurate results [56].

2.3.2 New and Emerging Methods

Changes in skin colour can be observed in the videos that are not visible to the naked human eye. So researchers are using the principle of contactless PPG, video magnification techniques, etc., to track skin colour changes towards finding the heart rate measurement [30, 57]. A video camera is typically used in contactless PPG

techniques to collect pictures/videos, which are subsequently analysed by image processing algorithms. Remote photoplethysmography (rPPG) mechanics are similar to invasive PPG, also known as imaging PPG (iPPG or PPGI) or non-contact PPG (ncPPG). The LED is replaced by ambient illuminance, and the video camera is replaced by photodetector rPPG methods, as shown in Figure 2.4. Two types of light reach the camera sensor: static and dynamic. The static components such as bone, tissue, and static blood, whereas the dynamic component represents fluctuations in light absorption caused by changes in arterial blood volume. As a result, the reflectance of the fluctuation in light absorption owing to volumetric variations in blood volume pressure is used to calculate PPG. A subject's PPG signal is collected from their finger, arm, or face. The reflection of ambient light dispersed across the skin layer, cell layer, and capillaries underneath the cell layer generates the PPG-like pulsing signal. Capillaries are the tiniest blood veins in the body, circulating just 2 to 5% of total blood volume. Compared to the body's overall change in blood volume pressure, this is a very modest volumetric change. PPG signal is generated by light propagation through the skin, primarily through the stratum corneum, epidermis, and dermis.

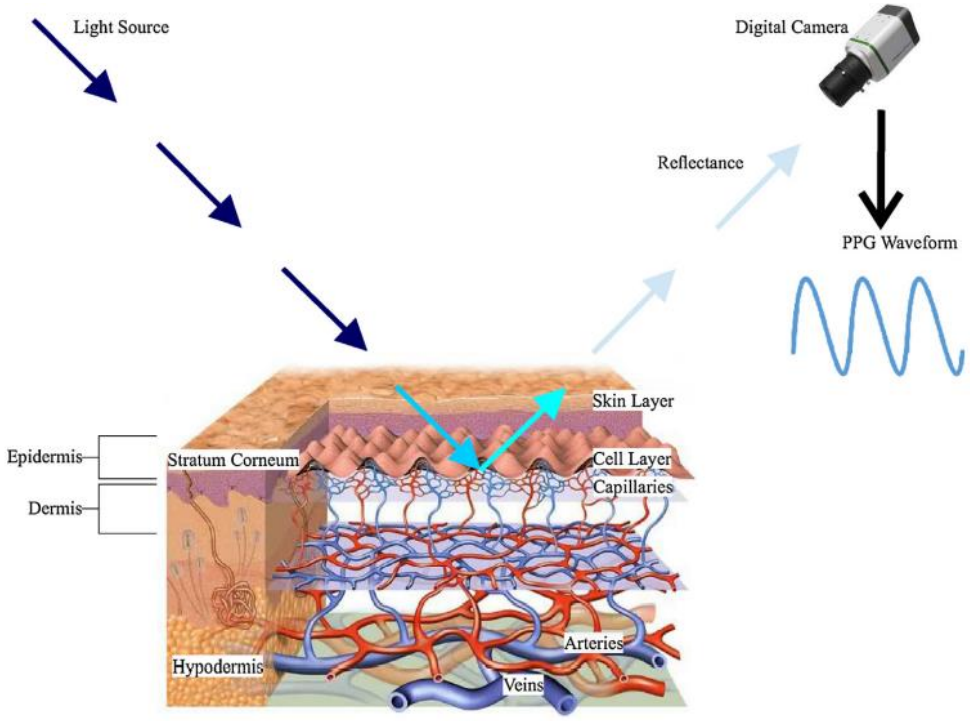


Figure 2.4 Simplified description of PPG extraction from facial tissue (cross-section view) [58]

As per Lambert's equation[6] of light intensity, light reflected through the skin might be considered diffusion and scattering [59]. Light is diffused by the skin's stratum corneum, whereas the epidermis and dermis scatter light. The epidermis layer is primarily made up of melanin. In contrast, the dermis layer consists of capillaries that flow haemoglobin across their microvascular network, making it more susceptible to light dispersion. The intense light scattering causes a tiny colour shift in the skins subsurface. As a result, shown in Figure 2.5, the camera sensor recorded a minor or slight change in light intensity values. Different researchers have made many contributions [52-54, 60-64] and published to improve HR measurement from the face image using PPG.

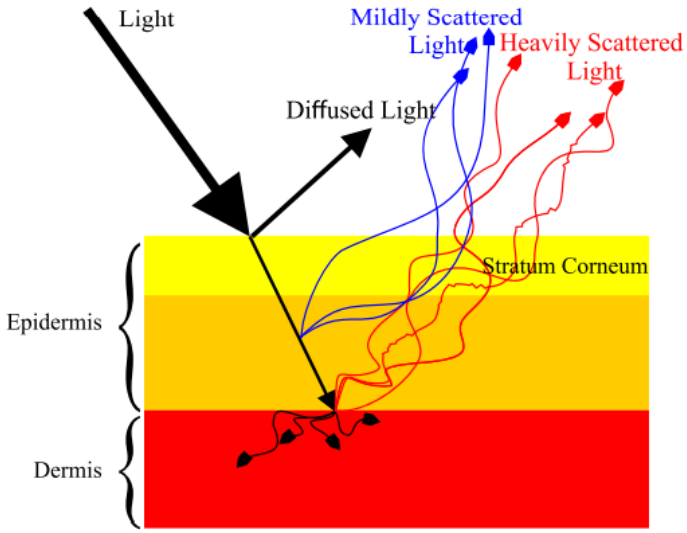


Figure 2.5 Shows light propagation through a simplified skin model [58]

De Haan, et al. [64] developed a PPG estimate technique based on chrominance characteristics extracted from the digital camera's RGB spectrum. A technique to calculate the ratio of two orthogonal chrominance signals was provided in the suggested method. Two orthogonal chrominance signals were extracted from the RGB traces of the RGB spectrum, which were then used to run the technique. Using empirically derived fixed coefficients, RGB colours were converted to chrominance colours. The PPG signal was calculated by dividing the orthogonal chrominance colours by their ratio. A normalising method based on the various absorption spectra was also proposed

[6] In optics, Lambert's cosine law says that the radiant intensity or luminous intensity observed from an ideal diffusely reflecting surface or ideal diffuse radiator is directly proportional to the cosine of the angle θ between the direction of the incident light and the surface normal; $I = I_0\cos(\theta)$

[64]. Researchers suggested a superior PPG extraction technique by analysing the optical filter's light spectrum and transfer characteristics in the camera. To correct the RGB colour spectrum, a normalised blood volume pulse vector was created using the Hulsbusch noise-free spectrum model [65]. McDuff, et al. [66] proposed utilising a five-band lens to enhance the optical characteristics of a digital camera. In addition to the standard bands like RGB, the five-band camera extracted alternate frequency bands like orange and cyan. The method was used to increase the overlap between the absorption spectra of haemoglobin and oxyhaemoglobin, which have a wavelength of about 520 to 580 nm. As a result, these chromophores substantially impact the pace of light absorption reflectance fluctuation, which activates the PPG signal.

Yan, et al.[67] presented a signal weighted analysis technique to extract the PPG signal. The authors employed a method involving obtaining a weighted average of the raw RGB colour spectrum traces. The major goal was to increase the signal while reducing the noise. After weighted signal analysis, the signal was treated for baseline drift removal and then de-noised using the wavelet transform to recover the PPG signal. Kumar, et al. [63] developed a technique for estimating PPG using a monochrome camera with a green spectrum filter in 2015. The PPG signal was recovered using a weighted average of many distinct region of interests (ROI's) in the face that was monitored independently.

2.4 Summary and Contribution of the Chapter

This chapter represents the structure of heart and how the blood flows within the human body, also shows the clinical and new emerging methods for heart rate measurement. It also gives the clear understanding of existing methods and techniques used for HR measurement and the limitations of those methods using electrical and optical signals.

Chapter 3
Heart Rate Measurement using Videos

This chapter presents a literature review of existing Heart Rate Measurement (HRM) methods in terms of their use in acquiring and analysing Heart Rate (HR) from videos before submitting the design and development process for the new algorithm for HRM used in this research.

3.1 Literature Review of Heart Rate Measurement Techniques

A review of peer-reviewed journal articles published since 2010 was carried out in renowned worldwide journals with a high impact factor [68-70] on contact-free techniques for heart rate measurement using videos. These criteria were satisfied in the four journals selected:

- IEEE Sensors with impact factor 3.57
- IEEE Transactions on Biomedical engineering with impact factor 4.53
- IEEE Transactions on Instrumentation and Measurement with impact factor IF 4.01
- Biomedical Signal Processing & Control with impact factor 3.88

The first search phrase, *'heart rate measurement'*, returned 260 results across international conferences and journals. The search phrase 'non-contact techniques for measuring heart rate' using videos is shown in Table 3.1.

Table 3.1 Yearly distribution from 2010 to 2021 of the number of published international journal and conference papers relevant to this research

Journal Name	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	Total
IEEE Sensors								1				1	2
IEEE Journal of Biomedical Engineering		1	1							1			3
IEEE Transactions on Instrumentation and Measurement										1	2		3
Biomedical Signal Processing and Control					1	2		2					5
Other International Journals						1		2	3	1	5		12
International Conferences			2	3	2	1	1		1			1	11
Total	-	1	3	3	3	4	1	5	4	3	7	2	36

The distribution from 2010 to 2021 publications annually reveals that the hits have been reported, examined, and ranked as a non-contact cardiac assessment technique.

Of the 36 relevant papers [6, 27, 28, 30, 37, 42, 58, 71-99], two were published in the *IEEE Sensors*, three in the *IEEE Journal of Biomedical Engineering* and three in *IEEE Transactions on Instrumentation and Measurement*, five in *Biomedical Signal Processing & Control*, 12 in *different recognised International Journals* and 11 in *International Conferences*. These all papers were generally interested in contactless methods for heart rate measurement. These papers were mainly concerned with the contactless techniques for heart rate monitoring using videos.

The literature review further showed the HRM using videos and video magnification methods. The proposed work architecture shows the working of videos, and magnification is done to extract the information of pixel values that is combined with machine learning-oriented analysis; hence, HR monitoring using videos and video magnification architectures are discussed in the next section.

3.2 Heart Rate Measurement using Video processing

Both medical and affective computing applications can benefit significantly from remotely assessing physiological activity and measurements. Different approaches for the non-invasive detection of HR using human facial recordings have been proposed by recent researchers. These techniques rely on minor colour changes or facial gestures caused by cardiovascular activity, invisible to the naked eye but may be caught via digital images or videos. Signal processing and machine learning are two ways that have been used. The research papers based on HR measurement using videos are discussed in this section.

Poh, et al. [99] presents method for the measurement of physiological signals using webcam. Recorded videos used to analyse the various parameters like HR, HRV and respiratory rate using the methodology of Independent Component Analysis (ICA). They selected the face as region of interest (ROI) shown in Figure 3.1 and average the pixel values of three channels red, green, and blue. Then ICA used to remove the motion artifacts. The highest peak of component whose power spectrum was shortlisted

for analysis. The result of experiment shows the limitations in terms of fluctuations 15 frames per second (fps) as they used standard computer for data acquisition.

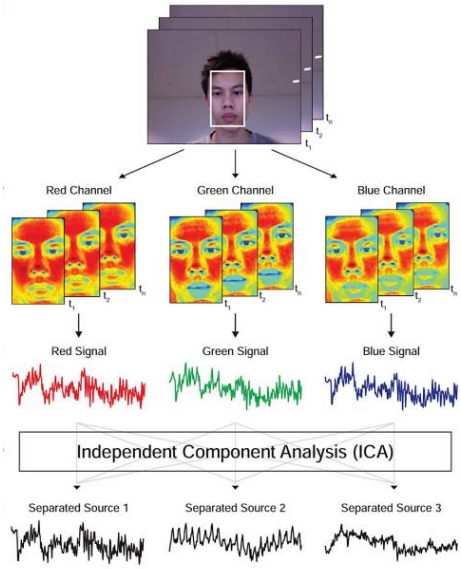


Figure 3.1 Region of interest selection and red, green and blue channels [99]

Kwon, et al. [97] investigated the possibility of remotely measuring a stable heart rate using a face video shot with a smartphone camera. First, a smartphone's front-facing camera was used to capture facial video, then used to recognize the facial area on each frame's picture using face recognition, yielding the raw trace signal from the image's green channel. The author used independent component analysis (ICA) to derive a more precise cardiac pulse signal on the raw trace data. Frequency analysis of the raw trace signal and the processed signal from ICA was used to obtain the heart rate. By comparing the estimated heart rate to the heart rate from a reference electrocardiogram (ECG) signal, the accuracy of the calculated heart rate was determined. Finally, based on this research, the author created FaceBEAT shown in Figure 3.2, an iPhone application for remote heart rate monitoring.

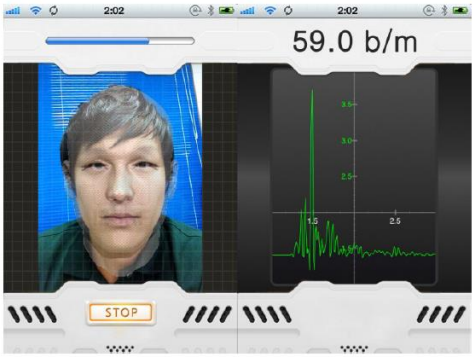


Figure 3.2 FaceBEAT: supports overlaid guide view on the recording(left) estimated heart rate and the FFT results (right) [97]

Scully, et al. [98] shows that physiological signals can be monitored using mobile phone. They use the video monitoring process for observing the skin colour changes due to the cardiac signal. They show that the montioration of HR can be done by using mobile phone as they validate the results with heart reading taken from Electrocardiogram (ECG). They used green signal instead of red and blue bands because of green shows the high absorption of haemoglobin. They mentioned that optimization can be used for improvement of detection of pulse signal.

Balakrishnan, et al. [37] used tiny head oscillations that followed the cardiac cycle to extract film heart activity information. This article's approach may be used to extract additional therapeutically valuable information about the cardiac activity, such as small variations in the duration of heartbeats related to the health of the autonomic nervous system and offering a discreet manner of detecting heart rate. The approach is a supplement to extracting pulse rate from video by analysing the minor colour changes in the skin induced by blood circulation. These approaches average all pixel values in the face area and apply a temporal filter to the signals. Then you can either analyse these signals directly or use ICA to isolate a single pulse wave. The pulse signal's frequency and variability were also taken into account. Head movement, on the other hand, might provide additional information regarding the heart cycle. If the force of blood being pumped by the heart is proportional to head movement, it may be a helpful measure for estimating blood stroke volume and cardiac output. In addition, the movement's direction may suggest asymmetries in blood flow into or out of the brain. This might help determine if the carotid arteries are stenosed or blocked. It can determine the component of motion corresponding to the pulse using a mix of frequency filtering and

PCA and then extract the peaks of the trajectory to identify individual beats. This approach yielded almost comparable heart rates to an ECG and caught certain aspects of InterBeat variability when tested on 18 participants.

Shan Li, et al. [96] suggested a simple and high-efficiency approach for measuring heart rate from a movie. The author derived trajectories in both the X-axis and the Y-axis by merely monitoring one feature point taken from a tiny Region of Interest (ROI) in the head region. The ICA generates a periodic signal after a series of operations such as signal filtering and interpolation, and then the heart rate may be computed. The author tested commercial heart rate measurement equipment (YUYUE YE680B) on ten participants and found a good agreement. A running time comparison experiment with the earlier suggested motion-based approach was conducted, and the results revealed that the new method's time cost was much lower.

Li, et al. [94] proposed framework in this article comprises three major procedures to decrease these impediments: initially, the researchers employ Discriminative Response Map Fitting (DRMF) to discover the accurate face ROI as well as use traceability to discuss the problem caused by rigid head movement; second, the authors use Normalized Least Mean Squares (NLMS) adaptive filter to rectify the interferences caused by illumination variations; and third, signal segments with large standard deviation values are discarded to reduce noise caused by sudden non-rigid movements. The paper shows that all three techniques increase HR measurement accuracy in real-world human to computer interaction (HCI) scenarios. Figure 3.3 shows the framework used in this research to reduce interferences for HR measurement.

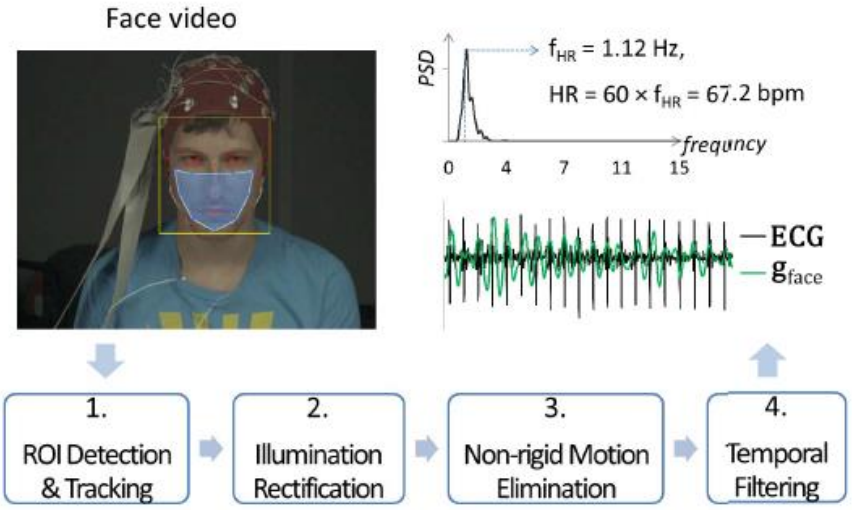


Figure 3.3 Framework for HRM from video [94]

Because all the interferences above are present in the movies, the MAHNOB-HCI database is utilised for testing. Because speech is seldom seen in MAHNOB-HCI films, the mouth area was incorporated in the ROI detection to cover more skin pixels in the ROI. However, the mouth area might be eliminated to prevent motion noise in conversation circumstances. Because the films of MAHNOB-HCI are dark and no other item is present in the scene, the backdrop was utilised as a reference for the step of lighting rectification. Other static items, such as a grey board put behind the subject's head, might be used to reference situations when the backdrop is not acceptable. On all 527 MAHNOB-HCI samples, the new technique significantly outperformed the four prior methods, with an average error rate of 6.87 %. Head rotations of a significant angle, particularly in the yaw direction, influence findings. Half of the face's feature points are lost in such cases, and the tracked ROI position may be incorrect, lowering HR measurement accuracy.

Kranjec, et al. [6] investigated the work related to non-contact HR and HRV parameters. They discussed the different methods for HR measurement includes HR from speech, thermal imaging, RGB imaging, capacitively coupled ECG (CCECG) and doppler effect based HR measurement. They conclude that the non-invasive methods for measuring the cardiac activity are not able to detect the parameters in majority of cases. Every method had limitations like for speech the noisy environment makes the values distorted and appropriate filters needed for that issue. The CCECG need high sensor sensitivity for accurate detection of parameters.

Yong-Poh, et al. [91] demonstrated how video sequences might be used to get dynamic heart rate readings, generally collected from sensors placed near the heart. A video camera is used to collect the face pictures of the seven individuals in two trials. The first experiment measures individuals' growing heart rates (79 to 150 bpm) while cycling, whereas the second involves measuring subjects' declining heart rates (153 to 88 bpm). An independent component analysis is paired with mutual information to guarantee that accuracy is not lost while using short video durations. Measurements of a heartbeat using the Polar heart rate monitor are obtained during both tests to compare with the results of the suggested approach. Overall, experimental findings demonstrated that the suggested approach might quantify dynamic heart rates, with Root Mean Square Error (RMSE) and correlation coefficients of 1.88 bpm and 0.99, respectively.

Christoph Hoog, et al. [92] had demonstrated that two modalities, skin colour change and head motion, can be retrieved from a webcam video stream. The author presented a new multimodal sensor fusion technique for unobtrusive vital sign monitoring. A webcam video stream extracts the motion and photoplethysmographic component arising from heart activity, then combined using a Bayesian technique. The method increased coverage of the beat-to-beat interval estimate from 25% (just motion) to 75% (only PPG) while retaining a low inaccuracy of 32 milliseconds (ms) compared to an ECG reference. Coverage and error were increased to 84 % and 25 ms, respectively, by merging an extra ballistocardiography (BCG) signal obtained in an inconspicuous manner with a mat put on the seat of a chair, as shown in Figure 3.4. By adopting a unique adaptive Gaussian prior, 90 %, coverage may be attained with just 24 ms average absolute error.

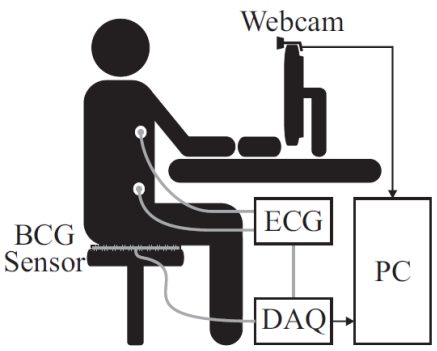


Figure 3.4 Overall set-up using a webcam and BCG mat sensor [92]

Hassan, et al. [58] have reported the advancement in heart rate monitoring utilising face videos has been critically analysed. The authors studied photoplethysmography (PPG) and ballistocardiography (BCG), two widely used heart rate measuring approaches. The ideas and philosophy underlying these approaches were addressed on heart rate monitoring utilizing face videos. The relevance and problems, as well as the recommended contributions to increase dependability, were highlighted. The author tests several approaches on a shared experimental platform depending on their overall importance. The techniques for measuring heart rate were cross-validated, and their dependability was assessed in actual settings. In a realistic context, the experiment results for heart rate measurement in a realistic context indicate that heart rate monitoring using facial videos is promising in simple circumstances (i.e. idle person and controlled lighting). The BCG-based strategy yielded encouraging outcomes. The methods in consistency jeopardized the method's trustworthiness in feature extraction. One of the major benefits of BCG-based approaches was their invariance to light. The motion artefact reduction approach, on the other hand, causes the method to underperform. The key driving trend is photoplethysmography-based heart rate monitoring.

The bulk of the study has been given to improving the trustworthiness. The total signal strength has increased because the direction for obtaining PPG data from various ROI was changed. Heart rate monitoring in actual scenarios, on the other hand, may need a lot more work. The correction of spatial and temporal lighting variation and motion variance produced by changes in body position should get much more attention. The PPG-based strategy might solve the constraints of the approaches related to motion variance and illumination variation by combining the qualities of the BCG-based approach. Using Lambert's rule of light intensity to develop a solution might be a promising step toward developing a dependable heart rate measuring system.

Furthermore, treating the digital camera as a sensor and holding it accountable for quantization noise will assist in reducing the consequences of noise. This article focuses only on heart rate measurement. On the other hand, physiological signal extraction may lead to the determination of oxygen saturation (pulse oximetry), respiration rate, blood pressure, heart rate, and the evaluation of automatic functioning and the diagnosis of peripheral vascular disorders. Because of its non-invasive nature,

this technology may be used in various sectors, including health care, telemedicine, rehabilitation, sports, ergonomics, and crowd analytics.

Jure, et al. [90] worked on measuring HRV which was based on ultrasound transducers which worked simultaneously on two frequencies 40kHz and 39kHz. They used the principal of Doppler effect using two ultrasound transducers. For recording of data, analysis and representation the LabVIEW was used. The experimental setup includes test signal experiments which measure the physiological signals, then laboratory experiment followed by clinical experiments with and without physical activity by participants. They conclude that HR can measure in controlled environment which might be temperature, noise or humidity. The robustness of the method limits the use of proposed work but that can be future work.

Huan, et al. [88] utilizing uncompressed face video capture is the topic of non-contact HR measurement. PPG based remote (non-contact) measurements of human cardiopulmonary physiological parameters may lead to fast and pleasant medical evaluation, which is significant in human healthcare. The fluctuation in human face blood volume throughout the cardiac cycle has been caught indirectly by conventional Red–Green–Blue (RGB) cameras. The authors proposed a novel non-contact HR measurement method that uses the joint blind source separation (J-BSS) method to analyse video data from different facial sub-regions jointly; they also proposed using a max-margin multi-label (M3L) classifier to predict the optimal correlation combination of input datasets from different facial sub-regions to achieve higher HR measurement accuracy; they tested the proposed method on a large public database (DEAP) to verify its performance. The overview of the method is shown in Figure 3.5. The suggested approach was evaluated on an extensive public database, which provided ground truth in the individuals' left-thumb plethysmograph signals. The suggested J-BSS approach outperforms existing ICA-based methods, according to experimental data.

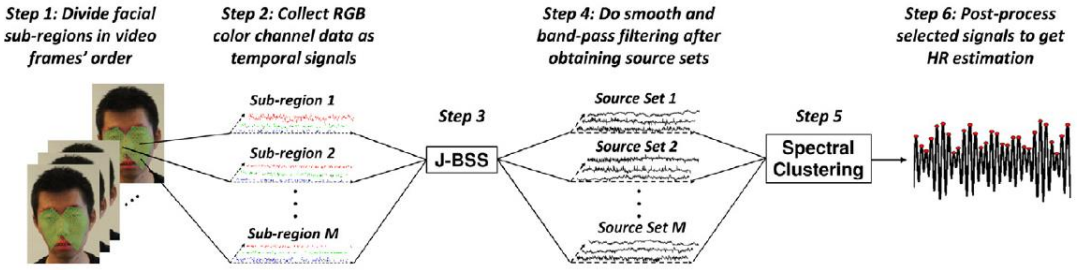


Figure 3.5 Overview of the method for HR measurement used facial landmark and J-BSS technique [88]

Sanyal, et al. [83] pioneered a non-invasive method for measuring pulse and respiration rate using a brief video of the subject's face. Rather than using typical imaging PPG (iPPG) methodologies to monitor the variation of a specific RGB colour space, the authors used the Hue channel in the HSV[7] colour space. This observable is a more accurate and robust technique for measuring vital signs using a video since it largely relies on the AC component of the pulsatile blood. According to this investigation, HR and RR calculated from iPPG acquired utilizing the Hue channel (range 0-0.1) deliver the highest co-related findings with conventional equipment. The HR and RR acquired from iPPG in movies filmed with an extra flash-based illumination are qualitatively superior to those obtained without it. This is further proved by the fact that in our present investigation, the Pearson's r and RMSE values achieved using Hue (0-0.1) at rest are 0.9201 and 4.1617, respectively, compared to 0.89 and 6 acquired utilizing the green channel. These results might be readily turned into a smartphone camera application that measures HR more precisely than existing market options utilizing a camera flash as a light source. Smartphone apps (or APIs) combined with such technology will have further uses in the video-based telemedicine sector as a Software As A Medical Device (SAAMD), enabling the typical user to monitor their HR and RR without purchasing extra equipment. Tele-medicine comprises tele-hospital care (where a consultant doctor may phone in to monitor patients) and tele-home care (where a patient can begin a remote healthcare connection with a network of clinicians, normally accessible 24/7 for non-emergency care). The emergence of tele-home care systems such as Babylon Health, MDLive, Doctor On Demand, Teladoc, and LiveHealth Online has expedited the clinical significance of telemedicine.

[7] HSV (for hue, saturation, value; also known as HSB, for hue, saturation, brightness) are alternative representations of the RGB colour model, In these models, colours of each hue are arranged in a radial slice, around a central axis of neutral colours which ranges from black at the bottom to white at the top.

Ernesto, et al. [84] provide a novel approach for estimating the instantaneous beat by beat pulse signal employing the Hermite transform and deep learning. First, the Hermite transform is used to magnify the variations in brightness, and then a convolutional neural network (CNN) is trained to predict the beat by the beat pulse signal. The authors used a regular camera to gather data for the tests. The results are confirmed using personal pulse monitoring equipment. To evaluate this method, image sequences and beat-by-beat signal pulses from a common RGB camera and a personal contact pulse monitoring device are used. In contrast to Gaussian based magnification, the RMS errors show that the Hermite magnification approach paired with the deep learning strategy performs better in estimating the beat by beat pulse signal.

Ghanadian, et al. [85] focused on heart rate measurement. It developed an algorithm that employs independent component analysis (ICA) to distinguish the source (physiological) signal from noise in the RGB channels of a face video. To enable subject mobility during video capture, the author generalized current approaches. Additionally, by implementing a light equalization scheme to reduce the effect of shadows and unequal facial light on HR estimation shown in Figure 3.6, a machine learning method to select the most accurate channel outputted by the ICA module, and a regression technique to adjust the initial HR estimate, the accuracy of existing methods was improved.

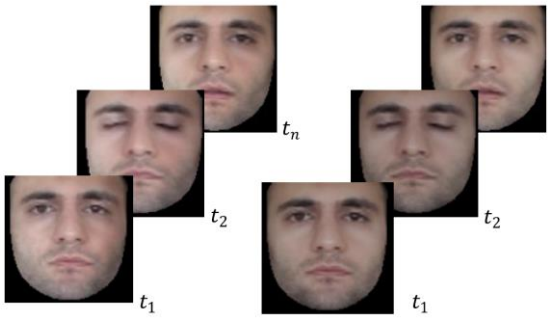


Figure 3.6 Face images: before light equalization (left) after light equalization (right) [85]

In the stationary situation, the suggested technique reduced the RMSE by 27% compared to the state-of-the-art ECG measurement ground truth. The proposed approach obtains an RMSE of 1.12 bpm while the subject is moving.

Zitong, et al. [80] suggested a two-stage, end-to-end strategy that uses hidden rPPG information improvement and attention networks to recover rPPG signals from highly compressed face movies and counter video compression loss. The author proposes an end-to-end architecture consisting of a video improvement module STVEN (Spatio-Temporal Video Enhancement Network) and a sophisticated signal recovery module rPPGNet for robust rPPG measurement directly from compressed movies. The rPPGNet can correctly assess HR and HRV levels because of its skin-based attention module and partition limitations. The proposed rPPGNet generates rPPG signals with richer curve forms and peak positions. Furthermore, even without employing the STVEN module, it outperforms state-of-the-art algorithms on a benchmark dataset's multiple video formats. The STVEN is the first video compression improvement network to improve rPPG measurement on highly compressed movies. It is a video-to-video translation generator with fine-grained learning. The author performs a cross-dataset test and finds that the STVEN can generalize well to improve unseen, highly compressed movies for rPPG measurement, implying that it has real-world potential.

Chen, et al. [82] present a complete overview of this promising technology, emphasizing recent improvements that address lighting changes and motion distortions. Estimating and monitoring heart rate is critical for determining a person's physiological and mental health. According to recent research, HR may now be extracted remotely from face video-based photoplethysmographic signals acquired with professional or consumer-level cameras. Many attempts have been made to increase the non-contact technique's detection accuracy. A broad methodology for measuring HR in each situation was shown. The most popular techniques for each condition were presented, contrasted, and analysed to disclose their principles, benefits, and drawbacks. The approaches used to remove motion-induced artefacts were divided into four categories: BSS-based, model-based, motion-compensated, and others. The future possibilities of this intriguing technology are examined, and other research avenues. Many healthcare applications would benefit from such a remote HR measuring technology, according to the authors, since it is inconspicuous while delivering comfort and ease.

Fernandes, et al. [79] First, the RGB colour space is transformed into LAB colour space, which is made up of three separate signals: L, A, and B, where L stands for picture lightness and A (red/green) and B (yellow/blue) stand for the combination of other colour channels. These two-colour channels were taken from the LAB colour

space, and HR was calculated using those colour channels. The suggested technique uses a web camera to provide a non-contact HR monitoring system based on LAB colour space. The illuminance of the surroundings has a major impact on the colour variety of the bodily blood. The suggested approach may reduce the illuminance impact during HR extraction from face pictures. The process for measuring heart rate from video records of the human face and implementation using a camera with ambient sunshine as lighting are presented here. The provided algorithm seems to be quite strong, successful, and simple to implement in the regular monitoring of home care patients. According to the findings, the heart rate in a normal state ranges from 60 to 90 beats per minute. It ranges from 80 to 110 beats per minute in a grin state. It ranges from 75 to 95 beats per minute in a sad state, and in an angry state, it ranges from 80 to 100 beats per minute.

Gupta, et al. [78] used MOMBAT, or monitoring utilizing Modelling and Bayesian Tracking, which is a new HR monitoring approach suggested in this research. It has used facial recordings obtained from a low-cost camera for HR monitoring contactless. Due to facial expressions, out-of-plane motions, camera specifications, and ambient conditions, HR monitoring via face recordings might be error-prone. MOMBAT has addressed these limitations and improved HR monitoring by incorporating pulse modelling and Bayesian HR tracking. By effectively reconstructing the bad quality pulse signal estimates utilizing the high-quality pulse signal estimates, the suggested Fourier basis-based modelling mitigates out-of-plane motions. Because of the noise, certain HR estimations may be erroneous, but the suggested Bayesian decision theory-based HR tracking mitigates such occurrences and improves HR monitoring. When pulse modelling and HR tracking are used, experimental findings show that HR monitoring may improve greatly. They further claimed that the MOMBAT monitors in near real-time, with an average absolute inaccuracy of 1.3293 bpm and a Pearson correlation of 0.9746 between projected and actual HR. This suggests that the suggested approach, MOMBAT, may be utilised to monitor HR successfully. When the facial video has motion that lasts for a long time, the proposed approach MOMBAT may perform spuriously.

Song, et al. [73] described a novel rPPG approach for remote HR estimate from face films based on CNN. Through a ResNet-18 network, the approach uses a feature-decoder architecture to convert the HR feature picture to the associated HR value. The

spatiotemporal feature pictures are created utilizing pulse signals derived from traditional rPPG techniques in a time-delayed manner. Synthetic feature pictures produced from ECG or BVP signals were initially used to train the CNN model. Real feature pictures created from noise tainted rPPG pulses are used to improve it further. To comprehensively explore the accuracy and generalization capabilities of the proposed technique, the authors conducted both within-database and cross-database investigations. Compared to traditional and DL-based approaches, experimental findings show that the method achieves overall state-of-the-art performance on the public MAHNOBHCI and ECG-Fitness datasets. The authors also looked at a few critical elements that may impact the method's effectiveness. It suggests that balancing training samples, using colour feature maps, pretraining with synthetic feature maps, and using high-quality rPPG signals as feature map sources are required to increase the proposed method's HR prediction accuracy.

Ronca, et al. [71] goal was to see whether a new video-based approach could be used to estimate neurophysiological parameters (such as eye blinks rate (EBR) and HR) while dealing with various tasks and if it may be a feasible option for patient telemonitoring. In terms of the NB and WEB tasks, the findings confirmed the video-based technique's dependability compared to laboratory technology, typically regarded as the gold standard. Furthermore, according to the repeated measure correlation analysis, the video-based approach captured the dynamics of the examined neurophysiological parameters with the same capacity as the laboratory equipment. The statistical studies confirmed the competence of the studied video-based approach in differentiating between nominal and non-nominal participant mental states, which is significant for future applications. The normalized EBR assessed in the NB and WEB tasks declined considerably during the non-nominal condition. Still, the normalized HR rose significantly in both tasks during the non-nominal condition. Subjective measurements, such as the NASA-TLX and SAM, revealed that the nominal and non-nominal circumstances were genuinely different in mental strain for the two experimental tasks, verifying the experimental hypothesis based on the provided analysis. Such data paves the way for video-based approaches for patient monitoring in distant places. In reality, such a procedure does not need any physical contact between the patient and the sensor, nor does it require the presence of a doctor or a facilitator to set the sensor. In addition, compared to current telemedicine platforms, which involve commercial and medical wearable gadgets, the video-based approach has relatively low costs since it simply

requires a commercial camera. The investigated video-based approach might be useful in operational and industrial applications and telemedicine and remote healthcare applications. In this regard, the ability of possible algorithms to automatically distinguish between the state in which the operator is active and learning, and the state in which the operator is resting, is a critical factor in triggering the activation of artificial intelligence (AI) or support system platforms. Furthermore, the findings revealed that the video-based approach for EBR and HR estimates is sensitive to visual attention and mental workload increases. As a result, such a strategy might be useful in situations where there is a minimum of interference between the subjects and the sensors, such as evaluating air traffic controllers' mental effort and attention (ATCOs) and monitoring vehicle drivers.

Table 3-2 shows the comparison of methods used for measurement of heart rate. It shows the process and the gap for further research work. The non-invasive nature of these applications and methods opens possibilities for health monitoring towards various fields such as health care, telemedicine, sports, ergonomics, and crowd analytics. From the comparison its clear that face video is the main input in any process and gives the enough information which researchers can use for HR estimation. Table also shows the gap in research, lack of comparison between different models, limited dataset, standardization in remote patient monitoring, these all limitations gives the clear picture to work on large dataset and comparison with various algorithms and machine learning methods can be used for improved results of heart rate measurement.

Table 3.2 Comparison of methods used to measure the heart rate

Reference	Inputs	Process	Outcome	Gap
Poh, et al. [99]	Face video	Independent Component Analysis	Physiological parameters/HRV	video sampling rate is much lower than the recommended rates (≥250 Hz) for HRV analysis
Monkaresi, et al.[100]	Face video	Machine learning/HCI/ICA	Heart rate	Lack of validation on a large dataset
Liu, et al. [101]	Video of a person's face	A multimodal quasi-contactless HR sensor /BCG and rPPG approach	Heart rate measurement	Limited evaluation in case of large motions
Ronca, et al. [71]	Videos of facial regions/ECG signal recordings and eye blink	PCA	Heart rate and Eye blink rate	Lack of comparison between different situations/person have to stand in front of camera
Kwon, et al. [97]	Facial video	FaceBEAT/iPhone application	PPG Signal	Limited evaluation on small dataset
Hassan, et al. [58]	Face video	PPG and BCG	Heart rate	Motion and illuminance variation effects the results
Song, et al. [73]	Face video	rPPG with CNN	Heart rate	high-quality rPPG signals required to improve the accuracy
Poh, et al. [91]	Face video	ICA	Heart rate	Gives better result for subject in rest position, motion artifacts effect the results

3.3 Video Magnification

Although a person's visual system has limited perceptibility in both spatial and temporal domains, technology may disclose many invisible signals beyond the range of human perception. These signals include vital information, such as a minor change in human skin colour due to blood circulation. Although this difference is not visible to the naked eye, it may be used to estimate the number of heartbeats. Similarly, magnification may show motion invisible to the human eye and has minimal spatial energy, allowing us to exploit fascinating behaviour.

Magnification examples for colour and motion changes are shown in Figure 3.7 (top) depicts a man's face as viewed by the naked eye, with no modifications visible. His pulse's colour fluctuation on his face is readily seen in the processed photos in Figure 3.7 (bottom). The so-called computerized microscope results from many methodologies and strategies for seeing these differences in either motion or colour. Computerized microscopes rely on computing rather than optical magnification to intensify small colour and motion changes in high-speed or conventional movies. Algorithm success has aided the development of current ways for detecting unnoticed signals in videos. The ability to amplify minute differences in an imaging video has opened up new biology, healthcare, and mechanical and materials engineering possibilities. A research group from the Massachusetts Institute of Technology's (MIT) computer science and artificial intelligence laboratory suggested a video magnification technique based on cluster trajectories in 2005 [29]. The technique enhanced tiny colour or motion changes over time, allowing previously unseen changes to be seen [30].

Figure 3.7 Colour Variation Magnification: Input frames (top) and Magnified frames (bottom) [30]

Video magnification methods effectively emphasize minute differences, which have wide uses. Consider detecting respiratory rate from chest movement or skin colour changes, estimating heart rate from blood flow in the human face or head wobbles, measuring a person's pulse from tiny blood vessel motion, recovering intelligible speech as well as music from high-speed videos of a vibrating potato chip bag or houseplant, magnifying geometric deviations, indicating material properties from small motions, and measuring the fluid depth or velocity from high-speed videos of a vibrating potato chip bag or houseplant. Unfortunately, these useful tiny signals are often found amid huge movements, distorting the findings of classic video amplification techniques [93].

3.3.1 Eulerian Video Magnification

The video magnification technology works like a microscope, exposing minuscule signals that would be invisible to the human eye otherwise. Most recently suggested approaches are based on a Eulerian viewpoint, a word taken from fluid mechanics. In contrast to Lagrangian approaches, Eulerian techniques may magnify tiny displacements or fluctuations over time without requiring explicit optical flow calculation. Eulerian video magnification (EVM) has demonstrated promising results in the colour viewing of face video and modest motion magnification. The linear EVM has the drawback of being able to handle only tiny magnification factors in regions with high spatial

frequencies. When the magnification factor is raised, it may also dramatically magnify noise. As a result, strategies for phase-based Eulerian motion amplification are presented. The Fourier shift theorem is used to create a link between phase fluctuations and movements in space, and these approaches are based on it. The techniques offer greater noise management qualities than the EVM and can handle higher magnification factors. Unfortunately, these approaches are ineffective in magnifying the tiniest differences in the face of substantial motion. Such motion will result in substantial blurring artefacts, overpower the modest temporal fluctuations emphasized in the video [28].

EVM's core technology takes a time series of pixels and amplifies any fluctuation in a specific temporal frequency region of interest. Figure 3.7 shows, for example, how believable human heart rates may be used to magnify a temporal band. As blood travels through the face, the amplification highlights the differences in redness.

EVM has been a popular research area in a variety of interesting applications since 2012, including extracting the depth and velocity of hot air, human feeling detection, Android smartphone software, plasma physics, sound reconstruction from a distance using the vibration measurement of an item in a high-speed video, rescue, biology, mechanical engineering, and civil engineering. EVM has the potential to be useful in medical diagnostic and monitoring applications. Most medical applications, such as assessing vital signs without touching patients and showing otitis media in babies, rely on the recovery of temporal aspects employing EVM's colour amplification capabilities. There are two types of EVM techniques: linear and phase based EVM. Motion in the video is proportional to intensity variation across a first-order Taylor series expansion in linear techniques. A video sequence is used as an input, and the frames are decomposed and filtered using a temporal filter. The temporal zone that is created is magnified to reveal concealed information. Although this approach is straightforward and identifies minor motion fluctuations quickly, it is susceptible to failure when the magnification factor is high because the Taylor approximation becomes erroneous. To solve this difficulty, phase-based magnification (PBM) substitutes Fourier decomposition with a complicated steerable pyramid for the linear approximation. The motion in various video frames is proportional to the differences in phase of the pyramid coefficients over time. To see motion, these variances may be temporally processed and then enhanced.

In contrast to the linear-based method, the phase-based technique is more sophisticated and takes longer to process, but it may allow more motion amplification. Eulerian linear and phase-based techniques are quicker and have less noise than Lagrangian-based video magnification. On the other hand, EVM techniques do not function well for big and arbitrary movements.

Small movements and changes in videos may be revealed and amplified with EVM. Because the purpose is to magnify the pyramid levels that contain the movement frequencies, EVM is applied to every level inside a pyramid rather than over the original photos. The required frequencies are then multiplied by a quantity known as the magnification factor, which the user sets. EVM enhances real motion, allowing us to perceive motions that would otherwise be undetected by the human eye. To create the final film with an exaggerated motion, the magnified values of the required frequency are added back to the not magnified ones of the same level.

Although several approaches operate on the same fundamental premise as EVM, they function differently. The pyramid used in the algorithm is one of the primary variations between these strategies. The linear video magnification (LVM) approach decomposes the video sequence according to spatial frequency using the Laplacian pyramid decomposition technique, followed by temporal filtering, as shown in Figure 3.8. The operation's output signals are then multiplied by a factor and added back to the signals sent into the temporal filter. On the other hand, low magnification factors are supported by this strategy. An Eulerian approach based on the 'complex steerable methodology,' which was inspired by phase-based optical flow, is presented to overcome this challenge. Compared to LVM, this approach accommodates huge magnification factors, has fewer artefacts, and generates less noise. However, due to the intricacy of the pipe representation of the steerable pyramid, this technique takes longer to process. This approach may take up to 21 times longer than the LVM method. Enhanced EVM can handle magnification factors higher and lower in noise than LVM. The efficient motion magnification system (EMMS) approach was created to boost processing speed based on wavelet decomposition. This approach allows for huge amplification factors, increases implementation speed, and lowers noise [28, 76].

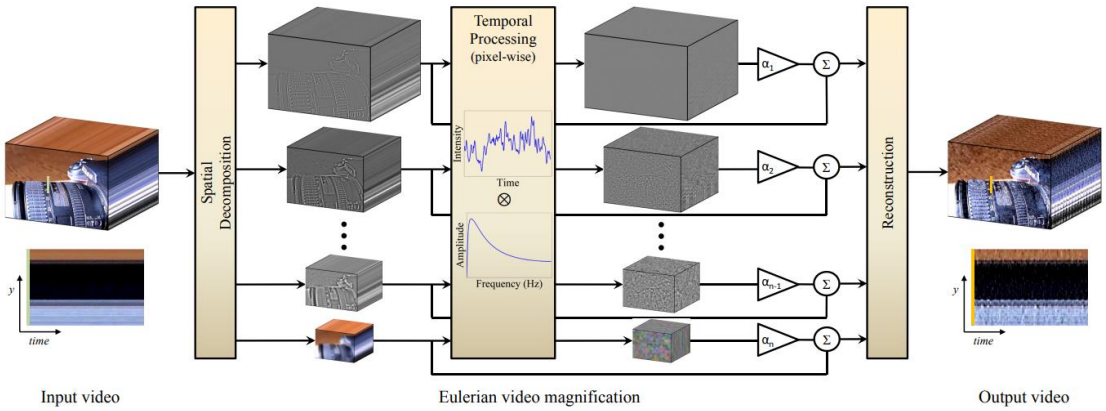


Figure 3.8 Linear based EVM [30]

The essential concept of EVM is to process each pixel's time sequence of colour values individually. This is accomplished by applying normal 1D temporal signal processing to each time series to magnify a band of temporal frequencies of interest, such as about 1 Hz (60 beats per minute) for colour changes and movements linked to heart rate. The resultant time series at each pixel produces an output video in which minute changes in the input. The notion of applying temporal signal processing to the colour values of each pixel is a simple one that has been explored before for ordinary movies. However, the findings are restricted since such processing is incapable of handling broad spatial phenomena such as massive movements involving complex space-time behaviour across pixels. When a big motion occurs, colour information flows across many pixels, necessitating a Lagrangian viewpoint to calculate motion vectors. One of the most important contributions of our research is the proof that Eulerian processing may accurately estimate the amplification of tiny movements in a particular situation of small motions. First-order Taylor arguments may be used to prove that linear, per-pixel amplification of colour fluctuations roughly approximates a bigger version of the motion since the movements are modest.

3.3.2 Phase-Based Video Magnification

The Eulerian technique of video magnification can process the time series of colour values at each pixel independently and without calculating movements explicitly. However, linear amplification increases noise power, and its dependence on first-order

approximations restricts its breadth. There was a representation for films that are global translations of a frame through time, which is the Fourier series. According to the Fourier shift theorem, its basic functions are complex-valued sinusoids that may be translated accurately by adjusting their phase. On the other hand, using the Fourier basis would restrict us to just being able to handle the same translation over the whole frame, preventing the amplification of complex spatially changing movements. Instead, they employ spatially local complex sinusoids realised via a wavelet-like representation termed the complex steerable pyramid to manage such movements. Images are decomposed into a sum of complex wavelets corresponding to distinct scales, orientations, and places in this representation. Like the amplitude and phase of a complex sinusoid, each wavelet has a local amplitude and phase concept.

Wadwa, et al. [95] worked on Phase-based video magnification techniques, which may enhance and reveal subtle motion changes that the human eye cannot detect. These techniques break down each video frame into an image pyramid, allowing for subtle motion changes to be recognised as phase shifts with varied orientations at each pixel and pyramid level. As a result, high-speed video magnification is made more difficult by the time required to calculate the local phase shifts. According to recent research, the Riesz pyramid is a decomposition strategy which can only detect changes in the predominant orientation at a local level. As a result, high-speed processing may be achieved by reducing over-completion and decreasing the arbitrariness of orientations. Nevertheless, when the input video quality improves, an increasing amount of data must be analysed, resulting in a prolonged computation period.

Shahadi, et al. [75] research offers a phase-based-EVM (PB-EVM) upgrade to drastically minimize video magnification processing time. The suggested method decreases the processed data by scaling down the input source video and subsequently resizing up the PB-EVM output frames. The proposed resizing technique uses the Lanczos-3 algorithm to keep the quality of the enlarged video the same as it is in the traditional PB-EVM. The Lanczos-3 technique resizes the input video in the suggested method. Then, using a steerable pyramid, it decomposes video frames to produce a multi-scale frame with its orientation. The filtered frames are then multiplied by a magnification factor before being filtered by temporal filters for particular bands. Both the magnified and unmagnified portions are now put together for each frame. Finally, using the inverse steerable pyramid, reassemble the resulting enlarged multi-scale

frames. The experimental findings reveal that the suggested technique outperforms the traditional phase-based EVM in terms of processing time, with a 60-65% decrease in processing time. Furthermore, their method does not affect video quality, keeping it within the traditional Phase-based EVM's bounds.

Shahadi, et al. [74] strategy reduced amplified noise in the enlarged video for both colours and motion amplification. A linear or phase-based Eulerian method was used to process the resulting video to cancel out any background noise. The approach employs the wavelet denoising technique to locate the frequencies of dispersed noise in different frequency bands. Figure 3.9 shows the wavelet denoising transform. Digital video magnification, a computer-based microscope, was used to detect small changes in recorded films of human eyes. A wide range of applications is possible with this technology, including medical, mechanical, and physical ones. The Eulerian method dominates video magnification.

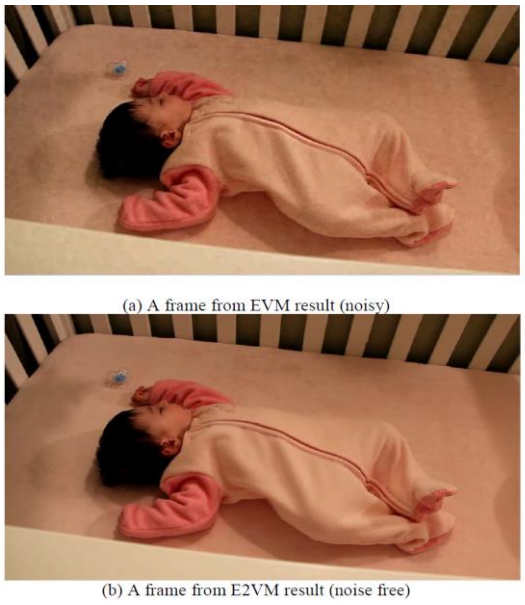
Figure 3.9 Process of image denoising using wavelets [74]

Contrary to popular belief, loudening the video's subtlety can lead to a higher noise level. By decreasing their amplitude, localized frequencies reduce the energy of the coefficients. This method exceeds typical linear and phase-based Eulerian television magnification techniques in terms of the quality of the enlarged movies according to experimental data. A larger amplification factor may be applied to the movies, allowing for a wider range of applications to be added to the list of Eulerian video magnifiers. Furthermore, the processing time does not significantly increase compared to standard EVM; the increase was approximately 3% of the total processing time.

3.3.3 Enhanced Eulerian Video Magnification

Le, et al. [27] introduced a post-processing method to improve the EVM technique based on spatiotemporal filtering. The authors used EVM as a video spatiotemporal motion analyser to achieve pixel-level mapping.

Video pixels were wrapped based on pixel-level motion mapping to amplify the motion input. E2VM supports larger magnification factors better than EVM and is less affected by frame noise because it does not involve modification of pixel values. But it is time-consuming, and some magnification specifications may be lost during image warping. Furthermore, because the E2VM uses the EVM as part of its algorithm, any failure in the EVM affects the quality of the results [86]. Figure 3.10 shows the difference between EVM and E2VM results regarding noise handling.



(a) A frame from EVM result (noisy)

(b) A frame from E2VM result (noise free)

Figure 3.10 Results of Eulerian video magnification and Enhanced Eulerian video magnification [27]

Mohsen, et al. [31] used four video magnification approaches (EVM, PBM, FPBM, and E2VM) and uncovered small findings that are impossible to detect with the naked eye. Observation of minute changes in a video source during the spatial-temporal phase of EVM. The Laplacian pyramid is used to analyse each video sequence frame at various frequencies and resolutions. However, alternative video magnification techniques are more effective than EVM in terms of results since they can enable higher magnification factors while reducing noise and improving video quality. In terms of performance, video quality parameters, and execution time, the E2VM approach is superior to other methods, as shown in Figure 3.11. Most video quality metrics examined on four source

videos were superior values for this video. E2VM has a decent execution time (15–20 % slower than EVM) than other approaches. However, compared to alternative approaches, the EVM is still faster. All approaches require a long time to process videos shot with high-resolution cameras at high frame rates.

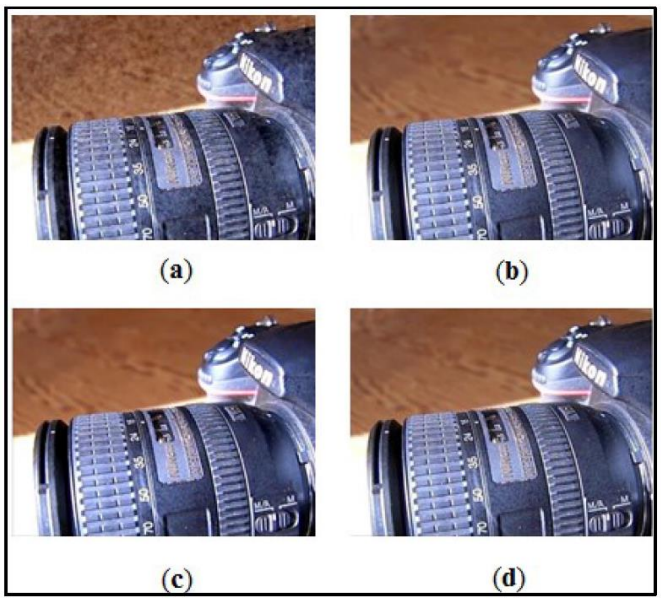


Figure 3.11 Magnified image using (a) Eulerian video magnification, (b) Phase based magnification, (c) Fast phase-based magnification, and (d) Enhanced Eulerian video magnification [72]

3.3.4 Efficient Motion Magnification System

Efficient Motion Magnification System (EMMS) outperforms other magnification methods in noise removal, output video quality, overall performance at large magnification factors and reduced time to execute. Ali, et al. [86] proposed EMMS, as shown in Figure 3.12.

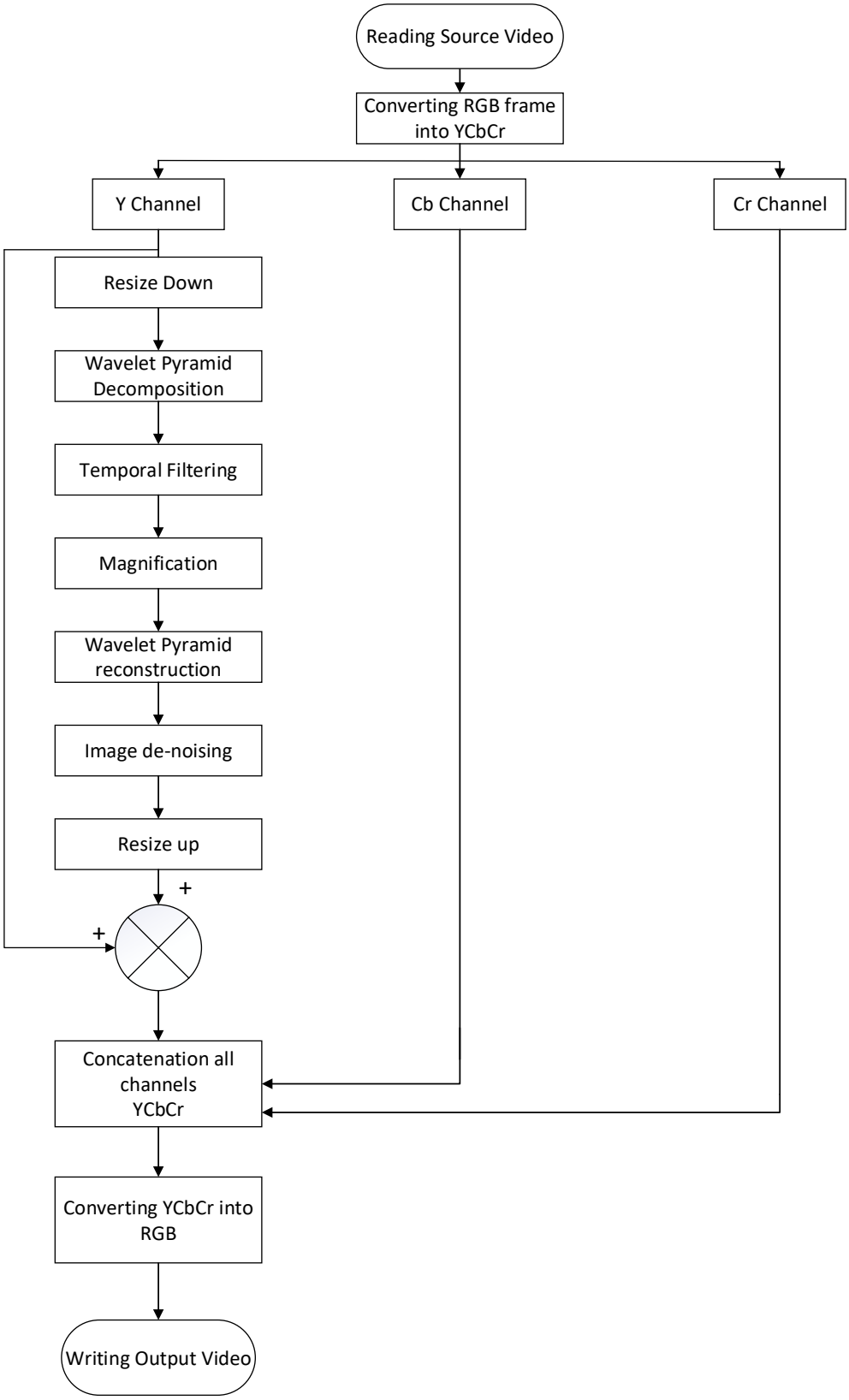


Figure 3.12 Block diagram of efficient motion magnification system[86]

The algorithm of EMMS considered three main steps as follows:

Step (1) – Convert RGB to YCbCr

Separate the intensity information from the colour information by converting the YCbCr colour space of RGB from the source video sequence.

Step (2) – Resampling

The intensity channel (Y) was resized down 50% by using the Lanczos resampling method to reduce processing time.

Step (3) – Decomposition

The Y channel is decomposed into different spatial frequency bands using a Multiresolution Wavelet Pyramid Analysis.

In EMMS method [86] used multiresolution pyramidal wavelet decomposition to decompose an image into a sequence of image transformations with a different spatial resolution to extract the information contained in an image. Wavelet decomposition has local support in the space and frequency domain, and fast algorithms are available.

Multiresolution pyramid analysis allows it to decompose an image into a sequence of image transformations with a different spatial resolution to extract the information contained in an image [86]. The multiresolution decomposition is based on a 'wavelet transform' to provide a time-frequency representation of an image and to analyse an image in horizontal, vertical and diagonal details with different resolution levels. Compared with Fourier-based methods, the wavelet decomposition has local support in the space and frequency domain. In contrast, Fourier-based methods are local in frequency but have global support in the space domain.

The inverse discrete wavelet transform (DWT) was used to reconstruct the wavelet pyramid. 2-D DWT based on the Haar family is considered, as this is the most straightforward to implement among all other wavelet families. Temporal filtering is applied on every single level of the wavelet pyramid based on a zero-phase Chebyshev type 1 band-pass filter. The Chebyshev filter provides a smaller root mean square error (RMSE) than the Butterworth filter during magnification. EMMS shows better results

as compared to other methods of video magnification. EMMS has a low execution time and better video quality.

In the magnification process, the extracted signal from temporal filtering is multiplied by the magnification factor to amplify the signal of interest. The magnified signals are then filtered by using the adaptive median filter algorithm. That algorithm was used because of the property to maintain the edge and detail information and increase the signal to noise ratio for the processed signal. The output signals are resized and added back to the input obtained from the Y channel and then converted into RGB to get the final output.

In EMMS Ali, et al. also compares the execution time in seconds with the methods discussed above and EMMS execution time is less than EVM, Phase-based, fast phase-based and E2VM. Table 3-3 shows the comparison of execution time in seconds. Researchers capture the video baby1 by digital camera with a resolution of 960 × 544 pixels and with a frame rate 30 frames per second (fps). The second video, baby2 captured with a resolution 960 × 540 with 60 fps.

Table 3.3 Execution time comparison in seconds for different magnification methods

Video Magnification Method	Baby1 Video	Baby2 Video
EVM	85.67	85.52
Phase-based method	213.17	211.74
Fast phase-based Method	117.45	115.73
E2VM	101.32	97.84
EMMS	50.68	48.54

The video quality metrics considered in this paper were divided into two categories: subjective and objective. The rate of image quality measures the subjective quality

metrics, and subjective metrics are a reliable way to measure the image quality. The objective metrics are computational, which are used to access the quality of the distorted image with the original image.

Table 3-4 shows the performance-based comparison of video magnification methods. Individual authors use different pyramids and minimize the noise from the resulting video. As shown in the table, EMMS is faster than the EVM process and reduces the noise.

Table 3.4 Comparison of video magnification methods

Parameter	EVM	Phase-Based	Fast Phase-Based	E2VM	EMMS
Decomposition	Gaussian or Laplacian Pyramid	Complex Steerable Pyramid	Riesz Pyramid	Laplacian Pyramid	Wavelet Pyramid
Noise	Magnified	Translated	Translated	Minimized	Minimized
Amplification Factor (α)	Small Value of α	Large Value of α	Large Value of α	Medium Value of α	Large Value of α
Limits of α	$(1+\alpha)\,\delta(t) < \lambda/8$	$\alpha\delta(t) < \lambda n/4$	Similar as complex steerable pyramid	Not stated	Not stated
Over completeness	3/4 image	3-4 times slower than EVM with octave bandwidth	20-80% faster than the phase-based method	15-20% slower than EVM	Faster than EVM

Video magnification has a broad scope in biomedical applications, e.g. for ultrasound [81] and respiratory rate measurement [89], because of its capacity to extract useful information in video and image sequences.

Below are the papers discussed which used video magnification for heart rate measurement.

Yongchao Yang, et al. [87] presented this research for video motion magnification, a new full-field, output only (video measurement) modal analysis approach. Multi-scale pyramid decomposition and representation, as well as an unsupervised learning blind source separation (BBS) technique, are used to model and manipulate the spatiotemporal pixel stages that encode this same local structural vibration in video measurements, allowing it to blindly extract a modal frequency, damping ratio and high-resolution mode shapes from video measurements of the structure. Comparatively speaking, it's far more efficient and self-sufficient than earlier vision-based systems since it doesn't need any kind of surface preparation beforehand. It does not need an explicit optical flow computation, as with the strongly linked phase-based motion magnification technique. Additionally, the phase-based motion magnification noise-resilience was driven and interpreted within identification and visualization of the weakly-excited mode, which was a common and challenging difficulty of operational modal identification. The novel approach was verified against the accelerometer measurement-based methodology in lab experiments. In comparison, the established technique has promise for operational modal analysis, and further research is needed before it can be used in practice.

Xiaochuan, et al. [42] used a digital camera to capture conventional wrist movies in this study. The wrist pulse signals are detected non-intrusively using the Eulerian video magnification approach. The signals are then analysed using the 2-Gaussian curve modelling approach. The frames of the captured films are then subjected to spatial decomposition and temporal filtering. The filtered signal was amplified to view the small motion and colour changes corresponding to the blood moving through the artery. The wrist pulse provides crucial information about a person's health. The wrist pulse's form and properties are important in assessing and detecting aberrant health situations. Measurement of the wrist pulse was considered an essential aspect of traditional Chinese medicine in ancient China. Recent research has sought to quantify this old diagnostic procedure, proposing numerous methods for extracting relevant data from the wrist pulse signal. The PPG pulse signal was measured to validate the efficiency of the Eulerian video magnification technique for identifying the wrist pulse signal. The two signals were captured at the same time and compared. The findings of the experiments show that the Eulerian video magnification approach may be utilised to capture the properties of the wrist pulse signal and can be used to anticipate major cardiovascular events.

Alessandra, et al. [77] used the magnification technique EVM to intensify skin colour variations caused by the cardiac cycle. This article described an imaging photoplethysmography-based oxygen saturation (SpO_2) monitoring approach that does not need physical touch with the patient. During surgery and in intensive care units, peripheral SpO_2 monitoring is essential. Several kinds of research on non-contact monitoring of vital signs have been published in recent decades, although few have focused on SpO_2. The created software runs on a low-cost, widely accessible hardware platform. Nine healthy people were included in the experiments. The suggested system for SpO_2 monitoring is a viable non-contact, low-cost alternative. A controlled hypoxia experiment in the future would be advantageous to allow for more fluctuation in the blood oxygen levels of all subjects.

3.4 Summary and Contribution of the Chapter

This chapter represents the methods of HRM using videos and video magnification from existing available literature. It has been identified that most of the research trend aims to train the system so precisely that the margin of error reduces at the time of classification. It has been concluded from the cited articles that; relevant feature selection plays a vital role in the appropriate training of the samples against the supplied ground truth values. It has also been concluded that big training engines like Neural Networks, Convolution Neural Networks etc., require a bulk amount of data to establish co-relation among the supplied ground truth identities. Hence, pre-processing also becomes crucial for the training of the data. The proposed algorithm extracts frames and magnifies them for the extraction of the forehead to co-related it with the heart rate. Hence, it becomes vital to present video magnification algorithms presented by researchers. This chapter shows the video magnification algorithms and their processing nature through different citations in the chapter. The proposed work contrasts with the prediction of heart rate, and hence the architecture of heart rate, measuring prospective have also been discussed.

Chapter 4
Data Acquisition and Pilot Study

4.1 Introduction

This chapter details the procedure followed for the data acquisition phase of this research using the newly developed algorithm described in Chapters 5 and 6. First, the overa.ll set-up for the data acquisition is outlined, along with the equipment used. The data collection involved participants sitting straight and still on the chair and a pulse oximeter was attached to their finger. Since data was obtained from human participants were part of this research, ethics approval by the Auckland University of Technology Ethics Committee (AUTEC) was required. An application for ethics approval was submitted explaining the research aim and procedures.

4.2 Ethics Approval

Ethics approval (Application Number 21/424) was granted by AUTEC on the 9th of December 2021. Appendix A includes a copy of the completed application forms for the ethics approval for this research.

4.3 Data Collection

The data collection is divided into two sections;

- The first section 4.3.1 describes the data collection methodology and equipment used for the pilot study.
- The second section 4.3.2 describes the data collection methodology and equipment used for the main study.

4.3.1 The Equipment and Set-up for Pilot Study

This section covers the overall set-up for testing each participant's heart rate and lists the equipment used. For the pilot study, two participants were the researcher and

primary supervisor for data collection. The overall layout of the equipment set-up involved in the pilot study is as follows:

- Chair
- Pulse Oximeter (Rossmax model SB100)
- Camera (Go Pro 6 and iPhone7)
- Tripod

The overall set-up for data collection using Iphone7 and Go Pro 6 camera is shown in Figure 4.1. The camera was placed at 1.5 metres on a tripod in front of the participant's face. Then record the videos for a maximum of 15 seconds. Videos were recorded of the participant's face whilst sitting upright in the chair. The pulse oximeter Rossmax SB100 was placed on the participant's finger of their dominant hand while recording the video.

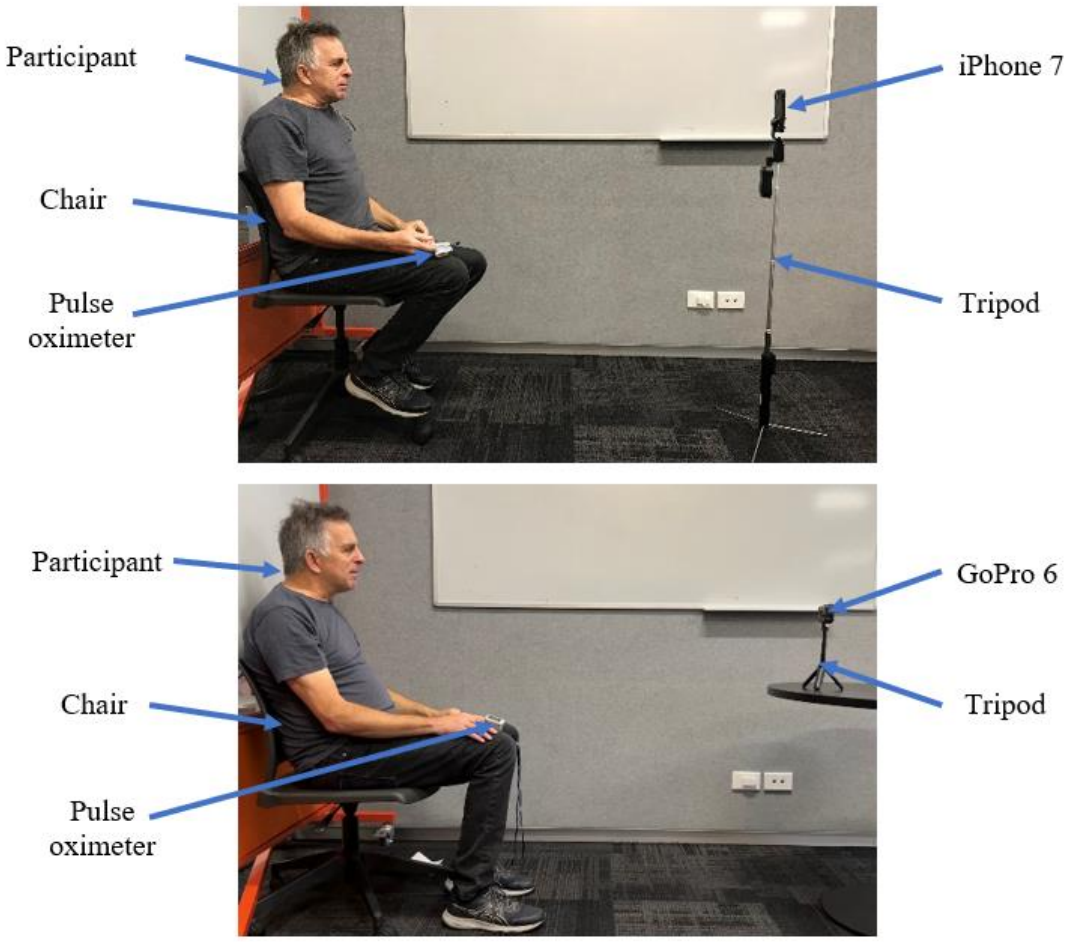


Figure 4.1 Experimental set-up: using iPhone 7 camera (top), using Go Pro 6 (bottom) for pilot study

Table 4.1 and Table 4.2 shows the video specifications used for each input video. Videos with the iPhone 7 camera were recorded, and GoPro 6 videos were in linear mode. The videos collected were analysed using MATLAB2020 running on a laptop with an i5 processor with 16GB RAM.

Table 4.1 iPhone7 camera specifications used for video recording

Frame Rate (fps[8])	Resolution
30	720P[9]
30	1080P[10]
60	1080P
30	4K[11]

Table 4.2 GoPro 6 camera specifications in linear mode for video recording

Frame Rate (fps)	Resolution
24	1520P[12]
30	1520P
60	1520P
24	1440P[13]
30	1440P
60	1440P
24	1080P
30	1080P
60	1080P
60	720P

[8] fps is Frame per second measurement for how many unique consecutive images a camera can handle each second.
[9] 720P camera provides images that are 1280 x 720 pixel (that is 921,600 pixels)
[10] 1080P provides 1920 x 1080 pixel (that is 2,073,600 pixels)
[11] 4K is 4000 x2000 pixel also called ultra-HD (that is 8,000,000 pixels)
[12] 1520P provides 1520 x 2704 pixel (that is 4,110,080 pixels)
[13] 1440P provides 1440 x 1920 pixel (that is 2,764,800 pixels)

The pilot study observed that high-resolution videos are unsuitable for this research because noise and artefacts increased with high resolution and high amplification factor for magnification. So 30 fps videos gave the required results, followed by the main study data collection. It was observed that pulse oximeter readings needed to be recorded manually as the used pulse oximeter could not save the data, so for the main study, a new pulse oximeter was used, which can save data and generate data a pdf file for each participant while recording.

4.3.2 Data Collection for Main Study

For the main study data collection first step was to get ethical approval from the AUT ethics committee. Then the process follows the advertisement, and the consent and personal details record of participants as below:

4.3.2.1 Consents and Personal Details

The following steps were taken before carrying out any testing of the participants:

- The principal researcher explained the purpose of the research and the protocols of the tests, which were designed to measure HR using video magnification.
- The data collector allowed each participant to read the Participant Information Sheet and sign the Consent Form, which is included in Appendix A. Participants were allowed to ask any questions at this stage.
- Before any preparation for the placement of the pulse oximeter, the data collector recorded each participant’s details, such as name and age.

For safety reasons because of COVID-19, the following steps are followed for the collection of data:

- The data were collected by giving the appointment to only one person at a time. Before the next participant's meeting, there was a 10-minute break for cleaning /wiping the pulse oximeter and chair with antibacterial wipes.
- There was a 2-metre distance.
- The data collector wore a face mask and gloves.

4.3.2.2 Equipment and set-up for the main study

The equipment used and the set-up involved for the main study were as follows:

- Chair
- Pulse Oximeter (EMAY EMO-80)
- DSLR[14] Camera (CANON 850 D)
- Tripod
- The laptop ran the application to start and stop the video recording
- Mobile for record data of pulse oximeter using the mobile application of pulse oximeter

The overall set-up for data collection is shown in Figure 4.2. The camera was placed at 2 metres on a tripod in front of the participant's face. Then record the videos for a minimum of 15 seconds and a maximum of 60 seconds. As in the pilot study, a pulse oximeter used was not store data and need to record data manually, but in the main study, the pulse oximeter used is EMAY EMO-80 because of its ability to store the data. As in the pilot study, this step was done manually. The DSLR camera Canon 850 D recorded the videos. Appendix B shows data recorded for the main study in which the participant's number, file name, age, heart rate, and the pulse oximeter datasheet.

Figure 4.2 Experimental set-up for main study

[14] DSLR is Digital Single-Lens Reflex

The pulse oximeter was placed on the participant's dominant hand's index finger and recorded the range of pulse using a mobile application. A laptop was used with the camera application to start and stop the video. The research used the reading to validate the results obtained after analysing the recorded videos. The videos collected were analysed using MATLAB2020 running on an HP laptop with an i5 processor with 16GB RAM.

Data was collected from 30 participants, and two videos were collected from each participant, total 60 videos were recorded for this research. Because of algorithm requirements of minimum movement and forehead selection, only 35 videos were finalised to be used for the main study. Each video was examined before being used for this research, and 25 videos were not selected for analysis because of the movement of participants, shadow, focus while recording and hairs were covering the forehead.

After collecting data (videos), then further processed and analysed offline later, using the image processing and newly developed algorithm is discussed in Chapters 5 and 6.

4.4 Pilot Study

In early stage of this research, a pilot study was carried out using the different methods for ground truth generation, feature extraction methods and ablation study using different algorithms.

4.4.1 The Ground Truth Generation

The ground truth generation mechanism follows statistical Machine Learning (ML) architecture. When there is no ground truth with the provided data set, ground truth can be generated using the co-relation between the existing data elements. To generate the ground truth, the following similarity measures were used.

4.4.2 Cosine Similarity

A measure of similarity between two numerical sequences used in data analysis is cosine similarity [102]. The cosine of the angle between the sequences, which would be defined as the dot product of the vector divided by the sum of their lengths, is used to calculate cosine similarity. Consequently, the angle of the vectors, rather than their magnitudes, determines the cosine similarity. The cosine similarity is usually shown as an interval. An inner product space's cosine similarity is a measure of how close two vectors are to each other. It examines the cosine of the angle between two vectors to determine whether they are travelling in the same general direction. Document similarity may be identified using this technique. As a result, two proportionate vectors have a trigonometric similarity of 1, whereas the same is true for two orthogonal vectors and two opposing ones. The similarity measure is extremely useful in a positive space when the result is neatly contained in display style {0, 1}.

A document may be represented as a series of the number of times each word is used, such as information retrieval and text analysis. The measure is a helpful metric if you're looking to compare two articles based on their topic matter.

In data mining, the approach is also employed to determine the cohesiveness of clusters. Similarity measure benefits from being simple, particularly when dealing with sparse vectors. Otsuka–Ochiai similarity is a kind of cosine resemblance used in binary data, also known as the Tucker coefficient for congruence [103].

Comparing various vector dimensions inside a relative instead of an absolute way is the most striking feature of the similarity measure. For any constant a and vector, vectors V to aV is maximum similar. Because the frequency is more relevant than absolute values, the measure works well for data such as the recurrence of phrases in texts. Semantic improvement has been proven to be possible using metrics based on information theory, such as Jensen–Shannon, and triangle divergence, in at least certain cases.

4.4.3 Euclidean Distance

In mathematics, the Euclidean distance is the length of a line segment connecting two points in Euclidean space. There are two ways to figure out the Pythagorean range: using the Pythagorean Theorem and utilizing the cartesian values of the locations. Even though Euclid did not use numbers to represent distances and the Pythagorean Theorem, these designations are drawn from the ancient Greek mathematicians' Euclidean as well Pythagoras and were not acknowledged until the eighteenth century. Distance between non-point objects is often referred to as the distance between points 1 and 2 from two objects. For example, the distance between a point and a line may be calculated using formulas. Abstract metric spaces and distances that aren't Euclidean have been studied in advanced mathematics. The square of the distance measure is used in several statistical and optimization applications rather than the distance itself [104].

4.4.4 Distance from Plane to Point

Distance between two points is defined as the length between the points of the two particles, although more intricate extensions like the Hausdorff metric[15] are typically used to calculate the shortest distance from points to sets. The formula below may be used to calculate distances between different objects.

- The separation of two points on the Euclidean plane
- The distance between a point and a plane in 3-dimensional Euclidean space
- The distance between two lines in 3-dimensional Euclidean space

The proposed architecture has utilised both cosine similarity and Euclidean distance to form the ground truths if they are not supplied with the ground truth value. Initially, the collection of real-time data was not available, and hence the method of ground truth generation was adopted. In the evaluation, the cosine similarity and the Euclidean distance are calculated from one frame to another. For example, in the case of the Artificial Neural Network (ANN) training sample, there are 342 frames, and cosine similarity and Euclidean distance is calculated from frame 1 to frame 342, as shown in Figure 4.3. The maximum cosine similarity value is 1, and the range of variation is 0 - 1. Cos 90 degrees represents no co-relation, and the value will be 0. In contrast, Cos 0

[15] Hausdorff metric or distance measures how far two subsets of a metric space are from each other.

degrees represents complete co-relation, and the value will be 1, as Euclidean distance is the cartesian distance between the supplied value.

```
        load HOG_Feature AllHOG

for i=1:numel(AllHOG)
    for j=1:numel(AllHOG)
        if i~=j
            EDist{i,j} = pdist2(AllHOG{i},AllHOG{j},'euclidean');
            CDist{i,j} = pdist2(AllHOG{i},AllHOG{j},'cosine');
            CERatio(i) = mean2(CDist{i,j})/mean2(EDist{i,j});
        end
    end
end
```

Figure 4.3 Evaluation of similarities frame by frame

The range of Euclidean distance in the case of the proposed case scenario is from 10-25 units. Figure 4.4 shows the Euclidean distance between frames.

	1	2	3	4	5	6	7	8	9	10
1	[]	12.2006	16.1870	13.2314	19.1286	12.8529	12.9233	14.4061	17.3677	21.1747
2	12.2006	[]	18.1346	12.6604	20.4635	14.7473	11.9594	16.1929	19.1483	21.8402
3	16.1870	18.1346	[]	16.6061	13.3104	13.5584	18.1337	12.5777	11.7020	18.7150
4	13.2314	12.6604	16.6061	[]	19.1847	13.7206	12.9670	14.2989	17.7979	21.1408
5						17.3386	20.2629	16.3645	12.5018	16.4380
6						[]	13.9645	10.8067	14.5278	20.8467
7	12.9233	11.9594	18.1337	12.9670	20.2629	13.9645	[]	15.4872	18.8845	21.7403
8	14.4061	16.1929	12.5777	14.2989	16.3645	10.8067	15.4872	[]	12.8803	20.0534
9	17.3677	19.1483	11.7020	17.7979	12.5018	14.5278	18.8845	12.8803	[]	18.5937
10	21.1747	21.8402	18.7150	21.1408	16.4380	20.8467	21.7403	20.0534	18.5937	[]
11	13.3793	14.0447	16.7347						17.0135	21.4790
12	20.2264	20.7275	18.0289						18.1796	14.7541
13	19.5429	20.5012	15.0698	19.5062	12.3933	18.0685	20.4132	17.0064	13.7596	15.3254
14	16.3918	13.8630	20.1593	15.2753	21.4563	18.2391	13.7597	18.9339	20.8806	22.1094
15	14.9861	16.0435	17.7221	14.3303	19.2784	15.7903	14.3751	16.3379	18.2378	21.0459
16	16.2434	17.7716	15.2583	15.3426	16.1929	15.2653	17.1917	14.6535	15.1764	19.0395
17	19.1730	19.9772	16.3231	17.9433	15.9702	18.3532	20.0948	16.7408	16.9368	16.5770
18	19.4260	20.1474	16.9555	18.1564	16.5127	18.7240	20.2462	17.1956	17.4641	16.4999

Euclidean distance of frame 2 to frame

Euclidean distance of frame 6 to frame

Figure 4.4 Euclidean distance in proposed case scenario

After data collection for the main study, the ground truth was available, so the generation of ground truth was not evaluated further for the results as pulse oximeter values were available and considered as ground truth.

4.5 Feature Extraction Algorithm and Architectures

The different feature extraction algorithms are discussed in this section, and also a comparison of methods is presented. In the proposed algorithm, feature extraction is done using the HOG.

4.5.1 Histogram of Oriented Gradients

HOG is a feature descriptor similar to canny edge detector and Scale Invariant Feature Transform (SIFT). Where SIFT descriptor uses a grid-based calculation of gradient orientation scatter plots inspired by local stance gradient orientation histograms. In contrast to HOG, SIFT is a local image descriptor. The HOG descriptor is defined across subregions of the picture domain, while regional visual cortex histograms were defined over the whole image domain [105].

- The HOG operator, which comprises a series of smaller histograms specified over sub-regions, includes a dependent on picture positions.
- For this reason, the HOG descriptor differs from the SIFT descriptor since it does not normalize orientation. As a consequence, the HOG descriptor does not exhibit rotational invariancy.
- Contrast is taken into account when the HOG operator adjusts histograms.

HOG has certain unique characteristics that set it apart from other feature descriptors.

- About an object's structure or shape, HOG is the descriptor of choice. The only information required is to know whether a pixel is an edge for edge characteristics. HOG may also be used to indicate the direction of the edge. Extracting the gradient and orientation of the edges is how this is done (or magnitude and direction).
- The 'localized' parts are also used to identify these orientations. There is an implied division into smaller portions, each with its slopes and direction. The HOG would generate a separate Histogram for each of these subsections of the image.
- Gradients and their orientations are represented in the phrase "Histogram of the Oriented Gradients" since the histogram comprises these values [106].

4.5.2 Scale Invariant Feature Transform

Image-based matching may be done using the SIFT, a picture descriptor. Image descriptors are often used when doing computer vision tasks like 3-D point matching and view-based object identification. In the picture domain, the SIFT description is not impacted by simple perspective or lighting changes, rotations, scaling, or any other form of alteration since it is resilient. In real-world photo pairing and object recognition investigations, the SIFT descriptor was beneficial.

The SIFT descriptor was first developed to discover interest spots in grey-level pictures by summing statistical data on local gradient directions for pixel intensity. This descriptor was intended to match corresponding interest points among different images. The classification and categorization of objects and textures may be improved by applying the descriptor to thick grids (dense SIFT). The SIFT descriptor can describe pictures in various formats, including grayscale, colour, and two-dimensional spatial images [107].

4.5.2.1 Interest points from scale-space extrema that are Scale-Invariant

When describing the original SIFT description, picture intensities around important areas or focus places in the provided image being used are shown in Figure 4.5 The differential of Gaussian pyramid interest points is located in the scale-space extrema of a Gaussians.

After smoothing and sampling an image for recurrences, the difference between neighbouring levels is utilised to build a Difference-of-Gaussian Pyramid (DoG). Then interest spots are discovered at locations where the difference-of-Gaussians values approach extremes in both spatial coordinates**.**

Figure 4.5 Scale-invariant feature transform keypoint extraction

4.5.2.2 Normalization of scale and orientation

A constant time detection scale is used to construct SIFT descriptor's estimation of an area around an interesting point, which is prompted by the interest point detector's capacity to deliver a distinctive size estimate for each interest point's scale selection process.

Gradient vectors calculate the local distribution of gradient directions at the interest point's detection scale. The accumulation window's size is proportional to that scale to get a preferred orientation estimate. The peaks in this orientation histogram indicate the prevalent orientation. Secondary peaks are permitted if their height exceeds 80% of the height of the highest peak in occurrences in which the interesting point is surrounded by more than one dominating direction. Visual representations for orientation estimates are created for each peak when there are numerous.

Gradient magnitude and a Gaussian window functional centred on the interesting point and scaled to match the detection scale are used to weight the increments when producing the orientation histogram. Large orientation samples are used, with 36 bins in the histogram to improve the orientation estimation. Furthermore, the apex is located using local parabolic interpolation in the vicinity of the histogram's peak [108].

4.5.2.3 Dense Scale Invariant Feature Transform

Scientists have shown that the SIFT descriptor is much more accurate than interest points derived by an interest operator for tasks such as the item positioning process in

scene classification in dense grids in the picture domain. An easy-to-understand rationale for this is that a dense grid of image descriptors, rather than a sparser collection of picture points, yields more information.

- Once SIFT descriptors have been calculated with a high degree of accuracy, clustering is commonly employed to reduce the vocabulary of word vectors that may be used to classify objects.
- To reduce the complexity of generating picture correspondences to a minimum, the identification of sparse interest spots is still a crucial pre-processing step in constructing image correspondences between originally unconnected separate photos of a 3-D object or a 3-D scene [109].

4.5.2.4 Colour Scale Invariant Feature Transform

An illumination model yields a collection of colour invariants. With distinct colour gradients invariant to a wide range of local levels of intensity, shadows, shadings, and highlights that may be generated, the technique develops a collection of colour SIFT descriptors. According to our evaluation, Both a grey-level SIFT operator and previously published colour SIFT descriptors are outperformed by a colour SIFT operator relying on the hue colour spaces [110].

4.5.2.5 Scale Invariant Feature Transform-like image descriptors for Spatio-temporal recognition

The choosing of a Spatio-temporal scale, local adaptation, and, consequently, scale invariance over geographical and temporal scales were observed. In the same manner that local spatial SIFT variables can recognize things and arrange them into categories, it was shown that this approach could identify human behaviours utilizing local Spatio-temporal image descriptors.

Spatiotemporal smoothing operations were adapted to local motions to normalise the possible unknown relative mobility of the world's objects and the observer. This method was shown to enable recognition of spatiotemporal events in crowded backgrounds as demonstrated.

This area of image-based matching and recognition research has been created by the SIFT and its related picture descriptor for receptive field-like images and their histograms. Operation or approximation in scale-space is theoretically well-founded. Therefore, these methods may be used to accurately compute image features and image descriptors from images of the actual world.

Because of its matching methods, the SIFT descriptor may create point matches across various views of a 3-D object or scene. Combining correspondences with multi-view geometry is necessary to create three-dimensional representations of objects and situations using this technique. Some algorithms can stitch together various images of a single item or scene into larger panoramas, such as the multi-view correspondence algorithms mentioned above [111].

4.5.3 Speeded Up Robust Features

Speeded Up Robust Features (SURF) is comparable with the SIFT descriptor because it is a feature vector formed from answers that are more likely to reply in the neighbourhood of an interesting point.

For example, instead of relying on a Laplacian operator for the SURF descriptor, Haar wavelets are used; equity points are approximations of magnitude extrema of a Hessian determinant; as well as the submissions in the feature vector have been computed as totals as well as absolute SURFs instead of histograms of coarsely subatomic gradient directions. As demonstrated in Figure 4.6, the SURF operator yields result identical to the SIFT operator.

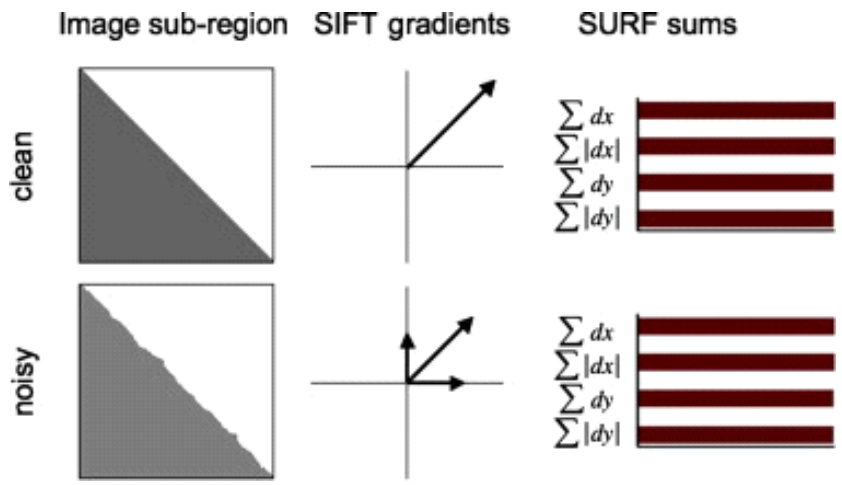


Figure 4.6 Speeded Up Robust Features [112]

It uses just a Hessian matrix [112] blob detector to search for interesting locations. Sites are chosen depending on the Hessian matrix's determinants, a measure of local change that is greatest at these locations. Both Mikolajczyk and Schmid and Lindeberg use the determinants of a Hessian in their Hessian-Laplacian detector. p and scale of the Hessian matrices: Concerning scale and location, the following is the equation (4.1) of Hessian matrix H(p):

$$H(p,\sigma) = \begin{pmatrix} L_{xx}(p,\sigma) & L_{xy}(p,\sigma) \\ L_{yz}(p,\sigma) & L_{yy}(p,\sigma) \end{pmatrix} \tag{4.1}$$

where $L_{xx}(p,\sigma)$ is the convolution of Gaussian second-order derivative with an image I and similarity for $L_{xy}(p,\sigma)$ and $L_{yy}(p,\sigma)$.

In an attempt to mimic Gaussian smoothing, SURF uses square-shaped filters. Scale-invariant characteristic spots are identified using cascaded filters, and the DoG is computed using consecutively rescaled pictures. Filtering a photo with a square is much quicker when utilizing the integral image:

$$S(x,y) = \sum_{i=0}^{x}\sum_{j=0}^{y} I(i,j) \tag{4.2}$$

where x and y are pixel counts in both horizontal and vertical dimensions.

4.5.4 Comparison of Feature Extraction Methods

This section examines the performance of HOG, SIFT, and SURF, three of the most used techniques for extracting image features. How well a method works is measured by how many accurate matches it can find between the original and noisy image. Gaussian, seasoning, and speckle noises have all been used to distort the images. HOG utilises the least time to calculate histograms by using the angle orientation of the edges [113].

4.5.4.1 Results of Histogram of Oriented Gradients

A set of results from the implementation of HOG for this research is presented in this section.

In a Histogram of Oriented Gradients (HOG) image, the dots represent the locations of interest points or key points[16] in the image where the HOG descriptors[17] have been calculated.

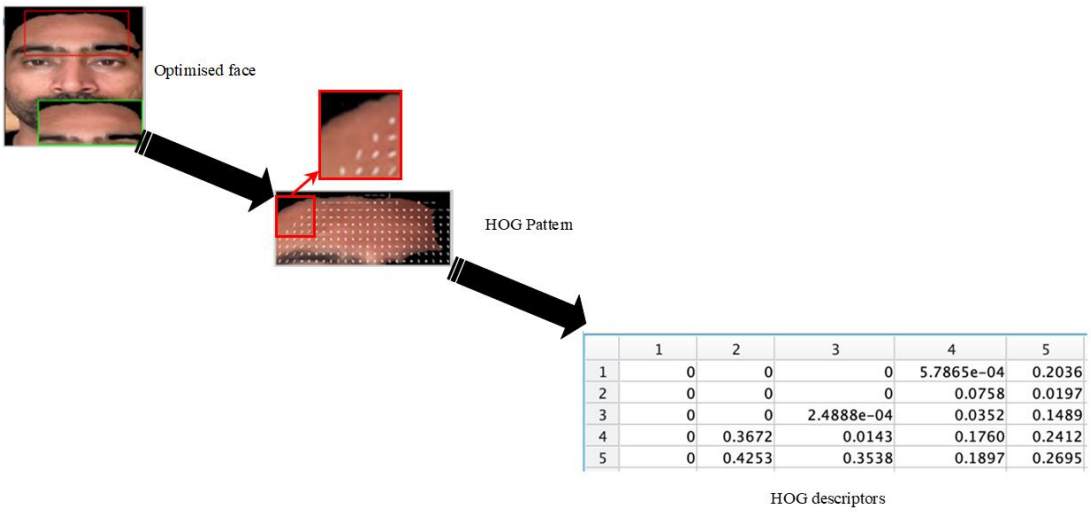
Each key point corresponds to a specific location in the image where local gradient information is analysed to generate the HOG descriptor. The HOG descriptor is

[16] Key points are important and distinctive points in an image that are used for various computer vision tasks, such as image matching, object recognition, and feature extraction.

[17] Descriptors are numerical representations or feature vectors that capture the essential characteristics of an object, region, or key point within an image.

computed in a local region around each key point. This process is repeated for multiple key points across the image, resulting in a set of HOG descriptors, one for each key point [114]. These descriptors capture the local shape and structure information in different image parts, making them useful for object detection and recognition.

In visual representations of HOG images, dots or markers overlaid on the original image, as shown in Figure 4.7, indicating the locations of these key points or interest points. These dots help visualize where the HOG descriptors have been computed and where the analysis of local gradients has occurred. The orientation and magnitude of gradients around these key points are used to construct the HOG descriptors, which are then used for object detection and other computer vision tasks.



	1	2	3	4	5
1	0	0	0	5.7865e-04	0.2036
2	0	0	0	0.0758	0.0197
3	0	0	2.4888e-04	0.0352	0.1489
4	0	0.3672	0.0143	0.1760	0.2412
5	0	0.4253	0.3538	0.1897	0.2695

Figure 4.7 Optimised face, HOG pattern and HOG descriptors

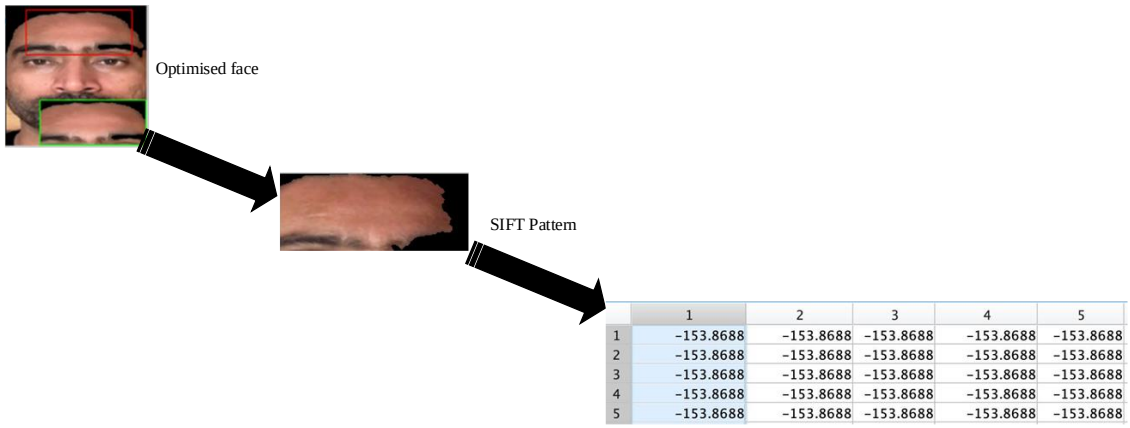
The values in the HOG descriptor represent the distribution of gradient orientations in the local image region. High values in particular bins of the descriptor indicate that specific gradient orientations are prominent in that region, while low or zero values indicate the absence of those orientations.

4.5.4.2 Results of SIFT

A set of results from the implementation of SIFT for this research is presented in this section.

In SIFT key points, a green circle is often used to visualize and indicate the location and extent of a detected SIFT key point. The green circle represents the region around the key point. Uniform regions where pixel values remain the same do not provide the

necessary information to serve as distinctive key points. As a result, feature detection algorithms like SIFT will not find patterns in such areas because they do not exhibit the characteristics that make them valuable for matching and recognition tasks. In this analysis, the value is constant, so no pattern is created, as shown in Figure 4.8.



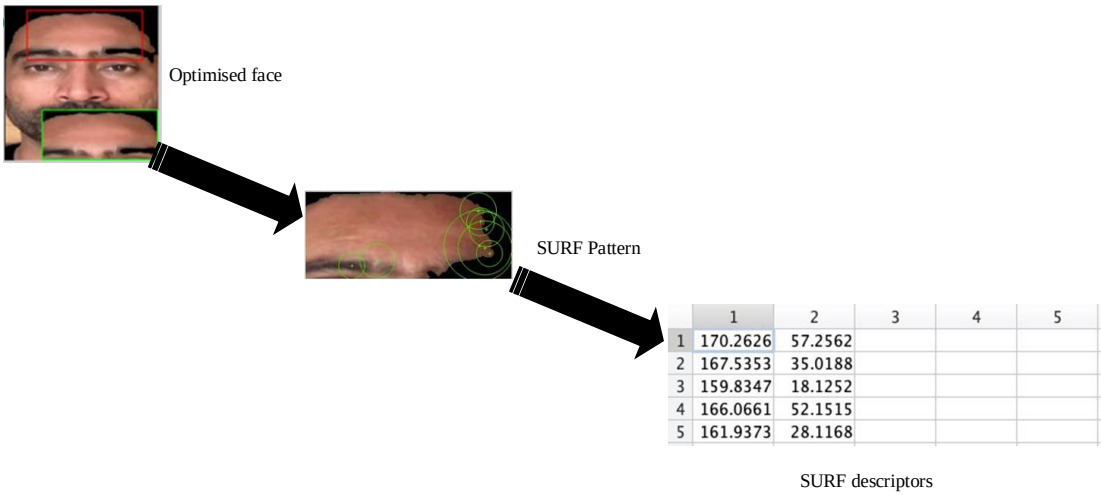
	1	2	3	4	5
1	-153.8688	-153.8688	-153.8688	-153.8688	-153.8688
2	-153.8688	-153.8688	-153.8688	-153.8688	-153.8688
3	-153.8688	-153.8688	-153.8688	-153.8688	-153.8688
4	-153.8688	-153.8688	-153.8688	-153.8688	-153.8688
5	-153.8688	-153.8688	-153.8688	-153.8688	-153.8688

Figure 4.8 Optimised face, SIFT pattern and SIFT descriptors

4.5.4.3 Results of SURF

A set of results from the implementation of SURF for this research is presented in this section.

In the SURF algorithm, a green circle is often used to visualize an image's detected SURF key points or interest points. These green circles serve as a visual representation of the locations in the image where SURF has identified key points. The green circle is drawn around each detected key point to highlight its position and make it visually distinguishable from the rest of the image, as shown in Figure 4.9.



	1	2	3	4	5
1	170.2626	57.2562			
2	167.5353	35.0188			
3	159.8347	18.1252			
4	166.0661	52.1515			
5	161.9373	28.1168			

Figure 4.9 Optimised face, SURF pattern and SURF descriptors

4.5.4.4 Overall Analysis

The HOG features show a range of values, indicating more variability in the features extracted. The HOG features are relatively consistent across rows, which suggests stability in the extracted features. The values in the SIFT key points figure seem to be primarily constant (-153.8688) for all the rows. There is some variation in a few columns, but the values are relatively small compared to the other methods. The SURF figure also displays some variation in the feature values, but not as much as the HOG features. The last two columns have a moderate range of values but are not as diverse as the HOG descriptors.

Based on the analysis of the results presented, the reason behind the selection of HOG features over SIFT and SURF is the higher variability and diversity in the extracted features. The HOG features show a more comprehensive range of values and more consistent descriptors than SIFT and SURF. This represents that HOG may capture more distinctive and discriminative information from the images, making it potentially more suitable for tasks that require robust feature representation, such as object recognition or image classification.

4.6 Machine Learning Algorithms

Machine Learning techniques are divided mainly into the following 4 categories:

- Supervised Learning
- Semi-supervised Learning

- Unsupervised Learning
- Reinforcement Learning

This chapter presents two learning methods supervised and semi-supervised learning.

4.6.1 Supervised Learning

Learning any function that transforms an input into such an output using instances of input-output pairs is referred to as reinforcement methods in supervised learning. It builds a functional from a labelled collection of training samples. There are two parts to the examples in supervised learning: the intended output value and the input object (also called the supervisory signal). It is possible to apply a supervised learning algorithm to new scenarios by analysing the training data [115]. In the best-case scenario, the algorithm would be able to identify the classifiers for unseen cases correctly. A "reasonably broad" generalization to unknown situations from the training data is thus required. Algorithms are judged on their statistical quality using the so-called generalization error.

Supervised Learning approaches are divided into two categories.

- Regression
- Classification

Regression: Regression aims to improve the co-relation among the training class objects by putting the most relevant identities altogether. Hypothetically, the value of regression lines between {-1,1}, but on the practical side, it lies between {0,1}, 1 stand for the highest correlation, and 0 stands for no correlation. Both the scenarios are completely hypothetical, and any training engine aims to reach a value closer to 1. In the case of the proposed work, three regression values have been evaluated, namely regression for training, regression for validation, and regression for testing. The overall regression has been calculated by taking the absolute mean of all the evaluated regressions [116].

Classification: The classification refers to identifying the class object against its supplied ground truth value. The classification engine must contain a trained repository that has to be mapped against the test data to perform the classification. For example, in

the case of the proposed work scenario, the heart rate is the classified object and is mapped against magnified face features.

Both the training and classification architecture are required in both kinds of learning, whether supervised or semi-supervised. To use supervised learning to solve a given issue, the stages listed below must be completed:

- Determine the training examples to be utilised. Before moving on to the next phase, a user must first decide what kind of training data will utilize. There are many ways to express yourself through handwriting, from the simplest of letters to the most complex of phrases and paragraphs.
- Put together a teaching set of tools. If feasible, the training set must mirror how the function will be utilised in the real world. The input and output elements are collected either by human specialists or by using measures to get the data.
- There must be a means to communicate one's thoughts on the learned function's input characteristic. The input item's representation strongly influences the trained function's accuracy. It is feature vectors created by transforming an object's features, which would be made up of several qualities that describe the item. To avoid the curse of dimensionality, the selection of attributes should not be excessive; yet there should be enough information in each attribute to forecast the outcome accurately.
- Learner function design and a proper learning strategy must be defined before the training can begin. A support vector machine or a decision tree, for example, might be used by the engineer.
- Finally, put the final touches on the design. Make use of the gathered data to train the algorithm. Specific control parameters need to be set for several supervised learning techniques. An element of the practice set (known as the validation set) may improve performance on these parameters, or cross-validation can be used.
- Observe how well the newly acquired function performs in practice. Finally, the final product must be tested on a set of data that is distinct from the training set[117].

4.6.2 Semi-Supervised Learning

Unlabelled and labelled data are used in conjunction to train an algorithm to learn about a certain subject. Two forms of learning are supervised and unsupervised (utilizing just

labelled datasets for training). Unlabelled data may be greatly enhanced when coupled with a small amount of tagged data. Obtain labelled data for a learning assignment often requires a human agent (for example, an audio clip could need to be transcribed manual), or physical experiments may be required (for example, to determine the 3D structure of a protein or to determine if there is oil in a particular location). As a result, labelling a huge training time set may be prohibitively costly, but gathering unsupervised learning is fairly affordable. It is possible that semi-supervised learning might be highly advantageous in some instances. Concerning both human and machine learning, semi-supervised learning is a popular approach [118].

A set of l Independently identically dispersed samples $x_1,...,x_l \in X$ with matching labels $y_1,...,y_l \in Y$ are processed and u unlabelled examples $x_{l+1},...,x_{l+u} \in X$. It is possible to improve classification performance by combining semi-supervised learning with unsupervised learning by eliminating the unlabelled data and executing supervised learning. Transductive or inductive learning are both examples of semi-supervised learning. The goal of transductive learning seems to be to predict the labelling of data that has not been labelled $x_{l+1},...,x_{l+u}$ alone. The goal of inductive learning is to arrive at the correct mappings X to Y.

Labelled data serves as multiple-choice questions in a learning problem that the instructor answers again for the class as a preparation for addressing another set of questions later on in the semester. Those issues that have not been answered serve as test questions in the transductive environment. As practice problems for the kind of questions on the test, they are used in an inductive environment. It is not necessary to classify a whole data set based on inferences drawn from the data to achieve transductive learning; yet, it is not uncommon in practical to utilize methods designed either transducer or duction in the same situation.

4.7 Ablation study

In the context of machine learning “ablation study” has been adopted to describe a procedure where certain parts of the network are removed, in order to gain a better understanding of the network’s behaviour [119]

In order to validate the proposed optimization using Artificial Bee Colony, two additional Swarm based algorithms have been also applied. The first algorithm is firefly algorithm, and second algorithm is Particle Swarm Optimization (PSO) algorithm. Both algorithms have significant capabilities to select and reject the pixel value or replace the pixel value based on the lower and upper bound threshold that are decided by the algorithmic architecture.

The collective behaviour of decentralised, self-organized systems, whether natural or artificial, is called Swarm Intelligence (SI). The notion is used in artificial intelligence research. SI systems are often made up of a population of simple agents or boids that interact with one another and their surroundings on a local level. Nature, particularly biological systems, is a frequent source of inspiration. Although there is no centralised control structure prescribing how individual agents should behave, local and, to some extent, random interactions between such agents create "intelligent" global behaviour, which is unknown to the individual agents [120]. Ant colonies, bee colonies, bird flocking, hawk hunting, animal herding, bacterial growth, fish schooling, and microbial intelligence are examples of swarm intelligence in natural systems. Swarm robotics is the application of swarm concepts to robots, whereas swarm intelligence refers to a broader collection of algorithms. In the area of forecasting difficulties, swarm prediction has been applied. In synthetic collective intelligence, genetically engineered organisms are being studied for techniques similar to those suggested for swarm robots.

Techniques based on swarm intelligence may be applied in a variety of situations. The US military is looking at swarm management systems for unmanned vehicles [121]. An orbital swarm for self-assembly and interferometry is being considered by the European Space Agency [122]. NASA is looking at using swarm technology to map all the planets [123]. Al-Rifaie and Aber, on the other hand, employed a stochastic diffusion search to find tumours [124, 125]. Data mining and cluster analysis have also used swarm intelligence. Modern management theory also includes ant-based models.

This research considered Artificial Bee Colony (ABC) due to its ability to adjust the judgment value based on the onlooker bee selection threshold and architecture. The ordinal measures of ABC are provided in the next section.

4.7.1 Artificial Bee Colony

The ABC algorithm is a meta-heuristic approach introduced for solving real-valued complex optimization problems [126]. The ABC algorithm consists of three phases: employee bees, onlooker bees and scout bees phase. The task of employee bees is to find a better food source by sharing the information with the neighbours of the present food source [127].

In this research, ABC algorithm uses the pixel distribution as the primary factor of the employed bee, and hence the proposed ABC algorithm is named Pixel Distributed-ABC (PD-ABC). It utilizes the multi-hive concept [128] and keeps two bees on the list, namely the Employed bee and the Onlooker Bee. The Scout Bee is not considered in the proposed algorithm as each pixel is evaluated only once in the distribution. It sounds unfair to provide each pixel with just one chance to prove itself worthy of being a part of the selected pixel [116]. Hence the PD-ABC uses a multi-hive concept in which each employed bee will be placed with multiple hive-group bees so that for each flight taken by one bee, it gets a reward score of .10 points. A total of 10 Levy flights have been applied to process one employed bee. Maximum 1 point and minimum 0 points can be awarded to one employed bee as a sum of all flight rewards.

To implement ABC algorithm the following algorithm steps have been implemented using MATLAB.

1. Algorithm Apply ABC
2. Inputs:PD Output:SPD, where PD is the pixel distribution, and SPD is the selected pixel distribution.
3. for i in PD
4. Eb=i;
5. Sp=R.sample(PD) // Select a random sample space by 70% distribution law
6. hv=kmean(Sp,3) // create hv as three hives from the sample space
7. Ei=Find(hv.Eb.index) // Employed bee index, extract from hive indexes
8. Initiate Reward=0 // Initiate reward for Levy flights
9. for j=1:10 // create 10 Levy flights
10. Create Ob_j // Create Onlooker bee for jth flight

11. Create Ei_j // Place employed bee into random groups and create employed bee group for jth flight
12. fx=bee_fitness (Obj,Eij) // Pass to bee fitness
13. If fx==1
14. Reward=Reward+.10
15. Else
16. Reward=Reward+0
17. End If
18. End for_j
19. If reward≥ .50
20. Accept bit value as forehead
21. Append to SPD
22. Else
23. Do nothing
24. End If
25. End for_i

4.7.2 Firefly Algorithm

The firefly algorithm is a metaheuristic optimization algorithm based on the flashing behaviour of fireflies. This algorithm is suitable for solving complex optimization problems that require a large search space. The algorithm works by simulating the flashing behaviour of fireflies to find the optimal solution. Below is the detailed description of the firefly algorithm and Figure 4.10 shows application of firefly algorithm in MATLAB.

```
[FFAThresh,~]=fireflyalgo(Data,nOfSelection,nOfFireFlies,itrationMax,fitfun);
[rcs,ccs]=size(data_to_optimize);
for kks=1:rcs
    for jks=1:ccs
        df=abs(data_to_optimize(kks,jks)-max(FFAThresh));
        if df<70
            data_to_optimize(kks,jks)=round(max(FFAThresh)+mean(FFAThresh)*ra
        end
    end
end
[validation_index,validation_centroid]=kmeans(data_to_optimize,2);

nbtrain=fitcsvm(data_to_optimize,validation_index');
```

Application of firefly

Figure 4.10 Application of firefly algorithm in MATLAB

Initialization: In this step, a swarm of fireflies is randomly initialized in the search space. Each firefly is represented by a position vector in the search space.

Objective Function: An objective function is defined to evaluate the fitness of each firefly based on its position in the search space.

Attraction and Repulsion: Fireflies are attracted to each other based on their brightness. The brightness of a firefly is determined by its fitness. Fireflies that are brighter (i.e., have a higher fitness) attract other fireflies towards them. Fireflies that are less bright (i.e., have a lower fitness) are repelled by other fireflies.

Movement: In this step, each firefly moves towards other fireflies that are brighter than itself. The movement of each firefly is determined by the following equation:

$$X_{i(t+1)} = X_{i(t)} + beta * \left(X_{j(t)} - X_{i(t)}\right) + alpha * e_i \quad (4.3)$$

where $X_{i(t)}$ is the position vector of the ith firefly at time t, $X_{j(t)}$ is the position vector of the jth firefly at time t, beta is the attraction coefficient, alpha is the randomness coefficient, and e_i is a random vector.

Brightness Update: After each firefly moves, its brightness (i.e., fitness) is re-evaluated based on its new position in the search space.

Termination: The algorithm terminates when a predefined stopping criterion is met. This criterion can be a maximum number of iterations, a minimum fitness value, or a maximum computation time.

4.7.3 Particle Swarm Optimization Algorithm

The PSO algorithm is a population-based optimization algorithm inspired by the social behaviour of birds flocking or fish schooling. In this algorithm, a swarm of particles moves around the search space to find the optimal solution. Below is the detailed description of the PSO Algorithm. Figure 4.11shows the application of PSO algorithm in MATLAB.

```
[PSOThresh,~]=ParticleSwarm(data_to_optimize,nVar,nPop,MaxIt,CostFunction);
[rcs,ccs]=size(data_to_optimize);
for kks=1:rcs
    for jks=1:ccs
        df=abs(data_to_optimize(kks,jks)-max(PSOThresh));
        if df<70
            data_to_optimize(kks,jks)=round(max(PSOThresh)+mean(PSOThresh)*ra
        end
    end
end
[validation_index,validation_centroid]=kmeans(data_to_optimize,2);
```

Application of PSO

Figure 4.11 Application of PSO algorithm in MATLAB

Initialization: In this step, a swarm of particles is randomly initialized in the search space. Each particle is represented by a position vector in the search space and a velocity vector.

Objective Function: An objective function is defined to evaluate the fitness of each particle based on its position in the search space.

Update Velocity: In this step, the velocity of each particle is updated based on its current position and its best position (i.e., the position that has the highest fitness value) and the best position of the entire swarm (i.e., the position with the highest fitness value among all particles in the swarm).

Update Position: In this step, the position of each particle is updated based on its velocity. The update equation for the position of the ith particle.

4.7.4 Implementation Details of PSO and Firefly algorithm

The proposed work has implemented both the PSO algorithm and the firefly algorithm along with proposed ABC algorithm using optimization system toolbox that is supported in MATLAB. Once the mask of the image is updated by all the three algorithms, the proposed work validates them by fitting them into a two-class segment viz. {1, 2}. The target is to divide the optimized data into two segments for higher value and lower value pixel. The segregated data is passed to Support Vector Machine (SVM) for model fitting and a supervised classification is applied. Same SVM is used for all cases with linear kernel.

To implement PSO and Firefly, the following algorithm steps have been implemented using MATLAB.

Particle Swarm Optimization (PSO):

1. Initialize the particle swarm:
    a. Generate a random population of particles.
    b. Each particle represents a point in the image.
    c. Initialize the velocity and position of each particle.
2. Evaluate the fitness:
    a. Define a fitness function for the PSO based on classification accuracy using the labelled training set.
3. Main loop:
    i. Repeat the following steps until reaching the maximum number of iterations:
    ii. Update the velocity of each particle using the PSO update equation.
    iii. Update the position of each particle based on its velocity.
    iv. Evaluate the fitness of each particle after the movement.
    v. Update the best position for each particle.
    vi. Update the global best position based on the best fitness value.
4. Select the valid points:
    a. Extract the points corresponding to the global best position as the selected valid points from the image.
5. Return the selected valid points.

Firefly :

1. Initialize the firefly population:
    a. Generate a random population of fireflies.
    b. Each firefly represents a point in the image.
2. Evaluate the fitness:
    a. Define a fitness function for the Firefly Algorithm based on classification accuracy using the labelled training set.
3. Main loop:
    i. Repeat the following steps until reaching the maximum number of iterations:
    ii. Calculate the attractiveness between fireflies based on their fitness values and distances.
    iii. Move the fireflies towards brighter fireflies with higher attractiveness.
    iv. Evaluate the fitness of each firefly after the movement
    v. Update the best position based on the best fitness value
4. Select the valid points:
    a. Extract the points corresponding to the best position as the selected valid points from the image
5. Return the selected valid points

In both the cases, the initial population is total number of pixel values in one data frame. The fitness function for both the algorithm and total number of independent runs is defined in the Table 4.3.

Table 4.3 Parameters of ABC, PSO and Firefly

Parameter	ABC	PSO	Firefly
Number of Independent runs	1000	1000	1000
Objective Function	Argmax(Classification Accuracy)	Argmax(Classification Accuracy)	Argmax(Classification Accuracy)
Fitness Function	Fitness *ABC* $f_x = 1$ $if\ Eb.cr.cg > Eb.cr.mg$ 0 $otherwise$	If $I_i > MI$ 1 $Otherwise$ 0 where, I_i is the current pixel and MI is the mean pixel value of the considered image frame segment	$If\ AI_i > \frac{\sum_{j=1}^{m} AI_j}{m}$ 1 $Otherwise$ 0 where, *AI* represents the attraction index, *i* be the current pixel of the considered segment of frame and *m* be the total number of pixels in the frame

Based on the classified samples, the fitting accuracy or prediction accuracy is calculated for a maximum of 1000 independent runs for all the samples that has been utilized in the results shown in Table 4.4 for each algorithm is based on average performance of the population.

Table 4.4 Fitting accuracy of ABC, firefly and PSO

Number of Simulations	**Fitting accuracy ABC**	**Fitting accuracy Firefly**	**Fitting accuracy PSO**
100	92.11813	91.79980	89.34890
200	90.13851	89.16719	81.90393
300	92.08448	88.91349	82.58226
400	90.10333	85.71589	86.28775
500	92.29655	84.34455	90.42782
600	91.46929	87.01343	85.00616
700	92.12809	84.58122	89.36784
800	92.03910	85.48812	90.41299
900	90.35699	85.37335	80.75955
1000	91.02115	85.16847	88.78303

Based on the given results, the proposed ABC algorithm outperformed the Firefly and PSO algorithms in terms of fitting accuracy. The proposed ABC algorithm achieved a fitting accuracy of 91.37556, while the Firefly algorithm achieved a fitting accuracy of 86.75655 and the PSO algorithm achieved a fitting accuracy of 86.48802.

The improvement of the proposed algorithm over other algorithm is calculated using the following formula:

$$\text{Improvement} = ((\text{ABC} - \text{Other}) / \text{Other}) \times 100\% \quad (4.4)$$

where ABC is the fitting accuracy of the proposed ABC algorithm and other is the fitting accuracy of two algorithms (Firefly or PSO).

For the improvement over Firefly algorithm:

Improvement = ((91.37556 - 86.75655) / 86.75655) × 100% Improvement = 5.32%

Therefore, the proposed ABC algorithm showed a 5.32% improvement over the Firefly algorithm in terms of fitting accuracy.

For the improvement over PSO algorithm:

Improvement = ((91.37556 - 86.48802) / 86.48802) × 100% Improvement = 5.62%

Therefore, the proposed ABC algorithm showed a 5.62% improvement over the PSO algorithm in terms of fitting accuracy.

4.8 Summary and Contribution of the Chapter

This chapter presents the data collection process for pilot and main study for the research presented in this thesis, also gives the detailed information of equipment and set up used for data collection. Chapter shows the study of different algorithms and methods for ground truth generation, feature extraction and ablation study shows that why hold out methods was suitable for this research.

The contribution of this chapter was to select the equipment and setting of camera which helps for decision of the procedure and methodology for video/image processing.

Chapter 5
Selection of Region of Interest and Optimization

This chapter covers the face detection, selection of a region of interest and video magnification techniques applied to the obtained forehead area acquired from the recorded video for this research. First, face detection methods are defined. The characteristics described before the other significant aspects involved in selecting the Region of Interest (ROI) – the video magnification and optimization using Artificial Bee Colony (ABC) – are outlined.

The theoretical principles underlying the optimization used in the proposed algorithm for this research are then described, including the morphological operation on images, polygon to region mask conversion, swarm intelligence and pixel distribution ABC.

5.1 Introduction

The proposed architecture explanation is divided into two parts, first part explained in this chapter, from upload the input video till the optimization of selected area. Chapter 6 will cover the training and classification using neural network. The Figure 5.1 shows the complete structure of proposed work.

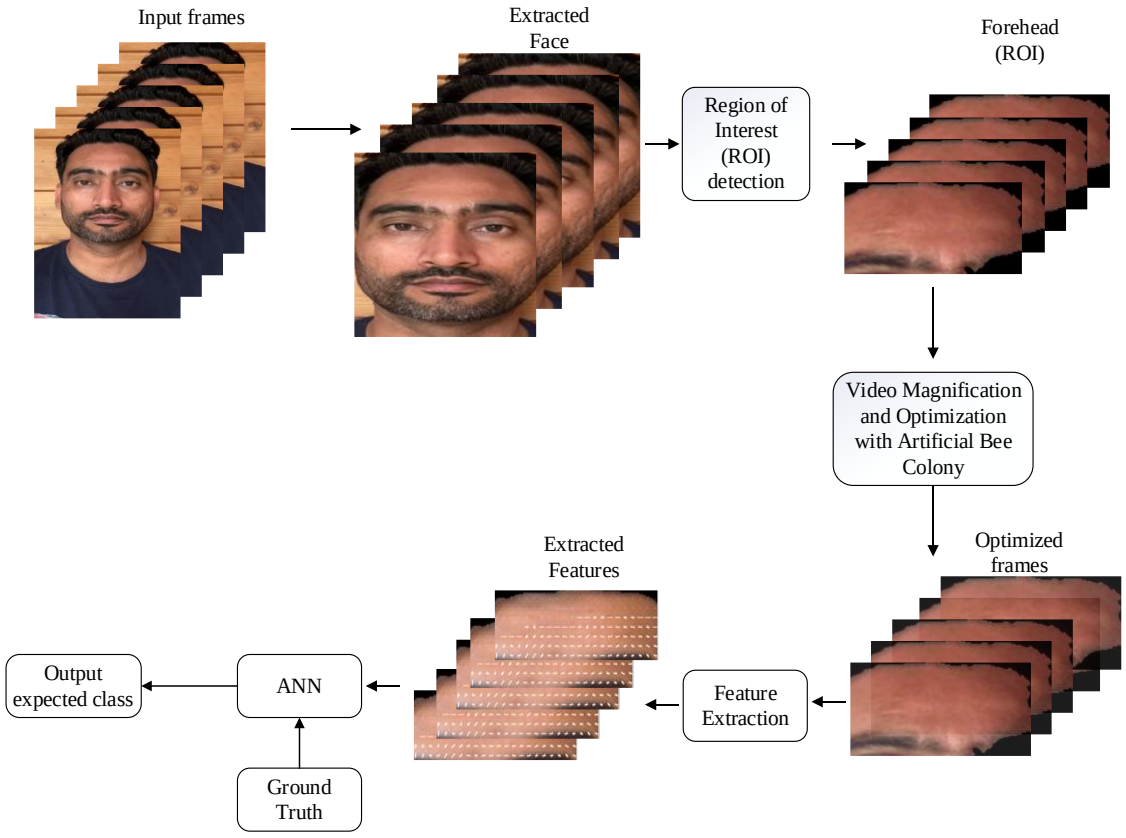


Figure 5.1 Block diagram of proposed work

MATLAB 2020 has been utilised to process the generated video, using the video reader libraries. Below are the steps used to select ROI and optimization.

Step 1: The video was uploaded.

Step 2: Frames are extracted from the video.

Step 3: Face detection, where cascaded object detection is used for selection of face from the frames. From the selected face frame next step is to select the forehead which is done based on pixel distribution of each frame and then magnify using Eular magnification [129].

Step 4: As the video frames are more centric on the red pixel distribution, the histogram equalization provides the pixel distribution of the frame.

Step 5: Optimization of selection of forehead using Artificial Bee Colony.

The development flowchart of this process is shown in Figure 5.2.

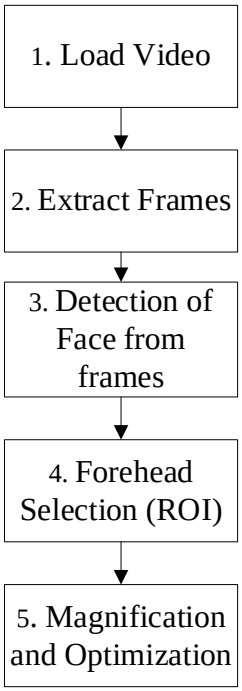


Figure 5.2 Development flowchart of the Forehead selection process and optimization

The video magnification architecture presented by Eulerian Magnification [76], as discussed in chapter 3, is applied to zoom in to the forehead section of the object. To apply Eulerian Magnification, 60 videos were taken using a developed protocol standard setup of cameras, and videos were taken to determine the heart rate. Due to the pandemic globally, it was not possible for any candidate to visit the university and collect the data; hence, the presented architecture is based on a specific collected and license-free dataset available on the internet.

The purpose is to enhance the forehead segment to train the system for its most efficient possible classification accuracy. To identify the forehead, the face has to be detected from each frame that will be processed. The face detection process is not complicated in the modern world of technology. There are several ways through which a face detection framework can be designed and defined [130]. Several types of detection are popular in algorithmic implementation designs and are generally termed the Upper Body Classification Model (UBCM).

5.2 Face Detection

In step 3, shown in Figure 5.2, what needs to be performed here is detecting the face of the subject from the video frames under study. Face detection is the process of recognizing the face, so it is essential to detect and extract the face region from the original image before attempting to recognise a face from it. Other features in the image that aren't part of a face can make the recognition process more difficult. Figure 5.3 shows the steps of face recognition.

Figure 5.3 Steps of Face recognition [131]

Some of the face detection methods popularly that have been used in the existing time are as follows [132]:

5.2.1 Appearance-Based Face Detection

The appearance-based technique uses a collection of delegate training face photos to find face models. Other methods of performance are inferior to the appearance-based approach. To uncover the significant aspects of face photos, appearance-based methods depend on statistical analysis and machine learning approaches [133]. This approach is also utilised in facial recognition feature extraction.

For face detection, the appearance-based model is further separated into sub-methods, which are as follows:

- Neural Network-Based Detection
- Distribution Based Detection
- Eigenface Based Detection
- Information Theoretical Approach
- Naïve Bayes

5.2.1.1 Neural Network-Based Detection

Neural Networks have effectively solved various detection challenges, including object detection, face detection, emotion detection, and face identification. The neural network architecture uses a training and classification mechanism to detect any object, whether a face or any other thing in the list. It is three-layer backpropagation architecture that includes an input layer, a hidden layer and the output layer.

5.2.1.2 Distribution Based Detection

The subspace representing face patterns may be defined using Principal Component Analysis (PCA) and Fisher's Discriminant. A trained classifier successfully distinguishes the target pattern class instances from the background picture patterns.

5.2.1.3 Eigenface Based Detection

Eigenface is a Face Recognition technique based on PCA, and it is a way of effectively modelling faces [134]. PCA is applied to a set of images to lower the dimension of the dataset, best describing the variance of data. In this method, a face can be modelled as a linear combination of eigenfaces (eigenvectors).

5.2.1.4 Information Theoretical Approach

Markov Random Fields (MRF) may be used for facial patterns and linked characteristics. Using Kullback-Leibler divergence, the Markov process maximizes class discrimination. As a result, this approach may be used to detect faces.

5.2.1.5 Naïve Bayes

Naïve Bayes is a classification algorithm that dignifies the separation between the class objects. Naïve Bayes uses a probabilistic model modelled by the Bayesian theorem of randomization. The simple classification architecture of Naïve Bayes can produce significant results in less computation time for a small set of data [135].

5.2.2 Template Matching Based Face Detection

Template matching refers to matching the detected regions based on the similarities between the templates of the face or if any other object is getting detected. Even in the case of template matching, ROI selection is performed using bounding boxes or rectangular boxes. The matching architecture can be applied over a specific colour pattern of the skin of the detected region. The matching algorithm must contain a mapper that can map one template to another. The template matching will involve patterns extracted from one template and patterns from another template that will be passed to the mapper. The mapper will use a rule set to match the templates [136]

5.2.3 Knowledge-Based Face Detection

The knowledge-based technique relies on a set of rules and is based on human understanding to recognise faces. A face, for example, must have a nose, eyes, and mouth that are all within particular distances and positions from one another. The difficulty in constructing an acceptable set of rules is a significant drawback of these systems. If the guidelines were too broad or too specific, there might be a lot of false positives. This method is inadequate and unable to locate numerous faces in a large number of photos [137].

5.2.4 Feature-Based Face Detection

It is possible to find faces using the feature-based technique by extracting structural information from the face. It is trained as a classifier first, and then it is used to distinguish between facial and non-facial parts of the body [138]. Ultimately, the goal is to surpass the limitations of our instinctual awareness of faces. This strategy is broken into numerous parts and includes photographs with several faces.

5.2.4.1 Viola-jones Algorithm

The Viola-Jones algorithm, created by Paul Viola and Michael Jones in 2001, is an object-recognition framework that enables the real-time identification of visual

properties [139]. Despite being an obsolete framework, Viola-Jones is highly effective, and its use in real-time face identification has proved to be particularly noteworthy.
The Viola-Jones Algorithm used two stages:

- Training
- Detection

Because Viola-Jones was created for frontal faces, it is better at detecting them than faces that are gazing sideways, upwards, or downwards. The picture is transformed to grayscale before identifying a face since it is simpler to work with and there is less data to analyse. The Viola-Jones method discovers the position of the coloured picture after detecting the face on the grayscale image [140].

The Cascade Object Detector (COD) uses the Viola-Jones algorithm to detect people's faces, noses, eyes, mouths, or upper body. COD is a bounding box algorithm that helps in drawing ROI based on the thresholds that incorporate the region of interest to be worked on. The region selection mechanism of COD is evident in the upper and lower sections of an image. COD is adaptive to select the threshold. COD sets the threshold based on the correlation between the pixels getting evaluated. The proposed work supports a two-way classification scenario, namely with an Artificial Neural Network (ANN) and a Convolutional Neural Network (CNN). In the case of CNN, the bounding boxes should be precise, and COD provides significant regions for CNN. If no optimization is performed, even in that case, COD can bring some results to the table. Still, the margin of accuracy can increase with improvement and hence the proposed algorithm incorporates optimization. The rest of the section discusses other algorithms that could have been a part of the proposed architecture. The proposed architecture has been used due to more preciseness in ROI selection.
The following Table 5.1 represents the selection comparison of detection methods to provide significance of selection.

Table 5.1 Comparison of face detection methods

Algorithm Name/Type	Basic Function	Selection Method	Number of Stages	Advantages	Disadvantages
Appearance Based Face Detection	Statistical analysis and machine learning are combined to find a face image's characteristics	Statistical analysis of regions	Count: 1 Iterative correlation analysis	The computational time for learning is less	Illumination condition
Template Matching	Images are compared to standard face patterns that have been previously stored.	Extracts patterns from divided regions termed templates	Count: 2 Pattern Extraction and Mapping	Easy to implement, Demonstrate good performance in tracking non-rigid features	These methods do not address variations in pose, scale and shape.
Knowledge-Based	A face is determined based on whether it meets a set of rules made by a human.	Works on rule mining architecture. It is complex for significant patterns.	Count: 2 Rule generation and rule mining	Reduce computational complexity	The problem with this method is to build an appropriate set of rules.
Feature-based Viola-jones	Detection done using features such as a person's nose or eyes are used to detect a face	Works on prior training of data types are supplied to identify objects/faces	Count: 2 Training and Detection	Detect fast; less data needed for training	Training time is slow, effective for a frontal view

5.3 Implementation of Cascaded Object Detection

The work has been implemented on the MATLAB simulation tool, which contains the image processing toolbox that provides the accessibility to alter and modify videos and images. For example, five different binarization methods, including Canny detectors, are integrated into the image processing toolbox. Specifically for visualising videos and frames, the vision system toolbox is integrated into MATLAB. As discussed earlier,

ROI has to be detected to extract the exact face to detect a face. To be precise, bounding boxes have always been the first choice for developers on ROI. COD provides relevance among the bounding boxes based on the extracted contrast of the image. Being recursive, it performs the bounding box creation and adaptation multiple times to draw the exact ROI that provides a significant improvement compared to other types of face detection architectures [141].

The implementation of the face detection architecture has been oriented in MATLAB using the vision system toolbox and the image processing toolbox. The vision system toolbox contains the object architecture “vision”, which contains the information on the pixel distribution of the image.

The proposed algorithm uses COD due to its accurate positioning of the face pixels. As shown in Figure 5.4, the COD method crops the outer segments and selects the inner segments for the processing. To implement it on code, a vision system toolbox incorporates the COD implementation module. The object detector is initially supplied with null minimum and maximum null values using the Frontal-Face CART algorithm. The reason behind the usage of frontal face cart is due to the type of video that has been supplied as an input, which has upper body architecture rather than having full body architecture. The scaling factor of COD is supplied to be 1.1000, which is also the by default value vision system toolbox. As the COD applied here uses the Viola-Jones algorithm, the toolbox does not require the enabling of ROI selection. The video frame is passed after resizing the frame that fits COD. The maximum and minimum bit size for all frames is 256-bit units.

Figure 5.4 Face detection using COD

The face detection process implements frontal cart processing. The ordinal measure is listed in Table 5.2.

Table 5.2 Detection measures

Classification Model	Frontal Face CART
Min Size	[0]
Max Size	[255]
Scale Factor	1.1000
Merge Threshold	4
Use ROI	false

To improve the contrast of the detected face image, Histogram Equalization (HE) has been used. Red pixel components have been extracted to perform HE.

5.3.1 Histogram Equalization

Histogram Equalization (HE) is a standard method for improving picture contrast. It is the most often used approach because of its simplicity and comparably greater performance on photos. HE operates by remapping the image's grey levels according to the probability distribution of the input grey levels. Many studies have been conducted on histogram equalisation, and several approaches have been presented [106].

Contrast enhancement is a critical image processing component for both human and machine vision. It's frequently utilised in medical image processing, voice recognition, texture synthesis, and a variety of other images/video processing applications [142].

Some of these approaches rely on basic linear/non-linear grey level transformation functions. In contrast, others rely on thoroughly examining various picture properties, including edge, related component information, etc.

5.3.1.1 Implementation steps for Histogram Equalization

The following steps and process designs have been used to perform histogram equalization. From the extracted histogram values of the detected face, a face mask is created containing only the red components of the frame, as shown in Figure 5.5 represents the histogram of the entire frame shown in Figure 5.6. The proposed architecture uses bin packing with 256-bit pattern packing that traces the locations of the packing. The packing is created using a distribution model either using Gaussian distribution or Lagrange distribution. Bin oriented histogram creation represents several bins through bar representation. The height of the bar represents the total number of detected or capitalized bins in the distribution.

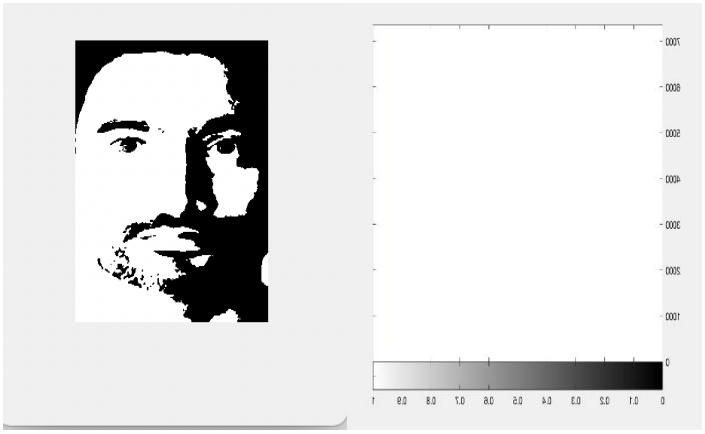


Figure 5.5 Face mask using red components and histogram of face mask

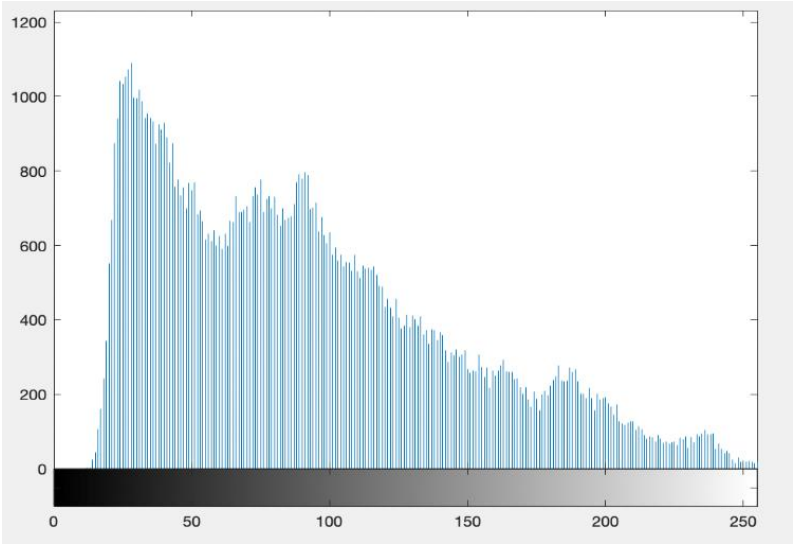


Figure 5.6 Histogram for the entire frame

The following steps have been implemented to process the mask image for the forehead selection. As proved in previous research, the selection of the forehead as ROI is performed better compared to the face area, and facial signals are stronger in the forehead area [143, 144]; in this research, the forehead is selected for HR estimation.

5.3.2 Morphological Operations on Images

Morphological image processing is a set of non-linear processes that deal with the form or morphology of picture features. Morphological procedures are well suited to the processing of binary pictures since they depend only on the relative ordering of pixel

values rather than their numerical values. Greyscale images may also be subjected to morphological treatments in which the light transfer functions are unknown, and the absolute pixel values are of little or minor importance [145]. Morphological approaches use a tiny form or pattern called a structuring element to explore a picture. The structuring element is placed in all available positions in the image and compared to the pixels in its immediate vicinity. Some operations determine if the element "fits" into the neighbourhood, while others decide whether or not it "hits" or intersects it.
Only if the test is successful at that position in the input picture does a morphological operation on a binary image result in a new binary image with a non-zero value for the pixel.

A little binary picture, i.e. a small matrix of pixels, each with a value of zero or one, serves as the structural element.

- The matrix dimensions determine the size of the structuring element.
- The arrangement of ones and zeros determines the form of the structuring element.
- The structuring element's origin usually is one of its pixels; however, it may sometimes be outside the structuring element.

The origin of the structuring matrix is usually specified as the centre of the matrix, which is a typical practice. Structuring components, like convolution kernels in linear image filtering, have a role in morphological image processing.

Each pixel of a structuring element is linked with the equivalent pixel of the neighbourhood beneath the structuring element when it is inserted into a binary picture. If each of the structuring element's pixels is set to one, the corresponding picture pixel is likewise one, and the structuring element is said to fit the image. A structuring element is said to touch, or intersect, an image if at least one of its pixels set to 1 corresponds to a matching picture pixel. The structuring element's zero-valued pixels are disregarded, indicating locations where the accompanying picture value is unnecessary.

5.3.2.1 Image Dilation

Dilation is the process of adding pixels to the edges of objects in a picture, while erosion removes pixels from the limits of things. Depending on the size and form of the

structuring element used to process the picture, the number of pixels added or deleted from the objects in an image varies.

In mathematical morphology, dilation is one of two fundamental operators, the other being erosion [146]. It's most often used on binary pictures, although there are grayscale variants. The operator's fundamental impact on a binary image is to progressively extend the borders of foreground pixels areas (i.e. white pixels, typically). As a result, foreground pixel areas expand in size while gaps within those regions shrink.

Concerning the morphological dilation and erosion procedures, the state of each given pixel in the output image is determined by applying a rule to the respective input image pixel and its neighbours to determine the state of any given pixel in the output picture. The procedure is classified as a dilation according to the rule that was employed to process the pixels. The rule mentions are as below:

- The most significant value of all pixels in the vicinity is the value of the output pixel. A pixel in a binary picture is set to 1 if any of its neighbours contain the value 1.
- Morphological dilatation increases the visibility of things and fills in tiny gaps. Fill forms seem more significant, and lines appear thicker.

Two bits of data are sent into the dilation operator. The picture that will be dilated is the first. The second is a structural element, a (typically tiny) group of coordinate points (also known as a kernel). This structural element determines the dilation's exact impact on the input picture.

Assume X is the set of Euclidean coordinates for pixels of input binary image, and K is the set of coordinates for the structuring element.

Let Kx stand for K's translation with its origin at x.

The set of all points x such that the intersection of Kx with X is non-empty is then the dilation of X by K.

Except for how the set of coordinates associated with the input picture is formed, the mathematical description of grayscale dilation is similar. Furthermore, these coordinates are three-dimensional rather than two-dimensional.

Consider the case of binary dilation, where the structural element is a 3×3 square with the origin in the middle, as shown in Figure 5.7. Foreground pixels are represented by 1's in this and the following diagrams. In contrast, background pixels are represented by 0's.

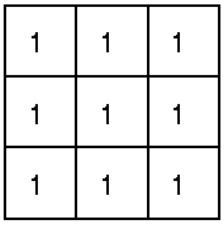


Figure 5.7 3×3 Square structuring element

Set of coordinate points = {(-1, -1), (0, -1), (1, -1), (-1, 0), (0, 0), (1, 0), (-1, 1), (0, 1), (1, 1)}

The research analysed each of the background pixels in the input picture, in turn, to calculate the dilation of a binary input image by this structuring element and superimpose the structuring element on top of the input picture for each background pixel (which we'll refer to as the input pixel) such that the structuring element's origin aligns with the input pixel location. The input pixel is set to the foreground value if at least one pixel in the structuring element corresponds to a foreground pixel in the picture below. The input pixel is left at the background value if all of the associated pixels in the image are background.

This action has the effect of setting the foreground colour of all background pixels with a nearby foreground pixel in the example 3×3 structuring element (assuming 8-connectedness). Because such pixels must be found at the margins of white zones, the result is that foreground regions expand (and holes inside a region shrink).

Dilation is the inverse of erosion, i.e. dilation of foreground pixels equals erosion of background pixels.

Most implementations of this operator assume a binary input picture, with foreground pixels set to pixel value 255 and background pixels set to pixel value 0. Using thresholding, such a picture may frequently be created from a grayscale image. It's critical to double-check that the input image's polarity is adjusted appropriately for the dilation implementation in use.

The structuring element may need to be given as a small binary image or in a particular matrix format. It may be hardwired into the implementation and not need to be specified. In this situation, a 3×3 square structural element is often considered, resulting in the expansion as mentioned earlier effect. Figure 5.8 depicts the impact of a dilation employing this structural element on a binary picture.

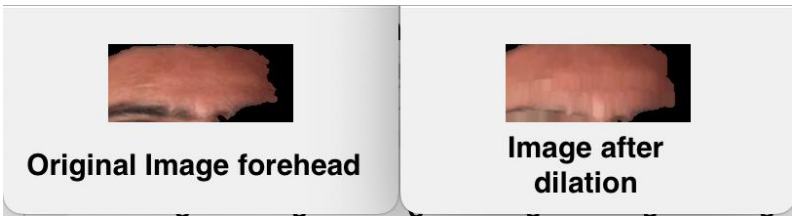


Figure 5.8 Effect of dilation

Although the 3×3 square is perhaps the most popular structural element used in dilation procedures, others may be employed. A more significant structuring element generates a more intense dilation effect, while repeated dilations with smaller but similarly shaped structuring elements may normally yield comparable results. It is typical to utilise a roughly disk-shaped structuring element rather than a square one with more significant structuring parts.

Two dilation passes to create this picture, using a disk-shaped structuring element with an 11-pixel radius, shown in Figure 5.9. It's worth noting that the corners have been rounded. Convex borders will become rounded when a disk-shaped structural element is dilated, but concave boundaries remain unchanged.

Figure 5.9 Binary dilation

Less symmetrical structural components may be used to create directed dilations. A structural element that is 10 pixels wide and 1 pixel high, for example, will only dilate horizontally. A 3×3 square structuring element with the origin in the middle of the top row rather than the centre will dilate the bottom of an area more than the top.

Grayscale dilation using a flat disk-shaped structural element brightens the picture in general. Dark areas surrounded by bright regions expand in size, whereas bright regions bordered by dark ones decrease. Small dark patches in photos will vanish when they are 'filled in' to the intensity value of the surrounding area. Small light spots will enlarge and grow bigger. The effect is most noticeable when the image's intensity fluctuates fast, whereas sections of more uniform intensity will remain primarily constant except at their borders.

5.3.3 Polygon to region mask conversion

Polygon to region masking refers to attaching a polygon template to the image to create a mask that could be supplied for processing. Polygon to region uses binary masking, and the process of binarization is again dependent upon the ROI that is being selected. After implementing the above-defined processes, the provided frame signifies the forehead. Most white pixel selection could lead to forehead selection from the upper side of the image, as shown in Figure 5.10. A similar process could be applied to identify faces from frames.

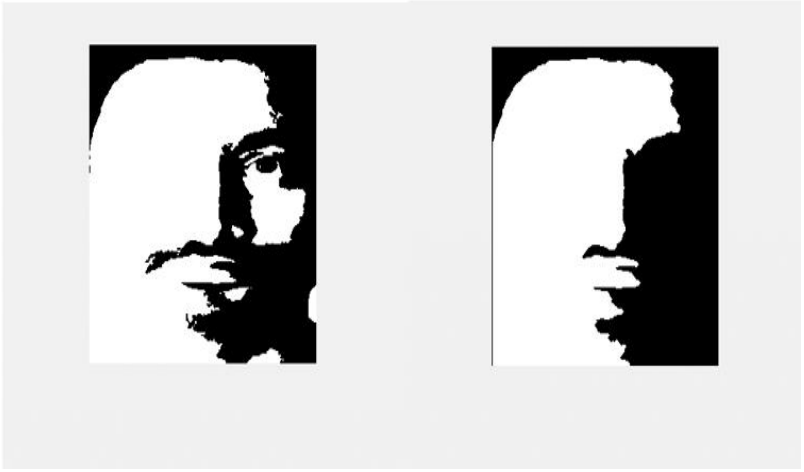


Figure 5.10 Logical Operation and Image dilation (left) Polygon to Region mask (right)

5.3.4 Implementation of Video Magnification

The extracted ROI is further provided for magnification using Euler magnification. Each frame is passed to the magnification function designed in MATLAB to do so. The ordinal measures of Eular video magnification have been provided in chapter 3 already.

To perform the magnification, the extracted face frame has been first converted into RGB contrast, and the green and blue pixels are masked with 0 before any other action step has to be performed. The masked image is passed for logical operations, including area opening and eccentricity calculation. The calculated eccentricity provides significance to boundary boxes used for the Eular magnification concept.

The proposed algorithm uses the developed magnify function that signifies the steps of Eular magnification. To magnify a specific frame, repetitive minimum and maximum outer boundaries are calculated, and the correlation toward red pixels has been evaluated. Figure 5.11 shows the results after the magnification and simulation.

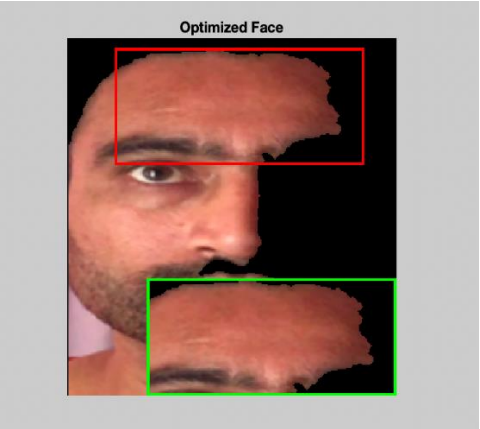


Figure 5.11 Forehead selection and Euler magnification

5.4 Implementation Architecture of Pixel Distribution-Artificial Bee Colony PD-ABC

It is aimed at forehead selection using meta-heuristic and Swarm Intelligence techniques. ABC algorithm uses the pixel distribution as the primary factor of the employed bee, and hence the proposed ABC algorithm is named Pixel Distributed-ABC (PD-ABC). It utilizes the multi-hive concept [128] and keeps two bees on the list, namely the Employed bee and the Onlooker Bee. The Scout Bee is not considered in the proposed algorithm as each pixel is evaluated only once in the distribution. It sounds unfair to provide each pixel with just one chance to prove itself worthy of being a part of the selected pixel [116]. Hence the PD-ABC uses a multi-hive concept in which each employed bee will be placed with multiple hive-group bees so that for each flight taken by one bee, it gets a reward score of .10 points. A total of 10 Levy flights have been applied to process one employed bee. Maximum 1 point and minimum 0 points can be awarded to one employed bee as a sum of all flight rewards. The overall work architecture of the proposed ABC algorithm can be understood well using Figure 5.12.

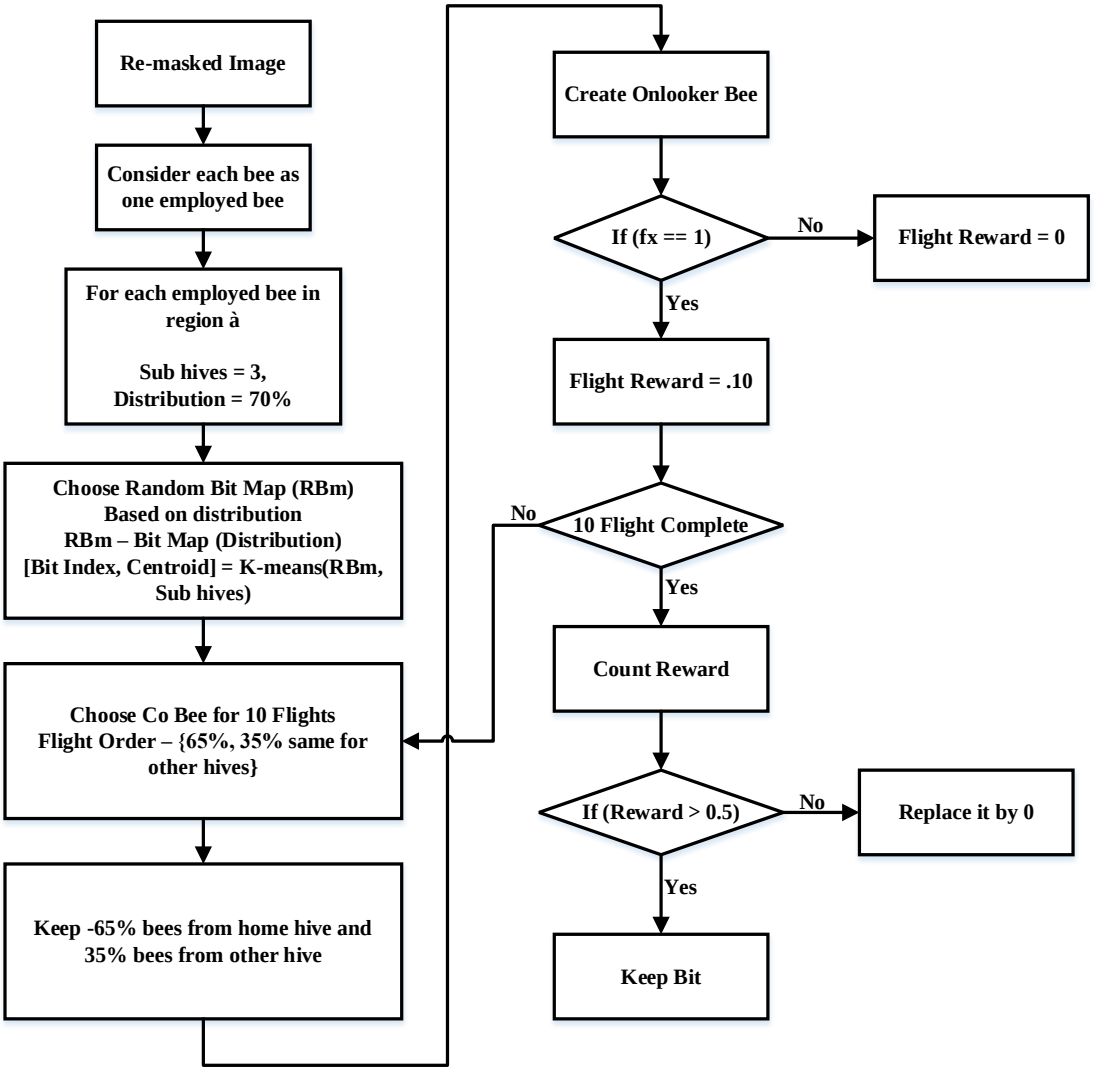


Figure 5.12 Workflow of proposed ABC algorithm

There are three stages of the proposed ABC algorithm. Each stage has its set of significance to provide a reward to the employed bee. The proposed algorithm generates a reward mechanism in the following order.

For each flight, one bee will be awarded a maximum score of .10; the total reward of 1 can be attained by one employed bee. If the employed bee gains more than a .50 value as a reward out of 10 Levy flights, the pixel is considered a suitable pixel for the forehead. To compose the employed bee, the pixel value of each distribution is considered. To create sub hives in the bee list, k-means is applied to a population of 70% of the total distribution. 65% of the population is from the same hive, and 35% distribution is taken from another hive group. The distribution is made so that each

employed bee gets a chance to coordinate with other bees working on different pixel distributions as well. The employed bee and the onlooker bee can be defined using equations (5.1) and (5.2).

$$Employed_{bee} = \sum_{i=1}^{t} P_{value_i}.BL.ohb \tag{5.1}$$

where BL is the Bee List which is ultimately the pixel distribution of the extracted image. P_{value_i} is the pixel value of the image, ohb is the other hive group formed by 70-30 % distribution of the pixels, and t is the total number of pixels in the extracted image. The onlooker bee is the arithmetic mean of the collected employed bees.

$$Onlooker_{bee} = \sum_{j=1}^{10} Ar\ (\ Employed_{beei)} \tag{5.2}$$

where Ar is the arithmetic progression calculated by Euclidian distance[18] and j represents levy flight.

The onlooker bee forms the bee food for 10 levy flights, and hence the iterative value of j starts from 1 and ends at 10. The fitness function of the proposed PD-ABC is expressed by equation (5.3)

$$\begin{gathered} Fitness\ ABC \quad f_x = 1 \\ if\ Eb.cr.cg > Eb.cr.mg \quad 0\ otherwise \end{gathered} \tag{5.3}$$

where Eb is the employed bee, cr is the co-relation, mg is the primary group/homegroup, and cg is the current group. The ABC algorithm decides the keeping threshold based on the evaluated co-relation among the group members and the co-relation of the bees with their centroid of the significant group that belongs to the employed bee.

[18] The Euclidean distance is the straight-line distance between two pixels. The city block distance metric measures the path between the pixels based on a 4-connected neighbourhood.

The ABC algorithm can be expressed as follows:

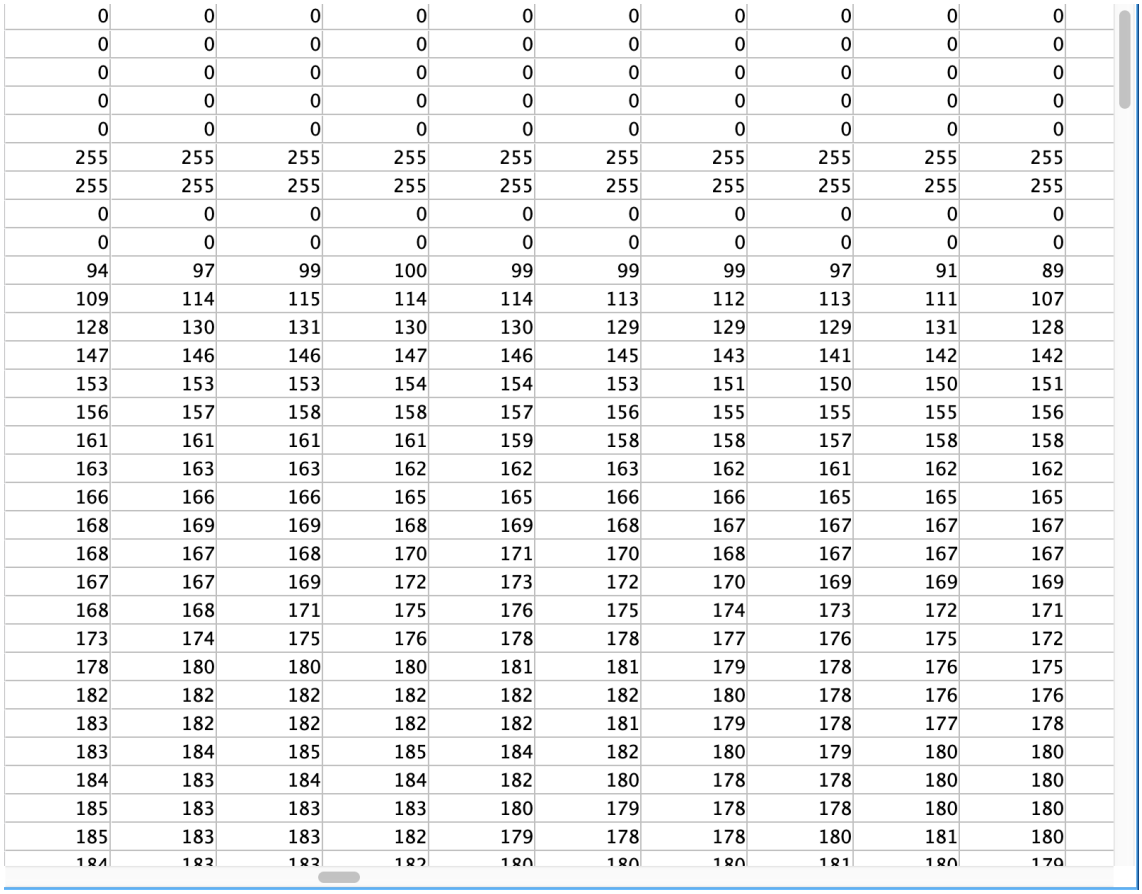
26. Algorithm Apply ABC
27. Inputs:PD Output:SPD, where PD is the pixel distribution, and SPD is the selected pixel distribution.
28. for i in PD
29. Eb=i;
30. Sp=R.sample(PD) // Select a random sample space by 70% distribution law
31. hv=kmean(Sp,3) // create hv as three hives from the sample space
32. Ei=Find(hv.Eb.index) // Employed bee index, extract from hive indexes
33. Initiate Reward=0 // Initiate reward for Levy flights
34. for j=1:10 // create 10 Levy flights
35. Create Ob_j // Create Onlooker bee for jth flight
36. Create Ei_j // Place employed bee into random groups and create employed bee group for jth flight
37. fx=bee_fitness (Obj,Eij) // Pass to bee fitness
38. If fx==1
39. Reward=Reward+.10
40. Else
41. Reward=Reward+0
42. End If
43. End for_j
44. If reward≥ .50
45. Accept bit value as forehead
46. Append to SPD
47. Else
48. Do nothing
49. End If
50. End for_i

Appendix C shows the results frames of input video, selection of face frames, forehead selection and frames after optimization, followed by the simulation panel in MATLAB. Table 5-3 shows the results of ABC, where first row represents the Considered Pixels indexes from forehead and data values against collected pixels are shown in following

columns and at the end it represents the group order. Figure 5.13 shows the normal pattern before ABC and Figure 5.14 shows updated pattern after ABC.

Table 5.3 Results of ABC

Considered Pixels indexes from forehead	70	42	98	42	27	7	10	20	81	34	68	59	28	55
Data Values against collected Pixels.	161	183	180	183	189	127	165	204	181	192	153	141	189	137
	181	187	191	187	179	0	0	104	188	187	187	188	184	192
	40	125	128	125	193	217	219	199	93	164	162	165	190	237
	181	187	191	187	179	0	0	104	188	187	187	188	184	192
	182	159	170	159	0	0	0	0	178	91	182	185	0	187
	172	0	160	0	0	0	0	0	165	0	174	165	0	157
	118	77	99	77	129	194	206	197	82	87	116	132	121	108
	184	181	170	181	124	0	0	0	181	169	183	187	141	188
	173	194	188	194	190	114	147	199	189	196	170	166	189	167
	193	197	202	197	199	0	88	175	206	199	198	208	199	205
Group Order	2	2	2	2	1	3	3	1	2	1	2	2	1	2



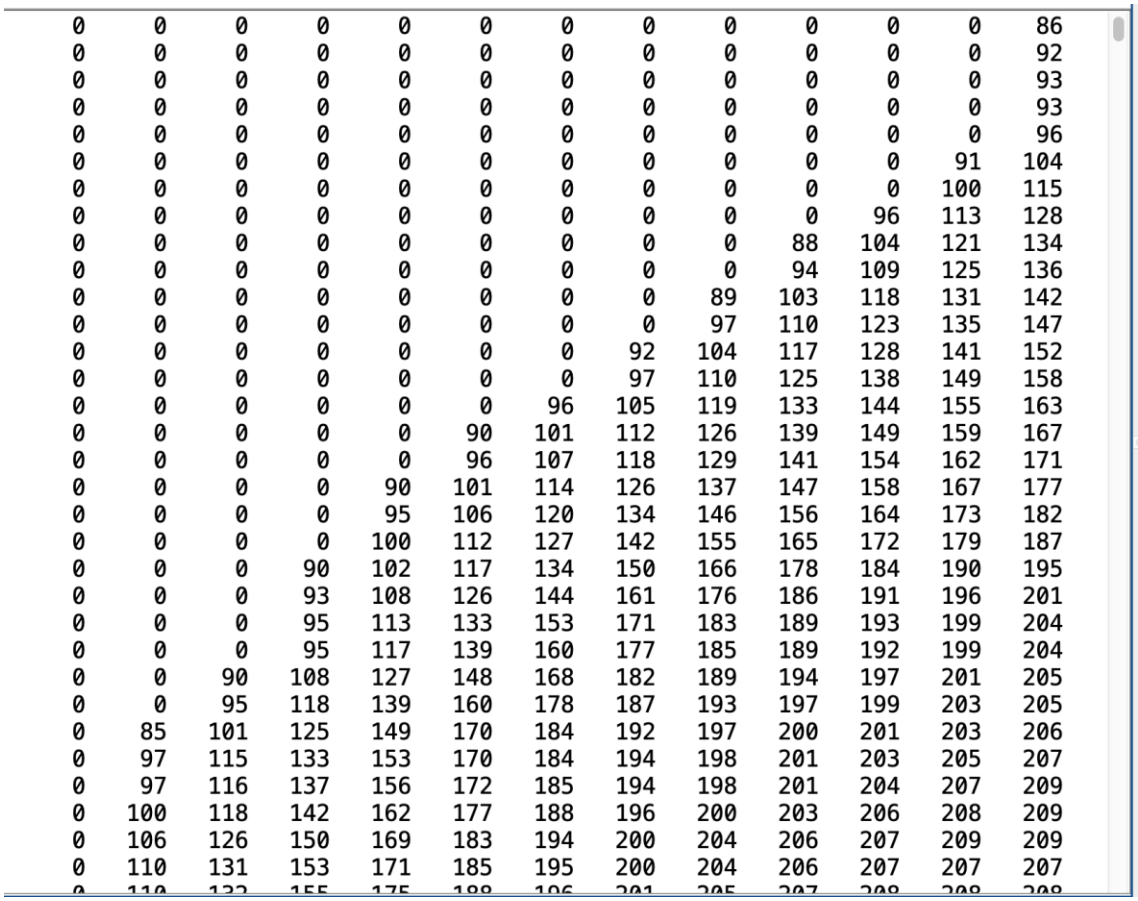
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
255	255	255	255	255	255	255	255	255	255
255	255	255	255	255	255	255	255	255	255
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
94	97	99	100	99	99	99	97	91	89
109	114	115	114	114	113	112	113	111	107
128	130	131	130	130	129	129	129	131	128
147	146	146	147	146	145	143	141	142	142
153	153	153	154	154	153	151	150	150	151
156	157	158	158	157	156	155	155	155	156
161	161	161	161	159	158	158	157	158	158
163	163	163	162	162	163	162	161	162	162
166	166	166	165	165	166	166	165	165	165
168	169	169	168	169	168	167	167	167	167
168	167	168	170	171	170	168	167	167	167
167	167	169	172	173	172	170	169	169	169
168	168	171	175	176	175	174	173	172	171
173	174	175	176	178	178	177	176	175	172
178	180	180	180	181	181	179	178	176	175
182	182	182	182	182	182	180	178	176	176
183	182	182	182	182	181	179	178	177	178
183	184	185	185	184	182	180	179	180	180
184	183	184	184	182	180	178	178	180	180
185	183	183	183	180	179	178	178	180	180
185	183	183	182	179	178	178	180	181	180
184	183	183	182	180	180	180	181	180	179

Figure 5.13 Pattern before ABC from MATLAB

The proposed ABC algorithm enhances the extracted ROI for improved feature extraction. As shown in figures, ABC arranges the ROI in much significant manner as compared to the raw data. It is clear from the figures that a pyramid of vector is created which contains 0 and pixel values with a much clear separation. This is due to the improved ABC algorithm that incorporates objective function to enhance the overall classification accuracy. The proposed ABC algorithm also incorporates a novel fitness function that ensemble co-related pixel values and discards pixel values with less collinearity.

To implement the proposed ABC algorithm, the pixel values are divided into two segments multiple segments. In each Lévy flight, each considered artificial bee is paired with multiple other bees. The paired bees belong to the home hive and away hive in which home hive is the current cluster of the employed bee. As for example, the home hive for the current collected sample is 2 and hence the group order for the paired bee

mostly contains group value 2. To check the rigidness of the proposed algorithm, 30% population is selected from other hives as demonstrated in 3rd row. For instance, pixel 27 belongs to 1st hive whereas 70,42,98 etc belongs to 2nd hive.



0	0	0	0	0	0	0	0	0	0	0	0	86
0	0	0	0	0	0	0	0	0	0	0	0	92
0	0	0	0	0	0	0	0	0	0	0	0	93
0	0	0	0	0	0	0	0	0	0	0	0	93
0	0	0	0	0	0	0	0	0	0	0	0	96
0	0	0	0	0	0	0	0	0	0	0	91	104
0	0	0	0	0	0	0	0	0	0	0	100	115
0	0	0	0	0	0	0	0	0	0	96	113	128
0	0	0	0	0	0	0	0	0	88	104	121	134
0	0	0	0	0	0	0	0	0	94	109	125	136
0	0	0	0	0	0	0	0	89	103	118	131	142
0	0	0	0	0	0	0	0	97	110	123	135	147
0	0	0	0	0	0	0	92	104	117	128	141	152
0	0	0	0	0	0	0	97	110	125	138	149	158
0	0	0	0	0	0	96	105	119	133	144	155	163
0	0	0	0	0	90	101	112	126	139	149	159	167
0	0	0	0	0	96	107	118	129	141	154	162	171
0	0	0	0	90	101	114	126	137	147	158	167	177
0	0	0	0	95	106	120	134	146	156	164	173	182
0	0	0	0	100	112	127	142	155	165	172	179	187
0	0	0	90	102	117	134	150	166	178	184	190	195
0	0	0	93	108	126	144	161	176	186	191	196	201
0	0	0	95	113	133	153	171	183	189	193	199	204
0	0	0	95	117	139	160	177	185	189	192	199	204
0	0	90	108	127	148	168	182	189	194	197	201	205
0	0	95	118	139	160	178	187	193	197	199	203	205
0	85	101	125	149	170	184	192	197	200	201	203	206
0	97	115	133	153	170	184	194	198	201	203	205	207
0	97	116	137	156	172	185	194	198	201	204	207	209
0	100	118	142	162	177	188	196	200	203	206	208	209
0	106	126	150	169	183	194	200	204	206	207	209	209
0	110	131	153	171	185	195	200	204	206	207	207	207

Figure 5.14 Pattern after ABC from MATLAB

The next chapter presents the proceeding steps after the implementation of ABC. To train a specific set of data, an execution pattern requires an inference engine and the data set itself and its related ground truth values. Chapter 6 describes how the optimized data is passed through a multiclass training inference engine and how the classification has been performed.

5.5 Summary and Contribution of the Chapter

The chapter describes the functional architecture of the selection of the forehead and magnifies the ROI. After all the studied literature and processing architectures of the video magnification, the proposed work architecture has identified that Euler video

magnification results in ample magnification that can be used for further processing. In addition to that, the forehead selection is applied to utilize the histogram equalization method. As the histogram equalization works on hard thresholds of selection, there is always a possibility of optimization. To enhance the current state of the art of the selection, the proposed work has utilised Swarm Intelligence. To be precise on the selection, the proposed architecture enhances the existing meta-heuristic oriented Artificial Bee Colony.

A new pixel distribution-based propagation behaviour has been incorporated and explained in different sections of this chapter. In addition, implementation steps and designs have also been presented using different workflow and other forms of pictorial representations.

Chapter 6
Training and Classification

6.1 Introduction

The proposed work algorithm was earlier planned to be executed so that a bulk amount of data was to be processed for the training of Convolutional Neural Networks (CNN). CNN requires bulk amount of data against the ground truth values to extract the most relevant feature. For CNN any input size of image can be used, but a similar type of data has to be provided in bulk so that CNN can establish co-relations between the object frames. Due to the global pandemic, universities were close to the student, and the scholars could not collect the data on their own due to a lack of standard equipment. With all the complications worldwide, this research was done on a limited set of samples. CNN requires a bulk amount of data and hence high accuracy with CNN was not a sure thing to be expected. So, the proposed work used Feed Forward Back Propagation Oriented Levenberg classification mechanism (FFBPNN-L), which requires a standard set of features to be processed against the provided set of classes or the ground truth value. The proposed algorithm has utilised the Histogram of Oriented Gradients (HOG).

The following are the two sections of this chapter:
Section 6.2: Feature extraction using HOG.
Section 6.3: The training and the classification using CNN and FFBPNN-L.

6.2 Feature Extraction Presented by Histogram of Oriented Gradients

HOG is a feature descriptor similar to the Canny Edge Detector and SIFT (Scale Invariant and Feature Transform). Using this approach, gradient orientations are counted in a specific region of the image. SIFT and Edge Orientation Histograms are two equivalent approaches. Structure or shape is the focus of the HOG description. When computing features, it uses both the magnitude and the angle of the gradient.

This makes it superior to other edge descriptors. It creates histograms for the image's areas again depending on the magnitude and directions of a gradient.
HOG has a few key characteristics that set it apart from other feature descriptors:

- HOG may also be used to indicate the direction of the edge. Efficiently doing this requires obtaining the gradient and the inclination of the edges [147].
- The 'localized' parts are also used to identify these orientations. There is an implied division into smaller portions, each with its slopes and direction.

The HOG would generate a separate Histogram for each of these subsections of the image. The fact that histograms are made up of gradients and orientations of pixel values gives rise to the phrase "Histogram of Oriented Gradients."

Pedestrian detection relies heavily on the HOG model's construction of a magnitude summation of gradient directions during an image region's distribution [148]. The HOG feature is used as a second additional feature in this investigation since the fracture region has striped features comparable to pedestrians. As shown in Figure 6.1, each normalizing picture window is divided into four tiny geographical units known as cells (a). Nine bins are used to discretize the gradient direction for each pixel in a CC pixel cell. There are nine potential directions in the [0, 2] range. Hence the gradient does have a real-valued magnitude since each pixel has a two-dimensional vector with discretized direction. Using the nine-bin histogram values for each cell, the cumulative distribution of the gradient’s direction across the pixel of cells is then shown. It is possible to generate 36-dimensional feature vectors for regions by combining the histogram data from different cells in a single image. Another example of the sliding window technique is presented in Figure 6.1, with the sliding step size set to C pixels. There will be four image windows and a 144-dimensional HOG feature vector for the 3C2C-pixel picture region.

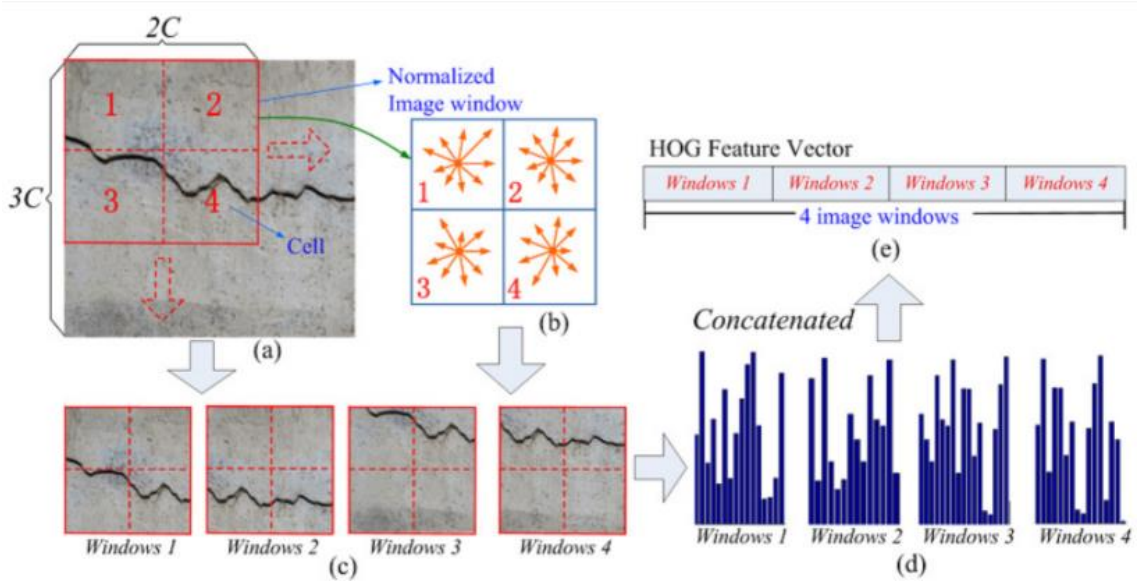


Figure 6.1 HOG feature extraction (a) input image region; (b) four overlapping image windows (c) HOG features of each image window (d) extracted HOG feature [149].

As discussed earlier in the section, the proposed algorithm utilizes HOG for feature extraction. The ordinal measures of utilised HOG architecture are demonstrated in Figure 6.2.

```
featureVisualization: 1x1 vision.internal.hog.Visualization =

Visualization

   Type <a href="matlab:plot(val)">plot(val)</a> to visualize.

   Read-only properties:
                CellSize: [8 8]
               BlockSize: [2 2]
            BlockOverlap: [1 1]
                 NumBins: 9
    UseSignedOrientation: 0
              BinCenters: [18×1 double]
```

Figure 6.2 Utilised ordinal measure of HOG

To implement HOG, the proposed algorithm utilised the MATLAB '*vision system toolbox'* that facilitates the users to pack the data into multiple bins to evaluate the oriented histograms. Figure 6.3 shows the HOG extracted pattern.

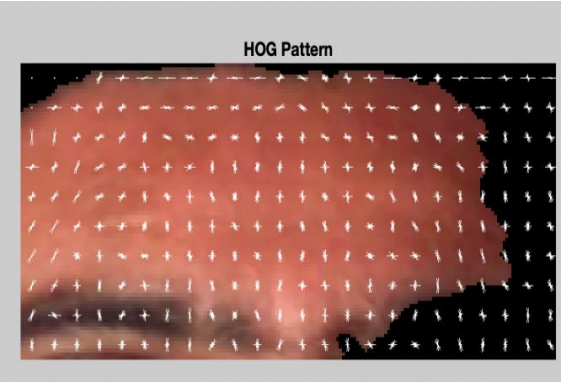


Figure 6.3 HOG extracted pattern

6.3 Training and Classification

Machine learning is a learning method that automates the acquisition of knowledge. It plays an important role in artificial intelligence research. Machine learning is an important way for computers to acquire knowledge and an important indicator of artificial intelligence. It is a discipline that studies how to use computers to simulate or realize human learning activities. It is generally believed that machine learning is a process of acquisition knowledge with a specific purpose. Its internal performance is a process of knowledge growth from unknown to known. Its external performance is the improvement of some performance and adaptability of the system, so that the system can finish the task that could not be completed or better completed [150].

Conventional machine learning is often referred to as an algorithm's ability to improve its ability to anticipate the outcomes of future events. Generally, there are four fundamental strategies to choose from:

- Supervised learning
- Unsupervised learning
- Semi-supervised learning
- Reinforcement learning

To assess the effectiveness of a supervised learning and a quasi-learning algorithm, it was constructed and verified the created and access data, and hence the suggested method architecture.

6.3.1 Convolutional Neural Network

To put it another way, Deep Learning and Convolutional Neural Networks (CNNs or ConvNets) are the most common form of Artificial Neural Networks (ANN). With the use of shared-weight convolution kernels, a filter that may be slid together with the input data, providing translations identical responses depending on feature maps, Shift Invariant/Space Invariant Convolutional Neural Networks (SIANN) were developed. Most convents are only rudimentary in translation, contrary to popular belief. Image and video identification and recommendation systems, as also image classification and segmentation (for medical imaging) and natural language processing, are just some of its numerous applications (for brain-computer interfaces) [151, 152].

Figure 6.4 shows the CNN sequence. CNNs are regularised forms of multilayer perceptron's, which are used in machine learning. Multilayer perceptron's are commonly referred to as fully linked networks, which means that each neuron in one layer is connected to all neurons in the next layer. This is because of their "full connectivity," making them more sensitive to overfitting data. To avoid overtraining, a range of strategies such as penalizing training parameters (such as weight decay) and regularization may be achieved via reducing connectivity (skipped connections, dropout, etc.). Rather than relying on a linear approach to regularization, CNN uses data's hierarchical structure to construct more complicated shapes from smaller ones that are etched into their filters.

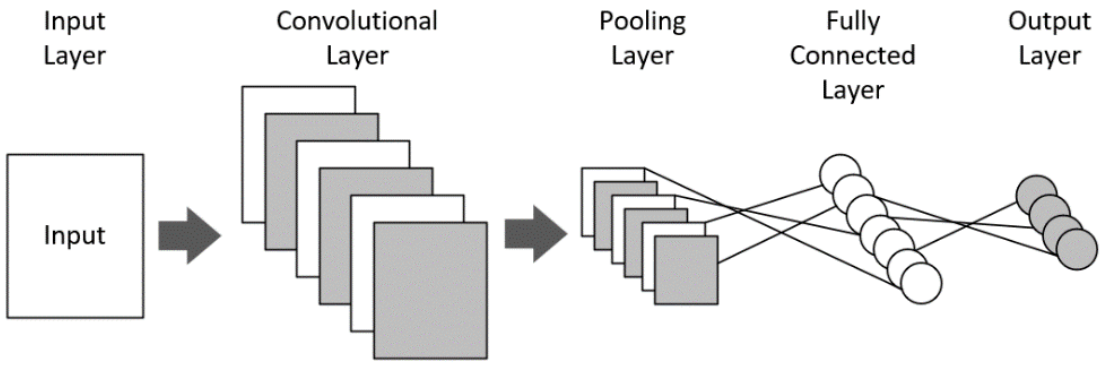


Figure 6.4 Convolutional Neural Network

This means that CNNs are at the lower end of the scale regarding connectivity and complexity. As a result of biological processes, convolutional networks were created

because their structure of connections resembles the animal visual brain. A tiny portion of the visual field, the receptive field is where individual cortical neurons react to visual input. Receptive fields in neurons may cross across, resulting in the neurons spanning the whole vision field [153].

With CNNs, there is a lot less pre-processing required than with conventional image categorization algorithms. Filters (or kernels) are improved automatically by the network rather than manually, as is the case with conventional approaches. This lack of dependency on past information or human interaction is a major plus when it comes to feature extraction.

A Convolutional Neural Network (CNN) is a type of deep neural network specifically designed for processing and analysing grid-like data, such as images or video frames [154-156]. It is widely used in computer vision tasks, such as image classification, object detection, and image segmentation. CNNs are inspired by the human visual system and excel at capturing hierarchical features within images, allowing them to learn intricate patterns and representations from the data.

A typical CNN architecture consists of several layers [157; chapter 4] each serving a specific purpose in the feature extraction and transformation process. The primary layers in a CNN include:

1. **Input Layer:** The input layer represents the raw data that is being fed into the network. For images, this layer would hold the pixel values of the image.
2. **Hidden layer:** The hidden layer is composed of convolutional layer, pooling layer and fully connected layer.
    - **Convolutional Layer:** Convolutional layers are the heart of CNNs. They apply various filters (also known as kernels) to the input data in order to extract different features, such as edges, textures, and more complex patterns. These filters slide over the input, performing element-wise multiplication and summation to create feature maps that highlight specific patterns in the data.
    - **Pooling Layer:** Pooling layers (such as max pooling or average pooling) reduce the spatial dimensions of the feature maps while retaining the most important information. They help in reducing computational

complexity, increasing the network's translation invariance, and promoting feature generalization.

- **Fully Connected Layer:** Fully connected layers, also known as dense layers, are similar to those in traditional neural networks. They take flattened feature maps from the previous layers and process them to make final predictions. These layers consolidate the learned features and combine them to classify or label the input data.

3. **Output Layer:** The output layer produces the final results of the network's computations. The number of neurons in this layer depends on the task, such as the number of classes in a classification problem. Activation functions here may vary based on the nature of the problem (softmax for classification, linear for regression).

For the research presented in this thesis consists of a CNN with seven layers namely; input layer, convolutional layer, batch normalization layer, ReLU activation layer, Fully connected layer, SoftMax layer and classification/output layer.

Each input has (number of inputs) × (height of inputs) × (width of inputs) × (input channels). The picture is transformed into an activation map using a convolution layer, also known as an activation map. An activation map is the sum of inputs multiplied by the feature map height divided by the width (feature map channels).

Furthermore, convolutional layer is created. A cell inside the visual cortex reacts to a particular stimulus in another part of the brain in a comparable way. A convolutional neuron only processes information relevant to its receptive field. The fact that neural networks could detect faces and categorize data is not a deterrent to using this method. The sheer amount of connections makes this architecture impractical for large inputs like high-resolution photographs. A narrow design would require many neurons, even if the picture's input size is very small since each pixel represents an essential input feature. Weights for each cell in the two layers of a linked layer for an image with a 100 x 100 resolution have 10,000 weights. While convolution reduces the number of design variables, it increases the network's complexity. For example, a 5×5 tiled area with identical shared weights may be used with any image size and only requires 25 parameters to be learned. Using regularized weights over a smaller number of parameters might alleviate some of the problems associated with traditional neural

networks' backpropagation, such as vanishing and growing gradients. On the other hand, convolutional neural networks are especially well-suited to grid-like data (such as photos) because convolution and/or pooling take into consideration spatial correlations between separate attributes [158].

In the case of CNN, the proposed algorithm uses a deep neural network toolbox that supports the propagation architecture of CNN through the softmax activation function, as shown in Figure 6.5.

```
layers: 7x1 nnet.cnn.layer.Layer =

  7x1 Layer array with layers:

     1   ''   Image Input             128x128x3 images with 'zerocenter' normalization
     2   ''   Convolution             20 10x10 convolutions with stride [1  1] and padding [0  0  0  0]
     3   ''   Batch Normalization     Batch normalization
     4   ''   ReLU                    ReLU
     5   ''   Fully Connected         3 fully connected layer
     6   ''   Softmax                 softmax
     7   ''   Classification Output   crossentropyex
```

Figure 6.5 Convolutional Neural Network propagation architecture

CNN intakes the optimized frame as input and propagates through the softmax activation function in three fully connected layers processed by batch normalization. A 10×10 convolution stride containing [1 1] range is padded against the padding values. The code is written using MATLAB script, and the training is shown in Figure 6.6.

```
% Code for image loading
imds = imageDatastore('BP','IncludeSubfolders',true,'LabelSource','foldernames');
numTrainFiles = 30;% No. of images to train (Should be less than maximum)
[imdsTrain,imdsValidation] = splitEachLabel(imds,numTrainFiles,'randomize'); % Split data into category
inputSize = [128 128 3];% Define image size
numClasses = 3; % Class in Dataset

layers = [
    imageInputLayer(inputSize)
    convolution2dLayer(10,20)
    batchNormalizationLayer
    reluLayer
    fullyConnectedLayer(numClasses)
    softmaxLayer
    classificationLayer];
options = trainingOptions('sgdm', ...
    'MaxEpochs',10, ...
    'ValidationData',imdsValidation, ...
    'ValidationFrequency',100, ...
    'Verbose',false, ...
    'Plots','training-progress');
net = trainNetwork(imdsTrain,layers,options);
```

Figure 6.6 CNN Training code

The hyperparameters for the CNN are provided in Table 6-1

Table 6.1 Hyperparameters for the CNN

Parameters	**Value**
Image Input	128 ×128 × 3
Maximum Epochs	10
MiniBatchSize	128
Verbose	0 (False)
Verbose Frequency	50
Validation Frequency	100
Validation Patience	5

There are three portions R, G, and B, each pattern has seven layers and every layer have 20 filters. So, the overall weight in CNN is 48000.

The architecture of the CNN is defined using a sequence of layers. The first layer is an image Input Layer that defines the input size of the images. The second layer is a convolutional layer with 10 filters of size 20 × 20. This layer performs the convolution operation on the input images, extracting features. The third layer is a batch normalization layer that normalizes the output of the convolutional layer to speed up training. The fourth layer is a ReLU activation layer that applies a non-linear activation function to the output of the batch normalization layer. The fifth layer is a fully connected layer that connects all the output features of the previous layer to the output neurons of the next layer. The last two layers are the softmax layer and classification layer, respectively. The softmax layer converts the output of the fully connected layer into a probability distribution over the three classes, and the classification layer outputs the predicted class label. The 'trainingOptions' function is used to define the training options for the network. The 'sgdm' solver is used, which is a stochastic gradient descent with momentum optimizer. The 'MaxEpochs' parameter sets the maximum number of training epochs to 10. The 'ValidationData' parameter specifies the validation data set, and the 'ValidationFrequency' parameter specifies how often the validation accuracy should be calculated. The 'Verbose' parameter is set to 'false' to suppress the training progress output, and the 'Plots' parameter is set to 'training-progress' to display the training progress plot. After the network is defined and the training options are set, the network is trained using the 'trainNetwork' function, which takes as input the training

data, the network layers, and the training options. The trained network is tested against the heart rate values that are categorised into three classes.

6.3.2 Feed-Forward Back Propagation Network using Levenberg-Marquardt

The Feed-Forward Back Propagation Network using Levenberg-Marquardt (FFBPNN-L) is a multi-layered architecture in which data passes through at least one hidden/middle layer from its way the input to the output. All of the cells in the surrounding layers communicate via a network of neurons. The connections are paired with numerical values (weights) that will be modified throughout the training phase.

An FFBPNN's primary function is to learn and map correlations between input and output. A system's weights and threshold values may also be altered using the FFBPNN learning rule to get the least potential error. In other words, it's a convoluted relationship between the extracted features of a network set. Each node or neuron's value is affected by its input from other communications system components. The value obtained from each input line amplifies the matching input signal.

Firstly, it checks the training conditions in forward direction as weight propagate one layer to other layer, without propagate the weight cannot be updated, once the training criteria is satisfied then calculate the mean square error (MSE) that is in backward direction. It was determined that training with 70% of the data was the most common use of the remaining 30% was shared evenly between verification and analysis. Equation (6.1) were used to train the model until it satisfied the criteria.

$$x_k = \sum_{i}^{n} W_{ki} x_i \tag{6.1}$$

where x_k is the new value of a variable, x_i is the starting value of the variable, and W_{ki} is the weight link value of the neuron/variable. With each epoch, the gradient of the generated weight in terms of minimum MSE is calculated, when the MSE becomes consistent and neither increasing nor decreasing, the trained weights with minimum MSE is considered as the trained neural network.

6.3.2.1 Feed Forward Back Propagation Neural Network Model Development

The FFBPNN-based model is trained to utilize historical data during the first training step. If the model fails to satisfy expectations, the Levenberg-Marquardt (LM) backpropagation method will enable the process to be re-propagated until it meets the optimal requirement. One-fifth of the data collected after the model train was utilised to verify the trained model during the validation step. The model was suggested throughout its paces and reviewed until it was confirmed to be accurate using the remaining 15% of the data sets. Accurate results were discovered, with a maximum Coefficient of determination[19] (R^2) value close to 1.0, where R^2 is the average regression value of training, testing and validation data. Input, covert, and exposed components of the network are clearly shown in Figure 6.7. In the network's hidden and output layers, Tang and Purelin were both employed [159].

Feed Forward Neural Network (FFNN) with a backpropagation algorithm is superior in precision and accuracy to other categorization algorithms. In forward processing, input data is sent to the hidden levels, and the hidden layers process the data. As depicted in Figure 6.7, the output layer provides the output. FFNN with backpropagation neural, the most frequent way to train an ANN to decrease the gradient is to use training.

[19]The coefficient of determination, or R^2, is a measure that provides information about the goodness of fit of a model. In the context of regression, it is a statistical measure of how well the regression line approximates the actual data.

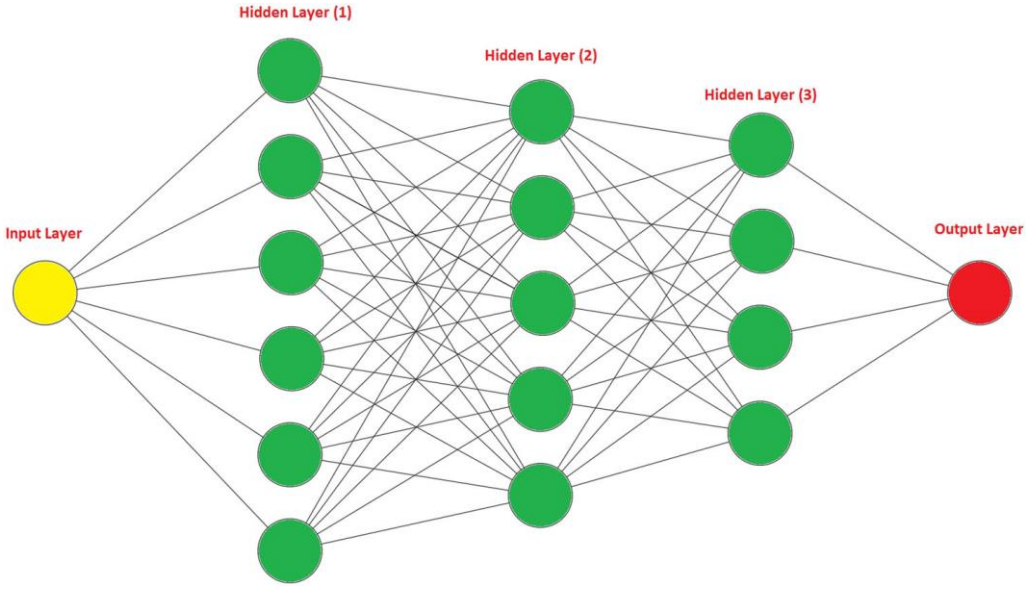


Figure 6.7 Feed-Forward Back Propagation Network Architecture [160]

The following are the steps of algorithm used to create this network for categorization:

Step 1. Generate the data
Step 2. Create the network
Step 3. Configure the network
Step 4. Initialize the weights as well as biases randomly
Step 5. Train the network
Step 6. Propagate the inputs to the forward
Step 7. Back propagate the error to the hidden layer
Step 8. Validate the network
Step 9. Use as well as analyse the performance of the network.

To test the network's ability to classify inputs, random data was generated and utilised to generate training data. To build and initialize the network, run the MATLAB code given in Appendix D-1. The weight values are generated at random. As a result of training, inputs are transmitted to forwarding, together with the output being computed in the forward direction utilizing the neural network toolbox used to train the network. The neural network training state at epoch 23 is represented in Appendix D-2 for gradient and validation check. The mean squared error is calculated in the reverse direction in this instance. Initially, it identifies how many neurons, layers, and activation functions are there, and then the performance error is computed after that. In

other words, the error histogram provides distribution of errors from the neural network during testing, training, and validation phases. The error histogram is also included in Appendix D-3 for neural network training at epoch 23.

The input layer receives the signals from the other source, and the hidden layer's job is to change them into something that the output signal can utilize. The suggested neural network design with 'tansig' function. A multi-layer perceptron is another name for this neural design. Using typical multi-layer perceptron's, any measurable function may be approximated to the appropriate accuracy rate [161].

As shown in Figure 6.8, the HOG feature is extracted from each ground truth category, which creates a data vector that is several frames into several HOG features. After that, weight adjustment starts, and if the gradient is attained, it stops training, but if not achieved, then again go to the back step and start to propagate.

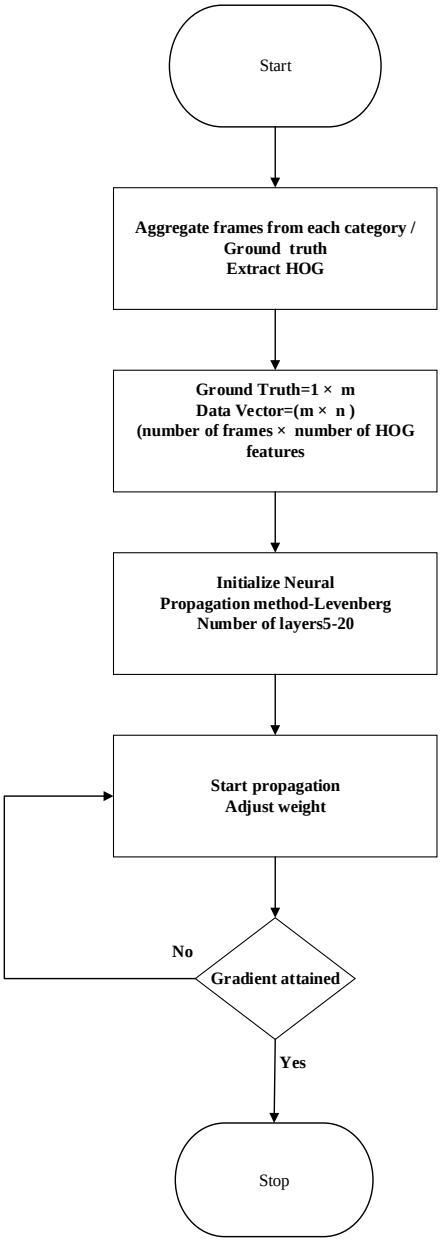


Figure 6.8 Flow chart of steps for training network

The feed-forward architecture has been attained using a neural network toolbox supported by the AI toolset in MATLAB. To train NN, the training data must be organized against its ground-truth value to get understood by the system. The training code architecture for the proposed work is shown in Figure 6.9.

```
[rows,cols]=size(AllNormal);
net=patternnet(10);
net.trainParam.epochs=1000;
net.divideParam.trainRatio = 70/100;
net.divideParam.valRatio = 15/100;
net.divideParam.testRatio = 15/100;
net=train(net,TrainingData',Target);
total_frames=numel(AllHOG);
```

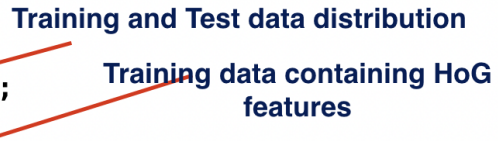


Figure 6.9 Training code architecture for the proposed algorithm

The hyperparameters for the FFBPN is provided in Table 6-2

Table 6.2 Hyperparameters for the FFBPN

Parameters	**Value**
Number of layers	3
Number of neurons in hidden layer	10
Training ratio	70%
Validation ratio	15%
Test ratio	15%
Maximum epoch	1000

The utilized neural network is a multi-neuron propagation architecture that follows the Levenberg mechanism for the feed forward movement and the back propagation movement. The training algorithm intakes the HoG features to train the system against the specified class labels of heart rate high, moderate and normal. The utilized dataset produces approximately 8100 HoG features for each image and hence the input set is consisted of n*8100 feature set where n represents the total number of images passed for the training purpose. Figure 6.10 shows the 8100 Features for 467 samples.

net TrainingData

467x8100 single

	1	2	3	4	5	6	7	8	9	10	11
1	0.0292	0.0159	0.0225	0.0655	0.1967	0.0523	0.0507	0.0220	0.0491	0.0388	0.0
2	0.4466	0.0187	0.0059	0.0052	0.0088	0.0034	0.0028	0.0125	0.4466	0.4466	0.0
3	0.4159	0.0105	2.7678e-...	2.7779e-...	6.6507e-...	0.0021	0.0040	0.0068	0.4159	0.4159	0.0
4	0.4184	0.0117	0.0018	5.0381e-...	7.8960e-...	0.0038	0.0013	0.0089	0.4184	0.4184	0.0
5	0.4159	0.0116	7.9124e-...	1.7141e-...	0.0011	1.6790e-...	0.0057	0.0126	0.4159	0.4159	0.0
6	0.4137	0.0127	7.7888e-...	8.5130e-...	5.6302e-...	6.2843e-...	0.0051	0.0114	0.4137	0.4137	0.0
7	0.0365	0.0136	0.0465	0.1180	0.1995	0.0659	0.0513	0.0504	0.0519	0.0319	0.0
8	0.4174	0.0133	8.4098e-...	6.8865e-...	4.7304e-...	0.0016	0.0051	0.0080	0.4174	0.4174	0.0
9	0.0370	0.0176	0.0351	0.1267	0.2056	0.0725	0.0510	0.0450	0.0522	0.0385	0.0
10	0.4153	0.0107	6.4113e-...	3.5792e-...	2.8965e-...	0.0011	0.0044	0.0106	0.4153	0.4153	0.0
11	0.4173	0.0137	3.7428e-...	2.2506e-...	1.1322e-...	0.0026	0.0045	0.0085	0.4173	0.4173	0.0
12	0.0294	0.0278	0.0381	0.0742	0.2068	0.0653	0.0287	0.0161	0.0260	0.0425	0.0
13	0.4180	0.0153	6.8145e-...	2.5208e-...	2.7791e-...	0.0025	0.0050	0.0111	0.4180	0.4180	0.0
14	0.4187	0.0134	9.0679e-...	1.7745e-...	2.4382e-...	0.0027	0.0048	0.0111	0.4187	0.4187	0.0
15	0.4166	0.0143	5.6734e-...	2.1388e-...	2.8597e-...	0.0028	0.0045	0.0088	0.4166	0.4166	0.0
16	0.4155	0.0099	1.7932e-...	4.0635e-...	5.0717e-...	0.0013	0.0058	0.0065	0.4155	0.4155	0.0
17	0.4147	0.0081	5.9475e-...	9.3269e-...	6.5567e-...	9.3400e-...	0.0044	0.0105	0.4147	0.4147	0.0
18	0.4180	0.0171	3.9025e-...	1.5534e-...	4.5071e-...	0.0025	0.0060	0.0094	0.4180	0.4180	0.0
19	0.0325	0.0277	0.0398	0.1209	0.2173	0.0881	0.0554	0.0376	0.0440	0.0331	0.0
20	0.4179	0.0096	0.0013	3.1735e-...	5.7070e-...	0.0012	0.0074	0.0076	0.4179	0.4179	0.0
21	0.4161	0.0128	0.0014	3.1530e-...	4.9044e-...	0.0013	0.0067	0.0093	0.4161	0.4161	0.0
22	0.0361	0.0221	0.0380	0.1106	0.2012	0.0815	0.0703	0.0250	0.0318	0.0365	0.0
23	0.4459	0.0246	0.0037	0.0019	0.0013	0.0034	0.0060	0.0137	0.4459	0.4459	0.0

Figure 6.10 8100 Features for 467 samples

The training algorithm divides the data into a distribution of 70-30 where it uses 70% of the data for the training purpose and the rest 30 % data is again divided into two segments namely 15% for the testing and rest 15% for the validation. Gradient and Epochs are the main stopping criteria for the training algorithm. If the gradient of the network is satisfied, the network will not complete the rest of the iterations. The gradient is based on Mean Squared Error. For each propagation the gradient varies as per the provided data.

The neurons are trained with multiple propagation layers and a 70-30 distribution of data. Three classes have utilised the ‘Target’ term for the ground truth values. The NN intakes HOG features that have been calculated from optimized forehead values. The proposed algorithm uses a bias connect architecture in which the layers are biased with each other using a propagation function. Figure 6.11 shows the network with functions representing three classes at the output layer (High, moderate and normal), number of features and propagating epochs other plots related to performance, training state, error histogram and confusion matrix are presented in Appendix D.

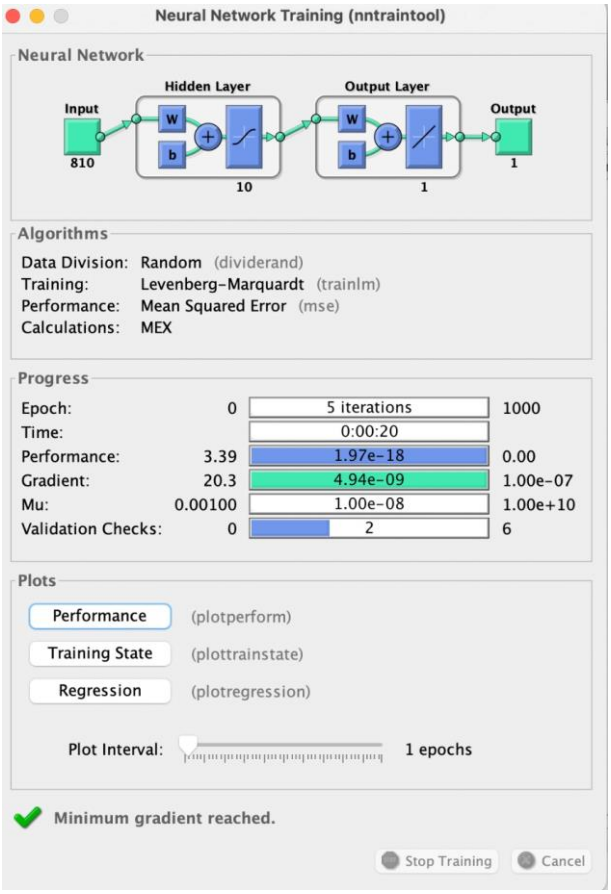


Figure 6.11 Network architecture with functions

6.4 Summary and Contribution of the Chapter

The chapter is dedicated to performing training and classification of the extracted features. It represents the feature extraction performed using HOG and two machine learning classifiers, namely, CNN and FFBPNN-L that were chosen to train and classify the designed work. The representation for the development of the Feed Forward Back Propagation Neural Network Model is discussed at the end of the chapter. The simulation outcomes of this multiclass classification-based training and classification are presented in the results and discussion chapter.

Chapter 7
Results and Discussion

The detailed simulation analysis results are discussed in this chapter. During processing, the video frames are extracted, and the performance analysis of the proposed work using Convolutional Neural Network (CNN) and Feed Forward Back Propagation Neural Network-Levenberg (FFBPNN-L) are evaluated separately. The evaluation was done utilizing the heart rate of the participants. A total of 60 videos were collected from 30 different participants and used the pulse oximeter for heart rate measurement.

As the data has been processed in frames, the monitored heart rate has been calculated by taking the average time interval for which the video has been shot. Hence, tagging the data with the same heart rate will not be a justifiable step to do for every frame. In order to dignify the proposed work, the data heart rate is further divided into three categories rather than putting the exact heart rate into action. The categories are termed as labels of the processing, namely “High”, “Moderate”, and “Normal”. The prediction results of this classification problem is summarized using confusion matrix. The correct and incorrect predictions are expressed using true positives, true negatives, false positives, and false negative. The representation of count values for each of the class is further provided in Appendix D-4 for training, testing and validation stage followed by the Neural Network training receiver operating characteristics (ROC) for all classes for training, validation and test ROC in Appendix D-5. As per the provided set of data that has been used for processing and has been collected in a real-time environment, the range for high would be (101-120) beats per minute. The normal label class will be identified with a range of (60-80) beats per minute, whereas the moderate will be (81-100). In order to perform the evaluation of the proposed algorithm, the extracted regions with their processing capabilities attributes have been passed to the training architecture with their associated labels. Each frame value is either associated with a high, moderate or normal ground-truth label.

True Positive Rate (TPR) is evaluated by calculating the ratio of the total number of frames that have matched its real ground truth value to the total number of frames that have been passed as the original class as in equation (7.1)

$$TPR = \frac{True_{positive}}{True_{positive} + False_{negative}} \quad (7.1)$$

In a similar fashion, the False Positive Rate (FPR) represent the ratio of the total number of false frames that have been detected against its ground truth or label to the total number of frames that have been supplied for the evaluation as represented in equation (7.2)

$$FPR = \frac{False_{positive}}{False_{negative} + True_{positive}} \quad (7.2)$$

The depth analysis in terms of accuracy, which is the percentage of correctly classifies instances out of all the cases. Equation (7.3) shows the formula for accuracy.

$$Accuracy = \frac{Correct\ Classifications}{All\ Classifications} \quad (7.3)$$

Kappa analysis is also subjected to variation in the training and testing sets. Kappa k can be calculated using (7.4).

$$k = \frac{2 \times (TP \times TN - FN \times FP)}{(TP+FP)\times(FP+TN)+(TP+FN)\times(FN+TN)} \quad (7.4)$$

where, TP is true positive, FP is false positive, and TN and FN are true and false negative, respectively.

Kappa coefficient, also known as Cohen's kappa, is a statistical measure that assesses the inter-rater agreement between two or more annotators. In the context of machine learning, kappa coefficient is often used as an evaluation metric for classification models, especially when dealing with imbalanced datasets [162].

Kappa coefficient ranges from -1 to 1, where a value of 1 indicates perfect agreement, 0 indicates no agreement, and -1 indicates perfect disagreement. When evaluating a classification model, the kappa coefficient can be used in conjunction with other metrics

such as accuracy and F-measure to provide a more comprehensive analysis of the model's performance.

Accuracy measures the proportion of correct predictions made by a classification model over the total number of predictions. While accuracy is a useful metric, it can be misleading when dealing with imbalanced datasets. For example, if a dataset has 90% of its samples belonging to one class, a model that predicts all samples as belonging to that class would achieve 90% accuracy, even though it would be an ineffective model.

F-measure, on the other hand, considers both precision and recall, and is particularly useful in situations where false positives and false negatives have different costs. It is the harmonic mean of precision and recall, and can be expressed as F1, F2, F0.5, depending on which metric is given more weight.

Kappa coefficient provides a measure of agreement that takes into account the possibility of agreement occurring by chance. It is particularly useful when dealing with imbalanced datasets as it provides a more accurate evaluation of a model's performance compared to accuracy alone.

When the kappa coefficient is high, it indicates a high level of agreement between the predictions made by a model and the ground truth labels. This is typically associated with high accuracy and F-measure values, as the model is correctly classifying a majority of the samples [162-164].

The ordinals of simulation analysis are summarized in Table 7.1.

Table 7.1 Ordinals of Simulation Analysis

Parameter	Description
Number of Frames	500 to 1000
Distribution of Training to Testing Sets	70:30, 80:20, 90:10
Classifiers	CNN, FFBPNN-L
Number of Layers in each neural architecture	5 to 20
Evaluation Parameters	TRP, FPR, Accuracy, Kappa

The graphs are displayed with the number of frames, and parameters are discussed with respect number of layers used in neural architecture.

The proposed algorithm is ABC with FFBPNN-L. To demonstrate the efficiency of the proposed algorithm, the model has been compared with FFBPNN-L, CNN and ABC with CNN as the neural network contains many hidden layers; the results are also extracted with various numbers of hidden layers. The number of hidden layer architectures chosen is 5 layered neural architecture, 15 layer neural architecture, and 20 layer neural architecture. This is to prove that the more hidden layers in the architecture, the more efficient the model. However, it was observed that the best HR predictions were obtained using 20-layered neural architecture. The detailed comparative analysis of various scenarios used with the implementation of a 20-layered neural architecture is discussed in this chapter, and the other results are added to Appendix E.

The performance parameters of proposed and existing algorithms are evaluated and displayed in accordance with the 20 hidden layers in the neural network architecture, considering different ratios of training and testing data sets.

Each distribution is made based on the signified frames that have been labelled using the heart rate. A distribution of 70:30 states that 70% of frames that have been generated from the videos would be applied for the training purpose. Those frames would be associated with the labels formed by heart rates. A distribution of 80:20 states

that 80% of frames that have been generated from the videos would be applied for the training purpose. Those frames would be associated with the labels formed by heart rates. In a similar fashion, if the distribution is 90:10, the set of training data will contain 90% of the frames that have been generated from the videos. A significant improvement in the accuracy will be observed as more training data will establish more correlation among the labelled frames from the heart rate. The analysis is presented for all the labels combined viz the analysis is the mean analysis for all the classes. If TPR is 0.78, for instance, it represents the average TPR of "high", "moderate", and "normal" heart rates. The same goes with FPR, accuracy and Kappa coefficient.

7.1 Analysis using 70:30 distribution

In this, the videos dataset is divided into a 70:30 ratio where 70% of the dataset is used for training and 30% for testing. The techniques used are Convolutional Neural Network (CNN), ABC with CNN, Feed Forward Back Propagation Neural Network using Levenberg-Marquardt training (FFBPNN-L), and ABC with FFBPNN-L.

7.1.1 True Positive Rate Analysis based on CNN

The frames from the video are extracted and analysed using CNN architecture with and without the involvement of an optimization algorithm. It has been observed that when only 500 frames are used in the analysis, the TPR using CNN alone is 0.79251 and using ABC with CNN is 0.80573. Further, when the number of frames is increased to 1000, a corresponding rise in the TPR is observed as 0.83633 using CNN alone and 0.86046 using ABC with CNN. The variation in the TPR values as the number of frames was increased from 500 to 1000 is summarized in Table 7.2.

Table 7.2 TPR analysis values of CNN and ABC with CNN

Number of Frames	CNN	ABC with CNN
500	0.79251	0.80573
600	0.79470	0.81052
700	0.80401	0.81886
800	0.81322	0.83307
900	0.82985	0.84998
1000	0.83633	0.86046

The graphical analysis of the variation in the TPR based on CNN architecture is shown in Figure 7.1. Overall, it is observed that with each variation in terms of the number of frames, the TPR remained relatively higher when optimized features were used for the training of CNN. Hence, the average TPR using CNN is observed to be 0.81177 and using ABC with CNN is 0.82977.

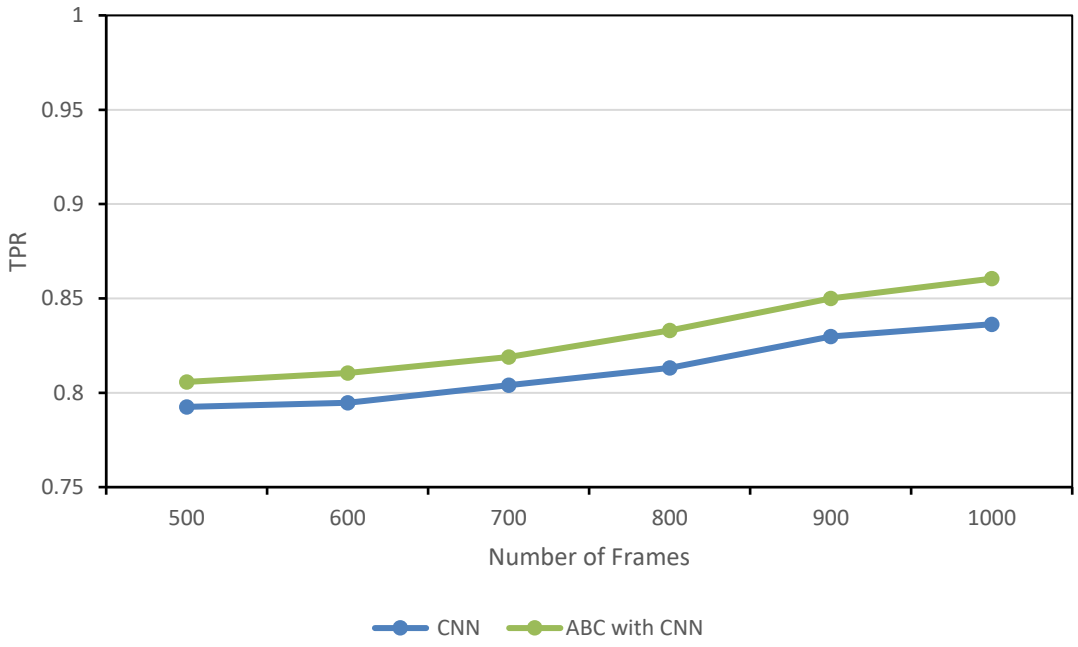


Figure 7.1 TPR analysis of CNN and ABC with CNN

7.1.2 True Positive Rate Analysis based on FFBPNN-L

The next neural architecture used in the analysis is FFBPNN-L. The TPR analysis performed using FFBPNN-L with and without integration of ABC based optimizations is summarized in Table 7.3.

Table 7.3 TPR analysis values of FFBPNN-L and ABC with FFBPNN-L

Number of Frames	FFBPNN-L	ABC with FFBPNN-L
500	0.86698	0.88020
600	0.86935	0.88496
700	0.88632	0.89715
800	0.90343	0.90915
900	0.91468	0.91954
1000	0.92598	0.93477

It is observed that when 500 frames are used for the analysis, the TPR of 0.86698 is achieved using only FFBPNN-L alone and the TPR of 0.88020 when ABC is integrated prior to FFBPNN-L. When the frames are increased to 700, the TPR using FFBPNN-L increases to 0.88632 and, using ABC with FFBPNN-L to 0.89715. When the number of frames is further increased to 1000, the TPR again show a slight rise of 0.92598 using FFBPNN-L and 0.93477 using ABC with FFBPNN-L. The graphical analysis of these variations represents using the line graph shown in Figure 7.2.

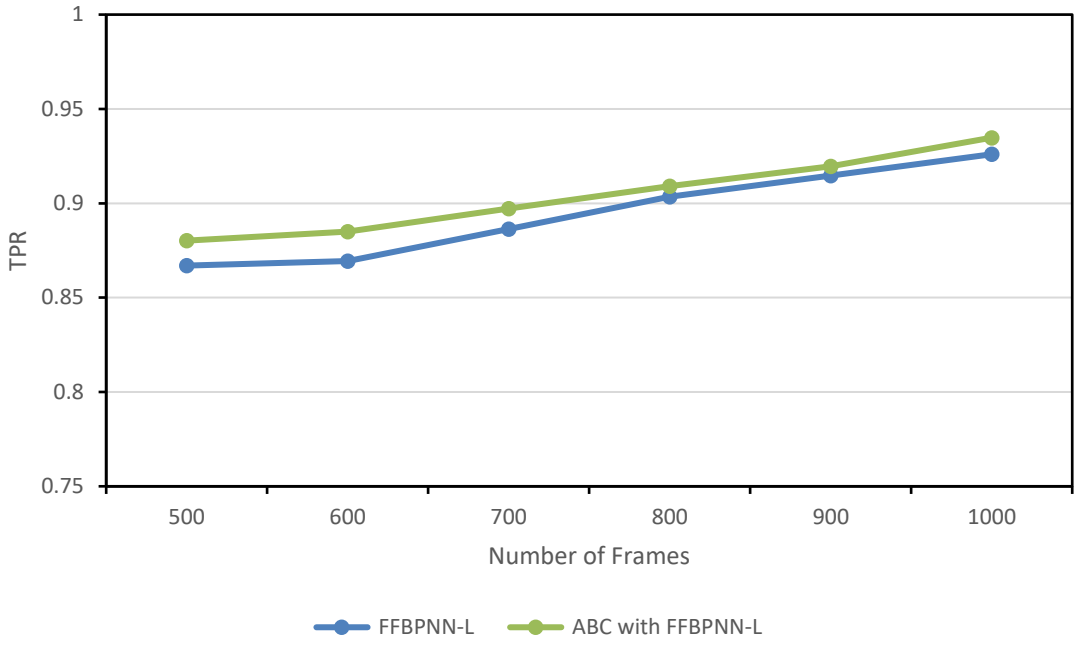


Figure 7.2 TPR analysis of FFBPNN-L and ABC with FFBPNN-L

The average values using various algorithms used for TPR analysis using 70:30 distribution is given in Table 7.4. Equation (7.5) shows the formula to calculate the average.

$$Average = \frac{sum\ of\ all\ TPR\ values\ at\ different\ frames}{total\ number\ of\ TPR\ values} \tag{7.5}$$

The average TPR using individual CNN and FFBPNN-L is found to be comparatively lower than the corresponding techniques when ABC based optimization strategy is integrated into the framework.

Table 7.4 Average TPR values

Parameter	Algorithm			
	CNN	ABC with CNN	FFBPNN-L	ABC with FFBPNN-L
Average TPR	0.81177	0.82977	0.89446	0.90429

The average TPR demonstrated over 1000 frames using CNN is 0.81177, ABC with CNN is 0.82977, FFBPNN-L is 0.89446 ABC with FFBPNN-L is 0.90429. This means

that ABC with FFBPNN-L resulted in 8.24082% higher TPR as compared to ABC with CNN using the 70:30 distribution of the dataset.

7.1.3 False Positive Rate Analysis based on CNN

The FPR analysis performed using CNN based architecture is described using a variable number of frames in Table 7.5. It is observed that when 500 frames are used in the analysis, the observed FPR using CNN is 0.13595 and using ABC with CNN is 0.12547. When the number of frames is 700, the corresponding FPR using CNN is observed as 0.15649 and using ABC with CNN is 0.14340. Further, when the number of frames is increased to 900 and 1000, the FPR using CNN increases to 0.175231 and 0.18856, while using ABC with CNN, it is 0.15987 and 0.16471, respectively. This shows that lower FPR is observed when ABC based optimization approach is integrated into the CNN based architecture.

Table 7.5 FPR analysis values of CNN and ABC with CNN

Number of Frames	CNN	ABC with CNN
500	0.13595	0.12547
600	0.14666	0.13381
700	0.15649	0.14340
800	0.16592	0.15334
900	0.17523	0.15987
1000	0.18856	0.16471

The graphical representation of the variation in the FPR for 500 to 1000 frames are shown in Figure 7.3. It is observed that with the increase in the number of frames, the FPR increases for both CNN as well as ABC with CNN. However, the FPR demonstrated using ABC with CNN is comparatively lower as compared to CNN alone. Hence, ABC significantly lowered the FPR and proved to be better than CNN alone.

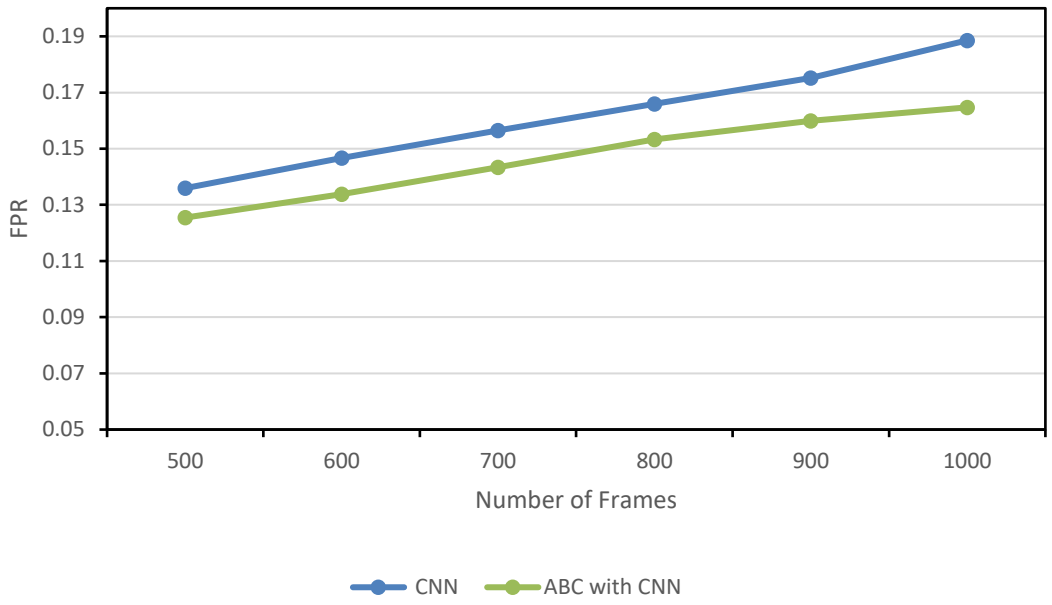


Figure 7.3 FPR analysis of CNN and ABC with CNN

7.1.4 False Positive Rate Analysis based on FFBPNN-L

The variation in the FPR observed using FFBPNN-L based architecture is described in Table 7.6. It is observed that when 500 frames are used in the analysis, the observed FPR using FFBPNN-L is 0.12228 and using ABC with FFBPNN-L is 0.11180. When the number of frames is 700, the corresponding FPR using FFBPNN-L is observed as 0.14690 and using ABC with FFBPNN-L is 0.13133. Further, when the number of frames is increased to 900 and 1000, the FPR using FFBPNN-L increases to 0.16370 and 0.17438, while using ABC with FFBPNN-L, it is 0.15070901 and 0.15794, respectively. This means that integration of ABC based optimization approach resulted in lower FPR associated during analysis.

Table 7.6 FPR analysis values of FFBPNN-L and ABC with FFBPNN-L

Number of Frames	FFBPNN-L	ABC with FFBPNN-L
500	0.12228	0.11180
600	0.13125	0.11951
700	0.14690	0.13133
800	0.15737	0.13525
900	0.16370	0.15070
1000	0.17438	0.15794

Figure 7.4 shows the graphical analysis of the variation in the FPR 1000 frames. It is observed that with the increase in the number of frames, the FPR increases for both FFBPNN-L as well as ABC with FFBPNN-L. However, the FPR demonstrated using ABC with CNN is comparatively lower as compared to FFBPNN-L alone. Hence, ABC significantly lowered the FPR and proved to be better than FFBPNN-L alone.

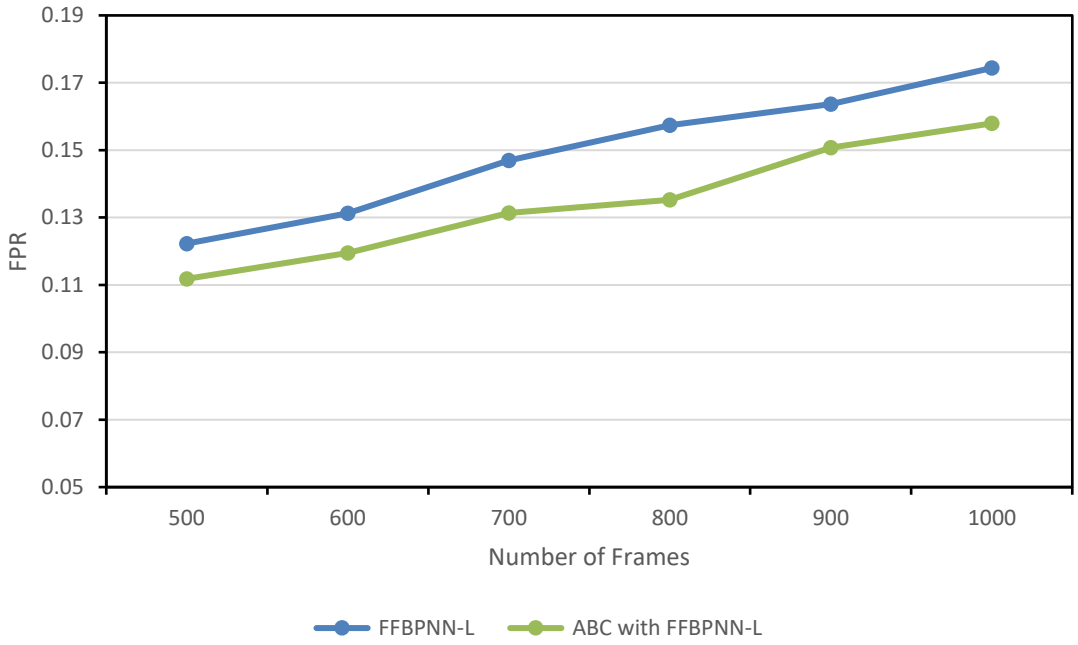


Figure 7.4 FPR analysis of FFBPNN-L and ABC with FFBPNN-L

The average FPR analysis using 70:30 distribution is given in Table 7.7. It is observed that the average FPR using individual CNN and FFBPNN-L is found to be

comparatively higher than the corresponding techniques when ABC based optimization strategy is integrated into the framework. The average FPR achieved over 1000 frames using CNN is 0.16147, ABC with CNN is 0.14676, FFBPNN-L is 0.14931, ABC with FFBPNN-L is 0.13442. This means that ABC with FFBPNN-L resulted in lower FPR as compared to ABC with CNN using the 70:30 distribution of the dataset.

Table 7.7 Average FPR values

Parameter	Algorithm			
	CNN	ABC with CNN	FFBPNN-L	ABC with FFBPNN-L
Average FPR	0.161471	0.14676	0.14931	0.13442

7.1.5 Kappa Coefficient Analysis based on CNN

The Kappa coefficient observed for different frames extracted from the video using CNN with and without ABC is given in Table 7.8. The number of frames selected for the analysis varied from 500 to 1000. It is observed that the Kappa analysis performed using 500 frames results in 0.82607 Kappa using CNN and 0.83929 using ABC with CNN. When 700 frames are used for the investigation, the Kappa increases to as 0.837281 using CNN and 0.85367 using ABC with CNN. When the number of frames is increased to 1000, the Kappa also increases and reaches 0.85996 using CNN and 0.89294 using ABC with CNN.

Table 7.8 Kappa analysis values of CNN and ABC with CNN

Number of Frames	CNN	ABC with CNN
500	0.82607	0.83929
600	0.82795	0.84099
700	0.83728	0.85367
800	0.84149	0.86192
900	0.84500	0.86752
1000	0.85996	0.89294

The graphical analysis of the variation observed in the Kappa over the 1000 frames is shown in Figure 7.5. In the line graphs, Kappa analysis using ABC with CNN had shown higher Kappa in comparison to using CNN alone on every frame. This indicates that using optimized features in CNN improves the overall analysis using CNN. The overall analysis shows that the average Kappa for 1000 frames is observed to be 0.83962 using CNN and 0.85939 using ABC with CNN.

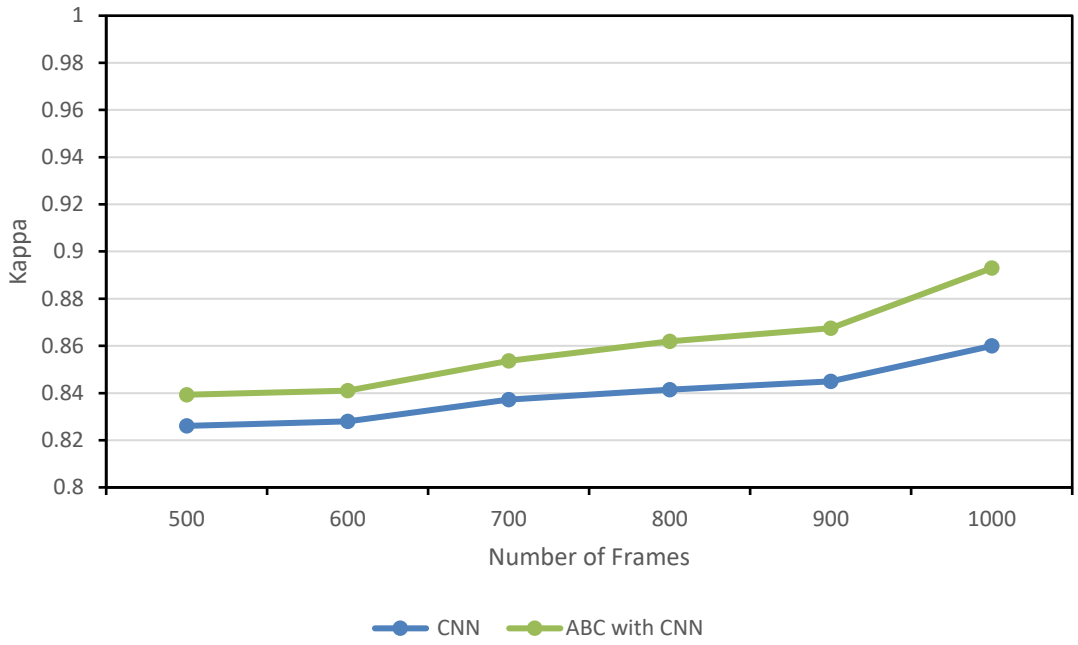


Figure 7.5 Kappa analysis of CNN and ABC with CNN

7.1.6 Kappa Coefficient Analysis based on FFBPNN-L

The Kappa analysis for the 1000 frames is tabulated using FFBPNN-L, and ABC optimized FFBPNN-L is given in Table 7.9. Initially, when the number of frames is set to 500 frames, the observed Kappa is 0.85856 using FFBPNN-L and 0.87178 using ABC with FFBPNN-L. When 800 frames are used for the analysis, the Kappa increases to 0.88092 using FFBPNN-L and 0.90825 using ABC with FFBPNN-L. When the number of frames is increased to 1000, the Kappa also increases and reaches 0.90882 using FFBPNN-L and 0.94103 using ABC with FFBPNN-L.

Table 7.9 Kappa analysis values of FFBPNN-L and ABC with FFBPNN-L

Number of Frames	FFBPNN-L	ABC with FFBPNN-L
500	0.85856	0.87178
600	0.86874	0.87586
700	0.87424	0.89037
800	0.88092	0.90825
900	0.89346	0.91671
1000	0.90882	0.94103

The graph shown in Figure 7.6 represents the Kappa analysis of FFBPNN-L and ABC with FFBPNN-L. It is observed that the Kappa represented by the curved line shows a steady improvement. The overall Kappa using FFBPNN-L is 0.88079 and 0.90067 using ABC with FFBPNN-L.

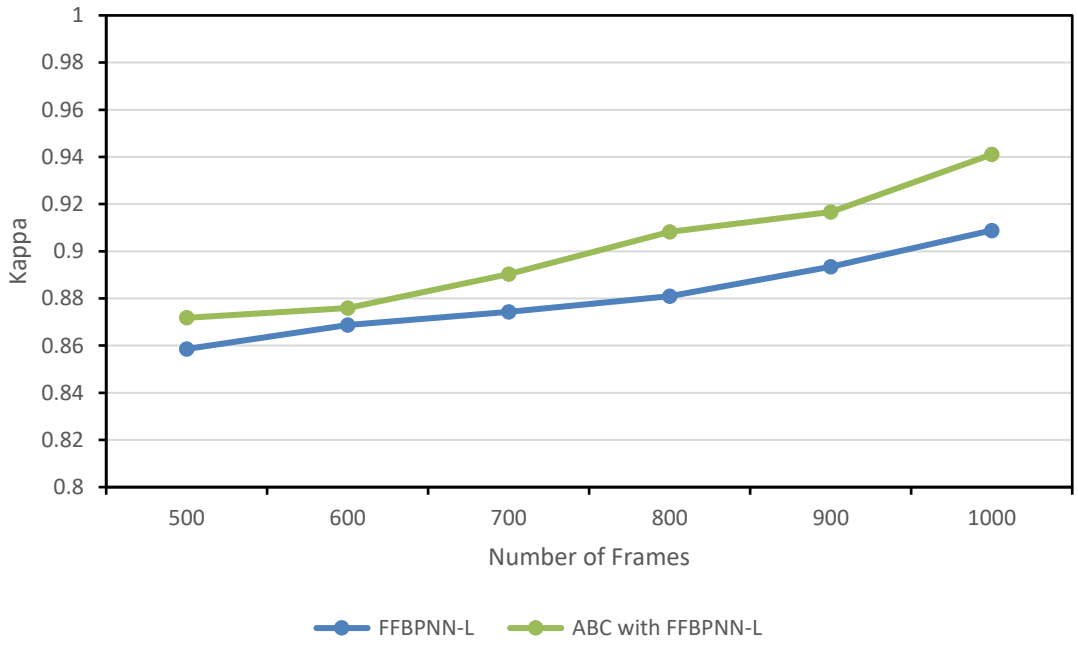


Figure 7.6 Kappa analysis of FFBPNN-L and ABC with FFBPNN-L

The average Kappa analysis of the two neural architectures with and without the involvement of ABC based optimization is summarized in Table 7.10. It is observed

that optimization had improved the Kappa values observed with the implementation of the neural architectures.

Table 7.10 Average Kappa analysis values

Parameter	Algorithm			
	CNN	ABC with CNN	FFBPNN-L	ABC with FFBPNN-L
Average Kappa	0.83962	0.85939	0.88079	0.90067

The average Kappa demonstrated over 1000 frames using CNN is 0.83962, ABC with CNN is 0.85939, FFBPNN-L 0.88079 and ABC with FFBPNN-L is 0.90067. This means that ABC with FFBPNN-L resulted in 4.58318% better Kappa as compared to ABC with CNN using a 70:30 distribution of the dataset.

7.1.7 Accuracy Analysis based on CNN

Accuracy analysis performed using CNN based architecture is summarized in Table 7.11. The number of frames extracted from the video used in the analysis varies from 500 to 1000. The corresponding accuracy values achieved using CNN and ABC with CNN are tabulated for each variation in the number of frames.

Table 7.11 Accuracy analysis values of CNN and ABC with CNN

Number of Frames	CNN	ABC with CNN
500	0.829268	0.84248
600	0.83575	0.84399
700	0.84342	0.85902
800	0.85819	0.86904
900	0.87285	0.89485
1000	0.88086	0.90520

The accuracy analysis performed for 500 frames is 0.8292 using CNN and 0.84248 using ABC with CNN. When 700 frames are used for the analysis, the accuracy

increases to as 0.84342 using CNN and 0.85902 using ABC with CNN. When the number of frames is increased to 1000, the accuracy shows a steady improvement to 0.88086 using CNN and 0.90520 using ABC with CNN.

The graphical analysis of the observed changes in the accuracy values using CNN architecture is given in Figure 7.7. The graph shows an unexpected rise in the accuracy with an increase in the number of frames used in the analysis. Further, it reflects an average accuracy of 0.85339245 using CNN and 0.869102977 using ABC with CNN.

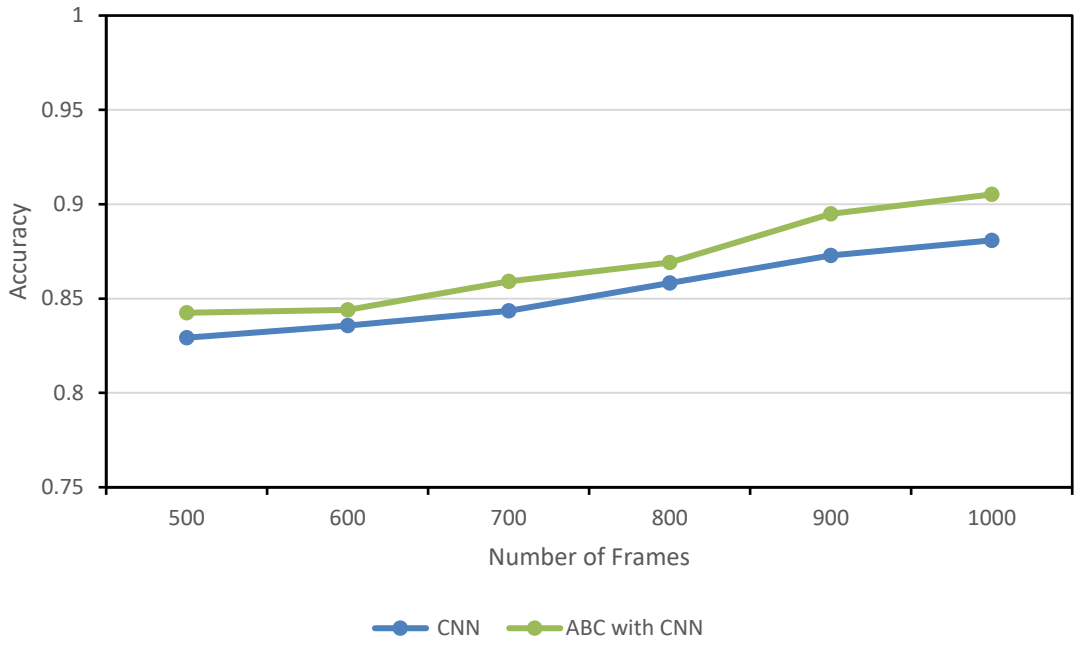


Figure 7.7 Accuracy analysis of CNN and ABC with CNN

7.1.8 Accuracy Analysis based on FFBPNN-L

The accuracy analysis using FFBPNN-L with and without ABC based optimization is summarised in Table 7.12. The variation used in the analysis ranges from 500 to 1000 frames extracted from the video. The accuracy values observed for 500 frames using FFBPNN-L are 0.84986 and 0.86611 using ABC with FFBPNN-L. When 800 frames are used for the analysis, the accuracy increases to as 0.88994 using FFBPNN-L and 0.91060 using ABC with FFBPNN-L. When the number of frames is increased to 1000, the accuracy shows a steady improvement to 0.921841 using FFBPNN-L and 0.96659 using ABC with FFBPNN-L.

Table 7.12 Accuracy analysis values of FFBPNN-L and ABC with FFBPNN-L

Number of Frames	FFBPNN-L	ABC with FFBPNN-L
500	0.84986	0.86611
600	0.85321	0.86815
700	0.87017	0.89622
800	0.88994	0.91060
900	0.90752	0.93867
1000	0.92184	0.96659

The graphical representation of Table 7.12 is given in Figure 7.8 for the accuracy analysis using FFBPNN-L and ABC with FFBPNN-L. Initially, when frames are increased from 500 to 600, there is a steady improvement in the accuracy values. However, a further increase in the number of frames significantly increased the accuracy of both scenarios.

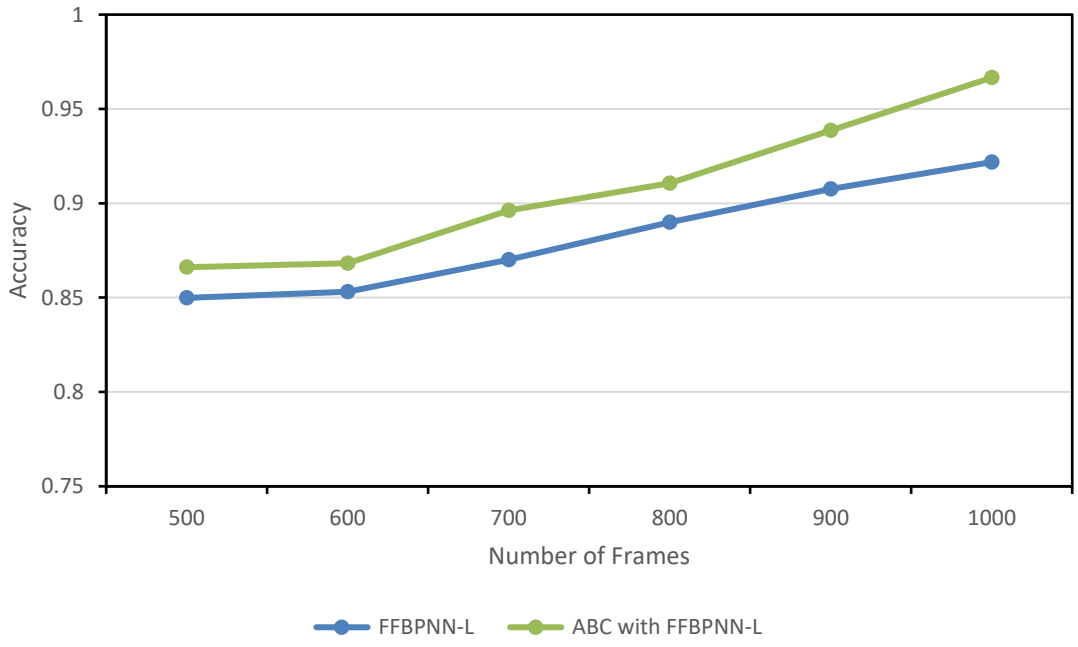


Figure 7.8 Accuracy analysis of FFBPNN-L and ABC with FFBPNN-L

The average accuracy analysis using both the neural network-based architectures is summarized in Table 7.13. The average accuracy demonstrated over 1000 frames using CNN is 0.8533, ABC with CNN is 0.86910, FFBPNN-L 0.88204 and ABC with FFBPNN-L is 0.09077. This means that ABC with FFBPNN-L resulted in 4.25498% better accuracy as compared to ABC with CNN using a 70:30 distribution of the dataset.

Table 7.13 Average Accuracy values

Parameter	Algorithm			
	CNN	ABC with CNN	FFBPNN-L	ABC with FFBPNN-L
Average Accuracy	0.85339	0.86910	0.88209	0.90772

7.2 Analysis using 80:20 distribution

In this part, the dataset is divided into an 80:20 ratio where 80% of the dataset is used for training, and 20% of the dataset is used for testing. Same as section 7.1, all the algorithms will be evaluated in accordance with performance parameters and the results for evaluation are as follows.

7.2.1 True Positive Rate Analysis based on CNN

The frames from the video are extracted and analysed using CNN architecture with and without the involvement of an optimization algorithm. It has been observed that when only 500 frames are used in the analysis, the TPR using CNN only is 0.7954 and using ABC with CNN is 0.80866. Further, when the number of frames is increased to 1000, a corresponding rise in the TPR is observed as 0.82907 using CNN alone and 0.86906 using ABC with CNN. The variation in the TPR values as the number of frames is increased from 500 to 1000 is summarized in Table 7.14.

Table 7.14 TPR analysis values of CNN and ABC with CNN

Number of Frames	CNN	ABC with CNN
500	0.79544	0.80866
600	0.79695	0.812295
700	0.80517	0.825539
800	0.80874	0.829454
900	0.824271	0.844657
1000	0.829073	0.86906

The graphical analysis of the variation in the TPR based on CNN architecture using the 80:20 scenario is shown in Figure 7.9. Overall, it is observed that for each variation in terms of the number of frames, the TPR remained relatively higher when optimized features were used for the training of CNN. Hence, the average TPR using CNN is observed to be 0.8099 and using ABC with CNN is 0.8316.

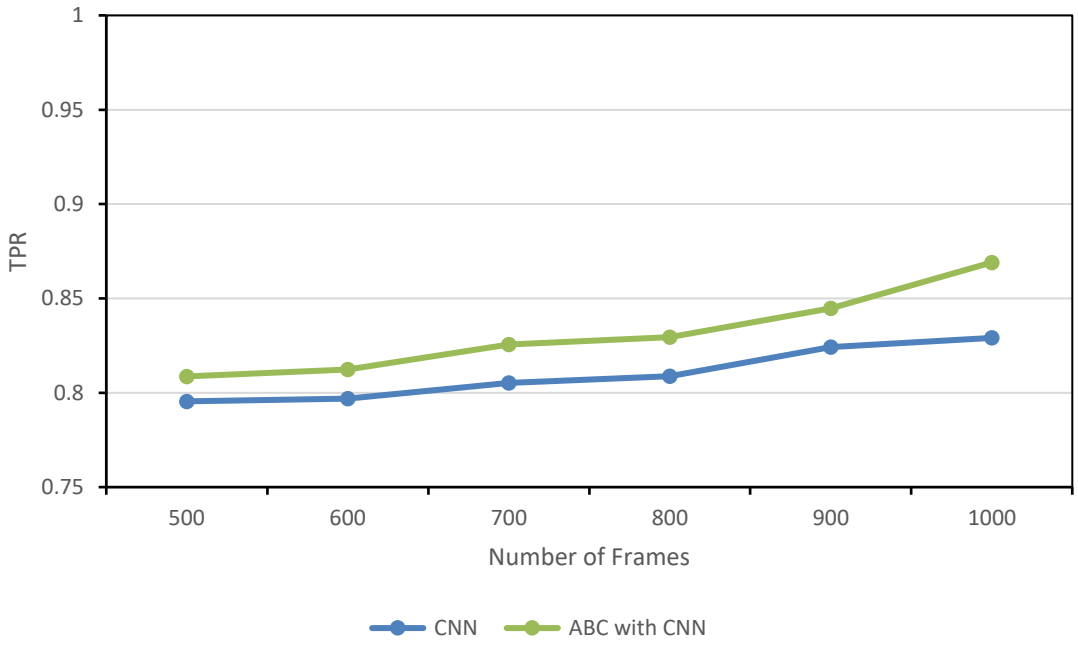


Figure 7.9 TPR analysis of CNN and ABC with CNN

7.2.2 True Positive Rate Analysis based on FFBPNN-L

The next neural architecture used in the analysis is FFBPNN-L. The TPR analysis performed using FFBPNN-L with and without integration of ABC based optimizations is summarized in Table 7.15.

Table 7.15 TPR analysis values of FFBPNN-L and ABC with FFBPNN-L

Number of Frames	FFBPNN-L	ABC with FFBPNN-L
500	0.88716	0.90038
600	0.89212	0.90624
700	0.91164	0.92777
800	0.91794	0.94313
900	0.93261	0.955102
1000	0.94491	0.97952

It is observed that when 500 frames are used for the analysis, the TPR of 0.88716 is achieved using only FFBPNN-L alone and the TPR of 0.90038 when ABC is integrated prior to FFBPNN-L. When the frames are increased to 700, the TPR using FFBPNN-L increases to 0.91164 and using ABC with FFBPNN-L to 0.92777. When the number of frames is further increased to 1000, the TPR again show a slight rise of 0.94491 using FFBPNN-L and 0.97952 using ABC with FFBPNN-L. The graphical analysis of these variations is shown using the line graph in Figure 7.10.

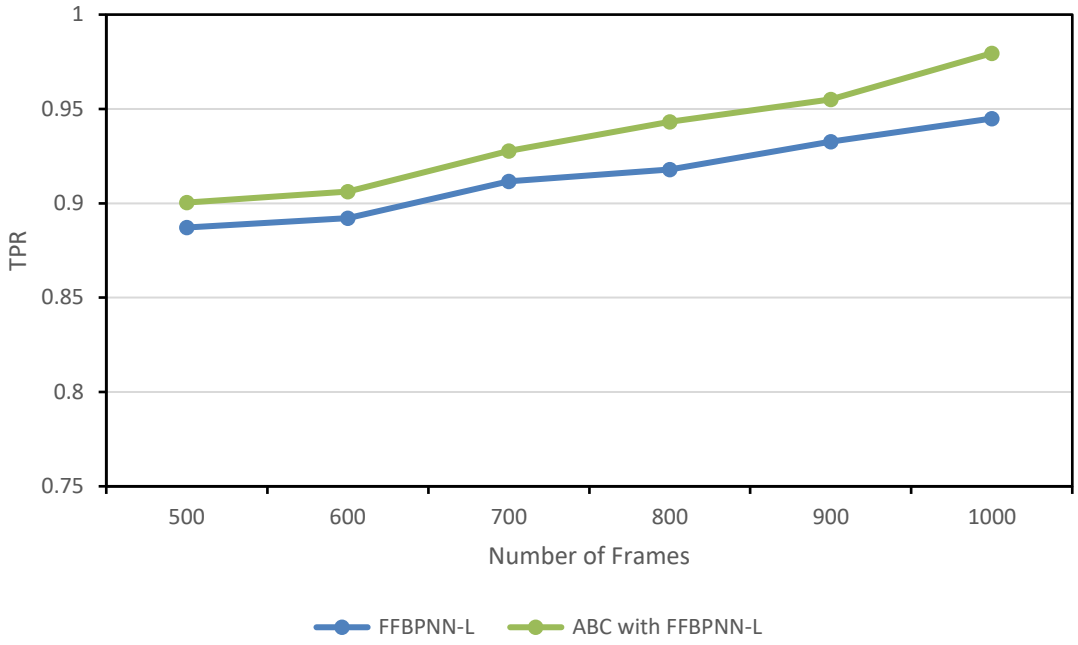


Figure 7.10 TPR analysis of FFBPNN-L and ABC with FFBPNN-L

The average values using various scenarios used for TPR analysis using 80:20 distributions are given in Table 7.16. The average TPR using individual CNN and FFBPNN-L is found to be comparatively lower than the corresponding techniques when ABC based optimization strategy is integrated into the framework.

Table 7.16 Average TPR values

Parameter	Algorithm			
	CNN	ABC with CNN	FFBPNN-L	ABC with FFBPNN-L
Average TPR	0.80994	0.83161	0.91440	0.93361

The average TPR demonstrated over 1000 frames using CNN is 0.81177, ABC with CNN is 0.83161, FFBPNN-L is 0.91440, ABC with FFBPNN-L is 0.93361. This means that ABC with FFBPNN-L resulted in 11.09183% higher TPR as compared to ABC with CNN using the 80:20 distribution of the dataset.

7.2.3 False Positive Rate Analysis based on CNN

The FPR analysis performed using CNN based architecture is described using a variable number of frames in Table 7.17. It is observed that when 500 frames are used in the analysis, the observed FPR using CNN is 0.12619 and using ABC with CNN is 0.11571. When the number of frames is 700, the corresponding FPR using CNN is observed as 0.15102 and using ABC with CNN is 0.13682. Further, when the number of frames is increased to 900 and 1000, the FPR using CNN increases to 0.16794 and 0.18033, while using ABC with CNN, it is 0.15288 and 0.15665, respectively. This shows that lower FPR is observed when ABC based optimization approach is integrated into the CNN based architecture.

Table 7.17 FPR analysis values of CNN and ABC with CNN

Number of Frames	CNN	ABC with CNN
500	0.12619	0.11571
600	0.13736	0.12807
700	0.15102	0.13682
800	0.15548	0.14182
900	0.16794	0.15288
1000	0.18033	0.15665

The graphical representation of the variation in the FPR for 500 to 1000 frames using 80:20 distribution is shown in Figure 7.11. It is observed that with the increase in the number of frames, the FPR increases for both CNN as well as ABC with CNN. However, the FPR demonstrated using ABC with CNN is comparatively lower as compared to CNN alone. Hence, ABC significantly lowered the FPR and proved to be better than CNN alone.

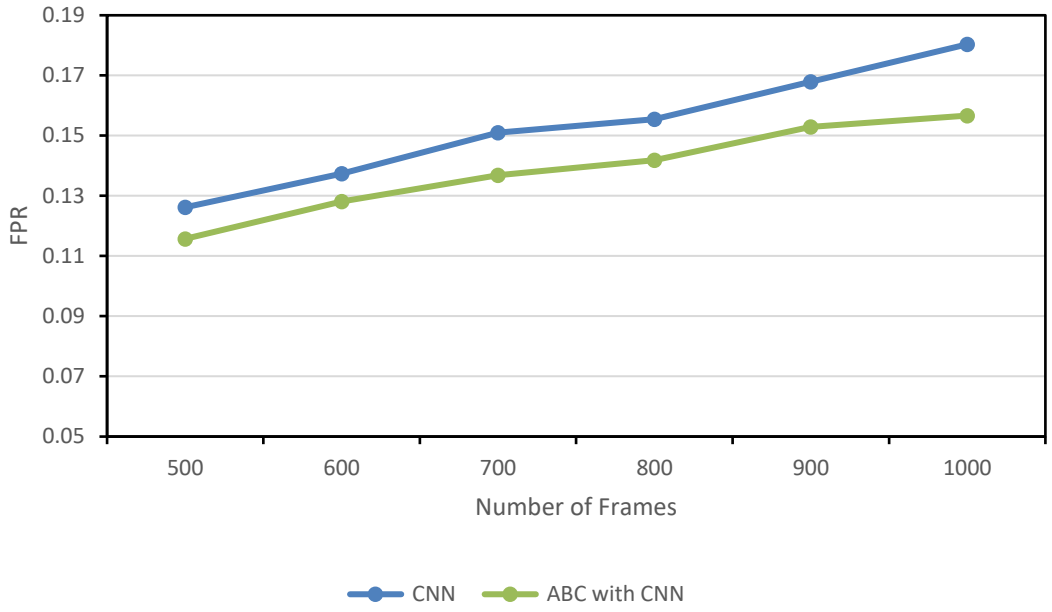


Figure 7.11 FPR analysis of CNN and ABC with CNN

7.2.4 False Positive Rate Analysis based on FFBPNN-L

Another technique used for the FPR analysis in the present work is FFBPNN-L. The variation in the FPR observed using FFBPNN-L based architecture is described in Table 7.18. It is observed that when 500 frames are used in the analysis, the observed FPR using FFBPNN-L is 0.12353 and using ABC with FFBPNN-L is 0.11004. When the number of frames is 700, the corresponding FPR using FFBPNN-L is observed as 0.14213 and using ABC with FFBPNN-L is 0.13169. Further, when the number of frames is increased to 900 and 1000, the FPR using FFBPNN-L increases to 0.15399 and 0.16334, while using ABC with FFBPNN-L, it is 0.14444 and 0.14225, respectively. This means that integration of ABC based optimization approach resulted in lower FPR associated during analysis.

Table 7.18 FPR analysis values of FFBPNN-L and ABC with FFBPNN-L

Number of Frames	FFBPNN-L	ABC with FFBPNN-L
500	0.12353	0.11004
600	0.13154	0.12263
700	0.14213	0.13169
800	0.15037	0.13951
900	0.15399	0.14444
1000	0.16334	0.14225

Figure 7.12 shows the graphical analysis of the variation in the FPR 1000 frames using an 80:20 ratio. It is observed that with the increase in the number of frames, the FPR increases for both FFBPNN-L as well as ABC with FFBPNN-L. However, the FPR demonstrated using ABC with CNN is comparatively lower as compared to FFBPNN-L alone. Hence, ABC significantly lowered the FPR and proved to be better than FFBPNN-L alone.

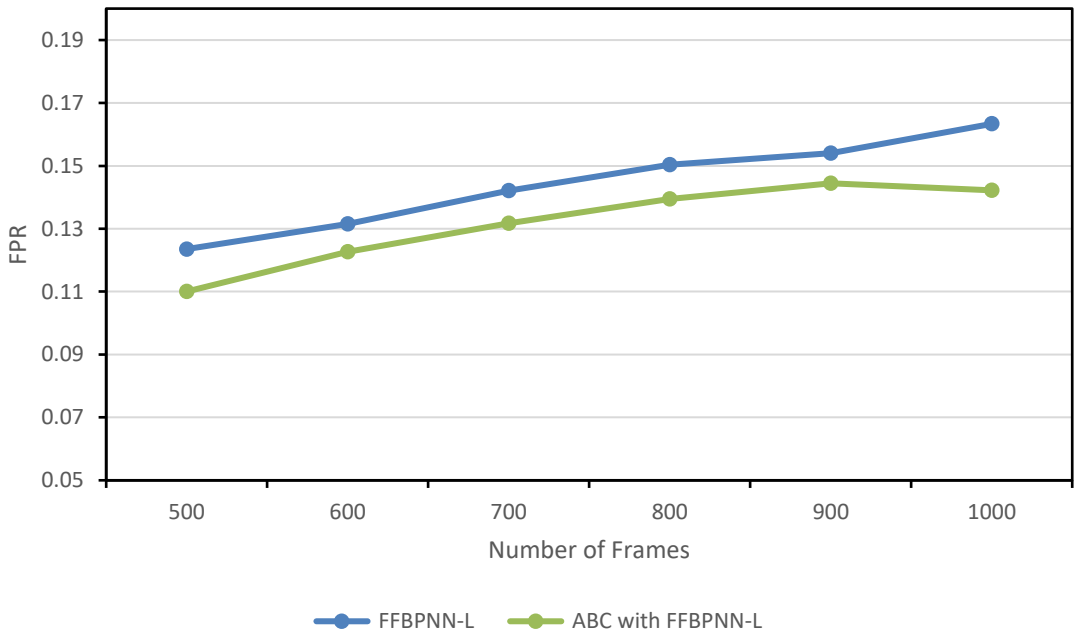


Figure 7.12 FPR analysis of FFBPNN-L and ABC with FFBPNN-L

The average FPR analysis using a 70:30 distribution is given in Table 7.19. It is observed that the average FPR using individual FFBPNN-L is found to be comparatively higher than the corresponding techniques when ABC based optimization strategy is integrated into the framework. The average FPR achieved over 1000 frames using CNN is 0.15305, ABC with CNN is 0.13866, FFBPNN-L is 0.14415 ABC with FFBPNN-L is 0.13176. This means that ABC with FFBPNN-L resulted in lower FPR as compared to ABC with CNN using the 80:20 distribution of the dataset.

Table 7.19 Average FPR values

Parameter	Algorithm			
	CNN	ABC with CNN	FFBPNN-L	ABC with FFBPNN-L
Average FPR	0.15305	0.13866	0.14415	0.13176

7.2.5 Kappa Coefficient Analysis based on CNN

The Kappa coefficient observed for different frames extracted from the video using CNN with and without ABC is given in Table 7.20. The number of frames selected for the analysis varied from 500 to 1000. It is observed that the Kappa analysis performed using 500 frames results in 0.84232 Kappa using CNN and 0.85554 using ABC with CNN. When 700 frames are used for the analysis, the Kappa increases to as 0.86356 using CNN and 0.86931 using ABC with CNN. When the number of frames is increased to 1000, the Kappa also increases and reaches 0.91738 using CNN and 0.93403 using ABC with CNN.

Table 7.20 Kappa analysis values of CNN and ABC with CNN

Number of Frames	CNN	ABC with CNN
500	0.84232	0.85554
600	0.84449	0.85902
700	0.86356	0.86931
800	0.87866	0.88439
900	0.89547	0.90426
1000	0.91738	0.93403

The graphical analysis of the variation observed in the Kappa over the 1000 frames is shown in Figure 7.13. In the line graphs, Kappa analysis using ABC with CNN had shown higher Kappa in comparison to using CNN alone on every frame used in the analysis. This indicates that using optimized features in CNN improves the overall analysis using CNN. The overall analysis shows that the average Kappa for 1000 frames is observed to be 0.87362 using CNN and 0.88442 using ABC with CNN.

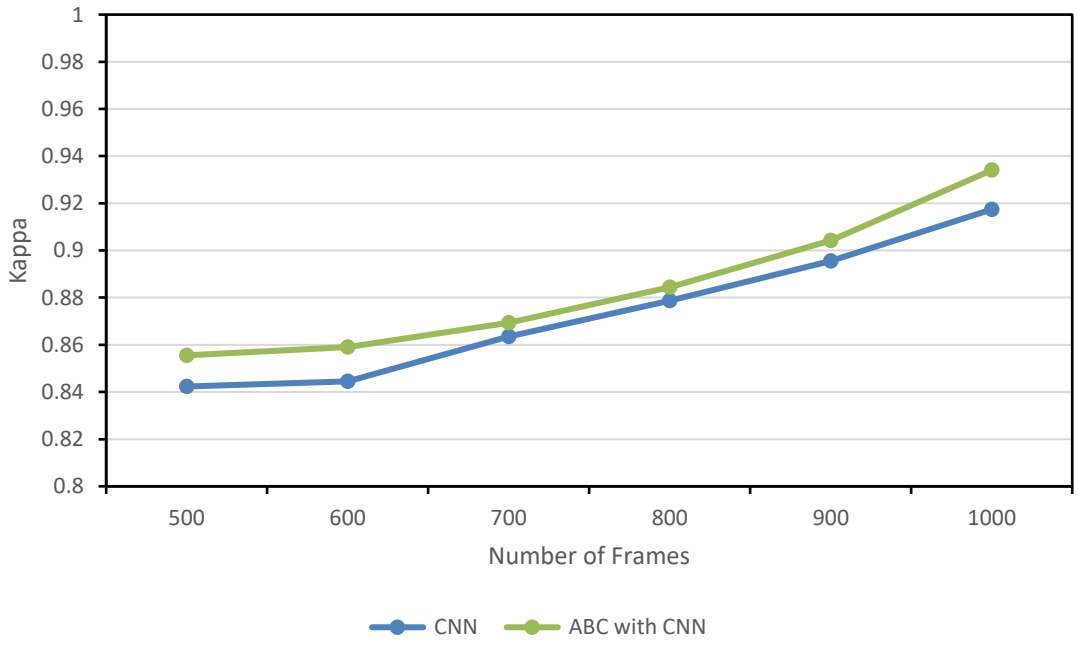


Figure 7.13 Kappa analysis of CNN and ABC with CNN

7.2.6 Kappa Coefficient Analysis based FFBPNN-L

The Kappa analysis for the 1000 frames is tabulated using FFBPNN-L, and ABC optimized FFBPNN-L is given in Table 7.21. Initially, when the number of frames is set to 500 frames, the observed Kappa is 0.88110 using FFBPNN-L and 0.89432 using ABC with FFBPNN-L. When 800 frames are used for the analysis, the Kappa increases to as 0.90596 using FFBPNN-L and 0.92147 using ABC with FFBPNN-L. When the number of frames is increased to 1000, the Kappa also increases and reaches 0.94512 using FFBPNN-L and 0.96668 using ABC with FFBPNN-L.

Table 7.21 Kappa analysis values of FFBPNN-L and ABC with FFBPNN-L

Number of Frames	FFBPNN-L	ABC with FFBPNN-L
500	0.88110	0.89432
600	0.88583	0.89829
700	0.90327	0.92081
800	0.90596	0.92147
900	0.92266	0.94620
1000	0.94512	0.96668

The graph shown in Figure 7.14 represents the Kappa analysis of FFBPNN-L and ABC with FFBPNN-L. It is observed that the Kappa represented by the curved line shows a steady improvement. The overall Kappa using FFBPNN-L is 0.90732 and 0.92463 using ABC with FFBPNN-L.

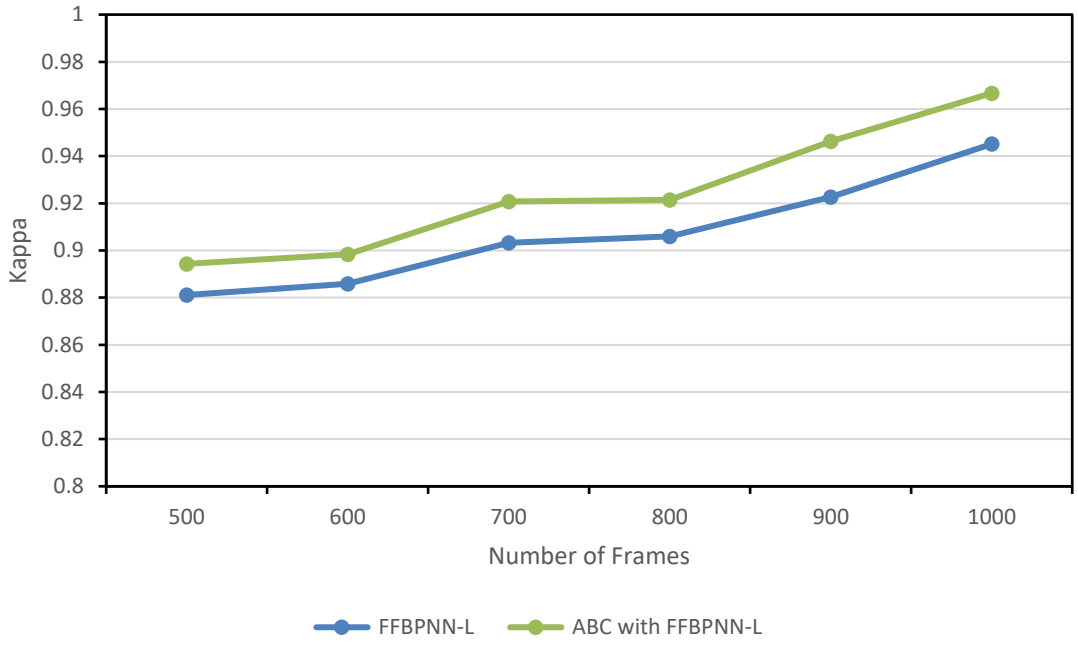


Figure 7.14 Kappa analysis of FFBPNN-L and ABC with FFBPNN-L

The average Kappa analysis of the two neural architectures with and without the involvement of ABC based optimization is summarized in Table 7.22. It is observed that optimization had improved the Kappa values observed with the implementation of the neural architectures.

Table 7.22 Average Kappa analysis values

Parameter	Algorithm			
	CNN	ABC with CNN	FFBPNN-L	ABC with FFBPNN-L
Average Kappa	0.87365	0.88442	0.90732	0.92463

The average Kappa demonstrated over 1000 frames using CNN is 0.87365, ABC with CNN is 0.88442, FFBPNN-L 0.90732 and ABC with FFBPNN-L is 0.92463. This means that ABC with FFBPNN-L resulted in 4.34825% better Kappa as compared to ABC with CNN using the 80:20 distribution of the dataset.

7.2.7 Accuracy Analysis based on CNN

Accuracy analysis performed using CNN based architecture is summarized in Table 7.23. The number of frames extracted from the video used in the analysis varies from 500 to 1000. The corresponding accuracy values achieved using CNN and ABC with CNN are tabulated for each variation in the number of frames using the 80:20 dataset.

Table 7.23 Accuracy analysis values of CNN and ABC with CNN

Number of Frames	CNN	ABC with CNN
500	0.84551	0.86175
600	0.85203	0.86412
700	0.86841	0.88328
800	0.88550	0.90073
900	0.89099	0.92029
1000	0.90505	0.94592

The accuracy analysis performed for 500 frames is 0.84551 using CNN and 0.86175 using ABC with CNN. When 700 frames are used for the analysis, the accuracy increases to as 0.86841 using CNN and 0.88328 using ABC with CNN. When the number of frames is increased to 1000, the accuracy shows a steady improvement to 0.90505 using CNN and 0.94592 using ABC with CNN. The graphical analysis of the observed changes in the accuracy values using CNN architecture is given in Figure 7.15.

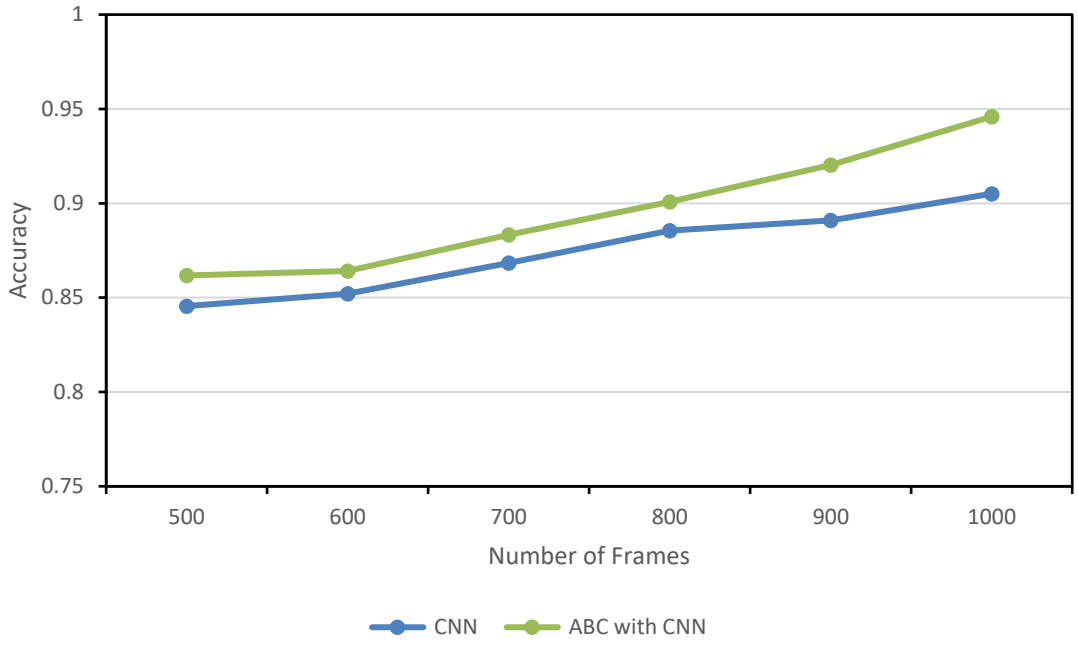


Figure 7.15 Accuracy analysis of CNN and ABC with CNN

The graph shows rise in the accuracy with an increase in the number of frames used in the analysis. Further, it reflects an average accuracy of 0.87458 using CNN and 0.89610 using ABC with CNN.

7.2.8 Accuracy Analysis based on FFBPNN-L

The accuracy analysis using FFBPNN-L with and without ABC based optimization is summarised in Table 7.24. The variation used in the analysis ranges from 500 to 1000 frames extracted from the video. The accuracy values observed for 500 frames using FFBPNN-L are 0.84205 and 0.87971 using ABC with FFBPNN-L. When 800 frames are used for the analysis, the accuracy increases to 0.87276 using FFBPNN-L and 0.93205 using ABC with FFBPNN-L. When the number of frames is increased to 1000, the accuracy shows a steady improvement to 0.90439 using FFBPNN-L and 0.99786 using ABC with FFBPNN-L.

Table 7.24 Accuracy analysis values of FFBPNN-L and ABC with FFBPNN-L

Number of Frames	FFBPNN-L	ABC with FFBPNN-L
500	0.84205	0.87971
600	0.85006	0.89475
700	0.86423	0.91975
800	0.87276	0.93208
900	0.89314	0.96306
1000	0.90439	0.99786

The graphical representation of Table 7.24 is given in Figure 7.16 for the accuracy analysis using FFBPNN-L and ABC with FFBPNN-L. Initially, when frames are increased from 500 to 600, there is a steady improvement in the accuracy values. However, a further increase in the number of frames significantly increased the accuracy of both scenarios.

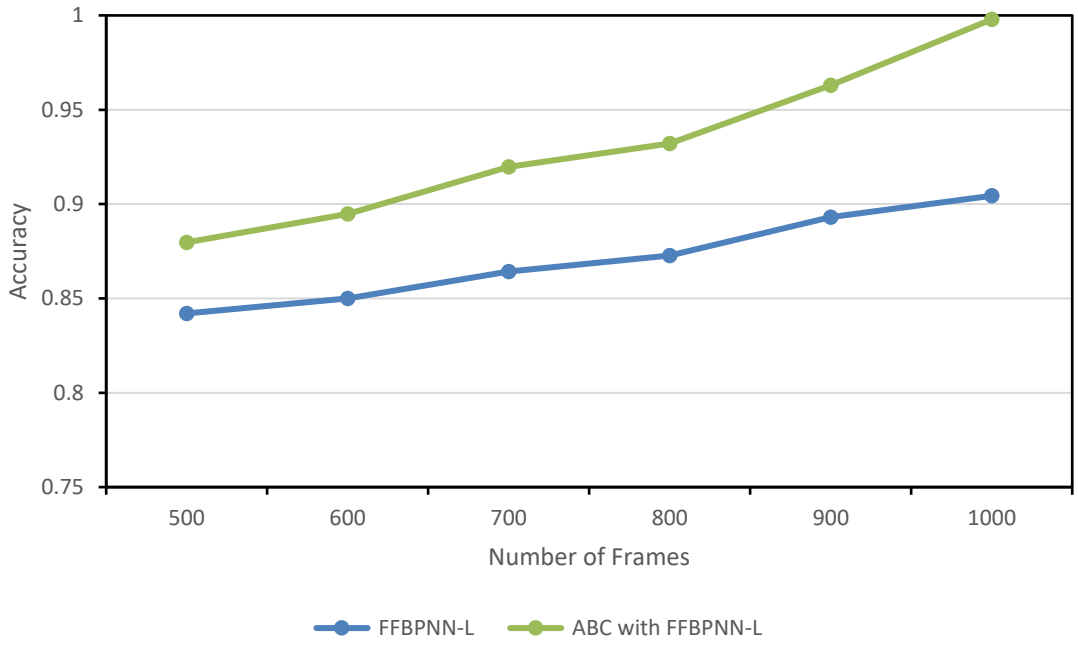


Figure 7.16 Accuracy analysis of FFBPNN-L and ABC with FFBPNN-L

The average accuracy analysis using both the neural network-based architectures is summarized in Table 7.25. The average accuracy demonstrated over 1000 frames using

CNN is 0.87458, ABC with CNN is 0.89602, FFBPNN-L is 0.87110 and ABC with FFBPNN-L is 0.93120. This means that ABC with FFBPNN-L resulted in 3.77832% better accuracy as compared to ABC with CNN using the 80:20 distribution of the dataset.

Table 7.25 Average Accuracy values

Parameter	Algorithm			
	CNN	ABC with CNN	FFBPNN-L	ABC with FFBPNN-L
Average Accuracy	0.87458	0.89602	0.87110	0.93120

7.3 Analysis using 90:10 distribution

In this part, the dataset is divided into a 90:10 ratio where 90% of the dataset is used for training, and 10% of the dataset is used for testing. As the training data increases, the training algorithm becomes more precise as it gets more relevant data to get trained. In real life scenario as the data fluctuation rate is high, 70:30 and 80:20 distribution is more suitable for 90:10 distribution which will provide more classification accuracy. Same as section 7.1 and section 7.2, all the algorithms will be evaluated in accordance with performance parameters and the results for evaluation are as follows.

7.3.1 True Positive Rate Analysis based on CNN

The frames from the video are extracted and analysed using CNN architecture with and without the involvement of an optimization algorithm. It has been observed that when only 500 frames are used in the analysis, the TPR using CNN alone is 0.79837 and using ABC with CNN is 0.81158. Further, when the number of frames is increased to 1000, a corresponding rise in the TPR is observed as 0.83550 using CNN alone and 0.87363 using ABC with CNN. The variation in the TPR values as the number of frames is increased from 500 to 1000 is summarized in Table 7.26.

Table 7.26 TPR analysis values of CNN and ABC with CNN

Number of Frames	CNN	ABC with CNN
500	0.79837	0.81158
600	0.79919	0.81685
700	0.81577	0.83507
800	0.82124	0.85050
900	0.82814	0.85999
1000	0.83550	0.87363

The graphical analysis of the variation in the TPR based on CNN architecture is shown in Figure 7.17. Overall, it is observed that with each variation in terms of the number of frames, the TPR remained relatively higher when optimized features were used for the training of CNN. Hence, the average TPR using CNN is observed to be 0.81637 and using ABC with CNN is 0.84127.

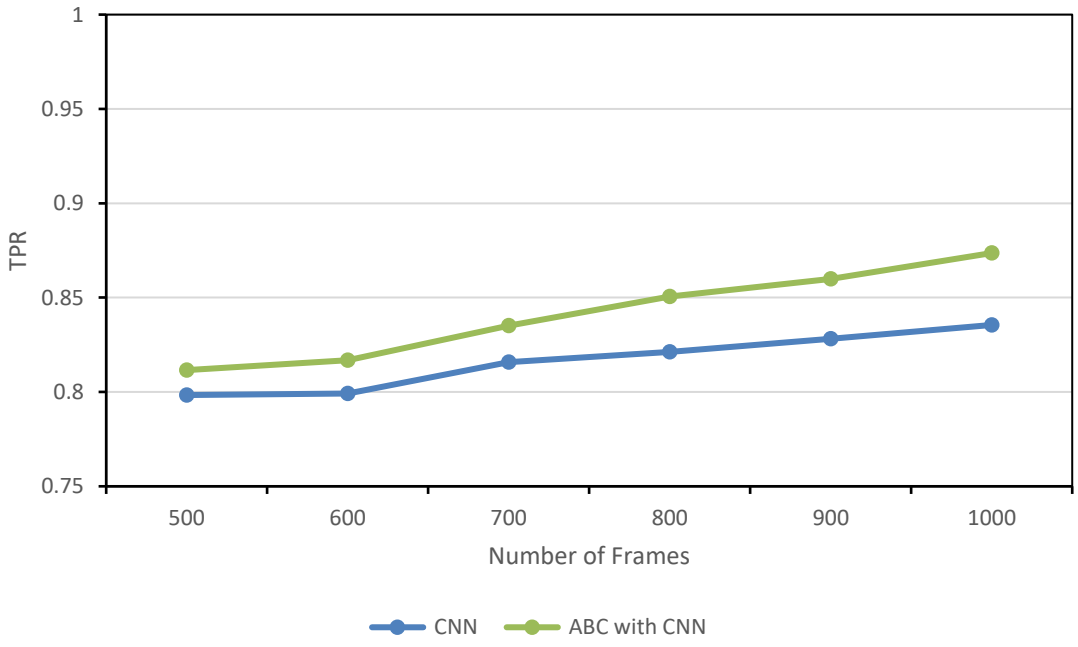


Figure 7.17 TPR analysis of CNN and ABC with CNN

7.3.2 True Positive Rate Analysis based on FFBPNN-L

The next neural architecture used in the analysis is FFBPNN-L. The TPR analysis performed using FFBPNN-L with and without integration of ABC based optimizations is summarized in Table 7.27.

Table 7.27 TPR analysis values of FFBPNN-L and ABC with FFBPNN-L

Number of Frames	FFBPNN-L	ABC with FFBPNN-L
500	0.90653	0.91975
600	0.90910	0.92468
700	0.91934	0.94503
800	0.93064	0.96384
900	0.94770	0.97854
1000	0.96762	0.99082

It is seen that when 500 frames are used for the analysis, the TPR of 0.90653 is achieved using only FFBPNN-L only and the TPR of 0.91975 when ABC is integrated prior to FFBPNN-L. When the frames are increased to 800, the TPR using FFBPNN-L increased to 0.93064 and, using ABC with FFBPNN-L to 0.96384. When the number of frames is further increased to 1000, the TPR again show a slight rise of 0.96762 using FFBPNN-L and 0.99082 using ABC with FFBPNN-L. The graphical analysis of these variations is conducted using the line graph shown in Figure 7.18.

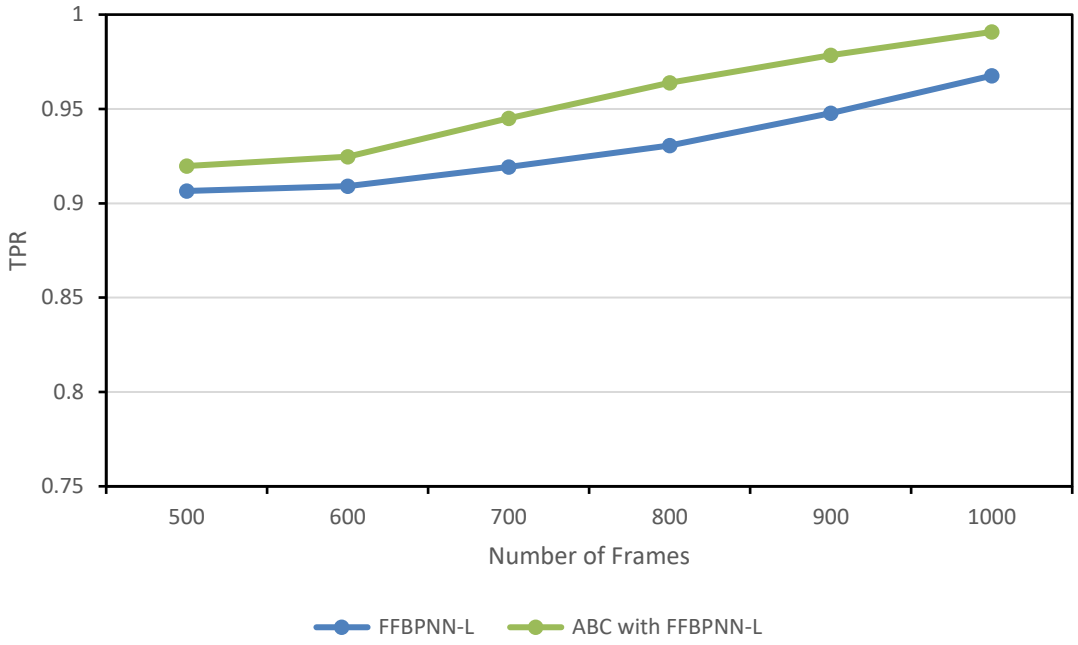


Figure 7.18 TPR analysis of FFBPNN-L and ABC with FFBPNN-L

The average values using various scenarios used for TPR analysis using 90:10 distributions are given in Table 7.28. The average TPR using individual CNN and FFBPNN-L is found to be comparatively lower than the corresponding techniques when ABC based optimization strategy is integrated into the framework.

Table 7.28 Average TPR values

Parameter	Algorithm			
	CNN	ABC with CNN	FFBPNN-L	ABC with FFBPNN-L
Average TPR	0.81637	0.84127	0.93015	0.95378

The average TPR demonstrated in Table 7.28 over 1000 frames using CNN is 0.81637, ABC with CNN is 0.84127, FFBPNN-L is 0.93015, ABC with FFBPNN-L is 0.95378. This means that ABC with FFBPNN-L resulted in 11.79593% higher TPR as compared to ABC with CNN using the 90:10 distribution of the dataset.

7.3.3 False Positive Rate Analysis based on CNN

The FPR analysis performed using CNN based architecture is described using a variable number of frames in Table 7.29. It is observed that when 600 frames are used in the analysis, the observed FPR using CNN is 0.12878 and using ABC with CNN is 0.11276. When the number of frames is 800, the corresponding FPR using CNN is observed as 0.14503 and using ABC with CNN is 0.12621. Further, when the number of frames is increased to 900 and 1000, the FPR using CNN increases to 0.15842 and 0.14389, while using ABC with CNN, it is 0.16322 and 0.14919, respectively. This shows that lower FPR is observed when ABC based optimization approach is integrated into the CNN based architecture.

Table 7.29 FPR analysis values of CNN and ABC with CNN

Number of Frames	CNN	ABC with CNN
500	0.11651	0.10603
600	0.12878	0.11276
700	0.14036	0.11804
800	0.14503	0.12621
900	0.15842	0.14389
1000	0.16322	0.14919

The graphical representation of the variation in the FPR for 500 to 1000 frames using an 80:20 distribution is shown in Figure 7.19. It is generalized that with the increase in the number of frames, the FPR increases for both CNN as well as ABC with CNN. However, the FPR demonstrated using ABC with CNN is comparatively lower as compared to CNN alone. Hence, ABC significantly lowered the FPR and proved to be better than CNN alone.

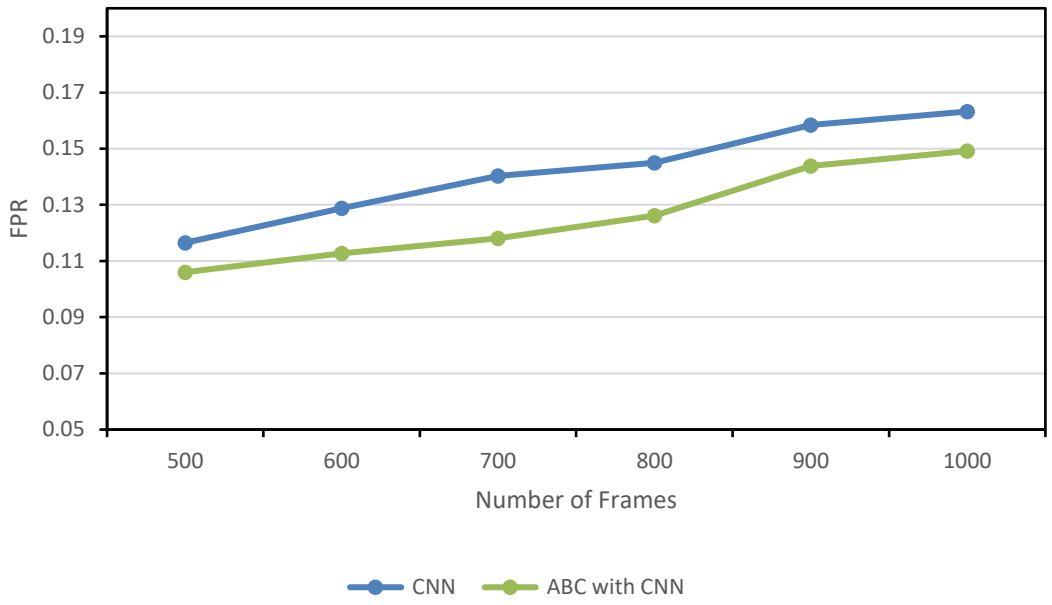


Figure 7.19 FPR analysis of CNN and ABC with CNN

7.3.4 False Positive Rate Analysis based on FFBPNN-L

Another technique used for the FPR analysis in the present work is FFBPNN-L. The variation in the FPR observed using FFBPNN-L based architecture is described in Table 7.30. It is observed that when 600 frames are used in the analysis, the observed FPR using FFBPNN-L is 0.12053 and using ABC with FFBPNN-L is 0.10985. When the number of frames is 800, the corresponding FPR using FFBPNN-L is observed as 0.13214 and using ABC with FFBPNN-L is 0.12237. Further, when the number of frames is increased to 900 and 1000, the FPR using FFBPNN-L increases to 0.14738 and 0.12855, while using ABC with FFBPNN-L, it is 0.15706 and 0.13388, respectively. This means that integration of ABC based optimization approach resulted in lower FPR associated during analysis.

Table 7.30 FPR analysis values of FFBPNN-L and ABC with FFBPNN-L

Number of Frames	FFBPNN-L	ABC with FFBPNN-L
500	0.11361	0.10036
600	0.12053	0.10985
700	0.12670	0.11763
800	0.13214	0.12238
900	0.14739	0.12855
1000	0.15707	0.13389

The graphical representation of the variation in the FPR for 500 to 1000 frames using a 90:10 distribution is shown in Figure 7.20. It is observed that with the increase in the number of frames, the FPR increases for both CNN as well as ABC with CNN. However, the FPR demonstrated using ABC with CNN is comparatively lower as compared to CNN alone. Hence, ABC significantly lowered the FPR and proved to be better than CNN alone.

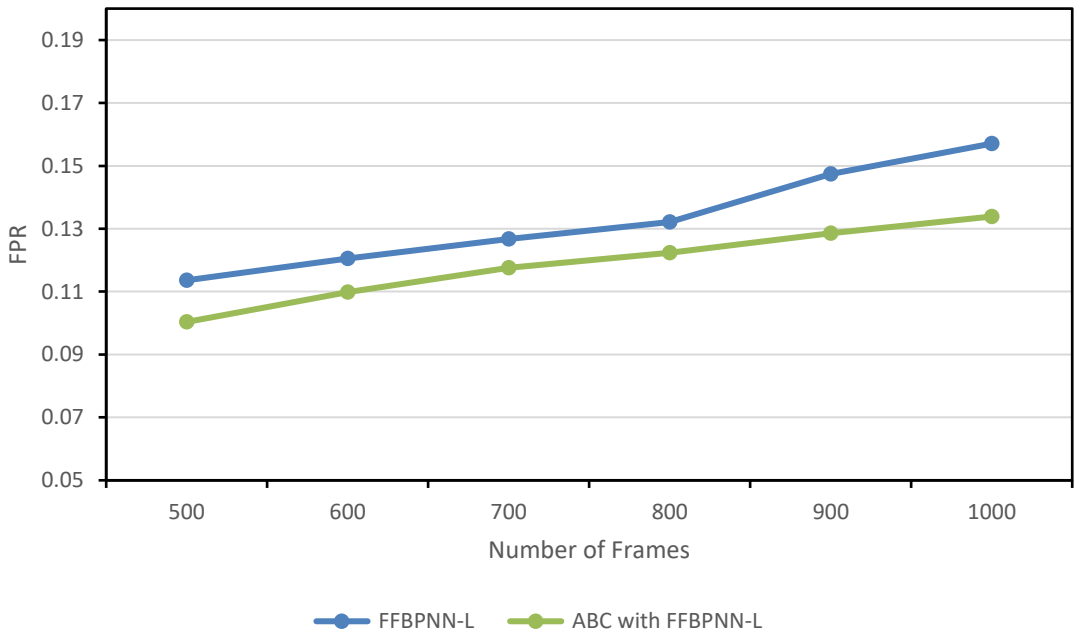


Figure 7.20 FPR analysis of FFBPNN-L and ABC with FFBPNN-L

The average FPR analysis using a 90:10 distribution is given in Table 7.31. It is observed that the average FPR using individual FFBPNN-L is found to be comparatively higher than the corresponding techniques when ABC based optimization strategy is integrated into the framework. The average FPR achieved over 1000 frames CNN is 0.14205, using ABC with CNN is 0.12602, FFBPNN-L is 0.12332 ABC with FFBPNN-L is 0.13290. This means that ABC with FFBPNN-L resulted in lower FPR as compared to ABC with CNN using the 90:10 distribution of the dataset.

Table 7.31 Average FPR values

Parameter	Algorithm			
	CNN	ABC with CNN	FFBPNN-L	ABC with FFBPNN-L
Average FPR	0.14206	0.12602	0.13290	0.13290

7.3.5 Kappa Coefficient Analysis based on CNN

The Kappa coefficient observed for different frames extracted from the video using CNN with and without ABC is given in Table 7.32. The number of frames selected for the analysis varied from 500 to 1000. It is observed that the Kappa analysis performed using 500 frames results in 0.85856 Kappa using CNN and 0.87178 using ABC with CNN. When 700 frames are used for the analysis, the Kappa increases to as 0.8767 using CNN and 0.89080 using ABC with CNN. When the number of frames is increased to 1000, the Kappa also increases and reaches 0.897421 using CNN and 0.93122 using ABC with CNN.

Table 7.32 Kappa analysis values of CNN and ABC with CNN

Number of Frames	CNN	ABC with CNN
500	0.85857	0.87179
600	0.86384	0.87429
700	0.87670	0.89080
800	0.87771	0.89611
900	0.88620	0.91181
1000	0.89742	0.93123

The graphical analysis of the variation observed in the Kappa over the 1000 frames is shown in Figure 7.21. In the line graphs, Kappa analysis using ABC with CNN had shown higher Kappa in comparison to using CNN only on every frame used in the analysis. This indicates that using optimized features in CNN improves the overall analysis using CNN. The overall analysis shows that the average Kappa for 1000 frames is observed to be 0.87674 using CNN and 0.89600 using ABC with CNN.

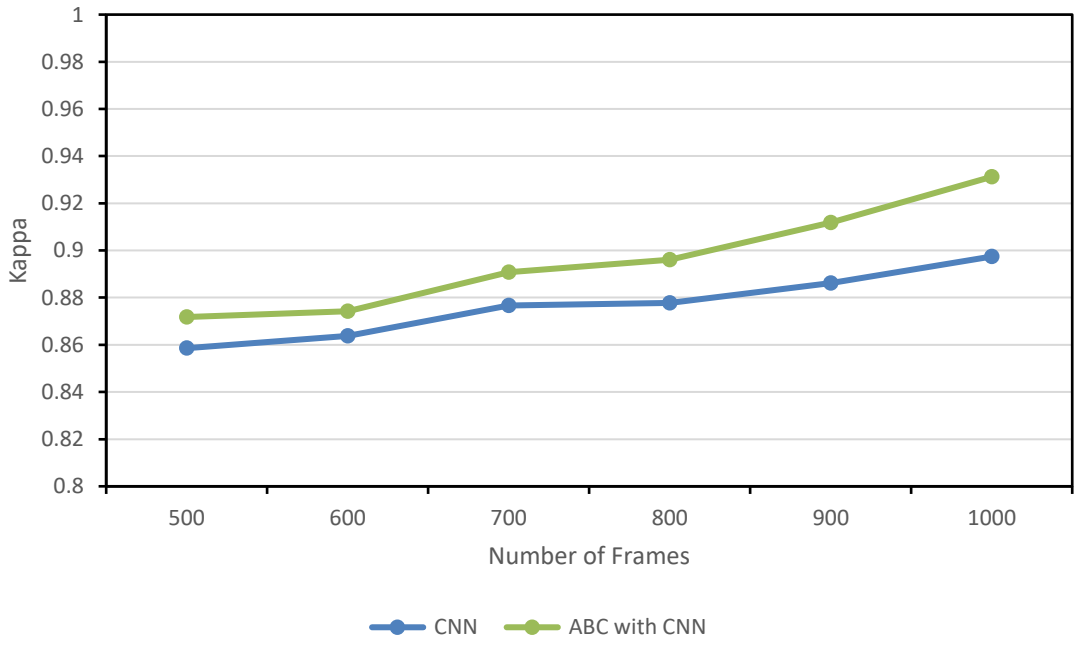


Figure 7.21 Kappa analysis of CNN and ABC with CNN

7.3.6 Kappa Coefficient Analysis based on FFBPNN-L

The Kappa analysis for the 1000 frames is tabulated using FFBPNN-L, and ABC optimized FFBPNN-L is given in Table 7.33.

Table 7.33 Kappa analysis values of FFBPNN-L and ABC with FFBPNN-L

Number of Frames	FFBPNN-L	ABC with FFBPNN-L
500	0.91360	0.92682
600	0.91762	0.93107
700	0.92731	0.95563
800	0.94459	0.96064
900	0.94903	0.97846
1000	0.97273	0.99789

Initially, when the number of frames is set to 600 frames, the observed Kappa is 0.91762 using FFBPNN-L and 0.93106 using ABC with FFBPNN-L. When 800 frames are used for the analysis, the Kappa increases to as 0.94459 using FFBPNN-L and 0.96063 using ABC with FFBPNN-L. When the number of frames is increased to 1000, the Kappa also increases and reaches 0.97273 using FFBPNN-L and 0.99789 using ABC with FFBPNN-L.

The graph shown in Figure 7.22 represents the Kappa analysis of FFBPNN-L and ABC with FFBPNN-L. It is observed that the Kappa represented by the curved line shows a steady improvement. The overall Kappa using FFBPNN-L is 0.93748 and 0.95841 using ABC with FFBPNN-L.

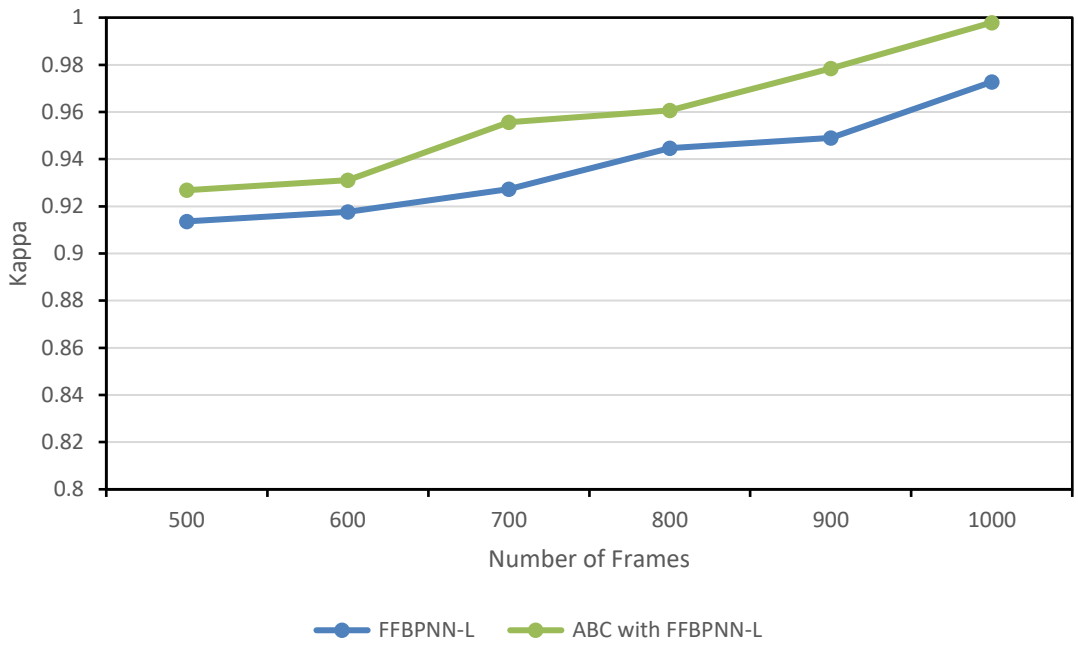


Figure 7.22 Kappa analysis of FFBPNN-L and ABC with FFBPNN-L

The average Kappa demonstrated shown in Table 7.34, over 1000 frames using CNN is 0.87674, ABC with CNN is 0.89601, FFBPNN-L is 0.93748 and ABC with FFBPNN-L is 0.95842. This means that ABC with FFBPNN-L resulted in 4.42406% better Kappa as compared to ABC with CNN using the 90:10 distribution of the dataset.

Table 7.34 Average Kappa analysis values

Parameter	Algorithm			
	CNN	ABC with CNN	FFBPNN-L	ABC with FFBPNN-L
Average Kappa	0.87674	0.89601	0.93748	0.95842

7.3.7 Accuracy Analysis based on CNN

Accuracy analysis performed using CNN based architecture is summarized in Table 7.35. The number of frames extracted from the video used in the analysis varies from 500 to 1000. The corresponding accuracy values achieved using CNN and ABC with CNN are tabulated for each variation in the number of frames.

Table 7.35 Accuracy analysis values of CNN and ABC with CNN

Number of Frames	CNN	ABC with CNN
500	0.85258	0.86882
600	0.85689	0.87399
700	0.87401	0.89874
800	0.88880	0.91401
900	0.90659	0.92963
1000	0.90614	0.93853

The accuracy analysis performed for 600 frames is 0.85688 using CNN and 0.87399 using ABC with CNN. When 800 frames are used for the analysis, the accuracy increases to as 0.88880 using CNN and 0.91400 using ABC with CNN. When the number of frames is increased to 1000, the accuracy shows a steady improvement to 0.90613 using CNN and 0.93853 using ABC with CNN.

The graphical analysis of the observed changes in the accuracy values using CNN architecture is given in Figure 7.23. The graph shows a random rise in the accuracy with an increase in the number of frames used in the analysis. Further, it reflects an average accuracy of 0.88083 using CNN and 0.90395 using ABC with CNN.

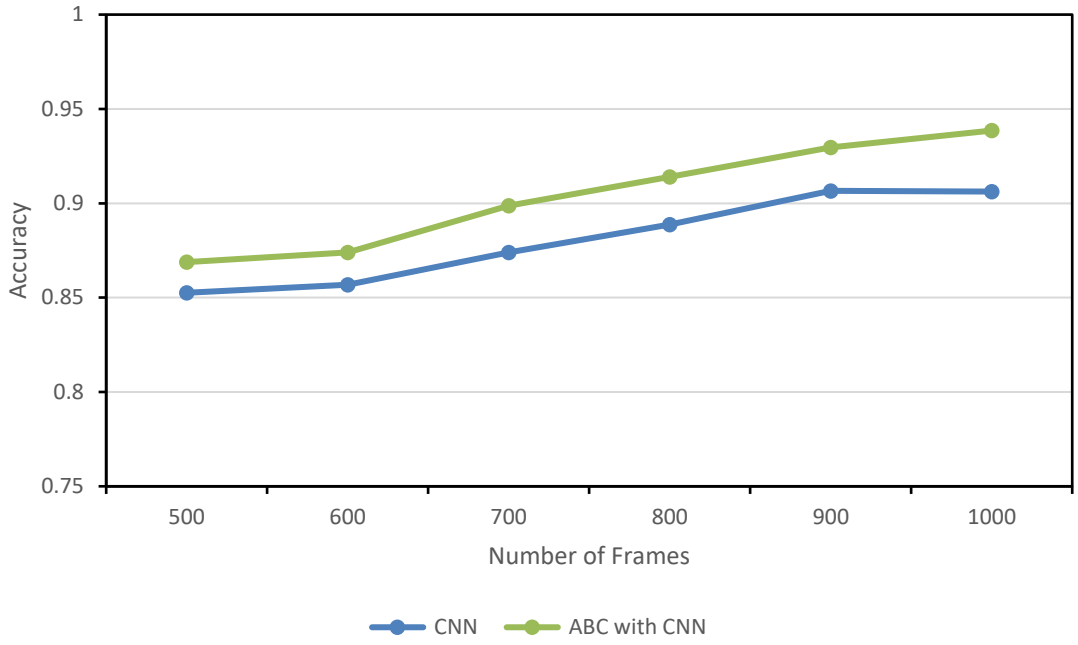


Figure 7.23 Accuracy analysis of CNN and ABC with CNN

7.3.8 Accuracy Analysis based on FFBPNN-L

The accuracy analysis using FFBPNN-L with and without ABC based optimization is summarised in Table 7.36. The variation used in the analysis ranges from 500 to 1000 frames extracted from the video. The accuracy values observed for 600 frames using FFBPNN-L are 0.92950 and 0.93654 using ABC with FFBPNN-L. When 800 frames are used for the analysis, the accuracy increases to as 0.94535 using FFBPNN-L and 0.97066 using ABC with FFBPNN-L. When the number of frames is increased to 1000, the accuracy shows a steady improvement to 0.96038 using FFBPNN-L and 0.99339 using ABC with FFBPNN-L.

Table 7.36 Accuracy analysis values of FFBPNN-L and ABC with FFBPNN-L

Number of Frames	FFBPNN-L	ABC with FFBPNN-L
500	0.92230	0.93552
600	0.92950	0.93655
700	0.94009	0.95789
800	0.94535	0.97066
900	0.95810	0.98870
1000	0.96039	0.99339

The graphical representation of Table 7.36 is given in Figure 7.24 for the accuracy analysis using FFBPNN-L and ABC with FFBPNN-L. Initially, when frames are increased from 500 to 600, there is a steady improvement in the accuracy values. However, a further increase in the number of frames significantly increased the accuracy of both scenarios.

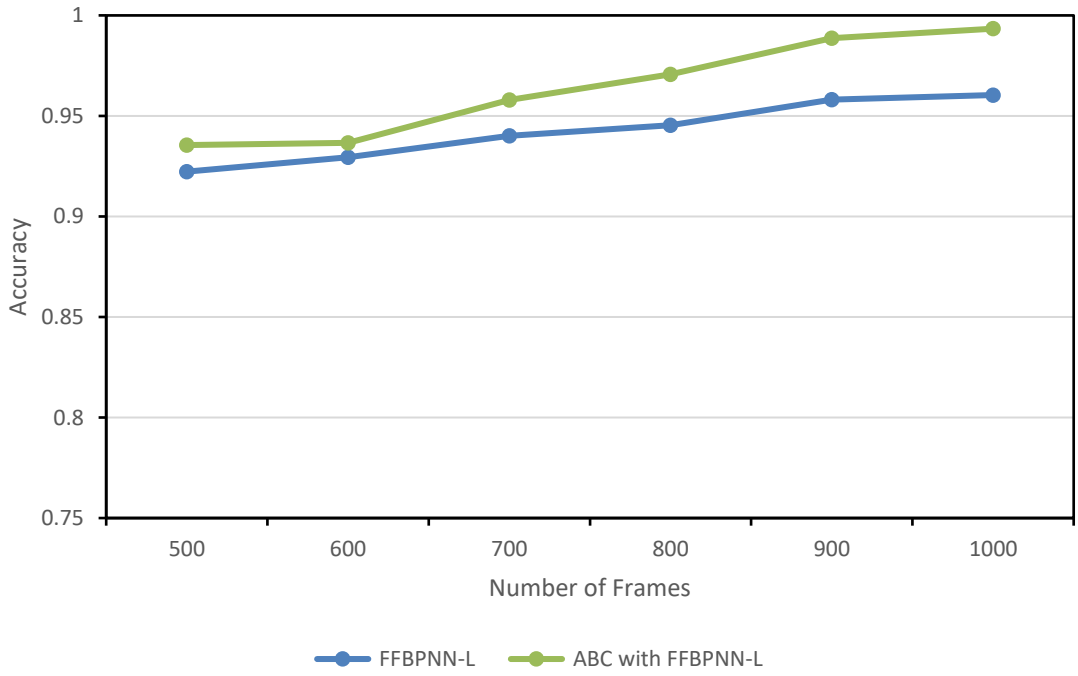


Figure 7.24 Accuracy analysis of FFBPNN-L and ABC with FFBPNN-L

The average accuracy analysis using both the neural network-based architectures is summarized in Table 7.37. The average accuracy demonstrated over 1000 frames using CNN is 0.88083, ABC with CNN is 0.90395, FFBPNN-L is 0.94262 and ABC with FFBPNN-L is 0.96378. This means that ABC with FFBPNN-L resulted in 6.20787% better accuracy as compared to ABC with CNN using 90:10 distribution of the dataset.

Table 7.37 Average Accuracy values

Parameter	Algorithm			
	CNN	ABC with CNN	FFBPNN-L	ABC with FFBPNN-L
Average Accuracy	0.88083	0.90395	0.94262	0.96378

7.4 Overall Findings

The below figures 7.25 to 7.28 show the average values of all the results (500 frames to 1000 frames) with ratios 70:30, 80:20 and 90:10 for TPR, FPR, kappa and accuracy using algorithms CNN, ABC with CNN,FFBPNN-L and ABC with FFBPNN-L.

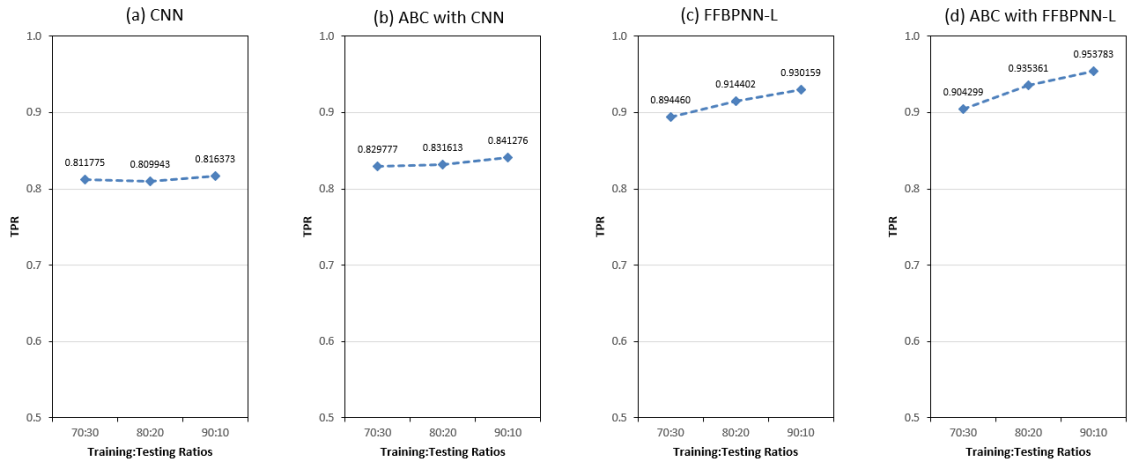


Figure 7.25 Plots show TPR values with 70:30, 80:20 and 90:10 ratios for (a) CNN, (b) ABC with CNN, (c) FFBPNN-L and (d) ABC with FFBPNN-L

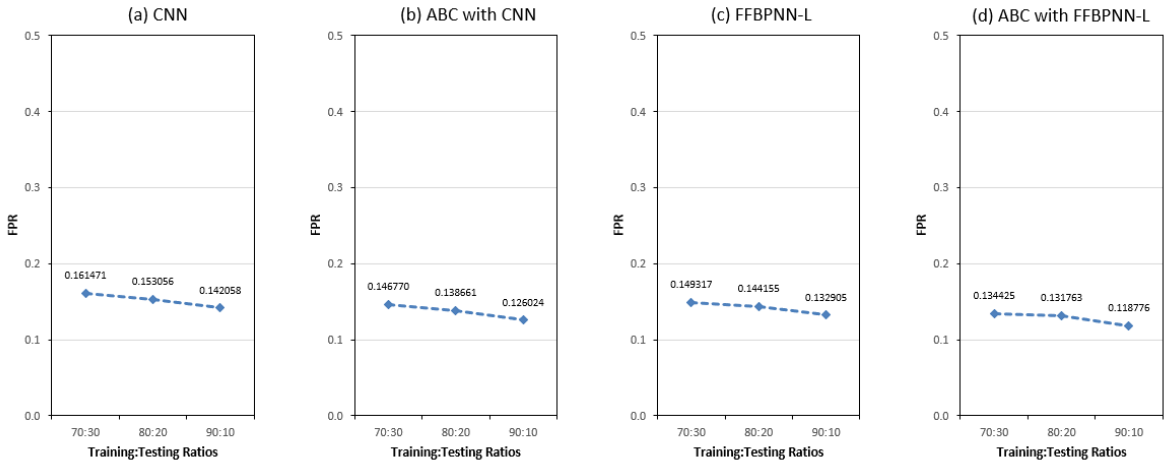


Figure 7.26 Plots show FPR values with 70:30, 80:20 and 90:10 ratios for (a) CNN, (b) ABC with CNN, (c) FFBPNN-L and (d) ABC with FFBPNN-L

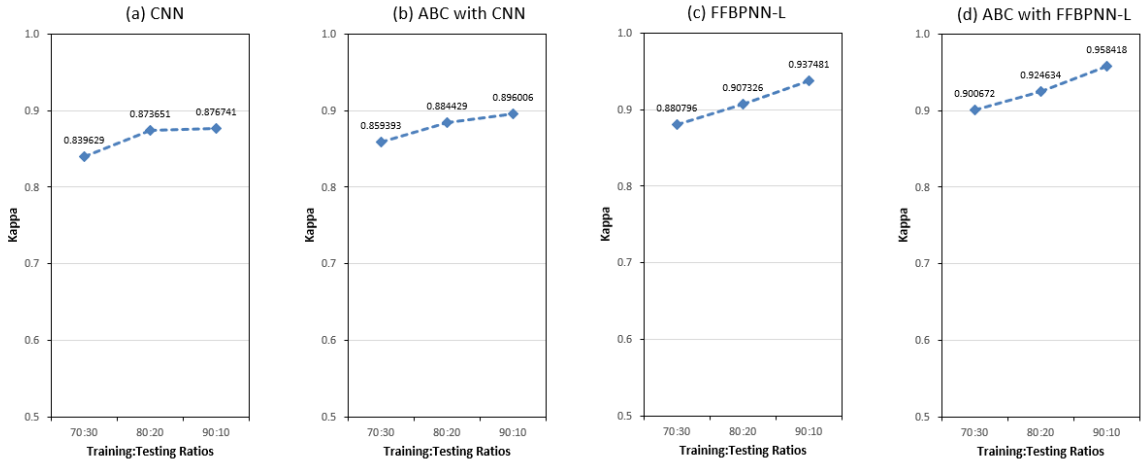


Figure 7.27 Plots show Kappa values with 70:30, 80:20 and 90:10 ratios for (a) CNN, (b) ABC with CNN, (c) FFBPNN-L and (d) ABC with FFBPNN-L

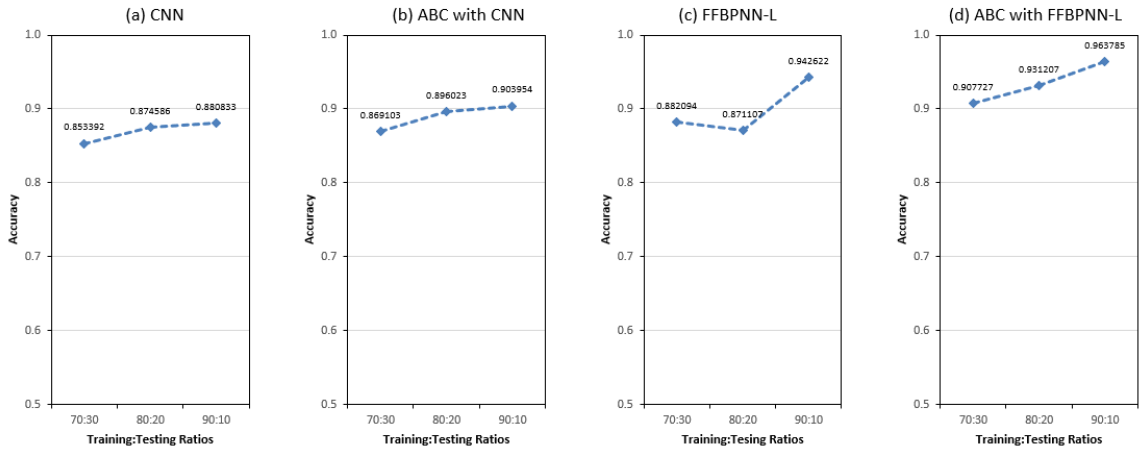


Figure 7.28 Plots show Accuracy with 70:30, 80:20 and 90:10 ratios for (a) CNN, (b) ABC with CNN, (c) FFBPNN-L and (d) ABC with FFBPNN-L

The plots in Figure 7.25 show that the results with a 90:10 ratio are better in all cases of the algorithm, but ABC with FFBPNN-L gives the TPR value 0.953783 which shows that the algorithm FFBPNN with ABC gives the better results with a ratio of 90:10.

A similar way Figure 7.26 shows the FPR values, which represent the FPR value 0.118776, which is better than other algorithms, and the 90:10 ratio in FFBPNN-ABC shows improved results. As the false positive rate represents the proportion of negative instances that are incorrectly classified as positive.

Figure 7.27 shows the value of Kappa in the case of CNN and ABC. CNN is decreased at a 90:10 ratio, but FFBPNN shows better results with a 90:10 instead of an 80:20 ratio.

Accuracy results shown in Figure 7.28, FFBPNN-L shows better results with 90:10.
The importance of this 90:10 split in this research is to train a CNN effectively with limited data. The use of original data, which highlights the significance of the study's conclusions, is its key component. In contrast to secondary data (online available data), primary data (data collected for research) is taken straight out of real, actual situations.

The findings from this chapter shows the best result with the 90:10 distribution of data with optimization.

Table 7.38 shows the comparison of 80:20 and 90:10 ratio results. The proposed algorithm FFBPNN-L with ABC compared to CNN with ABC shows that;

- TPR has improved by 11.79%,
- FPR has improved by 5.17%,
- Kappa has improved by 4.42%
- Accuracy has improved by 6.20%.

Table 7.38 Comparison of average values of 80:20 ratio with 90:10 ratio

Average	**80:20 ratio**				**90:10 ratio**			
	using CNN	using ABC with CNN	using FFBPNN-L	using ABC with FFBPNN-L	using CNN	using ABC with CNN	using FFBPNN-L	using ABC with FFBPNN-L
TPR	0.80994	0.83161	0.91440	0.93536	0.81637	0.84128	0.93016	0.95378
FPR	0.15306	0.13866	0.14416	0.13176	0.14206	0.12602	0.13290	0.13290
Accuracy	0.87459	0.89602	0.87111	0.93121	0.88083	0.90395	0.94262	0.96378

7.5 Summary and Contribution of the Chapter

The chapter provides detailed simulation analysis to support the proposed multiclass classification model based on CNN and FFBPNN-L architecture. The performance of the designed model is supported by a two-fold comprehensive analysis conducted by varying the training to testing dataset size as well as number of hidden layers. The metrics used for the comparison were TPR, FPR, Kappa and accuracy of prediction and classification of HR which are computed based on the mathematical relation between the confusion matrix parameters. The analysis shows that when more data is used for the training of the system it yields best prediction outcomes. This is due to the fact that increased amount of data will be available for the system to learn from to refine it prediction capabilities.

Chapter 8
Conclusion and Future Work

The fields of computer vision and computational photography have made it possible to measure physiological and other vital signs using cameras. This effort brings together knowledge from these fields and signals processing, machine learning, biomedical engineering, optics, and medicine to develop technologies that allow the environment's scalable and accessible physiological monitoring and monitoring. However, these changes in the colour contrast cannot be viewed directly from the naked eye, and hence the extracted portion needs to be magnified to detect heart rate or any ground truth for the manner. In the case of the proposed work, it has been identified from the studied literature that Euler video magnification is one of the best suitable techniques for video magnification. Due to the rising number of deaths caused by heart attacks throughout the globe, it has been the entire duty of academics and industry to do everything possible to prevent these tragedies from occurring. During a change in heart rate that might also result in a condition that could be unpleasant for the user, all of the vital indicators for the detection could not be seen with the naked eye, making it challenging to identify the change.

The proposed work has utilised upper body video covering the face as the primary object, extracted forehead once the video is magnified and post magnification, Artificial Bee Colony, which is Swarm Inspired algorithm, has been applied to select the most relevant pixel distribution that is significant to a specific supplied heart rate. The heart rate is divided into three following categories in which the rate varies from 60 beats per minute to 120 beats per minute. The sole objective is to optimize the training attained by the Levenberg support Feed-Forward Back Propagation mechanism with a varying number of hidden layers. As described in the introduction Convolution Neural Network was expected to be a primary classifier. Still, due to the global pandemic situation, the required number of videos to be collected using the university's research facility was not collected. Hence to comprehend convolution networks, a propagation-based neural network is applied that also requires feature extraction. To support the training, neural networks have been supplied with histogram-oriented features. A varying validation of 70:30, 80:20 and 90:10 for training and test data is being utilised from the collected

frames that have been extracted from videos and are also shown in the proposed work sections of the thesis.

Many hidden layers are considered and examined in terms of the algorithm's performance in line with performance metrics such as True Positive Rate, False Positive Rate, Kappa analysis, and Accuracy analysis. Models such as the Feed Forward Back Propagation Neural Network-Levenberg (FFBPNN-L) algorithm, the Convolutional Neural Network (CNN) algorithm, and the Artificial Bee Colony (ABC) algorithm with CNN are compared to other algorithms, such as the FFBPNN-L algorithm. In addition, several configurations of concealed layers were selected for the investigation. Comparing the suggested method to all the other algorithms, the findings demonstrate that the Artificial Bee Colony algorithm using Feed Forward Back Propagation Neural Network algorithm with a 20-layer architecture is very efficient.

8.1 Performance Metric Analysis based on CNN

The average outcomes of the performance analysis conducted based on CNN architecture with and without ABC, and three distributions of data using 20 layers are summarized in Table 8.1.

Table 8.1 Average Performance based on CNN architecture.

Training to Testing Dataset Variations	Average TPR		Average FPR		Average Kappa		Average Accuracy	
	CNN	ABC with CNN	CNN	ABC with CNN	CNN	ABC with CNN	CNN	ABC with CNN
Using 70:30 distribution	0.811	0.829	0.161	0.146	0.839	0.859392	0.853	0.869
Using 80:20 distribution	0.809	0.831	0.153	0.138	0.873	0.884	0.874	0.896
Using 90:10 distribution	0.816	0.841	0.142	0.126	0.876	0.896	0.880	0.903

8.2 Performance Metrics based on FFBPNN-L with ABC

The average outcomes of the performance analysis conducted based on FFBPNN-L architecture and three distributions of data using 20 layers are summarized in Table 8-2.

Table 8.2 Average Performance based on FFBPNN-L architecture

Training to Testing Dataset Variations	Average TPR		Average FPR		Average Kappa		Average Accuracy	
	FFBPNN-L	ABC with FFBPNN-L	FFBPNN-L	ABC with FFBPNN-L	FFBPNN-L	ABC with FFBPNN-L	FFBPNN-L	ABC with FFBPNN-L
Using 70:30 distribution	0.894	0.904	0.149	0.134	0.880	0.900	0.882	0.0907
Using 80:20 distribution	0.914	0.935	0.144	0.131	0.907	0.924	0.871	0.931
Using 90:10 distribution	0.930	0.953	0.133	0.132	0.937	0.937	0.943	0.963

Overall, the results compared with the proposed algorithm FFBPNN-L with ABC compared to CNN with ABC that TPR has been improved by 8.24%, Kappa coefficient has been improved by 4.58%. The limited dataset shows better results for FFBPNN-L, such that ABC with FFBPNN-L resulted in 4.25% better accuracy than ABC with CNN. Thus, FFBPNN-L with the integration of ABC shows improved results than when ABC is integrated with CNN. CNN architecture is a self-learning neural network that usually learns from a large dataset. This outcome with better performance of FFBPNN over CNN is mainly due to the smaller dataset size available for the research.

8.3 Contribution of the Research

This research work designed a Swarm Intelligence based improved algorithm that uses ABC with a novel fitness function. Validations over the proposed algorithm have been made based on quantitative parameter analysis. In addition, several configurations of concealed layers were selected for the investigation. Comparing the suggested method to all the other algorithms, the findings demonstrate that the Artificial Bee Colony algorithm using Feed Forward Back Propagation Neural Network algorithm with a 20-layer architecture is very efficient.

8.4 Limitations

- In the proposed work, neural network based machine learning demonstrated best performance. However, the unavailability of large data set for the training purposed limits its functionality.
- At present the number of neural layers is limited upto 20 and hyperparameter tuning is missing at present that can further refine the overall prediction analysis.
- The camera specifications, out of plane recording may also limit the accuracy of HR prediction.
- The illumination levels may also govern the colour intensity over the skin surface and leads to non-reliable results in such particular conditions.

8.5 Future Scope

The work provided here may be used with affective computing research to determine a person's heart rate without the need for the individual to wear a heart rate monitor on their person. It also has the potential to be used in hospitals to forecast heart rate as patients walk into the facility, without the need for the patient to remain in a fixed posture. This development will also detect heart rate when a person is doing an activity in the gym without scaring the subject by maintaining a close distance from them. Future research expansions could be devoted to predicting pulse rate in environments where subjects are required to walk longer distances, where gender difference is

considered an experimental design factor, and where activities other than motions performed before the subjects' laptops/desktop computers are devised as part of the experimental design, among other things.

Appendix A: Approval Forms for Data Collection

Section A–1: Completed Auckland University of Technology Ethics Committee Application Form (EA1)

Section A–2: Participant Information Sheet

Section A–3: Consent Form

Section A–4: Advertising Poster

Section A–5: Acceptance Letter Granting Ethical Approval

Section A–1: Completed Auckland University of Technology Ethics Committee Application Form (EA1)

EA1

APPLICATION FOR ETHICS APPROVAL BY AUTEC

Please print this application single sided in greyscale and do not staple. Once this application has been completed and signed, please read the notes at the end of the form for information about submission of the application for review.

NOTES ABOUT COMPLETION

- Ethics review is a community review of the ethical aspects of a research proposal. Responses should use clear everyday language with appropriate definitions being provided should the use of technical or academic jargon be necessary.
- The AUTEC Secretariat and your AUTEC Faculty Representative are able to provide you with assistance and guidance with the completion of this application which may help expedite the granting of ethics approval.
- The information in this application needs to be clearly stated and to contain sufficient details to enable AUTEC to make an informed decision about the ethical quality of the research. Responses that do not provide sufficient information may delay approval because further information will be sought. Overly long responses may also delay approval when unnecessary information hinders clarity.
- AUTEC reserves the right not to consider applications that are incomplete or inadequate. Please do not alter the formatting or numbering of the form in any way or remove any of the help text.
- Comprehensive information about ethics approval and what may be required is available online at http://aut.ac.nz/researchethics
- The information provided in this application will be used for the purposes of granting ethics approval. It may also be provided to the Graduate Research School, the Research and Innovation Office, or the University's insurers for purposes relating to AUT's interests.
- The form is focused around AUTEC's ethical principles, which are in accordance with the *Guidelines for the approval of ethics committees* in New Zealand.

To respond to a question, please place your cursor in the space following the question and its notes and begin typing.

A. *Project Information*

A.1. What is the title of the research?

If you will be using a different title in documents to that being used as your working title, please provide both, clearly indicating which title will be used for what purpose.

Non-Contact Measurement of Heart Rate using Video Recordings

A.2. Is this application for research that is being undertaken in stages? ☐ Yes ☒ No

If the answer is 'Yes' please answer A.2.1 and the following sections, otherwise please answer A.3 and continue from there.

A.2.1. Does this application cover all the stages of the research? ☐ Yes ☐ No

If the answer is 'No' please provide details here of which stages are being covered by this application, otherwise please answer A.3 and continue from there.

N/A

A.3. Who is the applicant?

When the research is part of the requirements for a qualification at AUT, then the applicant is always the primary supervisor. Otherwise, the applicant is the researcher primarily responsible for the research, to whom all enquiries and correspondence relating to this application will be addressed.

Dr Jeff Kilby

A.4. Further information about the applicant.

A.4.1. In which faculty, directorate, or research centre is the applicant located?

Faculty of Design and Creative Technologies – School of Engineering, Computing and Mathematical Sciences

A.4.2. What are the applicant's qualifications?

Doctor of Philosophy (PhD)

A.4.3. What is the applicant's email address?

An email address at which the applicant can be contacted is essential.

jeffrey.kilby@aut.ac.nz

A.4.4. At which telephone numbers can the applicant be contacted during the day?

09 921 9999 ext 8748 Mobile: 021 1764 933

A.5. Research Instruments

A.5.1. Which of the following does the research use:

☐ *a written or electronic questionnaire or survey* ☐ *focus groups* ☐ *interviews*
☐ *observation* ☐ *participant observation* ☐ *ethnography* ☐ *photographs*
☒ *videos* ☐ *other visual recordings* ☐ *a creative, artistic, or design process*
☐ *performance tests*
☒ *some other research instrument (please specify)*

A Canon EOS 850D DSLR Camera and a Rossmax SB100 Finger Pulse Oximeter.

Please attach to this application form all the relevant research protocols. These may include: Indicative questions (for interviews or focus groups); a copy of the finalised questionnaire or survey in the format that it will be presented to participants (for a written or electronic questionnaire or survey); a protocol indicating how the data will be recorded (e.g. audiotape, videotape, note-taking) for focus groups or interviews (Note: when focus groups are being recorded, you will need to make sure there is provision for explicit consent on the Consent Form and attach to this Application Form examples of indicative questions or the full focus group schedule. Please note that there are specific confidentiality issues associated with focus groups that need to be addressed); a copy of the observation protocol that will be used (for observations); full information about the use of visual recordings of any sort, including appropriate protocols and consent processes; protocols for any creative, artistic, or design process; a copy of the protocols for the instruments and the instruments that will be used to record results if you will use some other research instrument.

A.5.2. Who will be transcribing or recording the data?

If someone other than the applicant or primary researcher will be transcribing the interview or focus group records or taking the notes, you will need to provide a confidentiality agreement with this Application Form.

Gaganjot Kaur

A.6. Please provide a brief plain English summary of the research (300 words maximum).

Please provide a simple response to each of these three questions: What are you trying to find out? Who are you wanting to involve? and What would you like them to do for you?

This research aims to measure the heart rate without any physical contact with the participants; videos will be recorded using a camera placed 2m from the face of the participant. These recorded videos will be analysed off-line later to find the heart rate using advanced image processing techniques using MATLAB software.

A total of fifty healthy participants will volunteer to be part of this research aged between 18 and 35 years of age.

The participants will sit upright in a chair, where the camera records the face videos from 2-metres three videos for the period of 10 seconds, 20 seconds and 30 seconds.

The participants will be required to do the following tasks whilst remaining seated(a) keeping their head still, and (b) are free to move their head. The participants will be required to have a pulse oximeter placed on their index finger of their dominant hand to measure and record their pulse rate measurement while the video recording. This procedure will be performed three times for each participant.

A.7. Additional Research Information

A.7.1. Is this research an intervention study? ☒ Yes ☐ No

An Intervention Study is defined in NEAC's National Ethical Standards for Health and Disability Research and Quality Improvement. as "A study in which an investigator controls and studies an intervention(s) provided to participants for the purpose of adding to knowledge of the health effects of that intervention(s). The term 'intervention study' is often used interchangeably with 'experimental study'. Many intervention studies are clinical trials." (p.247)

A.7.2. Is this Health and Disability Research? ☒ Yes ☐ No

Broadly speaking, health and disability research should:

- *aim to answer a question or solve a problem and therefore generate new knowledge to prevent, identify and treat illness and disease*
- *have the ultimate purpose of maintaining and improving people's health – in the sense of a state of physical, mental and spiritual wellbeing, rather than simply the absence of disease or infirmity*
- *support disabled people to be included, participate more, exercise choice and control, and be more independent*
- *address health and disability disparities*
- *contribute to whānau ora.*

This description is necessarily broad; we acknowledge that people's health is influenced by a much wider range of social factors than their health care. (NEAC's National Ethical Standards for Health and Disability Research and Quality Improvement. p.28)

A.7.3. Does this research involve people in their capacity as consumers of health or disability support services, or in their capacity as relatives or caregivers of consumers of health or disability support services, or as volunteers in clinical trials (including, for the avoidance of doubt, bioequivalence and bioavailability studies)? ☐ Yes ☒ No

B. *The Ethical Principle of Research Adequacy*

AUTEC recognises that different research paradigms may inform the conception and design of projects. It adopts the following minimal criteria of adequacy: the project must have clear research goals; its design must make it possible to meet those goals; and the project should not be trivial but should potentially contribute to the advancement of knowledge to an extent that warrants any cost or risk to participants.

B.1. Is the applicant the person doing most of the research (the primary researcher)? ☒ Yes ☐ No

If the answer is 'No' please answer B.1.1 and the following sections, otherwise please answer B.2 and continue from there.

B.1.1. What is the name of the primary researcher if it is someone other than the applicant?

Gaganjot Kaur

B.1.2. What are the primary researcher's completed qualifications?

Master of Technology (Electronics and Communication Engineering)

B.1.3. What is the primary researcher's email address?

An email address at which the primary researcher can be contacted is essential.

gaganjot.kaur@aut.ac.nz

B.1.4. At which telephone numbers can the primary researcher be contacted during the day?

Mobile: 022 650 7399

B.2. Is the primary researcher

☐ an AUT staff member *☒ an AUT student*

If the primary researcher is an AUT staff member, please answer B.2.1 and the following sections, otherwise please answer B.3 and continue from there.

B.2.1. In which faculty, directorate, or research centre is the primary researcher employed?

If the response to this section is the same as that already given to section A.4.1 above, please skip this section and go to section B.2.2.

B.2.2. In which school or department is the primary researcher employed?

B.3. When the primary researcher is a student:

B.3.1. What is their Student ID Number?

17995196

B.3.2. In which faculty are they enrolled?

Faculty of Design and Creative Technologies

B.3.3. In which school, department, or Research Centre are they enrolled?

School of Engineering, Computing and Mathematical Sciences/Electrical and Electronic Engineering Department/BioDesign Lab.

B.4. What is the primary researcher's experience or expertise in this area of research?

Where the primary researcher is a student at AUT, please identify the applicant's experience or expertise in this area of research as well.

The primary researcher has previously used image processing techniques during their Master studies.

B.5. Who is in charge of data collection?

Gaganjot Kaur

B.6. Who will interact with the participants?

Gaganjot Kaur (Primary Researcher)

B.7. Is this research being undertaken as part of a qualification? ☒ Yes ☐ No

If the answer is 'Yes' please answer B.7.1 and the following sections, otherwise please answer B.8 and continue from there.

B.7.1. What is the name of the qualification?

Doctor of Philosophy (PhD)

B.7.2. In which institution will the qualification be undertaken?

Auckland University of Technology (AUT)

B.8. Details of Other Researchers or Investigators

B.8.1. Will any other people be involved as researchers, co- investigators, or supervisors? ☒ Yes ☐ No

If the answer is 'Yes' please answer B.8.1.1 and the following sections, otherwise please answer B.8.2 and continue from there.

B.8.1.1 What are the names of any other people involved as researchers, investigators, or supervisors?

Dr Hakilo Sabit

B.8.1.2 Where do they work?

Auckland University of Technology (AUT)

B.8.1.3 What will their roles be in the research?

Secondary Supervisor

B.8.1.4 What are their completed qualifications?

Doctor of Philosophy (PhD)

B.8.2. Will any research organisation or other organisation be involved in the research? ☐ Yes ☒ No

If the answer is 'Yes' please answer B.8.2.1 and the following sections, otherwise please answer B.9 and continue from there.

B.8.2.1 What are the names of the organisations?

B.8.2.2 Where are they located?

B.8.2.3 What will their roles be in the research?

B.9. Why are you doing this research and what is its aim and background?

Please provide the key outcomes or research questions and an academic rationale with sufficient information, including relevant references, to place the project in perspective and to allow the project's significance to be assessed.

This research study carried out to date has shown that a non-contact measurement of Heart Rate (HR) using video magnification is achievable. Where HR is a powerful indicator of the health and diagnosing medical conditions of a patient.

A healthy individual should have an HR of between 50 and 70 beats per minute (bpm) at rest and between 100 and 150 bpm when active. Most of the present techniques used for HR monitoring require sensors to be placed directly onto the patient's skin to acquire a reliable HR measurement. These may discomfort and soreness to the patient, and this non-contact measurement of HR may be an alternative method.

Due to the advances in digital photography, images and videos have become are easy to capture. There are many tools available to process the videos in terms of colour and motion [1-4], such as Video Magnification [5].

This research aims to process either low or high-quality videos using filtering to extract the information to detect the HR, as skin colour varies due to blood circulation, invisible to the naked human eye [4]. For HR measurement, computational time must be nearly real-time. The aim of this research is to process the videos in MATLAB and firstly, pre-processing will be done on the input video then face detection process will find the face region. After detection of face the region of interest will be forehead from that region of interest the skin colour changes will be detected by calculating the pixel values.

B.10. What are the potential benefits of this research to the participants, the researcher, and the wider community?

The participants will not benefit directly from taking part in this research. Still, all the participants (not by name) who contributed valuable time and effort to this research is to be acknowledged in the thesis or any publications.

This research provides the primary researcher with knowledge regarding the non-contact measurement of heart rate. It is a prerequisite for completing the PhD/Doctorate Degree plus conference and journal publications.

The research will provide academic groups within the biomedical imaging sector (not limited to) to facilitate the advancements in research and development involving heart rate measurement.

B.11. What are the theoretical frameworks or methodological approaches being used?

The data will be analysed on later date using MATLAB software, which is recognised software for research work in engineering field. The input video will be uploaded in software and using the designed algorithm it will detect only face from the video. Then the region of interest which is forehead will be selected. From the selected part the skin pixel values will be calculated and select the red pixel values only. Then magnify those values and detect the heart rate according to the variation in those values.

For validation of results the pulse oximeter reading will compare with the results from the newly developed algorithm.

B.12. How will data be gathered and processed?

Please provide your data collection protocols, describing step by step how you will be interacting with participants when collecting data.

Participants will be provided with a detailed explanation of the procedure and set upon arrival. Once this step is complete, they will then complete and sign a consent form if they wish to proceed. Participants will then prepare for the recording of the videos using a video camera. The camera will then be placed at 2-metres on a tripod from in front of the participant's face, shown in Figure 1. It will record the videos for a maximum of three minutes. Videos are recorded of the participant's face whilst sitting upright in the chair with two states: (a) one with where the head is kept steady as possible and (b) the second where there is some head movement.

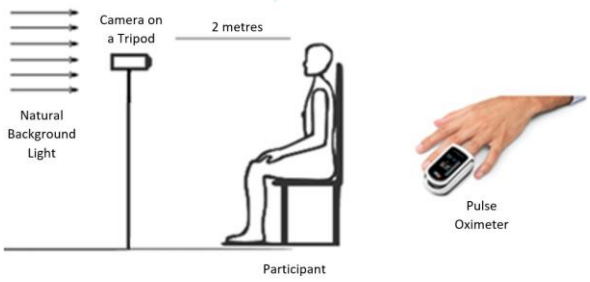


Figure 1: Experimental setup(left) and pulse oximeter (right)

To validate the results of the HR values obtained from the videos for this research. A reference heart rate measurement is required, taken using a commercially available pulse oximeter during the recording of the participant's video. The pulse oximeter will be placed on the participant's index finger of their dominant hand. The reading is used later in the research to validate the results obtained after analysing the videos.

For safety reason because of COVID-19 the following steps must be followed by the researcher for collection of data:

- Before giving the appointment to participant, will confirm about the COVID vaccination under AUT guidelines.
- The data will be collected by giving the appointment to only one person at a time and there will be 10 minutes break for cleaning /wipe pulse oximeter and chair with antibacterial wipes before next participant's appointment.
- There will be 2 metre distance

Researcher who will collect the data will wear face mask and gloves.

B.13. How will the data be analysed?

Please provide the statistical (for quantitative research) or methodological (for qualitative or other research) justification for analysing the data in this way.

The data will be analysed as follows:

- The recorded videos will be uploaded later into MATLAB software, which is a recognised software for image and video processing in engineering studies.
- Then developed algorithm will work in different steps firstly,
  1. detect the face from the video and
  2. from there the region of interest which is forehead will be selected. From the selected part the pixel values will be calculated and select the red pixel values only.
  3. then magnify those values and detect the heart rate according to the variation in those values.

For validation of results the pulse oximeter reading will compare with the results from algorithm.

B.14. Has any peer review taken place? ☒ Yes ☐ No

If your answer is 'Yes', please specify and provide evidence e.g. a letter of confirmation.

☐ *AUT Competitive Grant* ☐ *External Competitive Research Grant*
☐ *PGR1* ☐ *PGR2* ☒ *PGR9* ☐ *Independent Peer Review**

Optional exemplars for evidencing peer review are available from the Ministry of Health (HDEC) website (http://ethics.health.govt.nz/) or from the Forms section of the Research Ethics website (http://aut.ac.nz/researchethics)

C. *General Project Details*

C.1. Likely Research Output

C.1.1. What are the likely outputs of this research?

☒ *a thesis* ☐ *a dissertation* ☒ *a research paper* ☒ *a journal article*
☐ *a book* ☒ *conference paper* ☐ *a documentary* ☐ *an exhibition*
☐ *a film* ☐ *some other artwork* ☒ *other academic publications or presentations*
☐ *Some other output, please specify*

C.2. Research Location and Duration

C.2.1. In which countries and cities/localities will the data collection occur?

AUT City Campus, Auckland, New Zealand

C.2.1.1 Exactly where will any face to face data collection occur?

If face to face data collection will occur in participants' homes or similarly private spaces, then a Researcher Safety Protocol needs to be provided with this application.

In room WZ402

C.2.2. In which countries and cities/localities will the data analysis occur?

Auckland, New Zealand

C.2.3. When is the data collection scheduled to commence?

January 2022, once AUTEC grants approval

C.3. Research Participants

C.3.1. Who are the participants?

Participants will be students from AUT City Campus. Request notices will be posted on notice boards around the city campus. Participants can be of any gender between the age of 18 and 35.

C.3.2. How many participants are being recruited for this research?

If you are unsure, please provide an indicative range.

Fifty participants will participate in the study.

C.3.3. What criteria will be used to choose who to invite as participants?

The characteristics of the skin will need to vary in terms of the skin tone of each participant. Participants can be of any gender between the age of 18 and 35.

C.3.3.1 How will you select participants from those recruited if more people than you need for the study agree to participate?

A first-come, first-serve basis will be used to select participants, providing that they satisfy the criteria for the research. Inclusion criteria will require the physical characteristics of the participant's skin tone to vary in terms of light, medium and dark colour.

C.3.4. Will any people be excluded from participating in the study?

☒ Yes ☐ No

Exclusion criteria apply only to potential participants who meet the inclusion criteria. An exclusion criterion is any characteristic that ought to disqualify any potential participant from recruitment into the study. Consider exclusion criteria when there are heightened risks due to power differences in the relationship, recent injury, or other characteristics that might place potential participants at unreasonable risk of harms.

If the answer to this question is 'Yes' please answer C.3.4.1 and the following sections, otherwise please answer C.3.5 and continue from there.

C.3.4.1 What criteria will be used to exclude people from the study?

If any volunteer having facial tattoos or scars, they will be excluded for this research.

C.3.4.2 Why is this exclusion necessary for this study?

The developed algorithm is not that sensitive for facial tattoos and scars.

C.3.5. Recruitment of participants.

Please describe in detail the recruitment processes that will be used. If you will be recruiting by advertisement or email, please attach a copy to this Application Form

C.3.5.1 How will the initial contact with potential participants occur?

By placing a poster invitation around the AUT University City Campus.

C.3.5.2 How will the contact details of potential participants be collected and by whom?

The potential participants will contact the primary researcher who will record their details.

C.3.5.3 How will potential participants be invited to participate?

The information sheet will send to participant in order to understand about research and what they need to do in order to participate in the research and an invitation will be sent via email

C.3.5.4 How much time will potential participants have to consider the invitation?

Potential participants will have 20 days to consider the invitation. A mail will be sent if no reply is received confirming their participation.

C.3.5.5 How will potential participants respond to the invitation?

They will respond to the primary researcher via email or phone to make an appointment.

C.3.5.6 How will potential participants give consent?

The potential participants can provide consent by completing the applicant consent form attached.

C.3.5.7 How and when will the inclusion criteria and exclusion criteria given in sections C.3.2 and C.3.3 be applied?

The criteria will be applied when the participants show up to the lab for testing on the day.

C.3.5.8 Will there be any follow up invitations for potential participants?

One follow up invitation via email.

D. *Partnership, Participation and Protection*

D.1. How does the design and practice of this research implement the principle of Partnership in the interaction between the researcher and other participants?

How are the researcher and the participants working together? How will your research design and practice encourage a mutual respect and benefit and participant autonomy and ownership? How will you ensure that participants and researchers will act honourably and with good faith towards each other? Are the outcomes designed to benefit the participants and/or their social or cultural group? How will the information and knowledge provided by the participants be acknowledged?

After the first point of contact from the researcher, the participants will be informed and demonstrate how the data collection process will take place. The participants are free to ask as many questions as they wish and can withdraw themselves from the research at any time, and this will provide the participants with a better understanding of the research.

By participating in this research, you will not have any direct benefit. However, you will be acknowledged for the contribution of your valuable time and effort towards this study. If you wish, you will also be provided with a copy of your individual results against the overall results for the study. This research will give academic groups advancements in research and development involving non-contact measurement of heart rate. The outcomes will helpful for the healthcare professionals.

D.2. How does the design and practice of this research implement the principle of participation in the interaction between the researcher and other participants?

What is the actual role of participants in your research project? Will participants be asked to inform or influence the nature of the research, its aims, or its methodology? Will participants be involved in conducting the research or is their principal involvement one of sharing information or data? Do participants have a formal role as stakeholders e.g. as the funders and/or beneficiaries of the research? What role will participants have in the research outputs (e.g. will they be asked to approve transcripts or drafts)?

The principal involvement of participants in this research is to assist with data collection. They will have the simple task of following instructions given by the primary researcher to collect the data. They will have every right to withdraw themselves from the research if they feel uncomfortable at any time.

The participants are then be asked to sit on a chair. The camera facing towards them records the face video. Participants will be required to follow simple directions, which the primary researcher will give. They will be required to attach a pulse oximeter on the index finger of their dominant hand and need to stay stationary. Once the data has been recorded, the pulse oximeter will remove from the finger before they stand up. They are free to leave after this has been done. All the participants will be acknowledged in the final write up for giving up their valuable time in assisting with the research.

D.3. How does the design and practice of this research implement the principle of Protection in the interaction between the researcher and other participants?

How are the researcher and the participants protecting each other? How will you actively protect participants from deceit, harm and coercion through the design and practice of your research? How will the privacy of participants and researchers be protected? How will any power imbalances inherent in the relationships between the participants and researchers be managed? How will any cultural or other diversity be respected?

The participants will be seated on a chair before they attach a pulse oximeter on a fingertip. The camera will be in front of their face.

So to ensure privacy, the names and personal information of the participants will not be used in the research. Participants will be identifiable when the overall summary of the results are produced. For the identification of participants the consent form will be signed,if they will be agree with that. However, the privacy of the data collected will be kept protected and secured by the personnel in charge of managing the data. The data will be held for six years, after which it will be destroyed.

The potential participant will be able to withdraw from the research any time before and during the data collection process. During the data collection, the prime focus at all times will be the safety and comfort of the participants. If they wish to withdraw themselves from the data collection process, they can simply ask or raise their hand.

E. *Social and Cultural Sensitivity (including the obligations of the Treaty of Waitangi)*

E.1. What familiarity does the researcher have with the social and cultural context of the participants?

Participants are chosen as a response to an advertisement within the university. The primary researcher is a current PhD student. The primary researcher has been at AUT for the past three years, providing a common ground of being a university student, like participants that volunteer. Furthermore, the participants chosen for this research would be within an age group of 18-35 years, consisting of the primary researcher's age (35) group. Being a student at the university, the primary researcher is familiar with the variety of cultures within the university students, which would better the interaction between the participants and the primary researcher.

E.2. What consultation has occurred?

Research procedures should be appropriate to the participants. Researchers have a responsibility to inform themselves of, and take the steps necessary to respect the values, practices, and beliefs of the cultures and social groups of all participants. This usually requires consultation or discussion with appropriate people or groups to ensure that the language and research approaches being used are relevant and effective. Consultation should begin as early as possible when designing the project and should continue throughout its duration.

All researchers are encouraged to make themselves familiar with Te Ara Tika: Guidelines for Maori Research Ethics: A framework for researchers and ethics committee members which is able to be accessed through the Research Ethics website. Researchers may also find Te Kaahui Maangai a directory of Iwi and Maaori organisations to be helpful. This may be accessed via the Te Puni Kookiri website (http://www.tkm.govt.nz/). As well as these documents, the Health Research Council has published Pacific Health Research Guidelines, and Guidelines on research involving children. (see http://www.hrc.govt.nz). There are also guidelines by various organisations about researching with other populations that researchers will find helpful.

Consultation of data collection procedure, the nature of the research and its impacts was taken place with a group of individuals that have varying cultural and social backgrounds. The consultation consisted of a friendly casual discussion of how comfortable they would be if they were to participate in the research of such nature. The views of the individuals varied along with different cultures and ethnic backgrounds, which consists of Maori, Samoan, Indian, Asian and European backgrounds that count towards the majority of the university students.

The group of individuals, consisting of male and female, had no issue with the data collection method and were quite intrigued about the entire process. The individuals showed no concern about having their personal information collected when advised of its safe storage.

E.2.1. With whom has the consultation occurred?

Please provide written evidence that the consultation has occurred.

Consultation has occurred with a group of fellow university students, compromising male and female students of different social and cultural backgrounds.

E.2.2. How has this consultation affected the design and practice of this research?

Developing the new algorithm has not affected the design and practice of this research in any way.

The potential participants will be given full authority to their right to participate/withdraw from the study at any time.

Before data collection there will be discussion with participant if they feel uncomfortable in front of camera or using pulse oximeter any time they can ask to stop.

E.3. Does this research target Māori participants? ☐ Yes ☒ No

All researchers are encouraged to make themselves familiar with Te Ara Tika: Guidelines for Maori Research Ethics: A framework for researchers and ethics committee members

If your answer is 'No', please go to section E.4 and continue from there. If you answered 'Yes', please answer the next question.

E.3.1. Which iwi or hapu are involved?

Not applicable.

E.4. Does this research target participants of particular cultures or social groups? ☒ Yes ☐ No

AUTEC defines the phrase 'specific cultures or social groups' broadly. In section 2.5 of Applying for Ethics Approval: Guidelines and Procedures it uses the examples of Chinese mothers and paraplegics. This is to identify their distinctiveness, the first as a cultural group, the second as a social group. Other examples of cultural groups may be Korean students, Samoan husbands, Cook Islanders etc., while other examples of social groups may be nurse aides, accountants, rugby players, rough sleepers (homeless people who sleep in public places) etc. Please refer to Section 2.5 of AUTEC's Applying for Ethics Approval: Guidelines and Procedures (accessible in the Ethics Knowledge Base online via http://www.aut.ac.nz/about/ethics) and to the relevant Frequently Asked Questions section in the Ethics Knowledge Base.

If your answer is 'No', please go to section E.5 and continue from there. If you answered 'Yes', please answer the next question.

E.4.1. Which cultures or social groups are involved?

AUT students.

E.5. Does this research focus on an area of research that involves Treaty obligations? ☐ Yes ☒ No

All researchers are encouraged to make themselves familiar with Te Ara Tika: Guidelines for Maori Research Ethics: A framework for researchers and ethics committee members.

If your answer is 'No', please go to section E.6 and continue from there. If you answered 'Yes', please answer the next question.

E.5.1. Which treaty obligations are involved?

Not applicable.

E.6. Will the findings of this study be of particular interest to specific cultures or social groups? ☒ Yes ☐ No

If the answer is 'Yes' please answer E.6.1 and the following sections, otherwise please answer F.1 and continue from there.

E.6.1. To which iwi, hapū, culture or social groups will the findings be of interest?

Healthcare professionals and biomedical engineers.

E.6.2. How will the findings be made available to these groups?

Findings will be available in terms of written report on request.

F. *Respect for the Vulnerability of Some Participants*

"Vulnerable persons are those who are relatively (or absolutely) incapable of protecting their own interests. More formally, they may have insufficient power, intelligence, education, resources, strength, or other needed attributes to protect their own interests. Individuals whose willingness to volunteer in a research study may be unduly influenced by the expectation, whether justified or not, of benefits associated with participation, or of a retaliatory response from senior members of a hierarchy in case of refusal to participate may also be considered vulnerable." (Standards and Operational Guidance for Ethics Review of Health-Related Research with Human Participants, World Health Organisation).

F.1. Will your research involve any of the following groups of participants? ☐ Yes ☒ No

If your research involves any of these groups of participants, please clearly indicate which ones and then answer F.2 and the following section, otherwise please answer G.1 and continue from there.

☐ people unable to give informed consent? ☐ your (or your supervisor's) own students?
☐ preschool children? ☐ children aged between five and sixteen years?
☐ legal minors aged between sixteen and twenty years?
☐ people lacking the mental capacity for consent?
☐ people in a dependent situation (e.g. people with a disability, or residents of a hospital, nursing home or prison or patients highly dependent on medical care)?
☐ people who are vulnerable for some other reason (e.g. the elderly, persons who have suffered abuse, persons who are not competent in English, new immigrants)? – please specify

F.2. How is respect for the vulnerability of these participants reflected in the design and practice of your research?

F.3. What consultation has occurred to ensure that this will be effective?

Please provide evidence of the consultation that has occurred.

G. *Informed and Voluntary Consent*

G.1. How will information about the project be given to potential participants?

A copy of all information that will be given to prospective participants is to be attached to this Application Form. If written information is to be provided to participants, you are advised to use the Information Sheet exemplar. The language in which the information is provided is to be appropriate to the potential participants and translations need to be provided when necessary.

A written information sheet will be provided.

G.2. How will the consent of participants be obtained and evidenced?

AUTEC requires consent to be obtained and usually evidenced in writing. A copy of the Consent Form which will be used is to be attached to this application. If this will not be the case, please provide a justification for the alternative approach and details of the alternative consent process. Please note that consent must be obtained from any participant aged 16 years or older. Participants under 16 years of age are unable to give consent, which needs to be given by their parent or legal guardian. AUTEC requires that participants under the age of 16 assent to their participation. When the nature of the research requires it, AUTEC may also require that consent be sought from parents or legal guardians for participants aged between 16 and twenty years. For further information please refer to AUTEC's Applying for Ethics Approval: Guidelines and Procedures.

Written consent of the participants' will be obtained through the consent form that they will sign and date.

G.3. Will any of the participants have difficulty giving informed consent on their own behalf? ☐ Yes ☒ No

Please consider physical condition, cognitive status, age, language, legal status, or other barriers.

If the answer is 'Yes' please answer G.3.1 and the following sections, otherwise please answer G.4 and continue from there.

G.3.1. If participants are not competent to give fully informed consent, who will consent on their behalf?

Researchers are advised that the circumstances in which consent is legally able to be given by a person on behalf of another are very constrained. Generally speaking, only parents or legal guardians may give consent on behalf of a legal minor and only a person with an enduring power of attorney may give consent on behalf of an adult who lacks capacity.

G.3.2. How will these participants be asked to provide assent to participation?

Whenever consent by another person is possible and legally acceptable, it is still necessary to take the wishes of the participant into account, taking into consideration any limitations they may have in understanding or communicating them.

G.4. Is there a need for translation or interpreting? ☐ Yes ☒ No

If your answer is 'Yes', please provide copies of any translations with this application and any Confidentiality Agreement required for translators or interpreters.

H. *Respect for Rights of Privacy and Confidentiality*

H.1. How will the researchers respect the privacy and confidentiality of participants?

Please note that anonymity and confidentiality are different. For AUTEC's purposes, 'Anonymity' means that the researcher is unable to identify who the participant is in any given case. If the participants will be anonymous, please state how, otherwise, if the researcher will know who the participants are, please describe how the participants' privacy issues and the confidentiality of their information will be managed.

At the time of data collection participants will not be anonymous to the researcher as both will be interacting. However, no participants will be individually identified in the writing of the research results.

H.2. Will any participants be identifiable in the research outputs or findings? ☒ Yes ☐ No

If your answer is 'Yes", please answer H.2.1, otherwise please answer H.3

H.2.1. What level of confidentiality is able to be offered to participants and how will this be managed?

If the research involves small or distinctive groups of participants or procedures such as interviews conducted at the worksite, or focus groups with peers, researchers should identify the level of participant confidentiality that can be offered in the Information Sheet. If participants or groups will be identified, please state why this is appropriate, how this will happen, and how the participants will give consent.

N/A

H.3. What information on the participants will be obtained from third parties?

This includes use of third parties, such as employers or professional organisations, in recruitment.

No information will be given to third parties.

H.4. How will potential participants' contact details be obtained for the purposes of recruitment?

The primary researcher will collect the contact details of potential participants v

H.5. What identifiable information on the participants will be given to third parties?

Facial features will be used in terms of conference presentations and PhD thesis.

H.6. Who will have access to the data during the data collection and analysis stages?

Gaganjot Kaur (Primary Researcher)

Dr Jeff Kilby (Primary Supervisor)

H.7. Who will have access to the data after the findings have been produced?

Gaganjot Kaur (Primary Researcher)

Dr Jeff Kilby (Primary Supervisor)

H.8. Are there any plans for the future use of the data beyond those already described? ☐ Yes ☒ No

The applicant's attention is drawn to the requirements of the Privacy Act 1993 (see Appendix I of AUTEC's Applying for Ethics Approval: Guidelines and Procedures). Information may only be used for the purpose for which it was collected so if there are plans for the future use of the data, then this needs to be explained in the Information Sheets for participants. If you have answered 'Yes' to this question, please answer section H.8.1.1 and continue from there. If you answered 'No' to this question, please go to section H.9 and proceed from there.

H.8.1.1 If data will be stored in a database, who will have access to that information, how will it be used, for what will it be used, and how have participants consented to this?

H.8.1.2 Will any contact details be stored for future use and if so, who will have access to them, how will they be used, for what will they be used, and how have participants consented to this?

H.9. Where will the data be stored once the analysis is complete?

Please provide the exact storage location. AUTEC normally requires that the data be stored securely on AUT premises in a location separate from the consent forms. Electronic data should be downloaded to an external storage device (e.g. an external hard drive, a memory stick etc.) and securely stored. If you are proposing an alternative arrangement, please explain why.

All written Consent Forms are to be scanned, after which they are shredded, and the recorded videos are to be encrypted and stored on an external hard drive; for six years. The external hard drive is to be kept in a secured location (i.e. a locked cabinet) in Dr Jeff Kilby's office at AUT University City Campus. After six years, all data collected will be destroyed by erasing all the data from the hard drive.

H.9.1. For how long will the data be stored after completion of analysis?

AUTEC normally requires that the data be stored securely for a minimum of six years, or ten years for health data. If you are proposing an alternative arrangement, please explain why.

For six years

H.9.2. How will the data be destroyed?

If the data will not be destroyed, please explain why, identify how it will be safely maintained, and provide appropriate informed consent protocols.

After six years, all the digital copies of the scanned consent forms and the videos will be erased from the external device.

H.10. Who will have access to the Consent Forms?

Gaganjot Kaur (Primary Researcher)

Dr Jeff Kilby (Primary Supervisor)

H.11. Where will the completed Consent Forms be stored?

Please provide the exact storage location. AUTEC normally requires that the Consent Forms be stored securely on AUT premises in a location separate from the data. If you are proposing an alternative arrangement, please explain why.

All consent forms will be scanned into digital copies and stored on an external device. The device will be kept in a secure location in Dr Jeff Kilby's office (WS329) at AUT City Campus, separate from the data.

H.11.1. For how long will the completed Consent Forms be stored?

AUTEC normally requires that the Consent Forms be stored securely for a minimum of six years, or ten years in the case of research involving health data. If you are proposing an alternative arrangement, please explain why.

Six years

H.11.2. How will the Consent Forms be destroyed?

If the Consent Forms will not be destroyed, please explain why.

Hard copies of the consent form will be destroyed immediately after being scanned as digital copies and saved on an external device. After six years, the digitally scanned consent forms and videos will be erased from the external device.

H.12. Does your research involve the collection of personally identifiable and sensitive data? ☒ Yes ☐ No

Sensitive data can be used to identify an individual, object or location and has a risk of discrimination, harm or unwanted attention. Sensitive data potentially poses a substantial threat to those who are or who have been involved in it, especially if it is shared inappropriately, or if it falls into the wrong hands. If you have answered 'Yes' please identify what data is being collected and how it is sensitive and provide a Data Safety Management Protocol (see the Forms section of the Research Ethics website for a guide to drafting one). If the answer is 'No', please answer H.13 and continue from there.

Not much sensitive data.

H.13. Does your project involve the use of previously collected information or biological samples for which there was no explicit consent for this research? ☐ Yes ☒ No

If the answer is 'Yes' please answer H.13.1 and the following sections, otherwise please answer H.14 and continue from there.

H.13.1. What previously collected data will be involved?

H.13.2. Who collected the data originally?

H.13.2.1 Why was the information originally collected?

H.13.2.2 For what purposes was consent originally given when the information was collected?

H.13.3. How will the data be accessed?

H.14. Does your research involve the collection of information about organisational practices? ☐ Yes ☒ No

AUTEC applies a broad definition to the term 'organisations'. It could include for example, businesses, hospitals or clinics, schools, or sports clubs and teams If the answer is 'Yes' please answer H.14.1, otherwise please answer I.1 and continue from there.

H.14.1. How will the authorisation to access the organisation or its staff for research purposes be obtained?

H.14.2. Could disclosure of this information potentially disadvantage the oganisation or the participants? ☐ Yes ☒ No

If your answer is 'Yes", please answer H.14.2.1, otherwise please answer H.14.3

H.14.2.1 How will the risks associated with potential disadvantages be managed?

H.14.3. Will the participants or anyone else in the oganisation be identified in this information? ☐ Yes ☒ No

If your answer is 'Yes", please answer H.14.3.1, otherwise please answer I.1 and continue from there.

H.14.3.1 How will the potential risks involved be managed?

If the research involves procedures such as interviews conducted at the worksite, or focus groups with peers, researchers should identify the level of participant confidentiality that can be offered in the Information Sheet.

I. *Minimisation of risk*

I.1. Risks to Participants

Please consider the possibility of moral, physical, psychological or emotional risks to participants, including issues of confidentiality and privacy, from the perspective of the participants, and not only from the perspective of someone familiar with the subject matter and research practices involved. Please clearly state what is likely to be an issue, how probable it is, and how this will be minimised or mitigated (e.g. participants do not need to answer a question that they find embarrassing, or they may terminate an interview, or there may be a qualified counsellor present in the interview, or the findings will be reported in a way that ensures that participants cannot be individually identified, etc.) Possible risks and their mitigation should be fully described in the Information Sheets for participants.

I.1.1. How much time will participants be required to give to the project?

Approximately 20 minutes.

I.1.2. What level of discomfort or embarrassment may participants be likely to experience?

Might be minimal discomfort sitting on chair for period of time, participant can stop anytime.

I.1.3. In what ways might participants be at risk in this research?

None

I.1.4. In what ways are the participants likely to experience risk or discomfort as a result of cultural, employment, financial or similar pressures?

None

I.1.5. Will your project involve processes that are potentially disadvantageous to a person or group, such as the collection of information, images etc. which may expose that person/group to discrimination, criticism, or loss of privacy? ☐ Yes ☒ No

If your answer is 'Yes', please detail how these risks will be managed and how participants will be informed about them.

I.1.6. Will your research involve collection of information about illegal behaviour(s) which could place the participants at current or future risk of criminal or civil liability or be damaging to their financial standing, employability, professional or personal relationships? ☐ Yes ☒ No

If your answer is 'Yes', please detail how these risks will be managed and how participants will be informed about them.

I.1.7. If the participants are likely to experience any significant discomfort, embarrassment, incapacity, or psychological disturbance, please state what consideration you have given to the provision of counselling or post-interview support, at no cost to the participants, should it be required.

Adult research participants in Auckland are able to utilise counselling support from the AUT Counselling Team; otherwise, you may have to consider local providers for participants who are located nationwide, or in some particular geographical area or who are children. You may discuss the potential for participant psychological impact or harm with the Head of AUT Counselling if you require. Please check the relevant Frequently Asked Question on the research ethics website as well and ensure the appropriate wording is included in the Information Sheet when counselling opportunities need to be offered.

Very unlikely

I.1.8. Will any use of human remains, tissue or body fluids which does not require submission to a Health and Disability Ethics Committee occur in the research? ☐ Yes ☒ No

e.g. finger pricks, urine samples, etc. (please refer to section 13 of AUTEC's Applying for Ethics Approval: Guidelines and Procedures). If your answer is yes, please provide full details of all arrangements, including details of agreements for treatment, how participants will be able to request return of their samples in accordance with right 7 (9) of the Code of Health and Disability Services Consumers' Rights, etc.

I.1.9. Will this research involve potentially hazardous substances? ☐ Yes ☒ No

e.g. radioactive material, biological substances (please refer to section 15 of AUTEC's Applying for Ethics Approval: Guidelines and Procedures and the Hazardous Substances and New Organisms Act 1996).

If the answer is 'Yes', please provide full details, including hazardous substance management plan.

I.2. Risks to Researchers

If this project will involve interviewing participants in private dwellings, undertaking research in unfamiliar cultural contexts either in New Zealand or overseas, doing research in a place to which a travel warning applies, or going into similarly vulnerable situations, then a Researcher Safety protocol should be designed and appended to this application. This should identify simple and effective processes for keeping someone informed of the researcher's whereabouts and provide for appropriate levels of assistance. A guide to drafting one is provided in the forms section of the Research Ethics website.

I.2.1. Are the researchers likely to be at risk? ☒ Yes ☐ No

If the answer is 'Yes' please answer I.2.1.1 and then continue, otherwise please answer I.3 and continue from there.

I.2.1.1 In what ways might the researchers be at risk and how will this be managed?

Researchers might have Covid-19 risk, but the following procedure will follow to make sure the safety of all participants and researcher.

- Before giving the appointment to participant, will confirm about the COVID vaccination under AUT guidelines.
- The data will be collected by giving the appointment to only one person at a time and there will be 10 minutes break for cleaning /wipe pulse oximeter and chair with antibacterial wipes before next participant's appointment.
- There will be 2-metre distance
- Researcher who will collect the data will wear face mask and gloves.

I.3. Risks to AUT

I.3.1. Is AUT or its reputation likely to be at risk because of this research? ☐ Yes ☒ No

If the answer is 'Yes' please answer I.3.1.1 and then continue, otherwise please answer I.3.2 and continue from there.

I.3.1.1 In what ways might AUT be at risk in this research?

Please identify how and detail the processes that will be put in place to minimise any harm.

I.3.2. Are AUT staff and/or students likely to encounter physical hazards during this project? ☐ Yes ☒ No

If yes, please provide a hazard management protocol identifying how harm from these hazards will be eliminated or minimised.

J. *Truthfulness and limitation of deception*

J.1. How will feedback on or a summary of the research findings be disseminated to participants (individuals or groups)?

It is normally courteous to provide participants with a one or two page summary of the findings of the research. Please ensure that this information is included in the Information Sheet.

All participants will be provided with a brief report showing their results and how it was reflected and compares with the overall results. However, if they do not wish to receive the report, they can let us know on the consent form.

J.2. Does your research include any deception of the participants, such as non-disclosure of aims or use of control groups, concealment, or covert observations? ☐ Yes ☒ No

Deception of participants in research may involve deception, concealment or covert observation. Deception of participants conflicts with the principle of informed consent, but in some areas of research it may sometimes be justified to withhold information about the purposes and procedures of the research. Researchers must make clear the precise nature and extent of any deception and why it is thought necessary. Emphasis on the need for consent does not mean that covert research can never be approved. Any departure from the standard of properly informed consent must be acceptable when measured against possible benefit to the participants and the importance of the knowledge to be gained as a result of the project or teaching session. This must be addressed in all applications. Please refer to Section 2.4 of AUTEC's Applying for Ethics Approval: Guidelines and Procedures when considering this question.

If the answer is 'Yes' please answer J.2.1 and the following sections, otherwise please answer J.3 and continue from there.

J.2.1. Is deception involved?

J.2.2. Why is this deception necessary?

J.2.3. How will disclosure and informed consent be managed?

J.3. Will this research involve use of a control group? ☐ Yes ☒ No

If the answer is 'Yes' please answer J.3.1 and the following sections, otherwise please answer K.1 and continue from there.

J.3.1. How will the Control Group be managed?

J.3.2. What percentage of participants will be involved in the control group?

J.3.3. What information about the use of a control group will be given to the participants and when?

K. *Avoidance of Conflict of Interest*

Researchers have a responsibility to ensure that any conflict between their responsibilities as a researcher and other duties or responsibilities they have towards participants or others is adequately managed. For example, academic staff members who propose to involve their students as participants in research need to ensure that no conflict arises between their roles as teacher and researcher, particularly in view of the dependent relationship between student and teacher, and of the need to preserve integrity in assessment processes. Likewise researchers have a responsibility to ensure that any conflict of interest between participants is adequately managed for example, managers participating in the same research as their staff.

K.1. What conflicts of interest are likely to arise as a consequence of the researchers' professional, social, financial, or cultural relationships?

None

K.2. What possibly coercive influences or power imbalances are there in the professional, social, financial, or cultural relationships between the researchers and the participants or between participants (e.g. dependent relationships such as teacher/student; parent/child; employer/employee; pastor/congregation etc.)?

The participant will not be student of Dr Jeff Kilby

K.3. How will these conflicts of interest, coercive influences or power imbalances be managed through the research's design and practice and how will any adverse effects that may arise from them be mitigated?

None

K.4. Does your project involve payments or other financial inducements (including koha, reasonable contribution towards travel expenses or time, or entry into a modest prize draw) to participants? ☐ Yes ☒ No

If the answer is 'Yes' please answer K.4.1 and the following sections, otherwise please answer K.5 and continue from there.

K.4.1. What form will the payment, inducement, or koha take?

K.4.2. Of what value will any payment, gift or koha be?

K.4.3. Will potential participants be informed about any payment, gift or koha as part of the recruitment process, and if so, why and how?

K.5. Have any applications for financial support for this project been (or will be) made to a source external to AUT? ☐ Yes ☒ No

If the answer is 'Yes' please answer K.5.1 and the following sections, otherwise please answer K.6 and continue from there.

K.5.1. What financial support for this project is being provided (or will be provided) by a source external to AUT?

K.5.2. Who is the external funder?

K.5.3. What is the amount of financial support involved?

K.5.4. How is/are the funder/s involved in the design and management of the research?

K.6. Have any applications been (or will be) submitted to an AUT Faculty Research Grants Committee or other AUT funding entity? ☐ Yes ☒ No

If the answer is 'Yes' please answer K.6.1 and the following sections, otherwise please answer K.7 and continue from there.

K.6.1. What financial support for this project is being provided (or will be provided) by an AUT Faculty Research Grants Committee or other AUT funding entity?

K.6.2. What is the amount of financial support involved?

K.6.3. How is/are the funder/s involved in the design and management of the research?

K.7. Is funding already available, or is it awaiting decision?

K.8. Do the applicant or the researchers, investigators or research organisations mentioned in Part B of this application have any financial interests in the outcome of this project? ☐ Yes ☒ No

If the response is 'Yes', please provide full details about the financial interests and how any conflicts of interest are being managed, otherwise, please respond to section K.9 and continue from there.

K.9. Are the participants expected to pay in any way for any services associated with this research? ☐ Yes ☒ No

If the response is 'Yes', please provide full details about the charges and describe how any benefits will balance the burdens involved as well as how any conflicts of interest are being managed. Otherwise please respond to section L.1 and continue from there.

No.

L. *Respect for Property*

Researchers must ensure that processes do not violate or infringe legal or culturally determined property rights. These may include factors such as land and goods, works of art and craft, spiritual treasures and information.

L.1. Will this research impact upon property owned by someone other than the researcher? ☐ Yes ☒ No

If the answer is 'Yes' please answer L.1.1 and the following sections, otherwise please answer L.2 and continue from there.

L.1.1. How will this be managed?

L.2. How do contexts to which copyright or Intellectual Property apply (e.g. research instruments, social media, virtual worlds etc.) affect this research and how will this be managed?

Particular attention should be paid to the legal and ethical dimensions of intellectual property. Care must be taken to acknowledge and reference the ideas of all contributors and others and to obtain any necessary permissions to use the intellectual property of others. Teachers and researchers are referred to AUT's Intellectual Property Policy for further guidance.

Commercially available AUT equipment and licenced software are all available in public domain.

M. *References*

Please include any references relating to your responses in this application in the standard format used in your discipline.

[1] J. Kranjec, S. Beguš, G. Geršak, and J. Drnovšek, "Non-contact heart rate and heart rate variability measurements: A review," *Biomedical Signal Processing and Control,* vol. 13, pp. 102-112, 2014.

[2] N. Wadhwa, M. Rubinstein, F. Durand, and W. T. Freeman, "Riesz pyramids for fast phase-based video magnification," in *IEEE International Conference on Computational Photography (ICCP)*, 2014, pp. 1-10.

[3] W. N., R. M, D. F, and F. W. T., "Phase-based video motion processing," *ACM Transactions on Graphics,* vol. 32, no. 4, p. 1, 2013.

[4] Y. S. Dosso, A. Bekele, and J. R. Green, "Eulerian Magnification of Multi-Modal RGB-D Video for Heart Rate Estimation," in *IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 2018, pp. 1-6.

[5] H. Y. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, and W. Freeman, "Eulerian video magnification for revealing subtle changes in the world," 2012.

N. Checklist

Please ensure all applicable sections of this form have been completed and all appropriate documentation is attached as incomplete applications will not be considered by AUTEC.

Have you discussed this application with your AUTEC Faculty Representative, the Executive Secretary, or the Ethics Coordinator?	☒ Yes ☐ No
Is this application related to an earlier ethics application? If yes, please provide the application number of the earlier application.	☐ Yes ☒ No
Are you seeking ethics approval from another ethics committee for this research? If yes, please identify the other committee.	☐ Yes ☒ No

Section A	*Project information provided*	☒
Section B	*Research Adequacy information provided*	☒
Section C	*Project details provided*	☒
Section D	*Three Principles information provided*	☒
Section E	*Social and Cultural Sensitivity information provided*	☒
Section F	*Vulnerability information provided*	☒
Section G	*Consent information provided*	☒
Section H	*Privacy information provided*	☒
Section I	*Risk information provided*	☒
Section J	*Truthfulness information provided*	☒
Section K	*Conflict of Interest information provided*	☒
Section L	*Respect for Property information provided*	☒
Section M	*References provided*	☒
Section N	*Checklists completed*	☒
Section O.1 and 2	*Applicant and student declarations signed and dated*	☒
Section O.3	*Authorising signature provided*	☒

Spelling and Grammar Check (please note that a high standard of spelling and grammar is required in documents that are issued with AUTEC approval)

Attached Documents (where applicable)

Participant Information Sheet(s)	☒
Consent Form(s)	☒
Questionnaire(s)	☐
Indicative Questions for Interviews or Focus Groups	☐
Observation Protocols	☐
Recording Protocols for Tests	☐
Advertisement(s)	☒
Researcher Safety Protocol	☐
Hazardous Substance Management Plan	☐
Any Confidentiality Agreement(s)	☐
Any translations that are needed	☐
Other Documentation	☐

O.

Declarations

O.1. Declaration by Applicant

- *The information in this application is complete and accurate to the best of my knowledge and belief. I take full responsibility for it.*
- *In conducting this study, I agree to abide by all applicable laws and regulations, and established ethical standards contained in AUTEC's Applying for Ethics Approval: Guidelines and Procedures, the Auckland University of Technology Code of Conduct for Research,. and internationally recognised codes of ethics.*
- *I accept responsibility for ensuring that management approval for access for this research from any institution or organisation at which it will be conducted will be obtained. When the research is undertaken outside New Zealand, I agree to ensure that all ethical, legal, and locality obligations or requirements for those jurisdictions are met.*
- *I will continue to comply with AUTEC's Applying for Ethics Approval: Guidelines and Procedures, including its requirements for the submission of annual progress reports, amendments to the research protocols before they are used, and completion reports.*
- *I understand that brief details of this application may be made publicly available and may also be provided to the Graduate Research School, the Research and Innovation Office, or the University's insurers for purposes relating to AUT's interests.*

8/11/21

Signature *Date*

O.2. Declaration by Student Researcher

- *The information in this application is complete and accurate to the best of my knowledge and belief.*
- *In conducting this study, I agree to abide by all applicable laws and regulations, and established ethical standards contained in AUTEC's Applying for Ethics Approval: Guidelines and Procedures, the Auckland University of Technology Code of Conduct for Research,. and internationally recognised codes of ethics.*
- *I will continue to comply with AUTEC's Applying for Ethics Approval: Guidelines and Procedures, including its requirements for the submission of annual progress reports, amendments to the research protocols before they are used, and completion reports.*
- *I understand that brief details of this application may be made publicly available and may also be provided to the Graduate Research School, the Research and Innovation Office, or the University's insurers for purposes relating to AUT's interests.*

8/11/21

Signature *Date*

O.3. Authorisation by Head of Faculty/School/Programme/Centre

- *In authorising this study, I declare that the applicant is adequately qualified to undertake or supervise this research and that to the best of my knowledge and belief adequate resources are available for this research and all appropriate local research governance issues have been addressed.*
- *I declare that the applicant will ensure that management approval for access for this research from any institution or organisation at which it will be conducted will be obtained. When the research is undertaken outside New Zealand, I declare that the applicant will ensure that all ethical, legal, and locality obligations or requirements for those jurisdictions are met.*
- *I understand that brief details of this application may be made publicly available and may also be provided to the Graduate Research School, the Research and Innovation Office, or the University's insurers for purposes relating to AUT's interests.*

8/11/2021

Signature *Date*

Notes for submitting the completed application for review by AUTEC

- Please ensure that you are using the current version of this form before submitting your application.
- Please ensure that all questions on the form have been answered and that no part of the form has been deleted.
- Please provide either:
  - one printed, single sided, A4, and signed copy of the application and all related documents. This may be delivered or posted to the AUTEC Secretariat, room WU406, fourth floor, WU Building, City campus. The internal mail code is D-88. The courier address is 46 Wakefield Street, Auckland 1010. Alternatively, the application may be provided to the Research Ethics Advisor in person at one of the Drop In sessions at any of the four campuses (http://www.aut.ac.nz/researchethics/resources/workshops-and-drop-inns). OR

    ethics@aut.ac.nz. The application and documents must be scanned into a single .pdf file with the EA form at the beginning and the other documents in the order stated in the form. The application must have all the required signatures.

- Applications should be submitted once they have been finalised. For a particular meeting it needs to have been received in the AUTEC Secretariat by 4 pm on the relevant agenda closing day [AUTEC's meeting dates are listed in the website at http://www.aut.ac.nz/researchethics]. As many applications are reviewed under delegated authority, applicants are encouraged to submit their applications once they are ready rather than waiting for the closing date.
- If sending applications by internal mail, please post them at least two days earlier to allow for any delay that may occur.
- Late applications will be placed on the agenda for the following meeting.

MINIMAL RISK CHECKLIST

Your application may be eligible for delegated review if it poses no more than minimal risk of harm to participants. To assist AUTEC's Secretariat to screen the application for assignment to the correct review pathway, please complete the following checklist:

Does the research involve any of the following?

ANONYMOUS SURVEY ASSESSMENT

		Yes	No
1	The collection of anonymous and non-sensitive survey/questionnaire data only. *(If YES is checked, the application may receive an expedited review if the data is from adults and poses no foreseeable risks to participants OR where any foreseeable risk is no more than inconvenience – no further questions on this checklist need be answered.)*	☐	☐

MINIMAL RISK ASSESSMENT[1]

		Yes	No
2	Participants who are unable to give informed consent (including children under 16 years old), or who are particularly vulnerable or in a dependent situation, (e.g. people with learning difficulties, over-researched groups, people in care facilities, or patients highly dependent on medical care)?	☐	☐
3	A reasonable expectation of causing participants physical pain beyond mild discomfort, or that experienced by the participants on an every-day basis, or any emotional discomfort, embarrassment, or psychological or spiritual harm, (e.g. asking participants to recall upsetting events)?	☐	☐
4	Research processes which may elicit information about any participant's involvement in illegal activities, or activities that represent a risk to themselves or others, (e.g. drug use or professional misconduct)?	☐	☐
5	Collection of any human tissue, blood or other samples, or invasive or intrusive physical examination or testing?	☐	☐
6	The administration of any drugs, medicines, supplements, placebo or non-food substances?	☐	☐
7	An intervention of any form of exercise, or other physical regime that is different to the participants' normal activities (e.g. dietary, sleep)?	☐	☐
8	Participants who are being asked to give information of a personal nature about their colleagues, employers, teachers, or coaches (or any other person who is in a power relationship with them), and where the identity of participants or their organisation may be inferred?	☐	☐
9	Any situation which may put the researcher at risk of harm? (E.g. gathering data in private homes)?	☐	☐
10	The use of previously collected biological samples or identifiable personal information for which there was no explicit consent for this research?	☐	☐
11	Any matters of commercially sensitive information?	☐	☐
12	Any financial interest in the outcome of the research by any member(s) of the research team?	☐	☐
13	People who are not giving consent to be part of the study, or the use of any deception, concealment or covert observations in non-public places, including social media?	☐	☐
14	Participants who are in a dependent or unequal relationship with any member(s) of the research team (e.g. where the researcher is a lecturer/ teacher/ health care provider/ coach/ employer/ manager/ or relative etc.) of any of the participants?	☐	☐

[1] If "No" is checked to all items 2-14, the application's status as Minimal Risk will be checked by the Secretariat, and may be forwarded to expedited review. Applications with more than Minimal Risk (any one "yes" to questions 2-14 above), and applications where the checklist is not completed will appear on AUTEC's next agenda.

Section A–2: Participant Information Sheet

Participant Information Sheet

Date Information Sheet Produced:

xx November 2021

Project Title

Non-Contact Measurement of Heart Rate using Video Recordings

An Invitation

Kia Ora (Hello), I am Gaganjot Kaur, a student at AUT within the School of Engineering, and I am presently conducting this research to complete my PhD degree. I want to invite you to participate in my research investigating methods for the Non-contact measurement of Heart Rate. You are welcome to volunteer to participate in my research for any of you who have not got facial tattoos and scares.

What is the purpose of this research?

This research aims to measure the heart rate without any physical contact with the participants. Several videos are to be recorded by using a camera placed on a tripod set at a 2-metre distance from the participant's face. The recorded videos will be analysed offline later to find the heart rate using MATLAB software advanced image processing techniques. A total of fifty healthy participants will volunteer to be part of this research aged between 18 and 35 years of age.

One of the research areas in the School of Engineering within the Faculty of Engineering, Computing and Mathematical Sciences is the AUT Biodesign Lab (https://biodesign.aut.ac.nz/). The proposed research is in coherence with other ongoing research projects.

This research will theoretically and experimentally use video recordings and image processing to measure heart rate and towards the fulfilment of my PhD degree. The findings of this research are likely to be used for academic publications and presentations.

How was I identified, and why am I being invited to participate in this research?

The participants would have contacted the primary researcher (Gagnjot Kaur), and then they will be screened for inclusion and exclusion criteria. The criteria to successfully be selected as a participant for this research is to be in the 18-35 years age group with no face tattoos and scares. Before data collection, the primary researcher will give a live demo of what the participant is are required to do. The participant will be a chance to have any remaining questions answered before written consent is obtained.

Note: Any students of Dr Jeff Kilby may not participate in this research.

How do I agree to participate in this research?

Participation in this research is voluntary (your choice), and whether or not to choose to participate will neither advantage nor disadvantage the participant. The participant may withdraw from the study at any time. Suppose the participant decides to withdraw from the study. In that case, they will be offered the choice between having any identifiable data belonging removed or continuing to be used. However, once the findings have been produced, removal of the participant data may not be possible.

What will happen in this research?

The research will take place at the AUT city campus. The participant will be provided with an explanation and demo of the procedure upon arrival before data collection, i.e., recording any videos. Once this step is complete, the participant will be asked to complete a consent form to proceed. The participant will then prepare by sitting upright on a chair in front of a camera see Figure 1, and a pulse oximeter will be attached to the dominant hand's index finger.

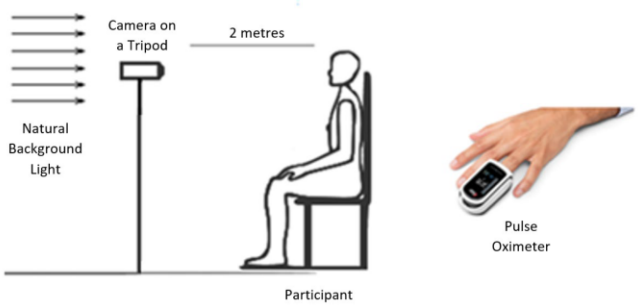


Figure 1: Experimental Setup

The participant will have the simple task of sitting straight on a chair; this will be demonstrated to the participant by the primary researcher after the explanation process. 2-3 test runs will be required to ensure that the data collected is consistent throughout.

What are the discomforts and risks?

There might be Covid-19 risk, but the following procedure will follow to ensure the safety of all participants and researchers.

- Before giving the appointment to the participant will confirm the COVID vaccination under AUT guidelines.
- The data is to be collected by giving an appointment to only one participant at a time. There will be 10 minutes break for cleaning/wiping the pulse oximeter and chair with antibacterial wipes before the next participant's appointment.
- There will be a 2-metre distance.
- Researcher who will collect the data will wear a face mask and gloves at all times.

How will these discomforts and risks be alleviated?

AUT Student Counselling and Mental Health can offer three free sessions of confidential counselling support for adult participants in an AUT research project. These sessions are only available for issues that have arisen directly due to participation in the research and are not for other general counselling needs. To access these services, you will need to:

- drop into our centre at WB203 City Campus, email counselling@aut.ac.nz or call 921 9998.
- let the receptionist know that you are a research participant and provide the title of my research and my name and contact details as given in this Information Sheet.

You can find out more information about AUT counsellors and counselling on https://www.aut.ac.nz/student-life/student-support/counselling-and-mental-health

What are the benefits?

The benefit of the research is to complete my PhD studies and develop research within the AUT Biodesign Lab in the School of Engineering, Computing and Mathematical Sciences.

What compensation is available for injury or negligence?

In the unlikely event of a physical injury resulting from your participation in this study, rehabilitation and compensation for injury by accident may be available from the Accident Compensation Corporation, providing the incident details satisfy the requirements of the law and the Corporation's regulations.

How will my privacy be protected?

All written Consent Forms are to be scanned, after which they are shredded, and the recorded videos are to be encrypted and stored on an external hard drive; for six years. The external hard drive is to be kept in a secured location (i.e. a locked cabinet) in Dr Jeff Kilby's office at AUT University City Campus. After six years, all data collected will be destroyed by erasing all the data from the hard drive.

What are the costs of participating in this research?

The cost of participating in this research is the participant's willingness to give up approximately 20 minutes of their time to take part in this research.

What opportunity do I have to consider this invitation?

The participant will have the opportunity to participate until the end of January 2022.

Will I receive feedback on the results of this research?

In the Consent Form, questions will be asked whether the participant wants feedback on their results obtained from this research. This will be given in a short report (one or two pages) to show their results are reflected and compared with the overall results.

What do I do if I have concerns about this research?

Participants may withdraw from this research at any time.

Any concerns regarding the nature of this project should be notified in the first instance to the Project Supervisor, Dr Jeff Kilby, jeffrey.kilby@aut.ac.nz, 09 921 9999 ext 8748

Any concerns regarding the conduct of the research should be notified to the Executive Secretary of AUTEC, *ethics@aut.ac.nz*, (+649) 921 9999 ext 6038.

Whom do I contact for further information about this research?

Would you please keep this Information Sheet along with a copy of the Consent Form for your future reference? You are also able to contact the research team as follows:

Researcher Contact Details:

Gaganjot Kaur, *gaganjot.kaur@aut.ac.nz*

Project Supervisor Contact Details:

Dr Jeff Kilby, *jeffrey.kilby@aut.ac.nz*, 09 921 9999 ext 8748

Approved by the Auckland University of Technology Ethics Committee on *type the date final ethics approval was granted*, AUTEC Reference number *type the reference number*.

Section A–3: Consent Form

Consent and Release Form

Project title: *Non-contact Measurement of Heart Rate using Video Recordings*

Project Supervisor: *Dr Jeff Kilby*

Researcher: *Gaganjot Kaur*

- ○ I have read and understood the information provided about this research project in the Information Sheet dated xx November 2021.
- ○ I have had an opportunity to ask questions and to have them answered.
- ○ I understand that taking part in this study is voluntary (my choice) and that I may withdraw from the study at any time without being disadvantaged in any way.
- ○ I understand that if I withdraw from the study, I will be offered the choice between having any identifiable data as belonging to me removed or allowing it to continue to be used. However, once the findings have been produced, removal of my data may not be possible. I permit the researcher to use the recorded videos that are part of this research project and/or any images from them and any other reproductions or adaptations.
- ○ I understand that the videos will be used for academic purposes only and will not be published in any form outside of this research project without my written permission.
- ○ I understand that any copyrighted material created by the video sessions is deemed to be owned by the researcher. I do not own the copyright of any of the photographs.
- ○ I agree to take part in this research project.

Participants Signature: .....

Participants Name: .....

Participants Contact Details (if appropriate):

.....

.....

.....

Date: .....

Approved by the Auckland University of Technology Ethics Committee on type the date on which the final approval was granted AUTEC Reference number type the AUTEC reference number

Note: The Participant should retain a copy of this form

Volunteers Needed

For Research in Non-Contact Measurement of Heart Rate

Be a part of AUT's latest research in Signal and Systems Group of the School of Engineering, Computer and Mathematical Sciences

This is a research investigated methods for Non-Contact Measurement of Heart Rate, using camera to record the face video and attach pulse oximeter for measurement of Heart Rate at the AUT City Campus.

Participants must be between 18 to 35 years of age, It will take up to 20 minutes of your time.

You are welcome to participate in this research

For further details please contact Gaganjot Kaur contact either by email - gaganjot.kaur@aut.ac.nz or mobile phone 022 650 7399.

AUT
Engineering, computer and mathematical sciences

AUT
Engineering, computer and mathematical sciences

Section A–5: Acceptance Letter Granting Ethical Approval

Auckland University of Technology Ethics Committee (AUTEC)

Auckland University of Technology
D-88, Private Bag 92006, Auckland 1142, NZ
T: +64 9 921 9999 ext. 8316
E: ethics@aut.ac.nz
www.aut.ac.nz/researchethics

9 December 2021

Jeffrey Kilby
Faculty of Health and Environmental Sciences

Dear Jeffrey

Re Ethics Application: **21/424 Non-Contact Measurement of Heart Rate using Video Recordings**

Thank you for providing evidence as requested, which satisfies the points raised by the Auckland University of Technology Ethics Committee (AUTEC).

Your ethics application has been approved for three years until 9 December 2024.

Non-Standard Conditions of Approval

the video protocol is inadequate and does not describe step by step how videoing will occur, where it will occur, and what is involved. Having said that, the Information Sheet contains a picture that says most of that. So approve it with a comment about doing a better protocol.

1.

Non-standard conditions must be completed before commencing your study. Non-standard conditions do not need to be submitted to or reviewed by AUTEC before commencing your study.

Standard Conditions of Approval

1. The research is to be undertaken in accordance with the Auckland University of Technology Code of Conduct for Research and as approved by AUTEC in this application.
2. A progress report is due annually on the anniversary of the approval date, using the EA2 form.
3. A final report is due at the expiration of the approval period, or, upon completion of project, using the EA3 form.
4. Any amendments to the project must be approved by AUTEC prior to being implemented. Amendments can be requested using the EA2 form.
5. Any serious or unexpected adverse events must be reported to AUTEC Secretariat as a matter of priority.
6. Any unforeseen events that might affect continued ethical acceptability of the project should also be reported to the AUTEC Secretariat as a matter of priority.
7. It is your responsibility to ensure that the spelling and grammar of documents being provided to participants or external organisations is of a high standard and that all the dates on the documents are updated.
8. AUTEC grants ethical approval only. You are responsible for obtaining management approval for access for your research from any institution or organisation at which your research is being conducted and you need to meet all ethical, legal, public health, and locality obligations or requirements for the jurisdictions in which the research is being undertaken.

Please quote the application number and title on all future correspondence related to this project.

For any enquiries please contact ethics@aut.ac.nz. The forms mentioned above are available online through http://www.aut.ac.nz/research/researchethics

(This is a computer-generated letter for which no signature is required)

The AUTEC Secretariat
Auckland University of Technology Ethics Committee

Cc: , gaganjot.kaur@aut.ac.nz; Hakilo.Sabit@aut.ac.nz

Appendix B: Record of Data Collection

B-1: Data for the main study

B-2: Data sheet of Pulse Oximeter

B-1: Data for the main study

Participant Number	File Name	Age	Heart Rate (bpm)		
			Maximum	Minimum	Average
1	V001	33	82	75	79
2	V002	30	70	65	68
3	V003	32	82	75	79
4	V004	28	96	90	93
5	V005	27	92	85	89
6	V006	24	72	65	69
7	V007	29	107	98	103
8	V008	24	71	65	68
9	V009	26	74	68	71
10	V010	31	100	87	94
11	V011	28	75	70	73
12	V012	25	79	65	72
13	V013	33	84	78	81
14	V014	35	77	69	73
15	V015	34	80	72	76
16	V016	22	78	70	74
17	V017	34	75	68	72
18	V018	35	72	65	69
19	V019	29	87	80	84
20	V020	22	80	75	78
21	V021	26	80	75	78
22	V022	31	71	65	68
23	V023	32	82	77	80
24	V024	29	79	70	75
25	V025	27	87	76	82
26	V026	26	80	73	77
27	V027	25	92	85	89
28	V028	32	90	83	87
29	V029	26	86	79	83
30	V030	29	87	80	84
31	V031	25	77	70	74
32	V032	28	75	68	72
33	V033	30	72	65	69
34	V034	33	68	61	65
35	V035	21	69	62	66

B-2: Data sheet of Pulse Oximeter

EMAY

Sleep Oxygen Report

Name:	Participant 1
Age:	35
Sex:	female
Duration:	00:01:20

	Max	Avg	Min
SpO2	98	97	95
PR	82	79	75

SpO2	Duration	%Total	Events (ODI4%)
94-100	00:01:20	100%	--
88-93	00:00:00	0%	--
80-87	00:00:00	0%	--
70-79	00:00:00	0%	--
<70	00:00:00	0%	--
Total	00:01:20	100%	--

ODI 4%	--
Total ODI 4% Events	--
Time in ODI 4% Events	--
% of Time in ODI 4% Events	--
Avg ODI 4% Event Duration	--

Definition : ODI 4% (Oxygen Desaturation Index) is the number of times per hour of sleep that the blood's oxygen level drop by at least 4% from baseline.

ODI 4% Event: a fall in oxygen saturation of at least 4% and persisting more than 10 seconds.

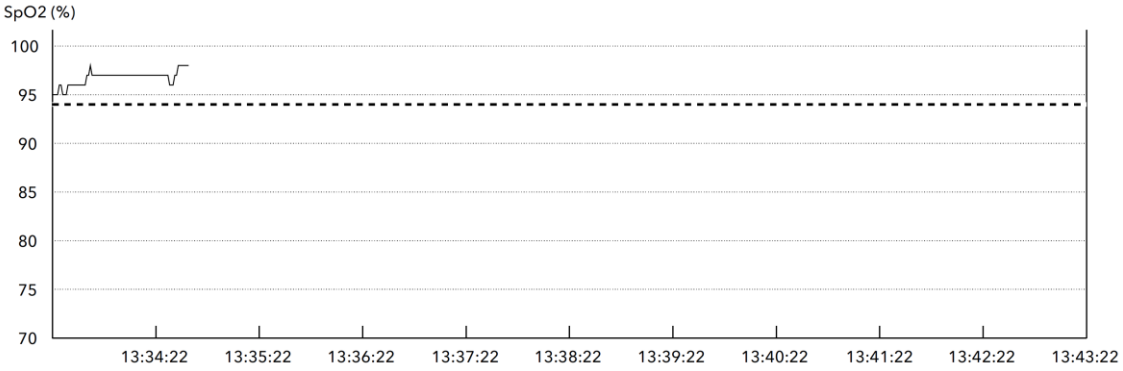
ODI 3%	--
Total ODI 3% Events	--
Time in ODI 3% Events	--
% of Time in ODI 3% Events	--
Avg ODI 3% Event Duration	--

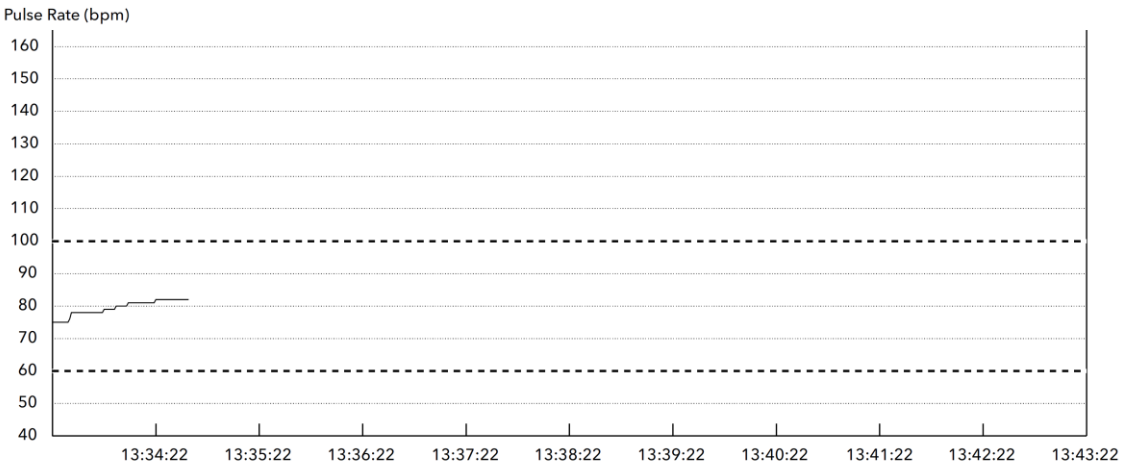
Definition : ODI 3% (Oxygen Desaturation Index) is the number of times per hour of sleep that the blood's oxygen level drop by at least 3% from baseline.

ODI 3% Event: a fall in oxygen saturation of at least 3% and persisting more than 10 seconds.

SpO2 Threshold:94	Duration	%Total
>=94	00:01:20	100%
<94	00:00:00	0%
Pulse Rate Threshold	**Duration**	**%Total**
Max:100 Min:60		
>100	00:00:00	0%
<60	00:00:00	0%

Definition : Threshold is used to differentiate normal and abnormal SpO2 or PR readings. Your can adjust threshold under "Settings" on the app





Appendix C: Detected Frames

C-1: Simulation results of Video V007

C-2: Simulation results of Video V015

C-1: Results of video V007 of participant number 07

Figure C1 shows the simulation panel, where (A) is the original input video (B) is the processing of detection of face from each frame (C) Detected face frame (D) selection of region of interest (forehead) and (E) Optimized Region of interest and (F) is the feature extraction using Histogram Oriented Gradients

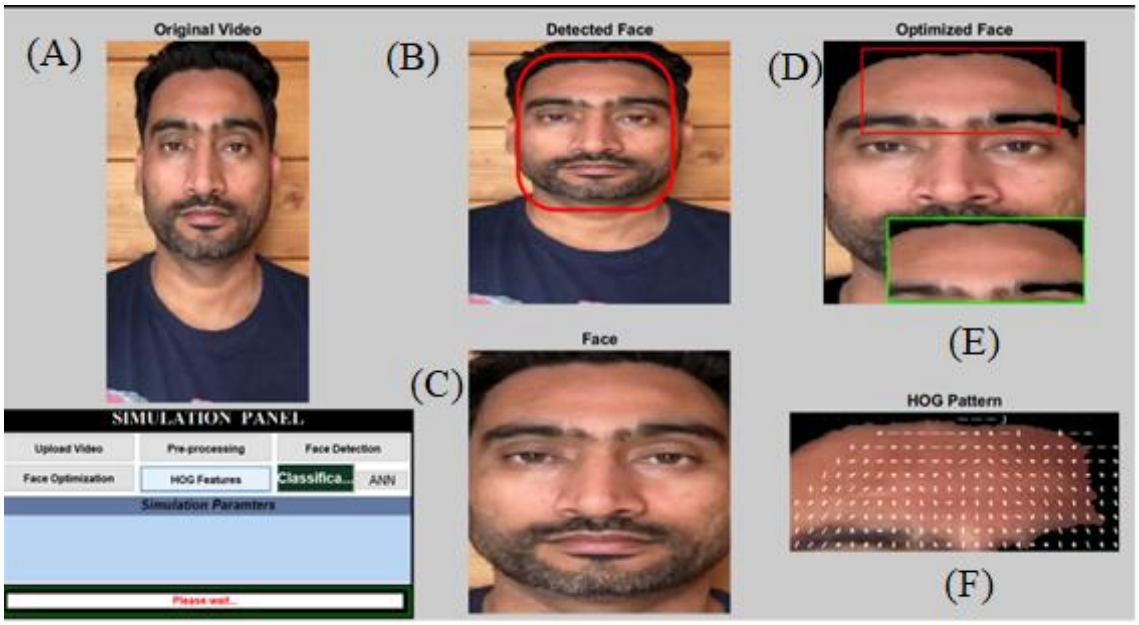


Figure C.1 MATLAB Simulation Panel

Figure C.2 Frames extracted from input video (Frame number 70 to 80)

Figure C.3 Face detection frames from extracted frames (Frame number 70 to 80)

Figure C.4 Detection of forehead from face frames (Frame number 70 to 80)

Figure C.5 Forehead frames after optimization (Frame number 70 to 80)

C-2: Results of video V015 of participant number 15

Figure C6 shows the simulation panel, where (A) is the original input video (B) is the processing of detection of face from each frame (C) Detected face frame (D) selection of region of interest (forehead) and (E) Optimized Region of interest and (F) is the feature extraction using Histogram Oriented Gradients

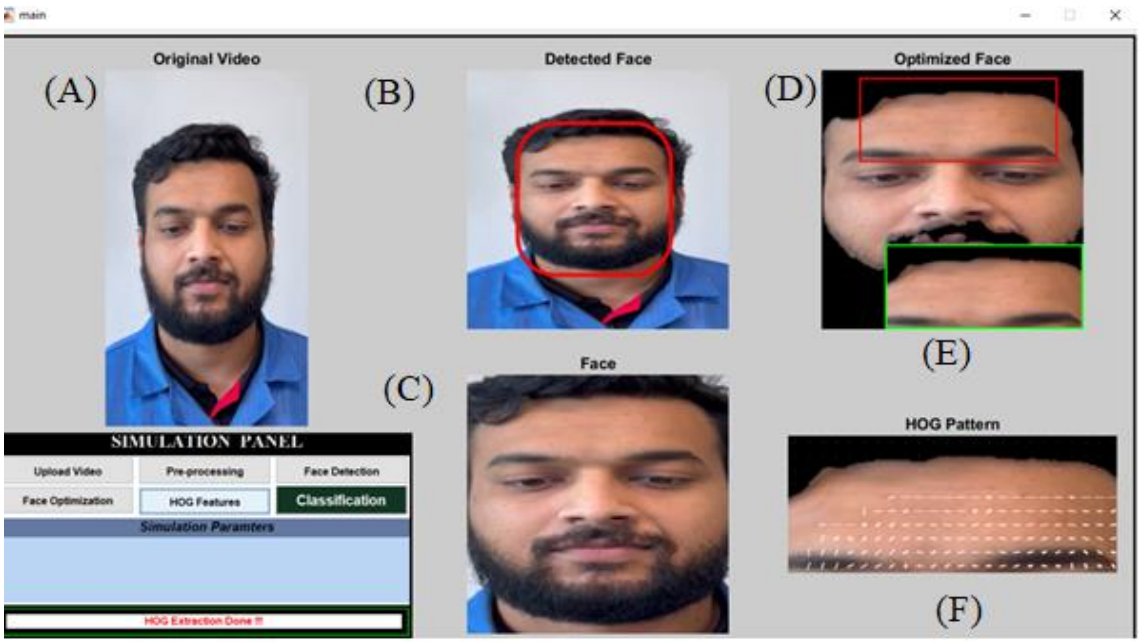


Figure C.6 Simulation Panel with HOG feature extraction

Figure C.7 Frames extracted from input video

Figure C.8 Face detection frames from extracted frames

Figure C.9 Detection of forehead from face frames

Figure C.10 Forehead frames after optimization

Appendix D: MATLAB Code for Training

D-1: ANN code

D-2: Training state

D-3: Error Histogram

D-4: Confusion Matrix

D-5: Receiver Operating Characteristics

D-1: ANN code

```
TrainingData=[AllNormal; AllModerate; AllHigh]; % normal 60-80, moderate 81-100
and high 101-120
Target=[];
for i=1:size(TrainingData,1)
   if i<=size(AllNormal,1)
      Target(1,i)=1;
      Label(i)=1;
   elseif i>size(AllNormal,1) & i<=size(AllNormal,1)+size(AllModerate,1)
      Target(2,i)=1;
      Label(i)=2;
   elseif i>size(AllNormal,1)+size(AllModerate,1)
      Target(3,i)=1;
      Label(i)=3;
   end
end
[rows,cols]=size(AllNormal);
net=patternnet(10);
net.trainParam.epochs=1000;
net.divideParam.trainRatio = 70/100; % 70 percent data for training
net.divideParam.valRatio = 15/100;   %15 percent for Validation
net.divideParam.testRatio = 15/100; % 15 percent for test
net=train(net,TrainingData',Target);
total_frames=numel(AllHOG);
TestData=[];
for i=1:total_frames
   cr=AllHOG{i};
   for j=1:cols
      TestData(i,j)=cr(1,j);
   end


end
%TestData=TrainingData';
Result=sim(net,TestData);
[MatchingRatio,Category]=max(Result);
[C_matrix,Result,RefereceResult]=ClassificationScore(Label,Category);
```

D-2: Training state

Figure D1 shows the neural network training state of at epoch 23.

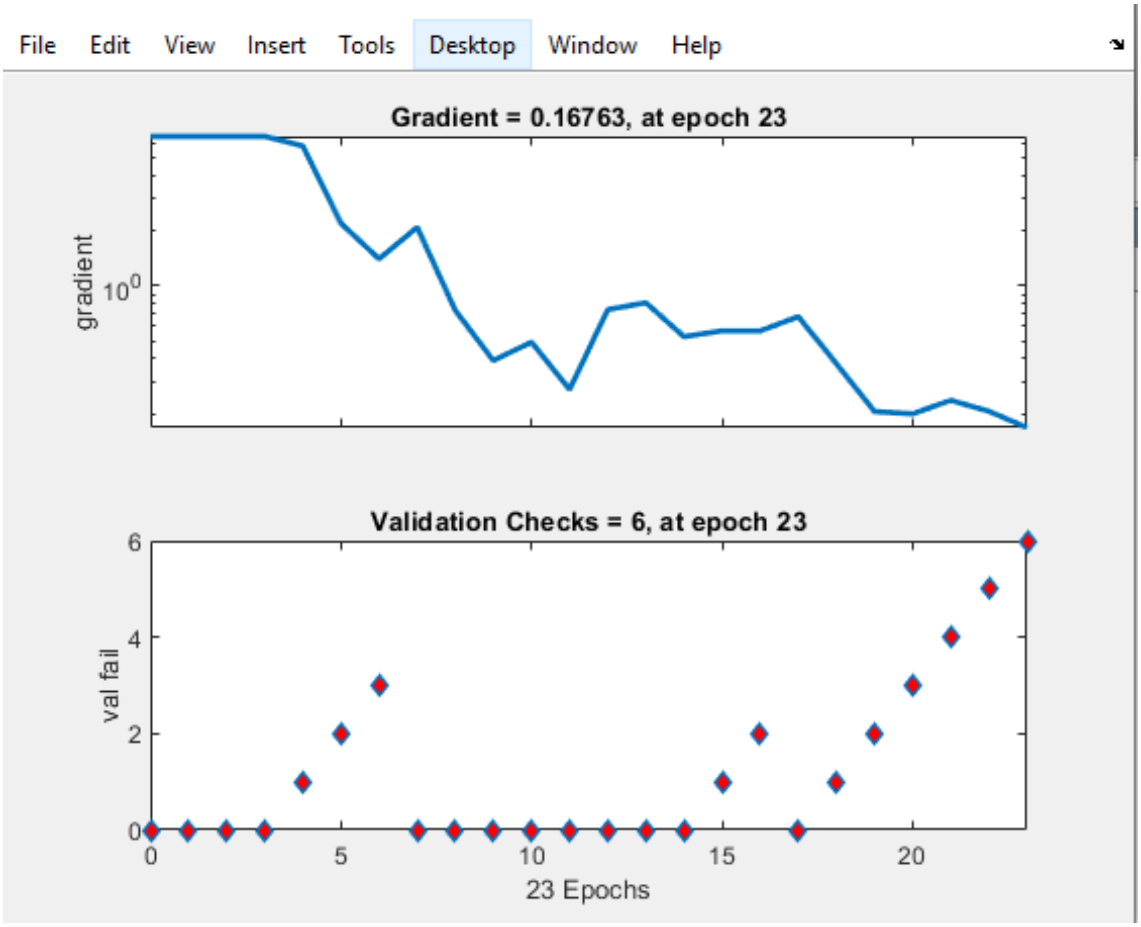


Figure D1 Neural Network training state

D-3: Error Histogram

Figure D2 shows the neural network training error histogram at epoch 23.

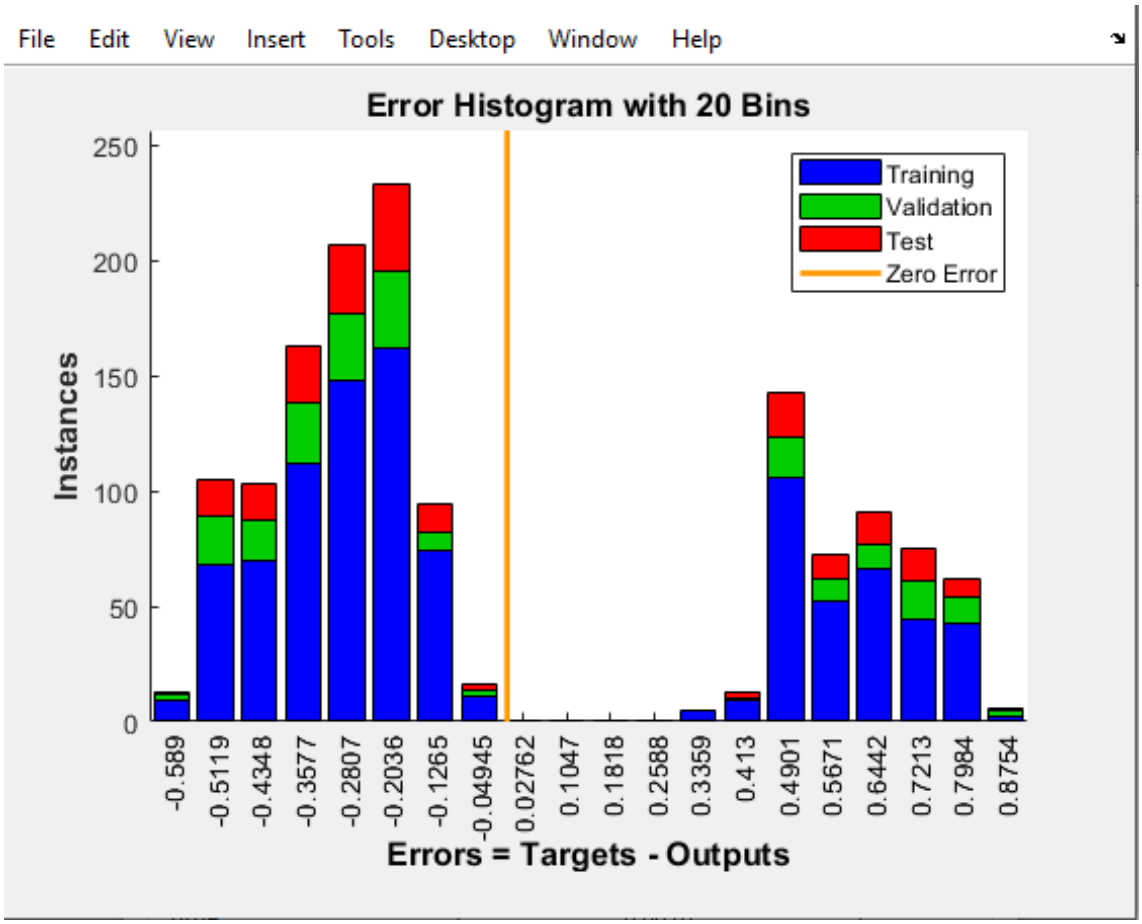


Figure D2 Neural Network training error histogram

D-4: Confusion Matrix

Figure D3 shows the neural network training confusion matrix which shows how the data is separated into classes and output and target class for training, validation and test are shown.

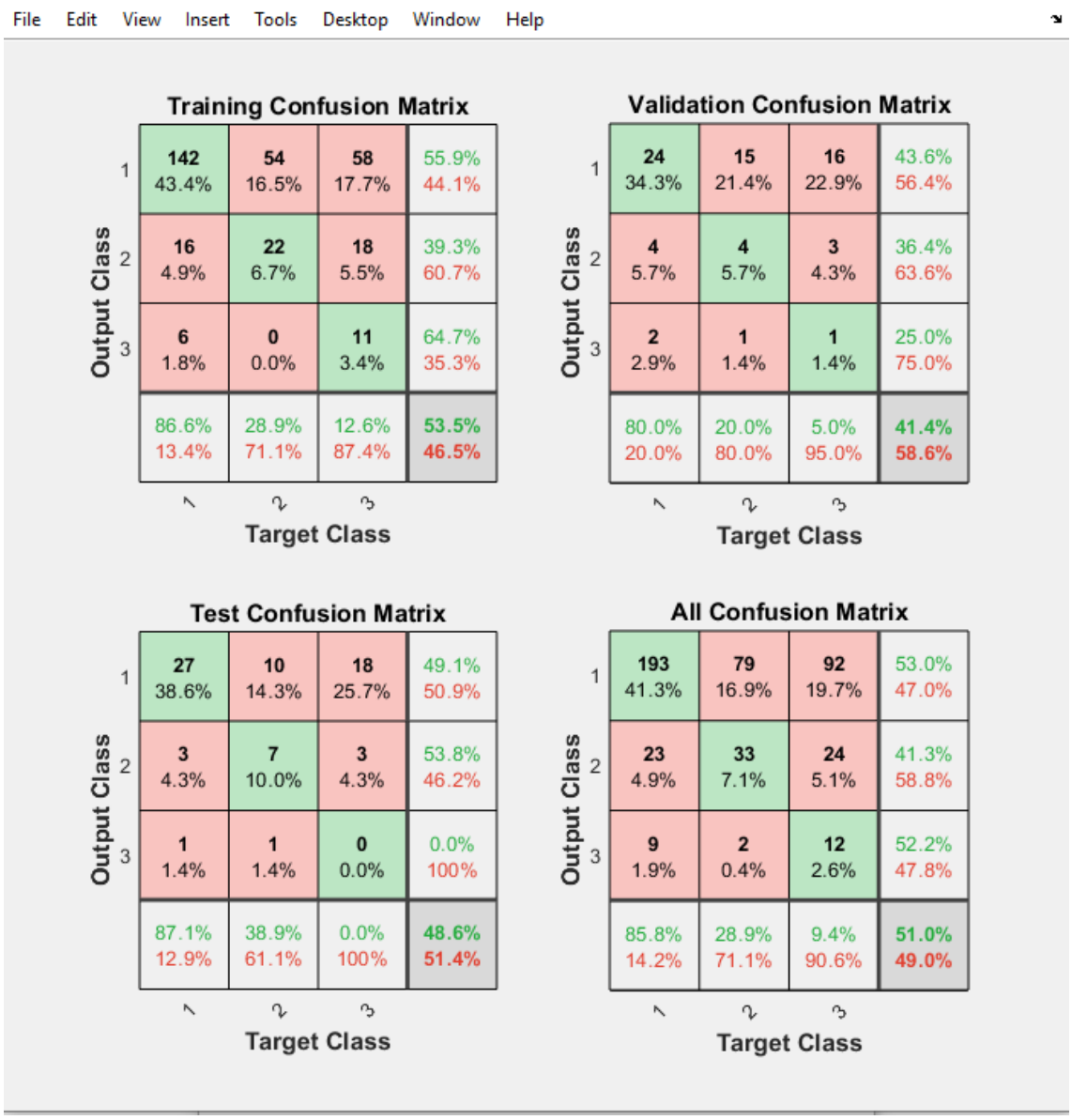


Figure D3 Neural Network training confusion matrix

D-5: Receiver Operating Characteristics

Figure D4 shows the Neural Network training receiver operating characteristics (ROC) for all classes for training, validation and test ROC for true positive rate.

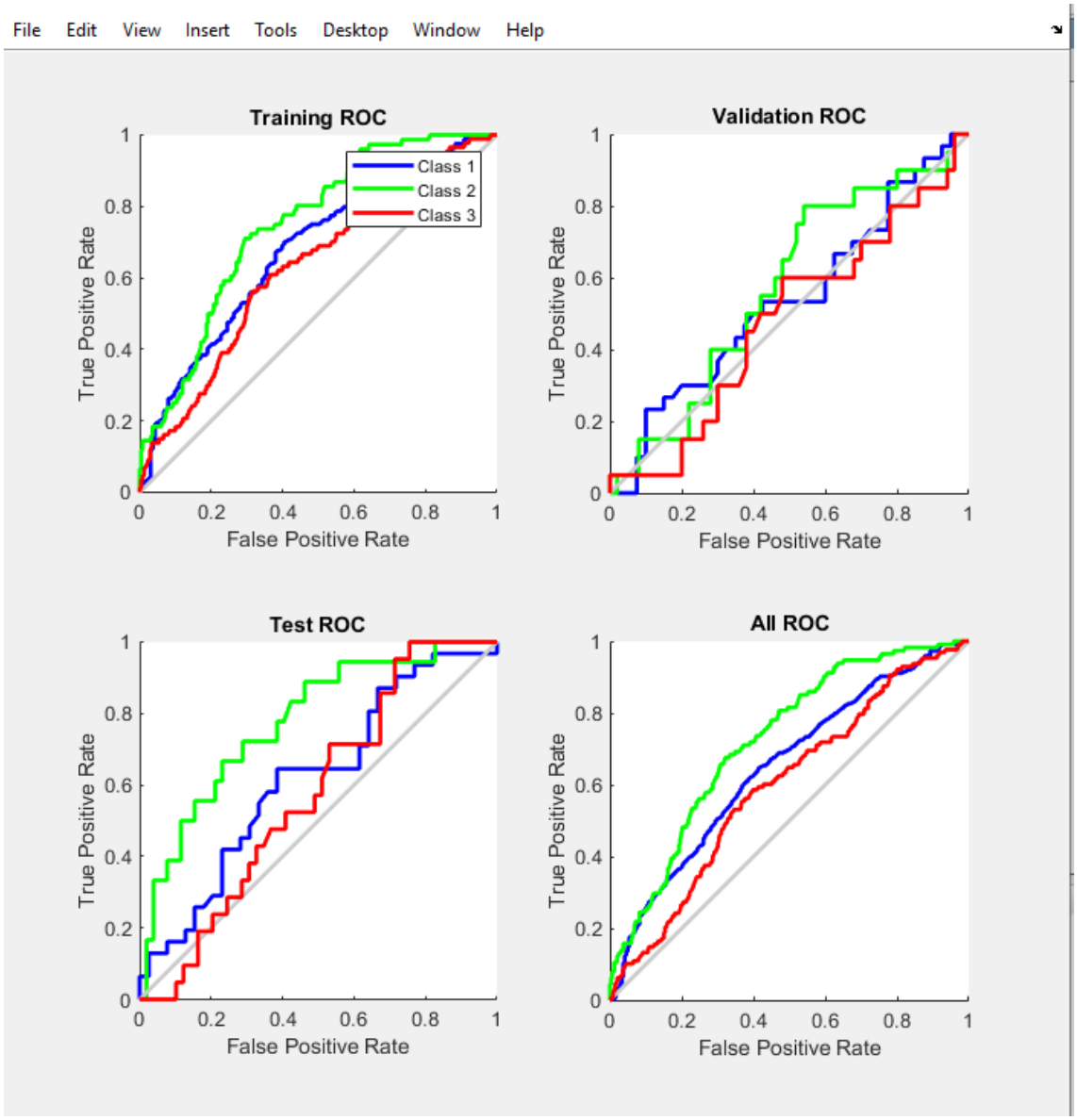


Figure D4 Neural Network training receiver operating characteristics

Appendix E: Results

Section E–1: Using 5 Layered Neural Architecture

Section E–2: Using 15 Layered Neural Architecture

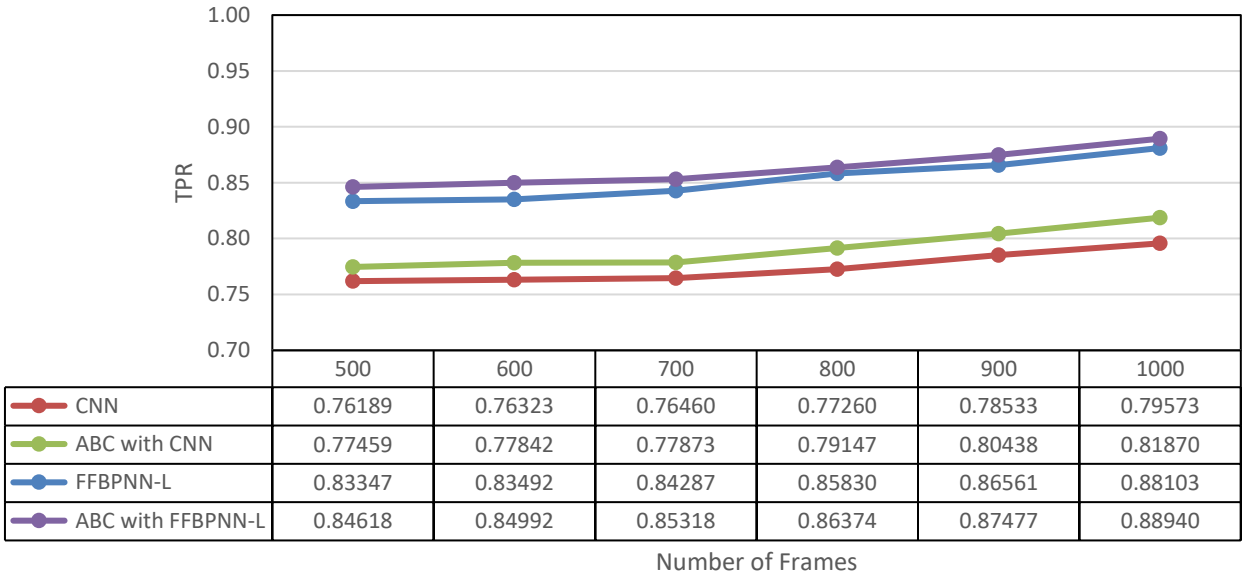
E-1: Using 5 Layered Neural Architecture

In this section the performance parameters of proposed algorithms are been evaluated and displayed in accordance with the 5 hidden layers in the neural network architecture considering different ratios of training and testing data sets.

E1.1 For 70:30 ratio

In this, the videos dataset is divided into 70:30 ratio where 70% of the dataset is used for training and 30% for testing. The techniques used are Convolutional Neural Network (CNN), Artificial Bee Colony (ABC) with CNN, Feed Forward Back Propagation Neural Network-Levenberg (FFBPNN-L), and ABC with FFBPNN-L.

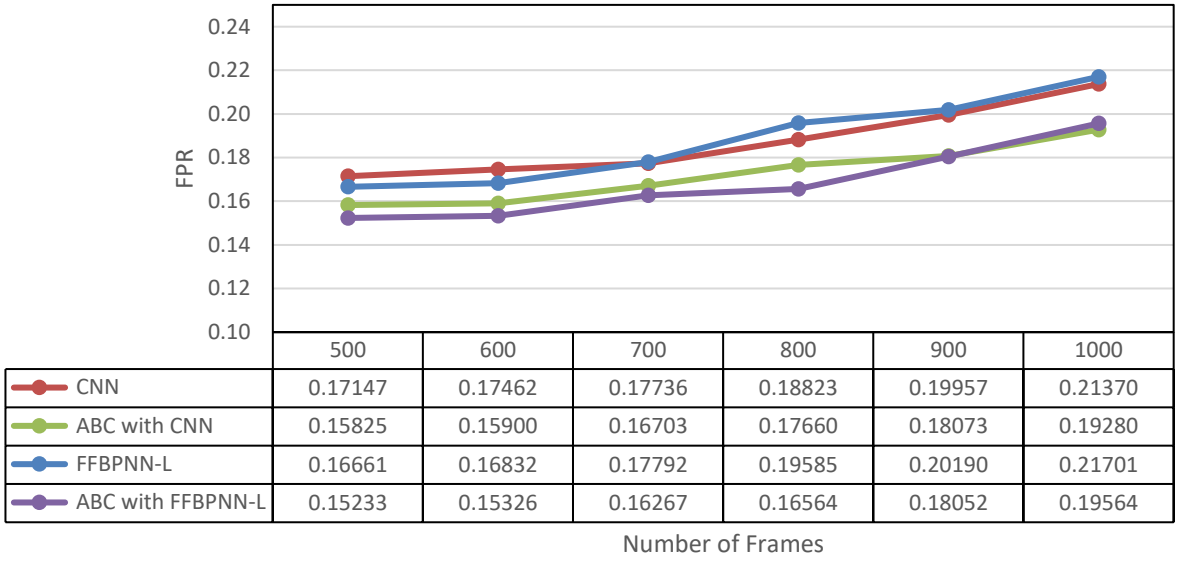
Figure E1 displays the True Positive Rate (TPR) analysis of CNN, ABC with CNN, FFBPNN-L and ABC with FFBPNN-L. It is observed that with increase in number of frames, the TPR of both the techniques has been gradually increased. The average TPR of CNN is 0.77, ABC with CNN is 0.79, FFBPNN-L is 0.85 and ABC with FFBPNN-L is 0.86. It is obsevered that proposed method FFBPNN-L have more TPR as compare to other algorithms.



	500	600	700	800	900	1000
CNN	0.76189	0.76323	0.76460	0.77260	0.78533	0.79573
ABC with CNN	0.77459	0.77842	0.77873	0.79147	0.80438	0.81870
FFBPNN-L	0.83347	0.83492	0.84287	0.85830	0.86561	0.88103
ABC with FFBPNN-L	0.84618	0.84992	0.85318	0.86374	0.87477	0.88940

Figure E1 TPR analysis of CNN, ABC with CNN, FFBPNN-L and ABC with FFBPNN-L

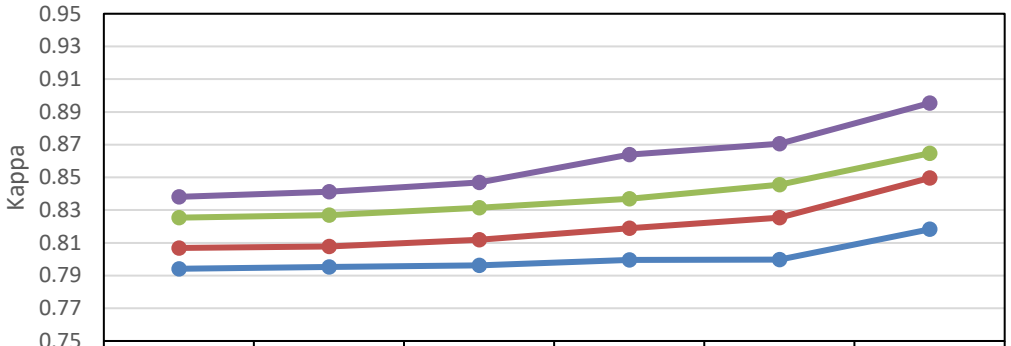
Figure E2 displays the False Positive Rate (FPR) analysis of CNN, ABC with CNN, FFBPNN-L and ABC with FFBPNN-L. The average FPR of CNN is 0.18, ABC with CNN is 0.17, FFBPNN-L is 0.18 and ABC with FFBPNN-L is 0.16. It is observed that FPR is less in the proposed method FFBPNN-L.



	500	600	700	800	900	1000
CNN	0.17147	0.17462	0.17736	0.18823	0.19957	0.21370
ABC with CNN	0.15825	0.15900	0.16703	0.17660	0.18073	0.19280
FFBPNN-L	0.16661	0.16832	0.17792	0.19585	0.20190	0.21701
ABC with FFBPNN-L	0.15233	0.15326	0.16267	0.16564	0.18052	0.19564

Figure E2 FPR analysis of CNN, ABC with CNN, FFBPNN-L and ABC with FFBPNN-L

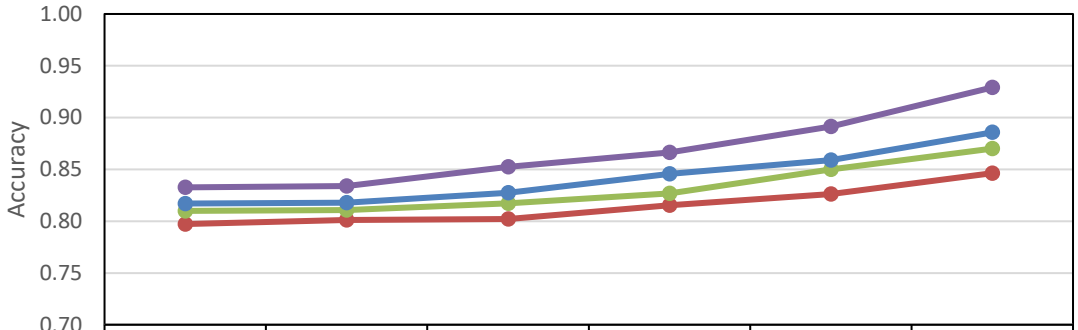
Figure E3 is the Kappa analysis of CNN, ABC with CNN, FFBPNN-L and ABC with FFBPNN-L. The curve of both the algorithms had increased gradually. The average Kappa of CNN is 0.80, ABC with CNN is 0.82, FFBPNN-L is 0.83 and ABC with FFBPNN-L is 0.85.



	500	600	700	800	900	1000
CNN	0.79415	0.79516	0.79624	0.79946	0.79966	0.81822
ABC with CNN	0.80686	0.80769	0.81183	0.81887	0.82529	0.84960
FFBPNN-L	0.82538	0.82697	0.83139	0.83692	0.84552	0.86471
ABC with FFBPNN-L	0.83809	0.84118	0.84694	0.86395	0.87064	0.89536

Figure E3 Kappa analysis of of CNN, ABC with CNN, FFBPNN-L and ABC with FFBPNN-L

Figure E4 is the Accuracy analysis of CNN, ABC with CNN, FFBPNN-L and ABC with FFBPNN-L. The curve of both the algorithms had increased gradually. The average accuracy of CNN is 0.81, ABC with CNN is 0.83, FFBPNN-L is 0.84 and ABC with FFBPNN-L is 0.86. Result compared to proposed algorithm ABC with FFBPNN-L has a more accuracy.



	500	600	700	800	900	1000
CNN	0.79721	0.80107	0.80208	0.81533	0.82602	0.84632
ABC with CNN	0.80992	0.81057	0.81712	0.82665	0.84988	0.87005
FFBPNN-L	0.81702	0.81781	0.82752	0.84549	0.85883	0.88569
ABC with FFBPNN-L	0.83264	0.83377	0.85250	0.86619	0.89150	0.92905

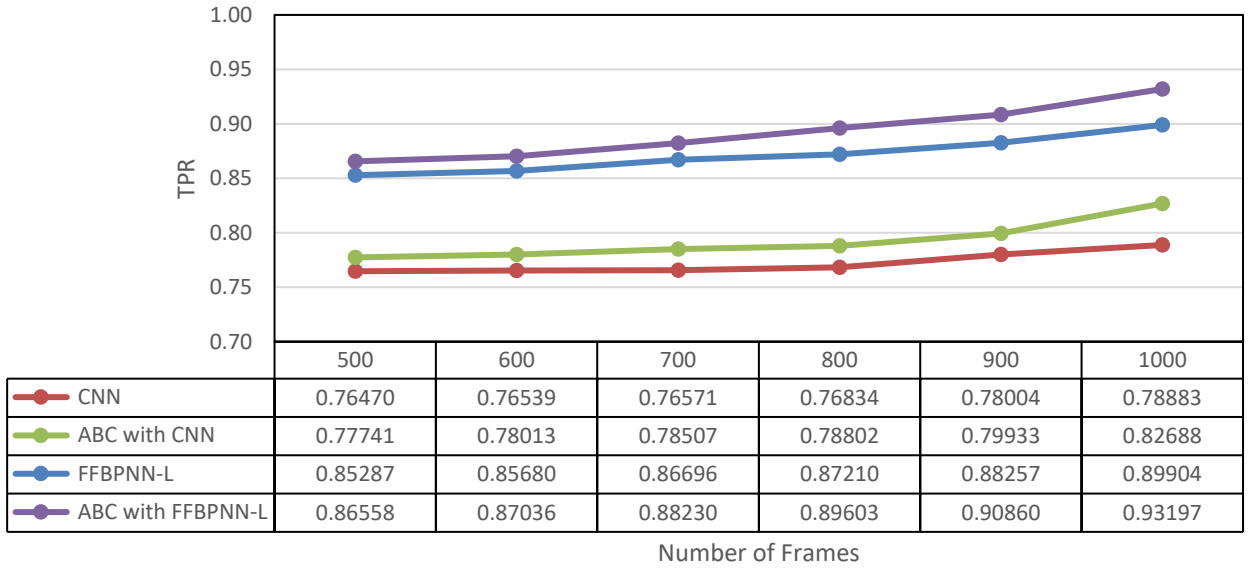
Number of Frames

Figure E4 Accuracy of CNN, ABC with CNN, FFBPNN-L and ABC with FFBPNN-L

E1.2 For 80:20 ratio

In this part, the dataset is divided into 80:20 ratio where 80% of the dataset is used for training and 20% of the dataset is used for testing

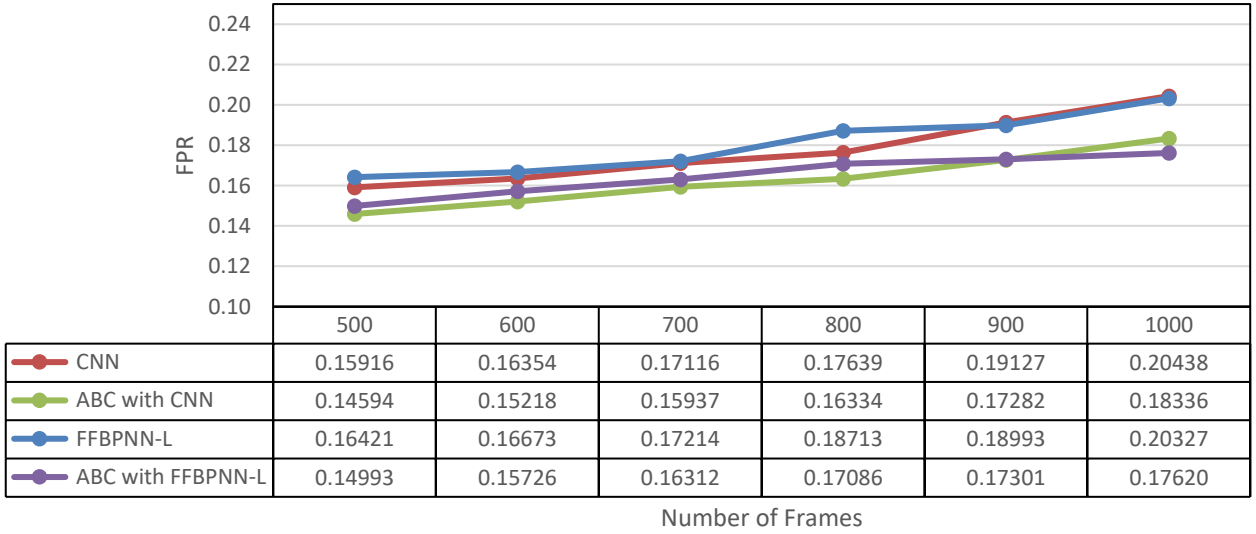
The TPR analysis of CNN and ABC with CNN, FFBPNN-L and ABC with FFBPNN-L is shown in the Figure E5. The TPR have gradually increased with the number of frames. The average TPR of CNN is 0.77 and ABC with CNN is 0.79 and FFBPNN-L is 0.87 whereas ABC with FFBPNN-L is 0.89.



	500	600	700	800	900	1000
CNN	0.76470	0.76539	0.76571	0.76834	0.78004	0.78883
ABC with CNN	0.77741	0.78013	0.78507	0.78802	0.79933	0.82688
FFBPNN-L	0.85287	0.85680	0.86696	0.87210	0.88257	0.89904
ABC with FFBPNN-L	0.86558	0.87036	0.88230	0.89603	0.90860	0.93197

Figure E5 TPR analysis of CNN and ABC with CNN and FFBPNN-L and ABC with FFBPNN-L

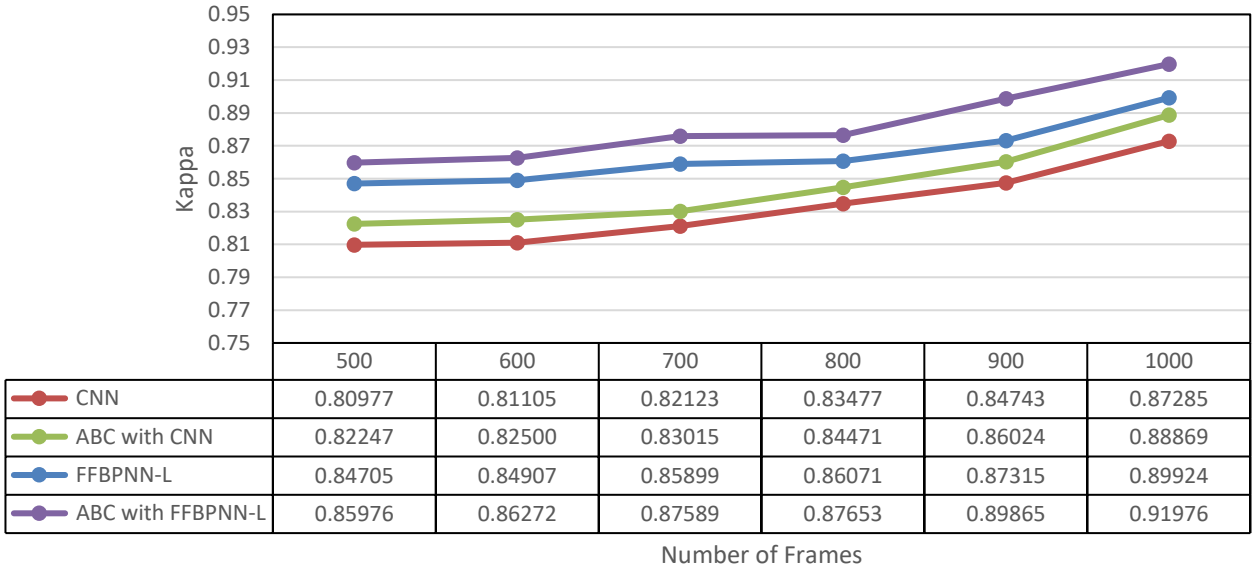
Figure E6 displays the FPR analysis of CNN and ABC with CNN, FFBPNN-L and ABC with FFBPNN-L. The curve of CNN started at 0.15 and ended at 0.20. The curve of ABC with CNN started at 0.14 and ended at 0.18. The average of CNN value of FPR is 0.17 and ABC with CNN is 0.16. Average of FPR of FFBPNN-L is 0.18 and ABC with FFBPNN-L is 0.16.



	500	600	700	800	900	1000
CNN	0.15916	0.16354	0.17116	0.17639	0.19127	0.20438
ABC with CNN	0.14594	0.15218	0.15937	0.16334	0.17282	0.18336
FFBPNN-L	0.16421	0.16673	0.17214	0.18713	0.18993	0.20327
ABC with FFBPNN-L	0.14993	0.15726	0.16312	0.17086	0.17301	0.17620

Figure E6 FPR analysis of CNN, ABC with CNN, FFBPNN-L and ABC with FFBPNN-L

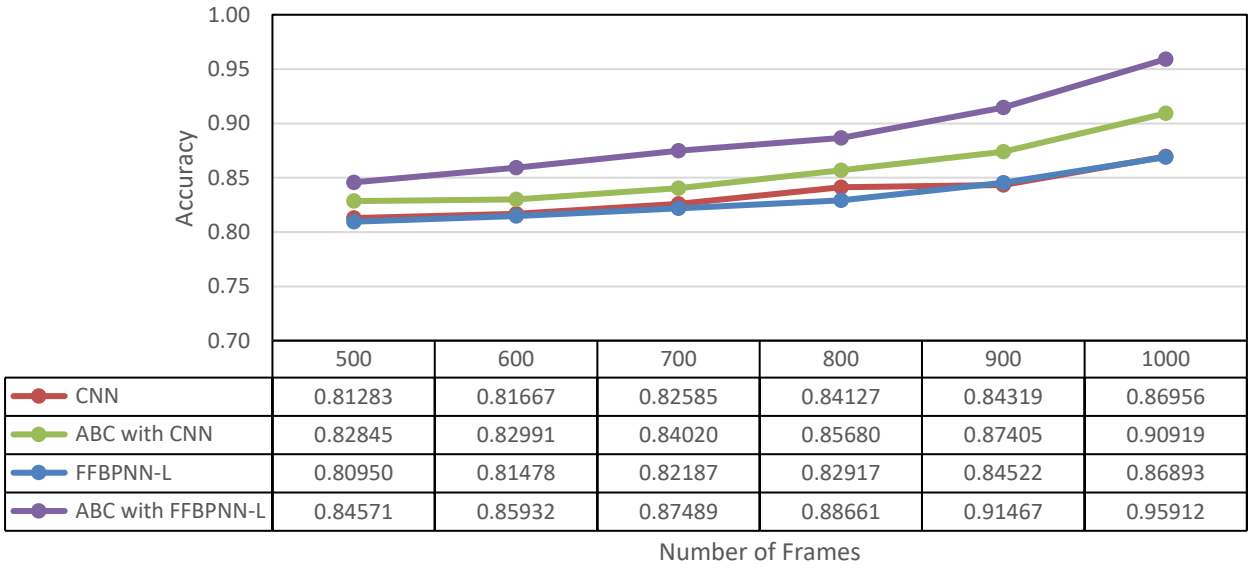
Figure E7 depicts the kappa analysis of algorithms CNN, ABC with CNN, FFBPNN-L and ABC with FFBPNN-L. It is observed that there is increase in curves with increase in number. The average values of Kappa for CNN 0.83, ABC with CNN 0.84, FFBPNN-L 0.86 and ABC with FFBPNN-L 0.88.



	500	600	700	800	900	1000
CNN	0.80977	0.81105	0.82123	0.83477	0.84743	0.87285
ABC with CNN	0.82247	0.82500	0.83015	0.84471	0.86024	0.88869
FFBPNN-L	0.84705	0.84907	0.85899	0.86071	0.87315	0.89924
ABC with FFBPNN-L	0.85976	0.86272	0.87589	0.87653	0.89865	0.91976

Figure E7 Kappa analysis of CNN, ABC with CNN, FFBPNN-L and ABC with FFBPNN-L

The accuracy of all algorithms is depicted in the line graph shown in Figure E8. The average accuracy of CNN is 0.83 and ABC with CNN is 0.85 its similar for FFBPNN-L 0.83 and 0.89 ABC with FFBPNN-L. It shows the accuracy values of all algorithms and ABC with FFBPNN-L had more accuracy.

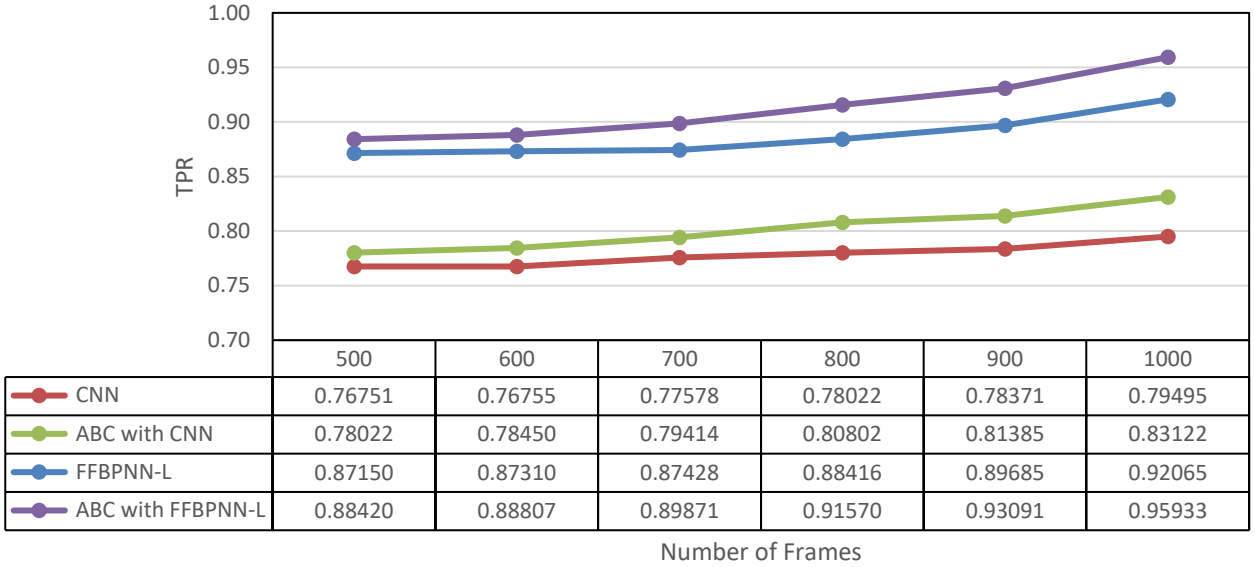


	500	600	700	800	900	1000
CNN	0.81283	0.81667	0.82585	0.84127	0.84319	0.86956
ABC with CNN	0.82845	0.82991	0.84020	0.85680	0.87405	0.90919
FFBPNN-L	0.80950	0.81478	0.82187	0.82917	0.84522	0.86893
ABC with FFBPNN-L	0.84571	0.85932	0.87489	0.88661	0.91467	0.95912

Figure E 8 Accuracy analysis of values of CNN, ABC with CNN, FFBPNN-L and ABC with FFBPNN-L

E1.3 For 90:10 ratio

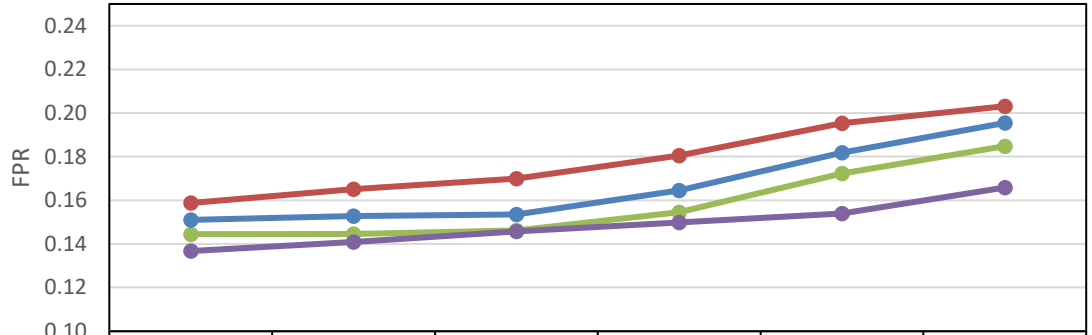
In this part, the dataset is divided into 90:10 ratio where 90% of the dataset is used for training and 10% of the dataset is used for testing. The Figure E9 displays the TPR analysis of all algorithms. It is observed that TPR had gradually increased. The average TPR of CNN is 0.77, ABC with CNN is 0.80, FFBPNN-L is 0.88 and ABC with FFBPNN-L is 0.91. As a result, as compared to the FFBPNN-L algorithm, ABC with FFBPNN-L has a TPR of 0.6% higher



	500	600	700	800	900	1000
CNN	0.76751	0.76755	0.77578	0.78022	0.78371	0.79495
ABC with CNN	0.78022	0.78450	0.79414	0.80802	0.81385	0.83122
FFBPNN-L	0.87150	0.87310	0.87428	0.88416	0.89685	0.92065
ABC with FFBPNN-L	0.88420	0.88807	0.89871	0.91570	0.93091	0.95933

Figure E 9 TPR analysis of CNN, ABC with CNN, FFBPNN-L and ABC with FFBPNN-L

The graph shown in Figure E10 is the FPR of all algorithms. It is observed that the curve of FPR have gradually increased with increase in number of frames. The average of FPR for CNN is 0.17, ABC with CNN is 0.15, FFBPNN-L is 0.16 and ABC with FFBPNN-L is 0.14.

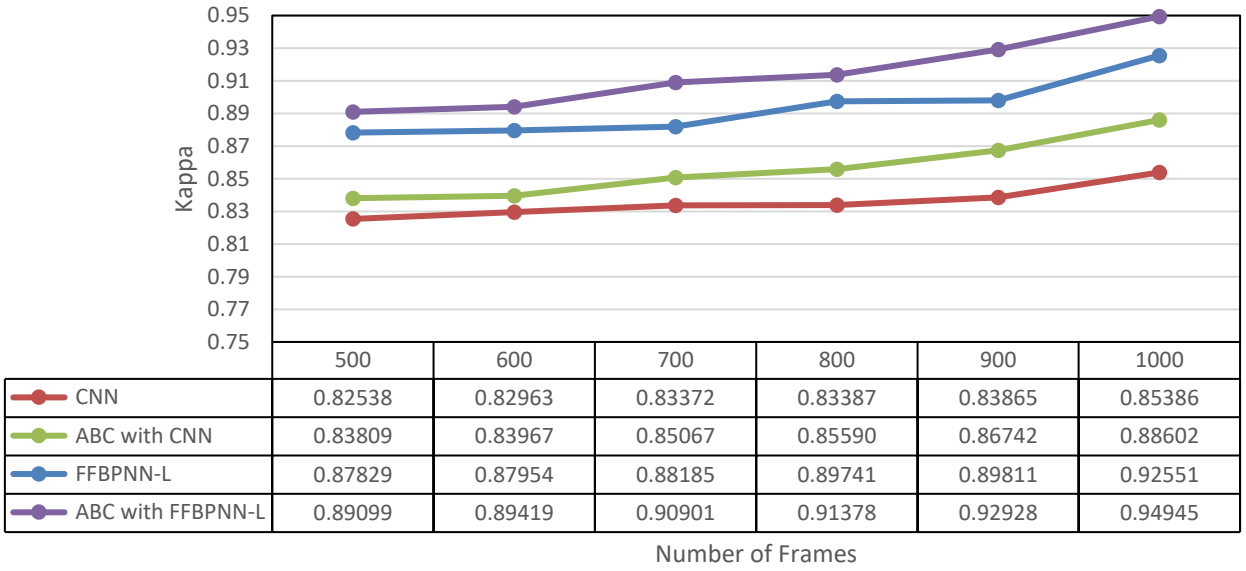


	500	600	700	800	900	1000
CNN	0.15875	0.16515	0.16999	0.18048	0.19539	0.20313
ABC with CNN	0.14447	0.14461	0.14622	0.15457	0.17235	0.18480
FFBPNN-L	0.15102	0.15277	0.15344	0.16444	0.18178	0.19546
ABC with FFBPNN-L	0.13674	0.14087	0.14569	0.14988	0.15398	0.16584

.

Figure E 10 FPR analysis of CNN, ABC with CNN, FFBPNN-L and ABC with FFBPNN-L

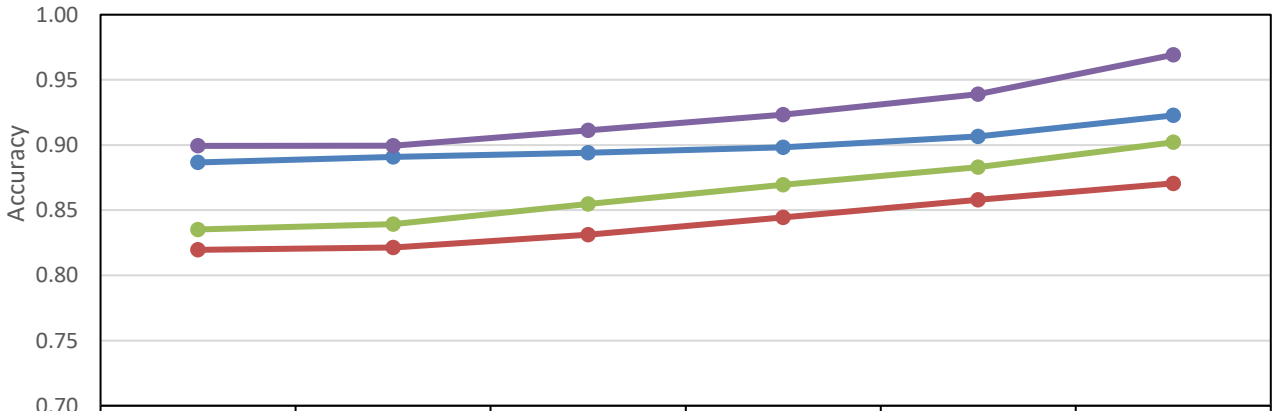
The Figure E11 depicts the kappa analysis of algorithms CNN, ABC with CNN, FFBPNN-L and ABC with FFBPNN-L. It shows the values of Kappa of all algorithms and average of CNN is 0.83, ABC with CNN is 0.85, FFBPNN-L is 0.89 and ABC with FFBPNN-L is 0.91.



	500	600	700	800	900	1000
CNN	0.82538	0.82963	0.83372	0.83387	0.83865	0.85386
ABC with CNN	0.83809	0.83967	0.85067	0.85590	0.86742	0.88602
FFBPNN-L	0.87829	0.87954	0.88185	0.89741	0.89811	0.92551
ABC with FFBPNN-L	0.89099	0.89419	0.90901	0.91378	0.92928	0.94945

Figure E 11 Kappa analysis of CNN and ABC with CNN

The accuracy of the algorithms CNN and ABC with CNN is depicted in the Figure E12 The curve had highly increased with increase in number of frames. The average of CNN is 0.84, ABC with CNN is 0.86, FFBPNN-L is 0.89 and ABC with FFBPNN-L is 0.92. Proposed algorithm had high accuracy as compared to other algorithms.



	500	600	700	800	900	1000
CNN	0.81962	0.82133	0.83116	0.84441	0.85795	0.87060
ABC with CNN	0.83524	0.83938	0.85490	0.86942	0.88291	0.90209
FFBPNN-L	0.88665	0.89093	0.89401	0.89813	0.90669	0.92272
ABC with FFBPNN-L	0.89936	0.89945	0.91117	0.92332	0.93901	0.96918

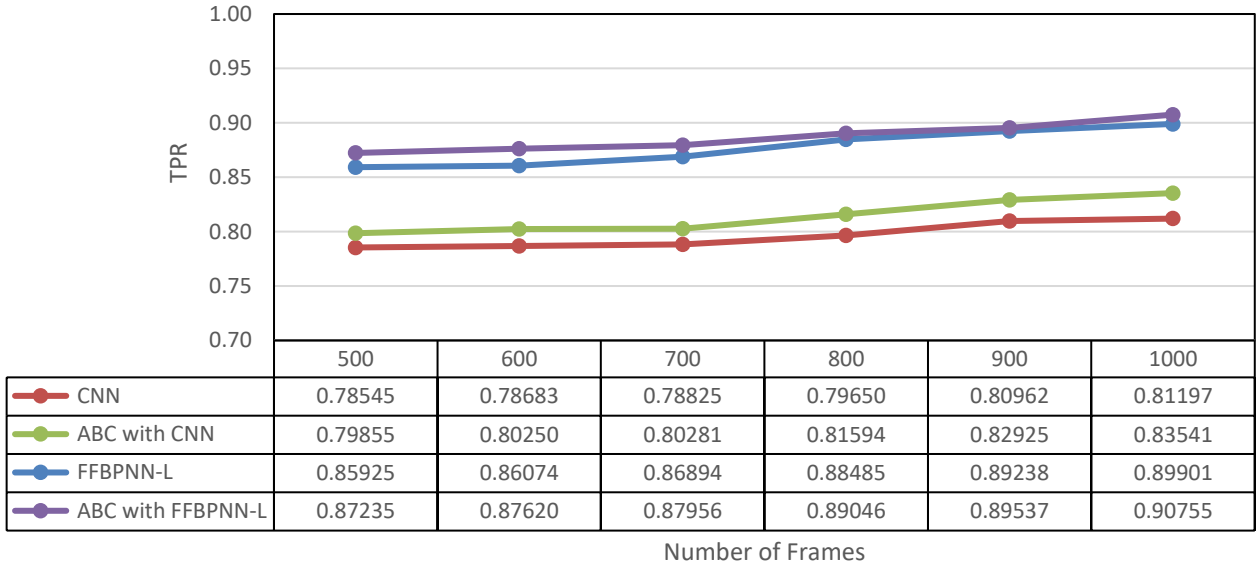
Figure E 12 Accuracy analysis of CNN and ABC with CNN

E-2: Using 15 Layered Neural Architecture

In this section the performance parameters of proposed and existing algorithms are been evaluated and displayed in accordance with the 15 hidden layers in the neural network architecture considering different ratios of training and testing data sets.

E2.1 For 70:30 ratio

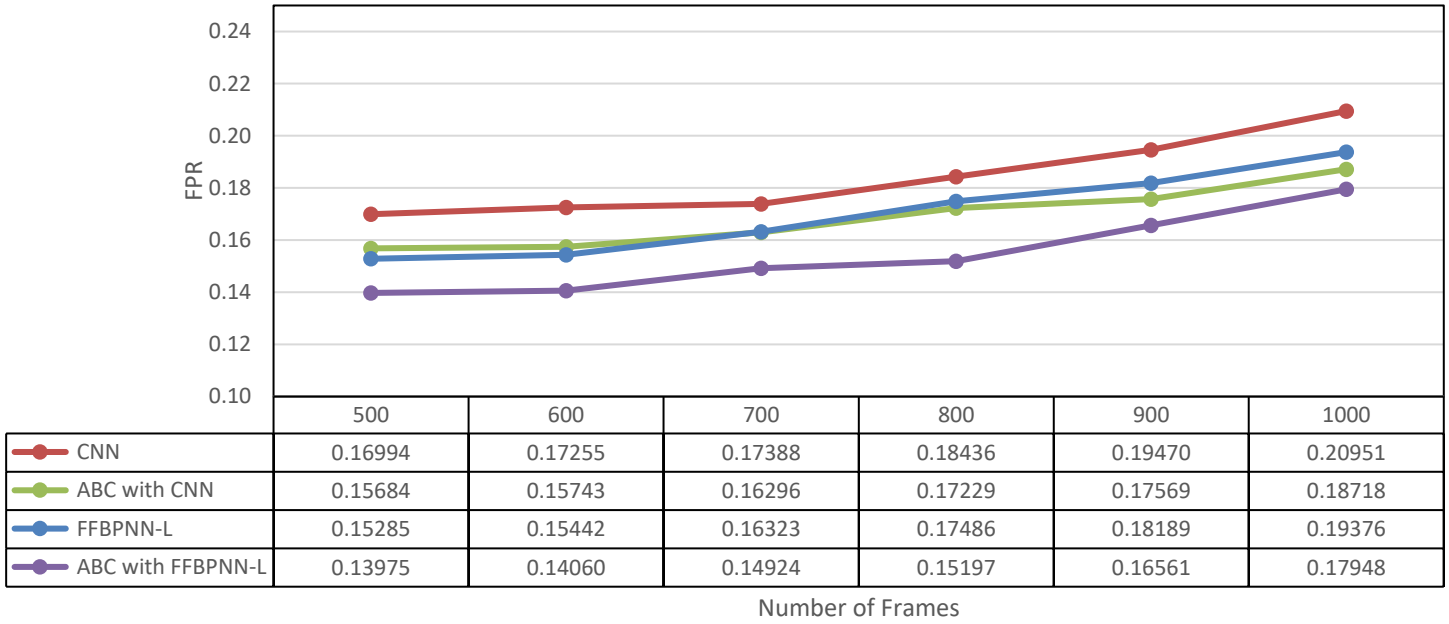
The Figure E13 shows the TPR analysis of all algorithms. The average TPR of CNN is 0.79, ABC with CNN is 0.81, FFBPNN-L is 0.87 and ABC with FFBPNN-L is 0.88. Hence, when compared with all the algorithms, the proposed algorithm is having high true positive rate.



	500	600	700	800	900	1000
CNN	0.78545	0.78683	0.78825	0.79650	0.80962	0.81197
ABC with CNN	0.79855	0.80250	0.80281	0.81594	0.82925	0.83541
FFBPNN-L	0.85925	0.86074	0.86894	0.88485	0.89238	0.89901
ABC with FFBPNN-L	0.87235	0.87620	0.87956	0.89046	0.89537	0.90755

Figure E 13 TPR analysis of CNN, ABC with CNN,FFBPNN-L and ABC with FFBPNN-L

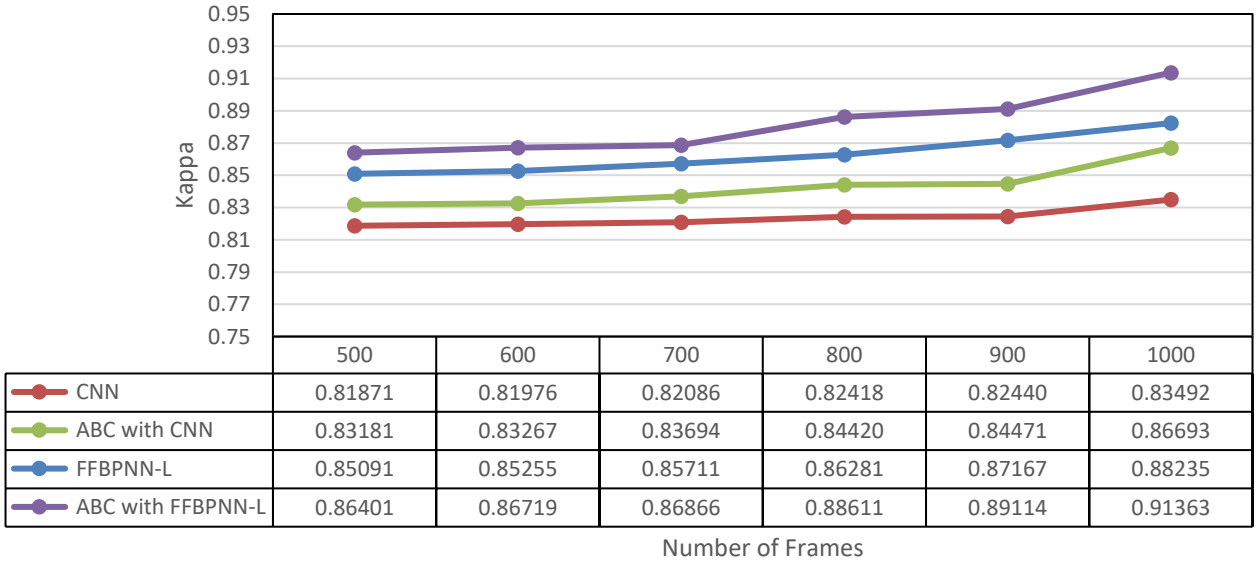
The Figure E14 shows the FPR analysis of all algorithms. The average FPR of CNN is 0.18, ABC with CNN is 0.16, FFBPNN-L is 0.17 and ABC with FFBPNN-L is 0.15. Hence, when compared with all the algorithms, the proposed algorithm is having less false positive rate.



	500	600	700	800	900	1000
CNN	0.16994	0.17255	0.17388	0.18436	0.19470	0.20951
ABC with CNN	0.15684	0.15743	0.16296	0.17229	0.17569	0.18718
FFBPNN-L	0.15285	0.15442	0.16323	0.17486	0.18189	0.19376
ABC with FFBPNN-L	0.13975	0.14060	0.14924	0.15197	0.16561	0.17948

Figure E 14 FPR analysis of CNN, ABC with CNN,FFBPNN-L and ABC with FFBPNN-L

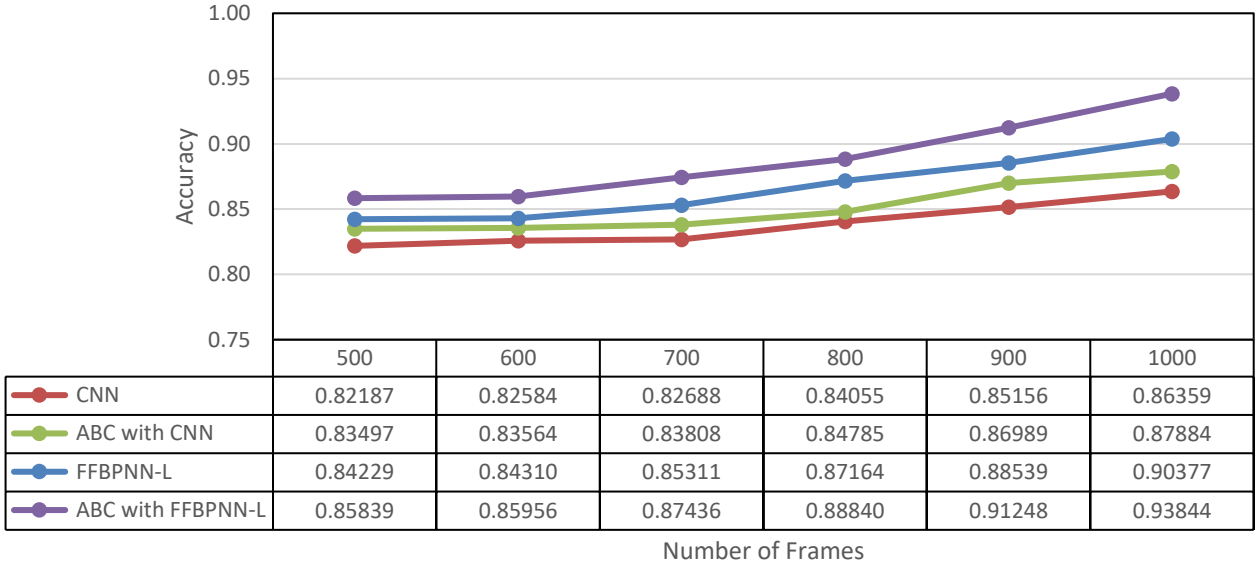
The Figure E15 shows the Kappa analysis of all algorithms. The average Kappa of CNN is 0.82, ABC with CNN is 0.84, FFBPNN-L is 0.86 and ABC with FFBPNN-L is 0.88. Hence, when compared with all the algorithms, the proposed algorithm is having more value of kappa analysis.



	500	600	700	800	900	1000
CNN	0.81871	0.81976	0.82086	0.82418	0.82440	0.83492
ABC with CNN	0.83181	0.83267	0.83694	0.84420	0.84471	0.86693
FFBPNN-L	0.85091	0.85255	0.85711	0.86281	0.87167	0.88235
ABC with FFBPNN-L	0.86401	0.86719	0.86866	0.88611	0.89114	0.91363

Figure E15 Kappa analysis of CNN, ABC with CNN,FFBPNN-L and ABC with FFBPNN-L

The Figure E16 shows the Accuracy analysis of all algorithms. The average accuracy of CNN is 0.83, ABC with CNN is 0.85, FFBPNN-L is 0.86 and ABC with FFBPNN-L is 0.88. Hence, when compared with all the algorithms, the proposed algorithm is having more accuracy.

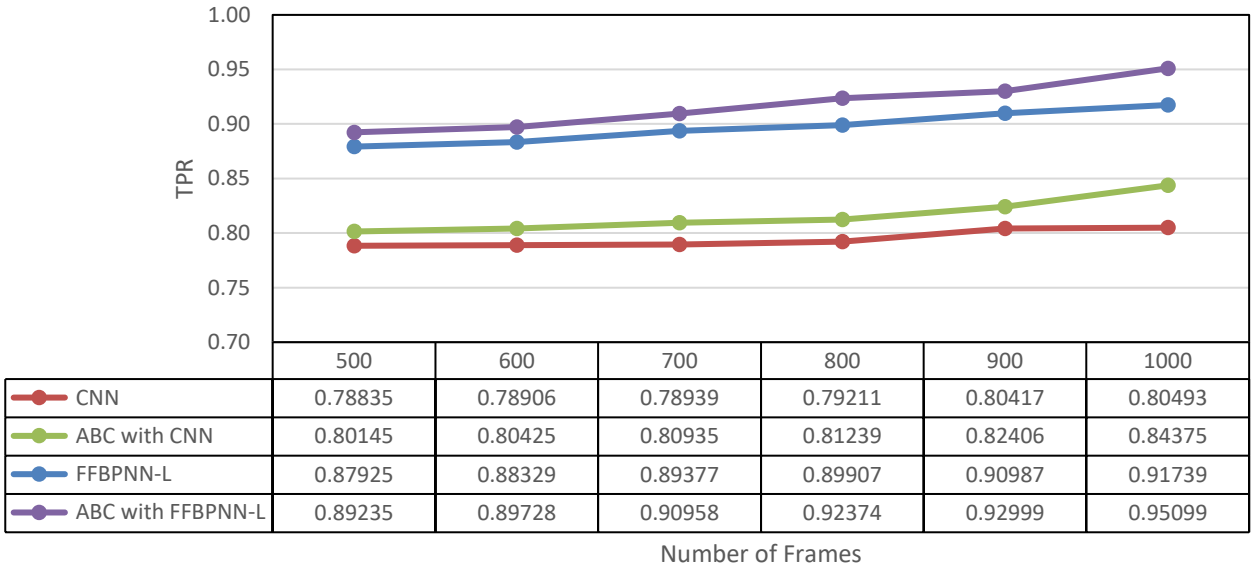


	500	600	700	800	900	1000
CNN	0.82187	0.82584	0.82688	0.84055	0.85156	0.86359
ABC with CNN	0.83497	0.83564	0.83808	0.84785	0.86989	0.87884
FFBPNN-L	0.84229	0.84310	0.85311	0.87164	0.88539	0.90377
ABC with FFBPNN-L	0.85839	0.85956	0.87436	0.88840	0.91248	0.93844

Figure E 16 Accuracy analysis of CNN, ABC with CNN,FFBPNN-L and ABC with FFBPNN-L

E2.1 For 80:20 ratio

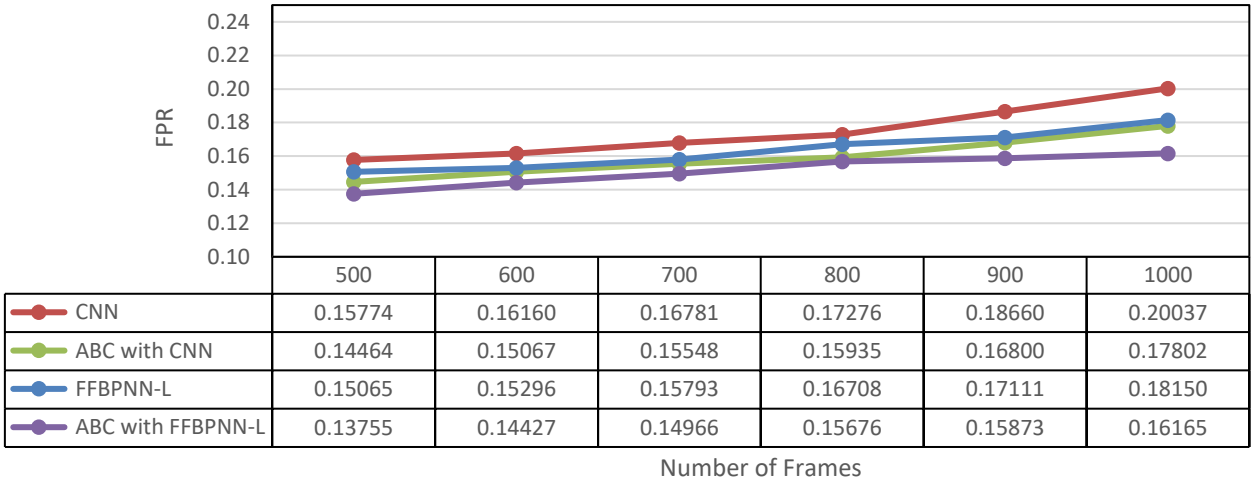
The Figure E17 shows the TPR analysis of all algorithms. The average TPR of CNN is 0.79, ABC with CNN is 0.81, FFBPNN-L is 0.89 and ABC with FFBPNN-L is 0.91. Hence, when compared with all the algorithms, the proposed algorithm is having more true positive rate.



	500	600	700	800	900	1000
CNN	0.78835	0.78906	0.78939	0.79211	0.80417	0.80493
ABC with CNN	0.80145	0.80425	0.80935	0.81239	0.82406	0.84375
FFBPNN-L	0.87925	0.88329	0.89377	0.89907	0.90987	0.91739
ABC with FFBPNN-L	0.89235	0.89728	0.90958	0.92374	0.92999	0.95099

Figure E 17 TPR analysis of CNN, ABC with CNN,FFBPNN-L and ABC with FFBPNN-L

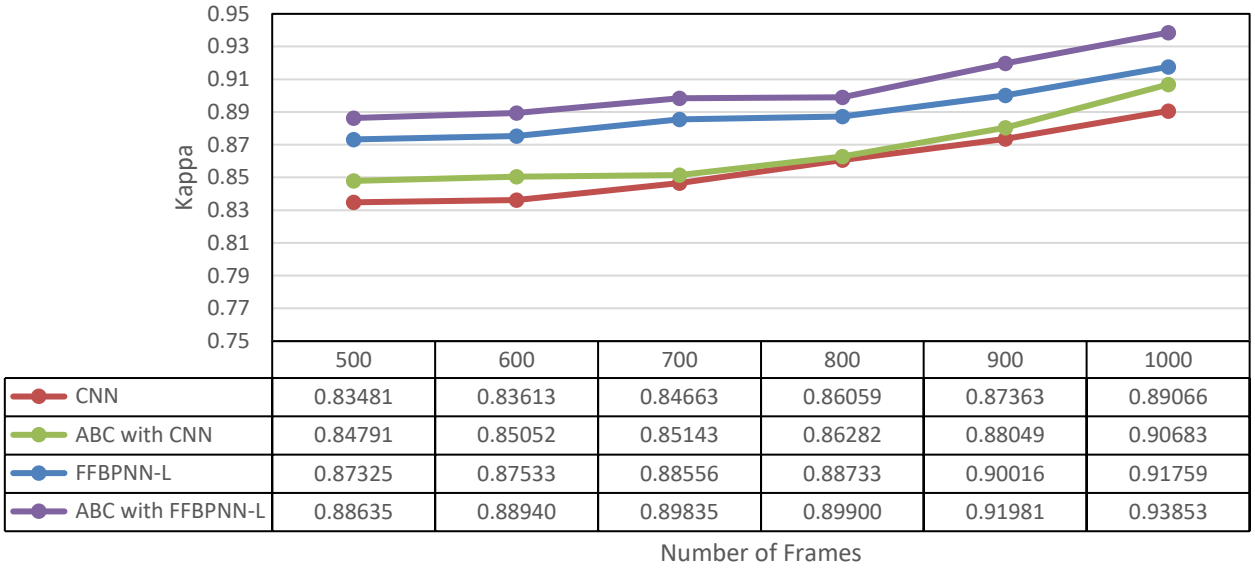
The Figure E18 shows the FPR analysis of all algorithms. The average FPR of CNN is 0.17, ABC with CNN is 0.15, FFBPNN-L is 0.16 and ABC with FFBPNN-L is 0.15. Hence, when compared with all the algorithms, the proposed algorithm is having almost same false positive rate with CNN.



	500	600	700	800	900	1000
CNN	0.15774	0.16160	0.16781	0.17276	0.18660	0.20037
ABC with CNN	0.14464	0.15067	0.15548	0.15935	0.16800	0.17802
FFBPNN-L	0.15065	0.15296	0.15793	0.16708	0.17111	0.18150
ABC with FFBPNN-L	0.13755	0.14427	0.14966	0.15676	0.15873	0.16165

Figure E18 FPR analysis of CNN, ABC with CNN,FFBPNN-L and ABC with FFBPNN-L

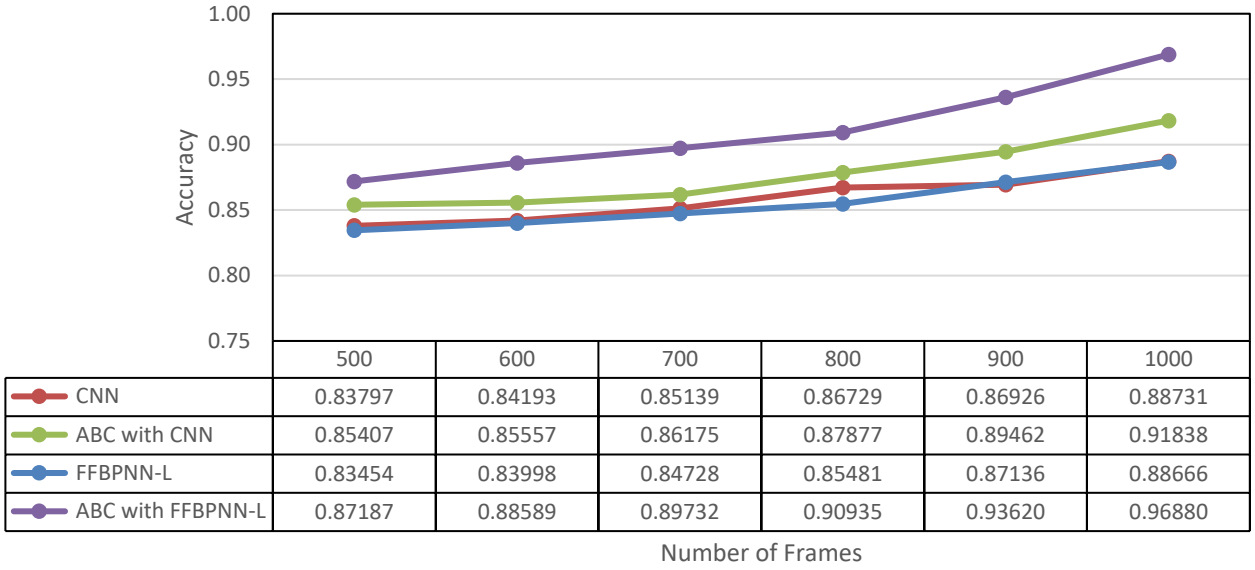
The Figure E19 shows the Kappa analysis of all algorithms. The average Kappa of CNN is 0.85, ABC with CNN is 0.86, FFBPNN-L is 0.88 and ABC with FFBPNN-L is 0.90. Hence, when compared with all the algorithms, the proposed algorithm is having more value of kappa analysis.



	500	600	700	800	900	1000
CNN	0.83481	0.83613	0.84663	0.86059	0.87363	0.89066
ABC with CNN	0.84791	0.85052	0.85143	0.86282	0.88049	0.90683
FFBPNN-L	0.87325	0.87533	0.88556	0.88733	0.90016	0.91759
ABC with FFBPNN-L	0.88635	0.88940	0.89835	0.89900	0.91981	0.93853

Figure E 19 Kappa analysis of CNN, ABC with CNN,FFBPNN-L and ABC with FFBPNN-L

The Figure E20 shows the Accuracy analysis of all algorithms. The average accuracy of CNN is 0.85, ABC with CNN is 0.87, FFBPNN-L is 0.85 and ABC with FFBPNN-L is 0.91. Hence, when compared with all the algorithms, the proposed algorithm is having more accuracy.

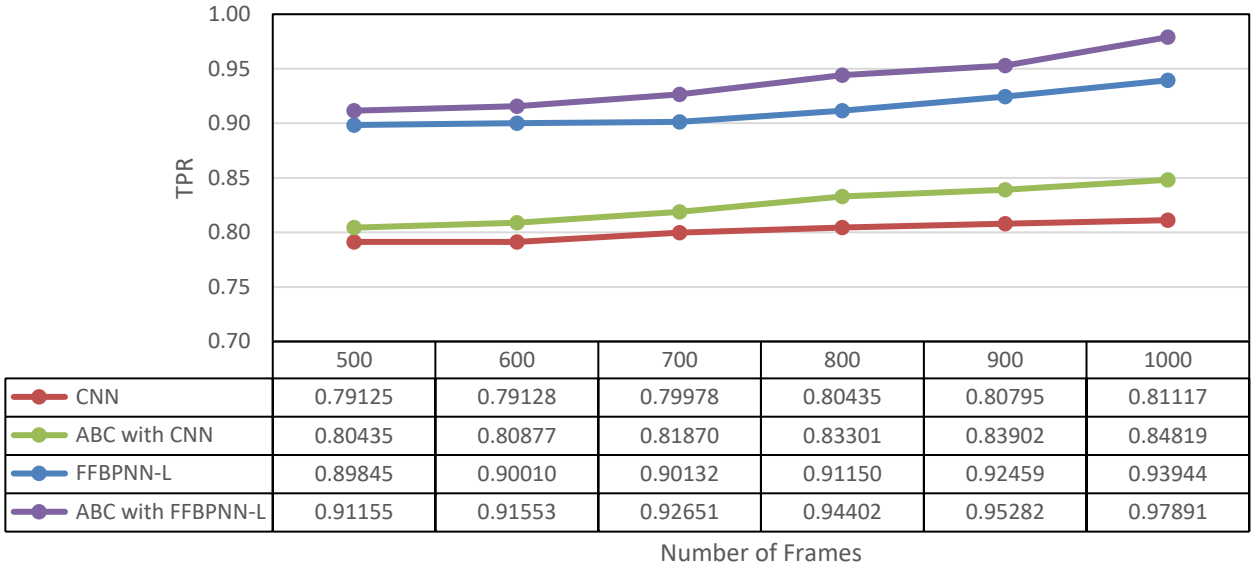


	500	600	700	800	900	1000
CNN	0.83797	0.84193	0.85139	0.86729	0.86926	0.88731
ABC with CNN	0.85407	0.85557	0.86175	0.87877	0.89462	0.91838
FFBPNN-L	0.83454	0.83998	0.84728	0.85481	0.87136	0.88666
ABC with FFBPNN-L	0.87187	0.88589	0.89732	0.90935	0.93620	0.96880

Figure E 20 Accuracy analysis of CNN, ABC with CNN,FFBPNN-L and ABC with FFBPNN-L

E2.3 For 90:10 ratio

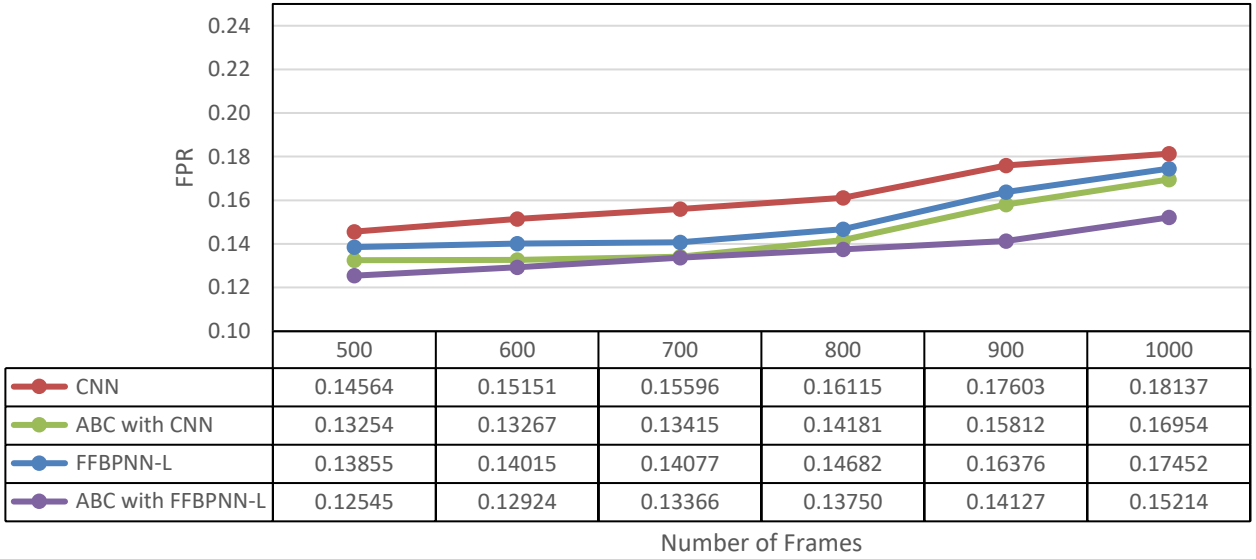
The Figure E21 shows the TPR analysis of all algorithms. The average TPR of CNN is 0.80, ABC with CNN is 0.82, FFBPNN-L is 0.91 and ABC with FFBPNN-L is 0.93. Hence, when compared with all the algorithms, the proposed algorithm is having high true positive rate.



	500	600	700	800	900	1000
CNN	0.79125	0.79128	0.79978	0.80435	0.80795	0.81117
ABC with CNN	0.80435	0.80877	0.81870	0.83301	0.83902	0.84819
FFBPNN-L	0.89845	0.90010	0.90132	0.91150	0.92459	0.93944
ABC with FFBPNN-L	0.91155	0.91553	0.92651	0.94402	0.95282	0.97891

Figure E 21 TPR analysis of CNN, ABC with CNN,FFBPNN-L and ABC with FFBPNN-L

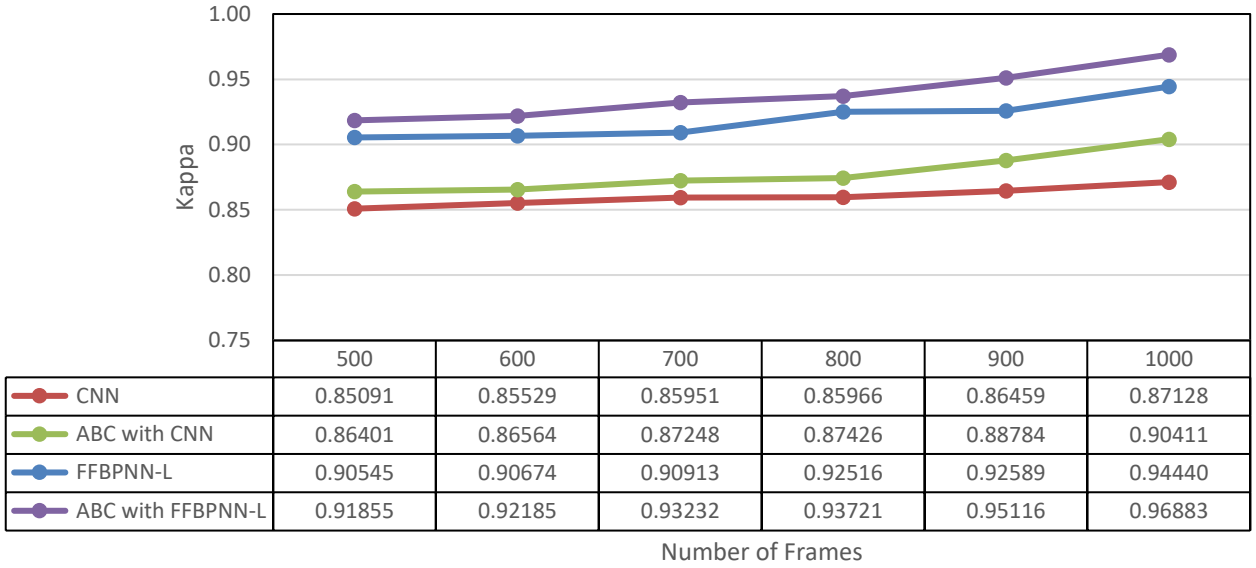
The Figure E22 shows the FPR analysis of all algorithms. The average FPR of CNN is 0.16, ABC with CNN is 0.14, FFBPNN-L is 0.15 and ABC with FFBPNN-L is 0.15. Hence, when compared with all the algorithms, the proposed algorithm is having almost same false positive rate with CNN.



	500	600	700	800	900	1000
CNN	0.14564	0.15151	0.15596	0.16115	0.17603	0.18137
ABC with CNN	0.13254	0.13267	0.13415	0.14181	0.15812	0.16954
FFBPNN-L	0.13855	0.14015	0.14077	0.14682	0.16376	0.17452
ABC with FFBPNN-L	0.12545	0.12924	0.13366	0.13750	0.14127	0.15214

Figure E 22 FPR analysis of CNN, ABC with CNN,FFBPNN-L and ABC with FFBPNN-L

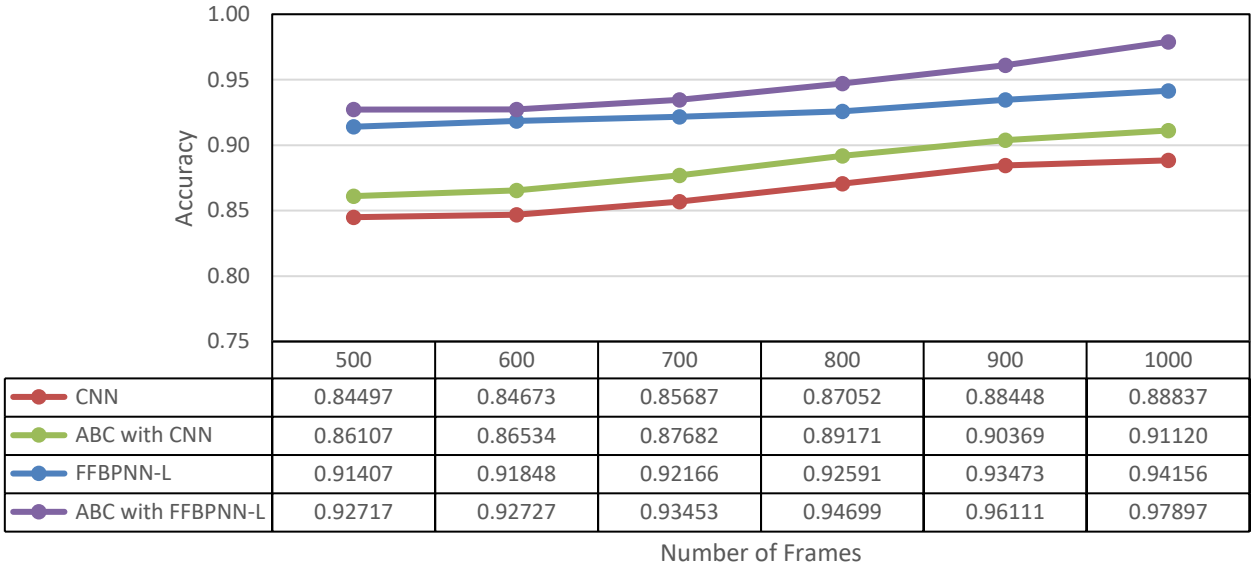
The Figure E23 shows the Kappa analysis of all algorithms. The average Kappa of CNN is 0.86, ABC with CNN is 0.87, FFBPNN-L is 0.91 and ABC with FFBPNN-L is 0.93. Hence, when compared with all the algorithms, the proposed algorithm is having more value of kappa analysis.



	500	600	700	800	900	1000
CNN	0.85091	0.85529	0.85951	0.85966	0.86459	0.87128
ABC with CNN	0.86401	0.86564	0.87248	0.87426	0.88784	0.90411
FFBPNN-L	0.90545	0.90674	0.90913	0.92516	0.92589	0.94440
ABC with FFBPNN-L	0.91855	0.92185	0.93232	0.93721	0.95116	0.96883

Figure E 23 Kappa analysis of CNN, ABC with CNN,FFBPNN-L and ABC with FFBPNN-L

The Figure E24 shows the Accuracy analysis of all algorithms. The average accuracy of CNN is 0.86, ABC with CNN is 0.88, FFBPNN-L is 0.92 and ABC with FFBPNN-L is 0.94. Hence, when compared with all the algorithms, the proposed algorithm is having more accuracy.



	500	600	700	800	900	1000
CNN	0.84497	0.84673	0.85687	0.87052	0.88448	0.88837
ABC with CNN	0.86107	0.86534	0.87682	0.89171	0.90369	0.91120
FFBPNN-L	0.91407	0.91848	0.92166	0.92591	0.93473	0.94156
ABC with FFBPNN-L	0.92717	0.92727	0.93453	0.94699	0.96111	0.97897

Figure E 24 Accuracy analysis of CNN, ABC with CNN,FFBPNN-L and ABC with FFBPNN-L

Appendix F: Published Papers

Paper 1: G. Kaur and J. Kilby, "Contactless Heart Rate Measurement using Image Processing", *43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) Oct 31 - Nov 4, 2021. Virtual Conference,* pp. 5184

Paper 2: G. Kaur and J. Kilby, "Contactless Heart Rate Measurement using Image Processing", *15th International Joint Conference on Biomedical Engineering Systems and Technologies BIOSTEC 2022,* Vol. 4, pp. 111-116

Paper 3: G. Kaur and J. Kilby, "Non-contact Methods for Heart Rate Measurement: A Review", *2021 International Conference on Computing Sciences (ICCS),* 978-1-6654-9445-8/2 IEEE, PP. 289-294 (Awarded Best paper)

Paper 4: G. Kaur and J. Kilby, "Non-Contact Heart Rate Measurement using Video Magnification: A Review" Springer *Nature Signal, Image and Video Processing* (submitted)

Paper 5: G. Kaur and J. Kilby, "An Optimized Artificial Bee Colony Algorithm for the Improvement of Heart Rate Detection from Video Frames" *International Journal of advanced Science and Technology (IJAST)* (submitted)

Paper 1:

2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Contactless Heart Rate Measurement using Image Processing

Gaganjot Kaur *Student Member, IEEE*, Jeff Kilby *Member, IEEE EMBS*

Abstract— **Non-contact methods of determining the human body's heart rate are of interest for clinical use. This research used a video magnification technique on the individual frames from a 15-second video taken using a digital single-lens reflex (DSLR) camera. It was possible to determine the beats per minutes of the heart rate by extracting the green spectrum from a region of interest information from the video frames. In this paper, three methods are presented using this colour change between the frames transform as a signal to find the heart rate.**

***Clinical Relevance*— Measurement of human physiological parameters such as heart rate using a non-contact method is becoming a vital technique compared to traditional monitoring systems in the clinical environment.**

I. INTRODUCTION

Heart rate (HR) variability is an essential parameter in the clinical environment for assessing the heart's function. It is helpful as an indicator of health status and diagnostics and assessing cardiovascular diseases and chronic diseases [1].

PPG became a popular method for measuring the heart rate: it is invasive, less expensive, and less complex [2]. A pulse oximeter is designed to fit over a finger or, in some cases, an ear, toe, wrist and send wavelengths of red and near-infrared through the body.

II. METHODS

The videos used were recorded from five participants using a mirror-less DSLR, Olympus Pen Lite E-PL5 16-megapixels CMOS sensor camera fitted with an Olympus M Zuiko Digital 14-150mm F4-5.6 lens. The videos used were recorded from five participants. All the videos were recorded at 30 frames per second with a resolution of 1920×1080 pixels. The camera was positioned at the required distance to have the participants face in the camera frame; this distance was 1.5 metres. Video capture took placein a large room with natural light and artificial light.

The video is separated into a sequence of image frames then passed through a series of image processing techniques. The spatial decomposition of each frame will characterize the variation over space; this process will reduce image noise and increase the temporal signal-to-noise ratio. Temporal image processing of signal followed to characterize the variation over time [3].

G. Kaur is a doctoral student within the Department of Electrical and Electronic Engineering at Auckland University of Technology, Private Bag 92006, Auckland, 1142, New Zealand (corresponding author to provide phone: +64-9-921-9999: e-mail: gaganjot.kaur@aut.ac.nz).

J. Kilby, Senior Lecturer in the Department of Electrical and Electronic Engineering at Auckland University of Technology, Private Bag 92006, Auckland, 1142, New Zealand (jeffrey.kilby@aut.ac.nz).

Region of interest (ROI) was identified by using object detection application; the algorithm is performed on each video frame.

Three methods were used: *Method 1* is the mean heart rate (HR) from the highest frequency occurrence. *Method* 2 is the mean HR from total peak count with respect to total time. *Method 3:* is the mean HR from the mean peak value across interval.

III. RESULTS

The results presented in this paper, which used video filename: P1001 gave HR values for (a) method 1 equal to 65.74 bpm, (b) method 2 a mean value HR of 65.67 bpm and (c) method 3 a mean HR of 67.73 bpm. A commercially purchased pulse oximeter was used to validate all the results.

Table I shows the experimental results from three recorded videos taken from participant one from five participants who taken part in this research.

TABLE I. VIDEO RESULT VALUES OF PARTICIPANT ONE

Video	Experimental HR Values (bpm)			Validation HR Values from Pulse Oximeter (bpm)	
	Method 1	*Method 2*	*Method 3*	*Range*	*Average*
P1001	65.74	65.74	67.73	68-65	66.5
P1002	63.69	63.69	64.93	60-63	61.5
P1003	87.19	79.92	86.68	108-80	94.0

IV. DISCUSSION & CONCLUSION

This research has demonstrated that it is possible to acquire HR measurement without physical contact with the participant by obtaining a signal through image processing of a video recording. However, precision drops under non-ideal conditions. Though the delivered product is promising, these limitations would be significant for real-world application.

REFERENCES

[1] G. Stouffer, M. S. Runge, C. Patterson, and J. S. Rossi, *Netter's Cardiology E-Book*. Elsevier Health Sciences, 2018.

[2] E. Feukeu and S. Winberg, "Photoplethysmography: Light Emitter Diode Wavelength Derivation from the Absorption Spectra of Haemoglobin," in *International Multidisciplinary Information Technology and Engineering Conference (IMITEC)*, 2019: IEEE, pp. 1-6.

[3] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation," Opt. Express, vol. 18, no. 10, pp. 10762-10774, 2010.

978-1-7281-1178-0/21/$31.00 ©2021 IEEE

Paper 2:

Contactless Heart Rate Measurement using Image Processing

Gaganjot Kaur[a] and Jeff Kilby[b]
Department of Electrical and Electronic Engineering, Auckland University of Technology, City Campus, Auckland, New Zealand
gaganjot.kaur@aut.ac.nz, jeffrey.kilby@aut.ac.nz

Keywords: Heart Rate (HR), Spatial and Temporal Processing.

Abstract: Non-contact methods of determining the human body's heart rate are of interest for clinical use. This research used a video magnification technique on the individual frames from a 15-second video taken using a digital single-lens reflex (DSLR) camera at 30 frames per second. It was possible to determine the heart rate beats per minute by extracting the green spectrum from a region of interest information from the video frames. In this paper, three methods are presented using this colour change between the frames transform as a signal to find the heart rate. While capturing the video's using the camera, a commercially available pulse oximeter was used to obtain the pulse rate from the participant's finger to validate the values calculated from the image processing techniques presented. The results show that it is possible to get a heart rate in terms of pulse rate reading using a camera and the developed MATLAB code.

1 INTRODUCTION

Heart rate (HR) variability is an essential parameter in the clinical environment for assessing the heart's function. It is helpful to indicate health status and diagnostics and assess cardiovascular diseases and chronic diseases (Stouffer et al., 2018). The frequency at which the heart beats, referred to as pulse rate, is measured in the number of these beats that occur per minute (bpm); the pulse rate changes depending on the body's need for oxygen (Schantz et al., 2019). Many factors can influence this, including but not limited to physical activity, emotion, illness, stress, and drugs (Allen, 2007). Regular resting heart rate for healthy adults falls within a range of 60 to 100 beats per minute (*The Gale Encyclopedia of Fitness*, 2012).

In 1938, Hertzman was the first to introduce the term photoplethysmography (PPG) as a description of a non-invasive optical technique for detecting the changes in blood volume in blood vessels (Kamshilin & Margaryants, 2017). It measures the light reflected from or transmitted through the body by using the principle that oxygenated haemoglobin absorbs more light than its surrounding tissue (McDuff et al., 2015; Verkruysse et al., 2008).

PPG became a popular method for measuring the heart rate: it is non-invasive, less expensive, and less complex (Feukeu & Winberg, 2019). Pulse oximeters first became available commercially in 1983 (Kamshilin & Margaryants, 2017). They usually contain two red and infrared lights (Aarthi et al., 2019). A pulse oximeter is designed to fit over a finger or, in some cases, an ear, toe, wrist and send wavelengths of red and near-infrared through the body.

Imaging PPG (iPPG) signal is similar to PPG signal recorded in a non-contact method using a camera. iPPG is an emerging technique that senses cardiovascular signals in the outer skin layers (Zaunseder et al., 2018).

2 THEORY

Non-contact measurement using videos is a popular research area, and many tools are available to process the videos in terms of colour and motion, such as Eulerian Video Magnification (Wu et al., 2012). The Eulerian method uses spatiotemporal filtering for the extraction of motion and colour variations in the video. Spatial decomposition is building image

[a] https://orcid.org/0000-0002-8379-4429
[b] https://orcid.org/0000-0001-5010-7170

Kaur, G. and Kilby, J.
Contactless Heart Rate Measurement using Image Processing.
In *Proceedings of the 15th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2022) - Volume 4: BIOSIGNALS*, pages 111-116
ISBN: 978-989-758-552-4; ISSN: 2184-4305

pyramids that decompose the video sequence into different spatial frequency bands. Spatial processing aims to increase the temporal signal-to-noise ratio by pooling multiple pixels, spatially low-pass filter the video frames, and down sample them for computational efficiency. In the general case, the entire Laplacian pyramid was computed (Burt & Adelson, 1983). The Laplacian pyramid is a sequence of images $L_0, L_1, \ldots, L_n$. Each is the difference between the two levels of the Gaussian pyramid is given by:

$$L_1 = g_l - \text{EXPAND}[g_2] \quad (1)$$

$$L_1 = g_l - \text{EXPAND}[g_3] \quad (2)$$

where g_l is the l^{th} level of Gaussian pyramid and g_2 and g_3 are second and third levels, respectively.

Temporal processing performs on each spatial band that considers the time series corresponding to a pixel value in a frequency band and applies a band-pass filter to extract the frequency band of interest (Wu et al., 2012).

3 METHODOLOGY

This section covers the procedure used to record and process videos to determine the heart rate in bpm. The results presented are from a pilot study where videos were recorded from two participants using a mirror-less DSLR, Olympus Pen Lite E-PL5 16-megapixels CMOS sensor camera fitted with an Olympus M Zuiko Digital 14-150mm F4-5.6 lens.

The videos were recorded at 30 frames per second with a resolution of 1920×1080 pixels. The camera was mounted on a tripod to eliminate any possible movement artefacts due to the camera during recording. The camera was positioned at the required distance to have the participants face in the camera frame; this distance was 1.5 metres. Video capture took place in a large room with natural light and artificial light throughout the room, so the participants were not under any direct or forced illumination. The participants sat upright on a chair with the back of the chair against a dark coloured wall to provide a clear difference in colour between the area that was the participant's face/head and the background, shown in Figure 1.

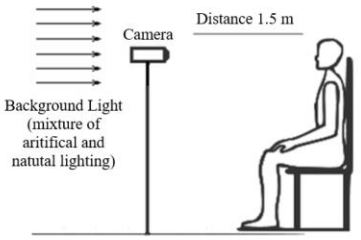


Figure 1: Setup of video recording.

Each video was recorded 15 to 20 seconds in length. During filming, the participants were asked to remain as still as possible but to breathe normally.

While videos were recording, the participants wore a Rossmax SB100 Finger Pulse Oximeter to validate the recorded videos' results, placed on their dominant hand's first finger.

3.1 Object Detection

Object detection can identify the video's region of interest (ROI) to mitigate background noise interference (Poh et al., 2010). Bush compared various ROIs of the face for HR detection: a standard bounding box ROI, a box with the eyes removed, a box section of the forehead, and the face cropped (Bush, 2016). Bush also reported that all four approaches performed equally in error rate whether the participant was still or moving; the forehead targeted area had a slightly lower error rate (Bush, 2016). However, the forehead ROI resulted in a significantly higher outlier percentage. Overall, the most efficient was the simple ROI as it is the fastest to compute and has equivalence performance to the more targeted ROIs. However, if noise interference from the background is a concern, then a segmented region cropping the face should be considered.

3.2 Image Processing

The video was separated into a sequence of image frames then passed through a series of image processing techniques.

The spatial decomposition of each frame will characterise the variation over space; this process will reduce image noise and increase the temporal signal-to-noise ratio (Poh et al., 2010). Wu et al. approach was to decompose each frame to different spatial frequency bands, applying Gaussian blur to pool multiple pixels together (Wu et al., 2012). The green colour channel's pixel values were pooled together to compute the average quantity of green across all

frames and the corresponding variation per frame; to obtain a signal representing the green colour channel's fluctuations within the ROI (Alghoul et al., 2017).

3.3 Video Processing

The MATLAB (release R2020b) function *'VideoReader'* was used for processing the videos, which extracted the required information: frame rate, number of frames, video height, video width and duration.

Using object detection application was used to identify an ROI; the algorithm is performed on each video frame to obtain an ROI. The MATLAB code detected the ROI once and applied the same ROI to all the frames. To improve the efficiency of the code by reducing the computation time. The ROI output is a four-element vector [*x*, *y*, *width*, *height*] bounding box square in shape see Figure 2.

Figure 2: ROI identified face detected.

To compute the average pixel value of the green colour channel, developed using MATLAB code was used, and the signal processing code was used for achieving temporal image processing.

3.4 Heart Rate Measurements

For the research presented using the signals, the heart rate beats per minute able to be found using one of the following three methods:

- *Method 1:* using the FFT of the filtered signal in the frequency domain and using the dominant peak frequency corresponds to the
- mean bpm measurement.
- *Method 2:* using the number of peaks over the video's total time to calculate the mean bpm. This method used the MATLAB' *findpeaks'* function with conditions applied to minimum peak height, minimum peak distance, and minimum peak prominence. The peak HR values were plotted against time, with each measurement identified with a marker.
- *Method 3:* uses results obtained in method 2; two consecutive peak values were used to calculate an HR value. These values were then plotted against time to show the variation of HR for the video

4 RESULTS

The results presented were obtained using video filename: P101.

4.1 Method 1

Figure 3 shows the selected colour channel fluctuations within the ROI, plotting the green colour channel signal.

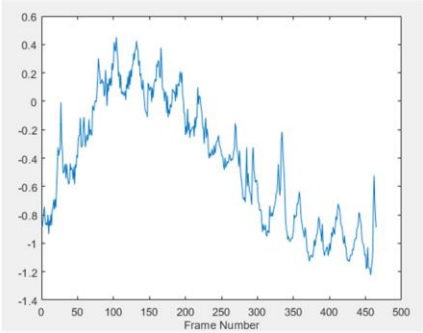


Figure 3: Signal for fluctuations in the green colour channel within ROI.

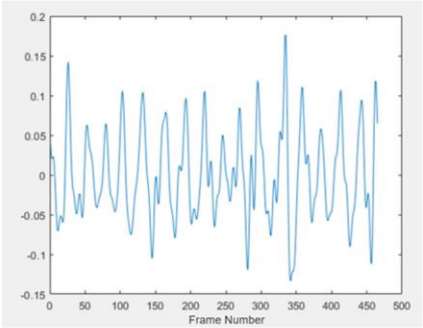


Figure 4: Filtered signal against frame number.

The filtered signal is plotted against to frame number, shown in Figure 4.

The FFT of the filtered signal shown in Figure 5 shows the maximum peak value was extracted and marked on the plot.

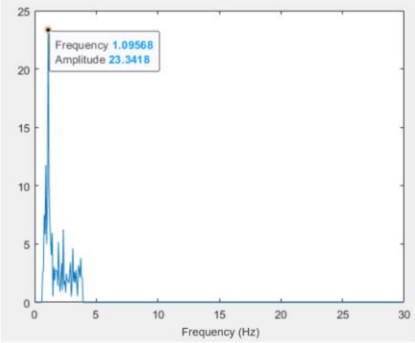


Figure 5: FFT plot of filtered signal in the frequency domain.

The dominant frequency is at 1.09568 Hz corresponds to a mean HR value of 65.74 bpm and which was calculated as follows:

$$T = \frac{1}{f} = \frac{1}{1.0956} = 0.9127 seconds$$

where T is the periodic time in seconds, and f is the frequency in Hz. So, the heart rate in beats per minute (bpm) is:

$$bpm = \frac{60\ seconds}{T} = \frac{60}{0.9127} = 65.74$$

4.2 Method 2

Figure 6 shows the filtered signal waveform where each peak associated with each heartbeat is marked and numbered and plotted against frame number.

The mean HR is 65.67 bpm, calculated from the total number of peaks against the number of frames and HR in bpm and was calculated as follows:

$$n = \frac{total\ frame\ number}{number\ of\ peaks} = \frac{466}{17}$$

$$= 27.41\ frames/peak$$

where n is the mean number of frame per peak, so:

$$T = \frac{n}{m} = \frac{27.41}{30} = 0.9136\ seconds$$

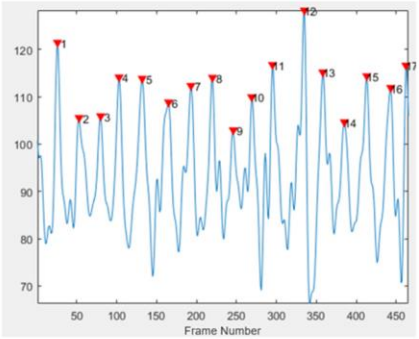


Figure 6: Peaks detected marked and numbered for the signal from Figure 4.

where m is the frames per second (fps), the heart rate in beats per minute (bpm) is:

$$bpm = \frac{60\ seconds}{T}\frac{60}{0.9136} = 65.67$$

4.3 Method 3

In Figure 7, the top plot is a repeat of Figure 6, and the bottom plot shows the result using two consecutive and finding the HR values in bpm; this is repeated for all the values and then plotted against time. The overall mean for the bottom plot HR values was calculated and had a value is 67.73 bpm.

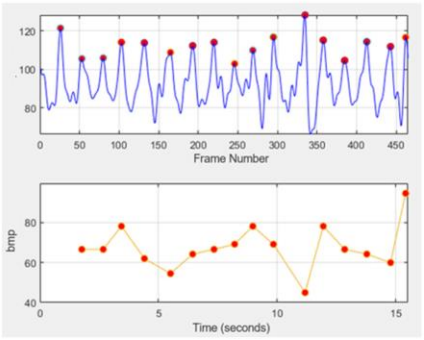


Figure 7: The signal is taken from Figure 6 (top), and HR values are plotted against time (bottom).

The results presented in this paper, which used video filename: P101 gave HR values for (a) method one equal to 65.74 bpm, (b) method two a mean value HR of 65.67 bpm and (c) method three a mean HR of 67.73 bpm. A commercially purchased pulse

oximeter was used to validate all the results taken; for video file P101, the HR range obtained was 68-65 bpm, an average of 66.5 bpm.

Table I shows the experimental results from six recorded videos taken using two participants. Both particants where female with an average age of 25 years with fair skin complextion.

Table 1: Video result values of Participant 1 and 2.

Participant 1					
Video No.	**Experimental HR Values (bpm)**			**Validation HR Values from Pulse Oximeter (bpm)**	
	Method 1	***Method 2***	***Method 3***	***Range***	***Average***
P101	65.74	65.67	67.73	68-65	66.50
P102	63.69	63.69	64.93	60-63	61.50
P103	87.19	79.92	86.68	108-80	94.00
Participant 2					
Video	**Experimental HR Values (bpm)**			**Validation HR Values from Pulse Oximeter (bpm)**	
	Method 1	***Method 2***	***Method 3***	***Range***	***Average***
P209	98.72	98.72	103.10	110-98	104.00
P210	98.08	94.45	95.75	104-99	101.50
P211	99.33	95.90	99.46	102-99	100.50

5 CONCLUSION

From the results obtained, three limitations have been identified.

Firstly, the recording of the pulse oximeter results parallel to the video recording was done by observing and writing the values down on paper. This approach introduced human error to the validation results. This error was eliminated by using a pulse oximeter with a wireless data logger.

Secondly, there is a weak correlation in the results at higher HR values and reduced accuracy from the experimental results.

The third limitation is the video recording were taken with the highest quality camera that was accessible, though these observations were also evident in the other standard consumer camera devices. When reviewing the video footage, video recording starts and stops; this can introduce unwanted interference in the image due to the camera sensor's light sensitivity. At times during the recording, the camera footage becomes blurry, and then the camera refocuses itself.

This research has demonstrated that it is possible to acquire HR measurement without physical contact with the participant by obtaining a signal through image processing of a video recording. Factors such as the lighting conditions, video recording settings, and ROI. All these variables require further investigation to see how they influence HR value accuracy. However, precision drops under non-ideal conditions. Though the delivered product is promising, these limitations would be significant for real-world application.

Further work is to create a more robust product and record videos from a more significant number of participants for data collection so results are validated.

REFERENCES

Aarthi, Y., Karthikeyan, B., Raj, N. P., & Ganesan, M. (2019). Fingertip Based Estimation Of Heart Rate Using Photoplethysmography. 5th International Conference on Advanced Computing & Communication Systems (ICACCS),

Alghoul, K., Alharthi, S., Al Osman, H., & El Saddik, A. (2017). Heart Rate Variability Extraction From Videos Signals: ICA vs. EVM Comparison. *IEEE Access, 5*, 4711-4719.

Allen, J. (2007, Mar). Photoplethysmography and its application in clinical physiological measurement. *Physiol Meas, 28*(3), R1-39. https://doi.org/10.1088/0967-3334/28/3/R01

Burt, P., & Adelson, E. (1983). The Laplacian Pyramid as a Compact Image Code. *IEEE Transactions on Communications, 31*(4), 532-540.

Bush, I. (2016). Measuring heart rate from video. In *Standford Computer Science, in press*.

Feukeu, E., & Winberg, S. (2019). Photoplethysmography: Light Emitter Diode Wavelength Derivation from the Absorption Spectra of Haemoglobin. International Multidisciplinary Information Technology and Engineering Conference (IMITEC),

The Gale Encyclopedia of Fitness. (2012). (First ed.). Gale, Cengage Learning.

Kamshilin, A. A., & Margaryants, N. B. (2017). Origin of Photoplethysmographic Waveform at Green Light. *Physics Procedia, 86*, 72-80.

McDuff, D. J., Estepp, J. R., Piasecki, A. M., & Blackford, E. B. (2015). A survey of Remote Optical Photoplethysmographic Imaging Methods. 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC),

Poh, M.-Z., McDuff, D. J., & Picard, R. W. (2010). Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Optics express, 18*(10), 10762-10774.

Schantz, P., Salier Eriksson, J., & Rosdahl, H. (2019). The heart rate method for estimating oxygen uptake:

Analyses of reproducibility using a range of heart rates from cycle commuting. *PLOS ONE, 14*(7), 1-20.

Stouffer, G., Runge, M. S., Patterson, C., & Rossi, J. S. (2018). *Netter's Cardiology E-Book*. Elsevier Health Sciences.

Verkruysse, W., Svaasand, L. O., & Nelson, J. S. (2008). Remote plethysmographic imaging using ambient light. *Optics express, 16*(26), 21434-21445.

Wu, H.-Y., Rubinstein, M., Shih, E., Guttag, J., Durand, F., & Freeman, W. (2012). Eulerian Video Magnification for Revealing Subtle Changes in the World. *ACM Transactions on Graphics - TOG, 31*.

Zaunseder, S., Trumpp, A., Wedekind, D., & Malberg, H. (2018, Oct 25). Cardiovascular assessment by imaging photoplethysmography - a review. *Biomed Tech (Berl), 63*(5), 617-634. https://doi.org/10.1515/bmt-2017-0119

Paper 3:

2021 International Conference on Computing Sciences (ICCS)

Non-contact Methods for Heart Rate Measurement: A Review

Gaganjot Kaur
School of Engineering, Computer and Mathematical Sciences
Auckland University of Technology
Auckland, New Zealand
ORCID 0000-0002-8379-4429

Jeff Kilby
School of Engineering, Computer and Mathematical Sciences
Auckland University of Technology
Auckland, New Zealand
ORCID 0000-0001-5010-7170

Abstract*— Health care is a main analysis area required instant results. Data storage as well as assessment have become more difficult as a result of the digitalization of information in all fields. As a result, the demand for skilled methodologies for analyzing health information is growing. Predictive analytics is an important issue from the health care field to computer technology researchers in ability to forecast as well as reduce potential health uprisings. Parallel research efforts are being made in many areas to forecast the disease's potential effect on multiple healthcare areas. Even so, those attempts are restricted and do not go as far to produce the desired outcomes. Lately, in the context of information systems, non-contact methodologies have been shown to make a positive contribution to the healthcare profession through improving the accuracy as well as speed of disease diagnosis. As a result, this research analyzed heart rate assessment utilizing non-contact technique to measure disease severity stages.

Keywords*— Heart Rate, ECG, PPG, Video magnification

I. Introduction

Non-communicable diseases presently kill more people per year as compared to the infectious diseases. The leading cause of mortality in the world is heart disease, which is one of the most common non-communicable illnesses. WHO research show that heart disease now causes 16 percent of all fatalities, and that since 2000, the number of people who die of heart disease has increased by more than two million to about 9 million. [1]. Indians are highly affected by heart disease in early age compared to European countries (in western countries, heart disease deaths are about 23 percent before 70 years, while in India it is about 52 percent). The WHO describes poor eating, physical inactivity, tobacco usage and heavy alcohol consumption are the main significant risk features for heart disease behavior. These "intermediate risk factors" could be evaluated in healthcare settings as well as signify a greater risk of having a heart attack and some other complications [1][2]. Some of this information can be provided immediately, while tests need to be carried out in the other cases. These can include an Electrocardiogram (ECG) or blood test. ECG is a medical instrument that is commonly utilized to calculate and record distinct cardiac electrical potentials. Willem Einthoven constructed the ECG technique in the early 1900s, and although it's a fairly easy experiment to conduct, ECG tracing analysis takes a substantial number of trainings [3]. ECG is an approach to monitoring the electrical activity of the heart, usually a non-invasive approach to the detection of heart defects. The ECG signal has different properties. A standard pulse or cycle begins with a wave P followed by a complex QRS. Then the beat ends with a wave in T. The U wave may sometimes be seen following the T wave. The U wave might emerge following the T wave on rare occasions. Cardiac arrhythmias, for example, alter the rhythm and shape of the electrocardiogram (ECG) waveform.

The most common way of measuring heart rate (HRM) is by touching the skin. This contact is critical in the measurement of skin electrical changes during an ECG. Skin contact was utilised to visually capture a plethysmogram for the Photoplethymography (PPG) available on numerous smart watches [5]. While non-invasive techniques like this one may be less intrusive than others, they nonetheless have the potential to cause injury to those with skin problems or newborns. Wearing a fitness tracker in a business situation may be a source of confusion or distress for the wearer. A less invasive, contactless measuring approach would be preferred in these cases as well. Over the last decade, a large body of work has been published on HRMs that do not come into contact with the skin. Non-contact computations have been tried using video and image methods. New developments in miniaturisation and minimal thermal imaging inventions, the Euler video magnification and the innovation normal camera for thermal physiology signal analysis have been established [6]. Skull important design with regular thermal camera for comfort/discomfort pose detection has also been established as one of the research directions.

Thus, this paper explores the recently proposed non-contact techniques for heart rate measurement.

The following are the article's big contributors:

1. Cardiac increased pulse rate variability are explained in depth in this article.
2. A detailed description on various non-contact approaches for HR measurement has been presented.

A comparative analysis of reviewed methods with emphasis on their signal type, signal extraction, and signal estimation (filtering) along with their heart rate estimation techniques has been provided.

II. Heart Rate And Heart Ratevariability

The number of heartbeats per minute is known as the heart rate. Variability of neighbouring heartbeats in time intervals (HRV) is defined as the fluctuation of HRV. Human heart rate variability (HRV) is a neurocardiac structural indicer generated by processes that need a vigorous non-linear autonomic nervous system (ANS). Interdependent regulatory systems that act across a wide variety of time periods help us adapt to genetic and behavioral challenges. HRV is an emerging resource. Blood Pressure (BP), gas exchange, the heart and vascular tone, which indicates the size of the blood vessels regulating BP, and perhaps the face muscles [2] are all represented by HRV.

Heart rate is a recognised physiological metric in a person's normal condition. The parameter is influenced by a wide range of parameters, including gender, age, physical

978-1-6654-9445-8/21/$31.00 ©2021 IEEE
DOI 10.1109/ICCS54944.2021.00064

and psychological limits, and pharmacological treatments. Consequently, the parameter's value might vary substantially across different patient populations. In post-MI patients, a wide range of changes, which might be impacted by many physiologic processes, has been observed to correspond with increased risk. It has also been discovered that patients with low HRV rates are more likely to die than those with normal HRV rates [8].

III. Non Contact Methods

In this part, you will learn about the numerous non-contact methods for measuring heart rate. Sections 2 through 4 go through in depth the non-contact techniques shown in Fig. 1 for measuring heart rate.

A. Discriminative Response Map Fitting with Camera

The Discriminative Response Map Fitting (DRMF) procedure, a new approach centered on discriminative regression for Constrained Local Models (CLMs) paradigm, has shown outstanding conductivity in the generic face fitting condition. In contrast to the comprehensive texture-based qualities employed in discriminative AAM techniques, a restricted collection of parameters may characterise the response picture, yet these variables can be utilised to reconstruct unseen reply pictures extremely successfully. [9]

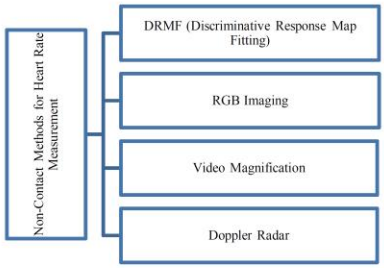


Fig.1: Non-Contact approaches for HR measurement

To limit the impact of numerous interferences, Li et al. [10] proposed a device for remote HR computation. Facial tracking has been utilised to correct stiff facial motions, and the contextual green value has been used to reduce the disruption caused by brightness variations. Shearing as well as segmentation were then utilised to reduce the influence of fast, non-rigid activity. Under real human-computer interaction (HCI) situations, the suggested technique will significantly decrease interference and enhance HR measurement accuracy. In order to minimise these artefacts, they propose a system that utilises four main processes: first, the DRMF method is used to accurately select the ROI for the face and to detect facial landmarks and to create a ROI mask in the first frame; second, the Kanade-Lucas-Tomasi (KLT) approach is used to map the ROI position led by rigid head movement; and third, the Normalized Least Mean Squares (NLMS) flexible filter is used to correct the interference of As with LMS, adaptive filters may be applied iteratively using NLMS, which is a variation of that algorithm. In a PPG investigation, it was shown that the LMS filter significantly reduced movement artefacts. Using Welch's technique of calculating the power spectral density, the HR frequency may be determined in the fourth step All three strategies have been proved by the authors to help enhance the precision of HR readings in real-world situations. MAHNOB-HCI specimens had a mean failure rate of 6.87 percent using the new approach, which greatly exceeded the conventional method.

RGB Imaging

For example, in Fig. 2, a data from a range of up to several metres may be obtained using the RGB imaging colour model [11]. rPPG is the common name given to these procedures because of their similarities to traditional PPG. The periodic stretching of the vessel walls is caused by the pulse wave travelling through the body. A PPG may be used to measure the volumetric changes in the human body caused by variations in the blood and trapped air composition. Light absorption through a given volume of tissue is affected by such oscillations. PPG generates noise by requiring a regular light source and also accounting for ambient light. Research shows that certain cardiovascular signals (such as HR and IBI) may be obtained wirelessly from a distance of several metres by using a video picture taken from a typical ambient light camera system as the light source for a video sequence of a human face. A mixture of the mirrored plethysmograph signal and fluctuations in the set of reflected ambient light is selected using the camera's RGB sensor. When volumetric changes in the facial blood volume (BV) occur during the cardiac cycle, this condition is activated, which implies the scheduling of cardiovascular events.

Using a blur detection and denunciation technique for every frame, as well as a minimum KF method for motion prediction and tracking features, Prakash et al. [13] suggested an algorithm to reduce motion artefacts including blurring and noise caused by head movement (uniform, random). Non-contact pulse rate prediction of patients with normal head and facial expressions has been shown in a test scenario. The approach outperforms state-of-the-art rPPG methods in HR detection, according to benchmarked results. Luminance variance's flaw is addressed by the suggested method. HSV colour space is used to further invariantly monitor characteristics utilising the rotation, size, and change invariance of dynamically changing pixel intensity values. The predominant wavelength of light reflected or acquired via ROIs is reflected in the HSV colour space hue parameter. For this approach, a device-dependent colour space known as HSV has been used since it has been shown to be successful in recording noise-free signals from rPPGs, making the algorithm illumination invariant. Using the OpenCV library's Back Projection function, we were able to perform skin segmentation after transforming the colour space. As a result, the BKF approach was utilised to track the areas of interest from one frame to the next using motion prediction modelling. Fast Fourier Transform was used to estimate HR (FFT). According to the data, the proposed strategy was subject to changes in illumination. For example, the rPPG signal identification technique suggested in this paper was not employed in cases when the light was coloured or multicoloured.

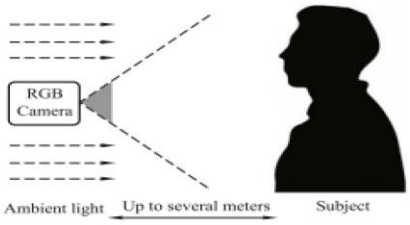


Fig.2: RGB Imaging

Fernandes et al. [14] devised a method for determining heart rate from LAB-color face footage. The writers must keep in mind that these variations in blood circulation cause variations in the coloring of the face skin and may be used to calculate heart rate. The pulse rate was then recorded and compared to a standard value. This approach offers tremendous potential for expanding telemedicine, a person's well-being, and a wide variety of other applications that rely on real-time data. The important idea was that cameras can detect the colour changes in the skin that occur when blood flows into the neck, even though they are invisible to the human eye. A new method for recognising human faces and specific parts of the skin on those faces is described. Exploring a signal's least differential volatility yields the ideal de-mixing angle. The band pass filter normalises the pulse signal, allowing it to choose frequencies between 0.5 Hz and 3 Hz. To determine the heart rate, it uses the heart rate interval. The band pass filter normalises the pulse signal and picks frequencies ranging from 0.5Hz to 3Hz.

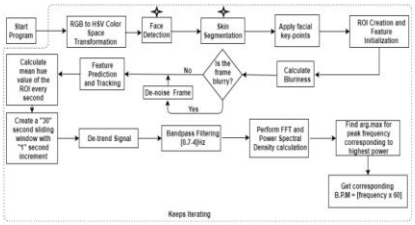


Fig.3: Block diagram of Face detection and skin segmentation

B. Video Magnification

For displaying the characteristics and attributes of magnified motion in a photograph, a microscope-like approach of magnifying visual motion was given [15]. This method linked the visual motion assessed with pixel resolution from many video pictures using Lagrangian methodology. To show small emotions and colour shifts the human eye can't notice, Euler video was formally recognised. Euler pre - process on video pyramids and spatial magnification processing rather than monitoring motion, unlike the Lagrangian technique. Time-series data of colour values at every pixel as well as the changes of people in a certain time band are often examined in this manner. Besides enlarging the laryngeal node and monitoring minor heart rate variability, pulses are employed for structural identification by Euler video amplitude, as well as human skin colour and blood flow. Researchers at Ume University in Sweden and Virginia Tech University inside the United States then employed Euler video magnification for skin temperature measurements that might indicate thermal comfort positions or provide control signals to assess HVAC systems.

To increase the size of moving images, Albor et al [17] introduced the Hermite transform, which is seen in Fig. 4. Additionally, a DL technique was used to estimate the beat-by-beat pulse message. Using an electrical pulse monitoring device, the certified software tested the promised performance. A Gaussian pyramid was used to get a magnification effect comparable to that of a traditional photograph. The HT was used to achieve a multi-resolution spatial decomposition of low-frequency components. The series includes an implementation of a temporal large-band filter. An adequate spectral range is covered by the cutting frequencies chosen. There is a gradual increase in amplification of the chromatic and temporal changes. The HT synthesis method is used to alter the picture sequence. The enlarged sequence is included in the original. Multispectral images from colour changes are improved by the Gaussian approach, enabling for a more precise measurement of the immediate beat via the beat pulse.

Using an EVM method developed by Bennett et al. [18], it is possible to boost input signals while at the same time reducing noise. Both thermal and iPad cameras caught footage of a person who was wearing an ECG-collecting textile sensor band. EVM was delighted by the iPad video, which used a huge BF and a lower magnification factor. The HR signal was then derived from the MI signals of five

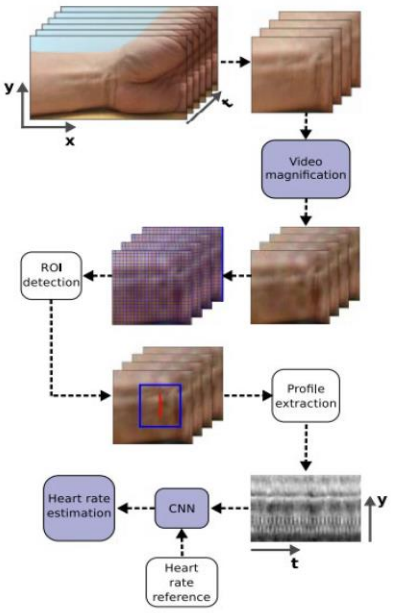


Fig.4: Block diagram of heart rate measurement system

ROIs. As a consequence of using an ECG signal to validate the ROI, HR is accurately represented by the mean intensity signal. A broad bandpass filter was used to expose thermal video to EVM and the post-processing level MI of the optimum ROI was established. In contrast to the iPad video MI signal that was manipulated, this signal was recovered in order to verify the content. The original thermal video was once again subjected to EVM storage and ROI measures post processing velocity, this time with an updated, targeted restricted bandpass filter.. Thermal video might expose physiological signs such as heart rate in line with the recommended customised EVM system with post-processing ROI, and could limit the possibility for disclosing an enhanced noise signal, according to the results of the study.

An algorithm developed by Albor et al. [19] used a patient's video clip to estimate HR. According to their plan, the first step is to collect a ROI; then, to use an Eulerian technique of video movement magnification utilising HT, to enhance changes in colour intensity; and finally, to extract the feature to create a raw pulse signal and estimate HR using an appropriately trained neural network. They used a conventional camera to record information videos for the studies and checked the results of a personal pulse monitoring device. Windowing the raw pulse signal was used to create the feature set. A 10-second window was moved every second, and the DFT was computed in each window. " DFT function vectors were used for every minute of the signal to represent the times / minute (bpm). Researchers came up with a novel method for predicting the HR pulsing signal from a conventional RGB sensor's picture sequence, employing the HT and NN. In order to monitor the patients, the NN utilised was a simple framework that only required a little amount of input. In the actual world of human resources management, this is a common scenario. Comparing Gaussian and Hermite HR estimations, they found that the proposed method is competitive from the data (RMSE=1.86 bpm) when compared to other relevant literature articles. This is further supported by a noise investigation where the HT technique's HR estimate is compared to the Gaussian approach. This assessment shows that the Hermite approach is able to cope with distortion as the initial approximation under actual acquisition settings.

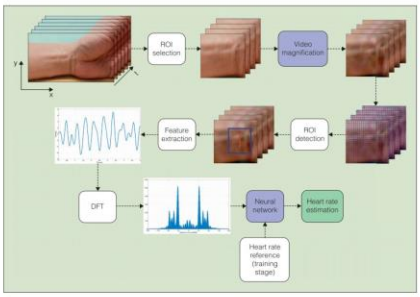


Fig.5: Block diagram of heart rate estimation system

A more comprehensive discussion of VidMag's preparedness for psychophysiological examination was provided by Sarkar et al.[20]. As part of the experiment, VidMag will be compared to blood volume pulse (BVP) signals generated by medical systems. VidMag does not employ colour indicators, and a new skin detection method was developed as a pre-processing step to enable that technique. The raw output of the visual magnification algorithm may be improved using a systematic post-processing method based on Savitzky Golay filtering.

Eulerian Video Magnification's HR estimate accuracy was to be improved by Huang et al. [21]. It has been shown that the Y module's signal in the YIQ colour spectrum is more reliable in HR estimation than the I element utilising the selected area as well as the peak detection technique. Heart rates obtained at fewer than 30 fps were shown to be more exact than those obtained at less than 60 fps, according to the findings of this study. While recording with a GoPro Hero 6 with a brightness level of more than 1500 lumens and a refresh rate of 30 frames per second, the HR extract error was just 5 percent. False peak identification and more data are need to accurately identify heart rate variability. Using a GoPro Hero 6 camera, designers tested the very same light output with different frame rates and filming times, and same frame rate and shooting time with different light levels. When it comes to determining heart rate, the Y element consistently outperformed the I element. In addition, to get an accurate estimate of heart rate, the illumination level must be higher than 1500 lx. If the environment is stable and the illumination level is at minimum 1500 lx, the HR prediction error percentage using Eulerian Video Magnification will surpass 5%.

C. Doppler Radar

There are several applications for Doppler radar systems in the field of unobtrusive surveillance of health care, including the detection of microscopic human chest displacement induced by cardiac exercise[22]. Doppler radar is able to detect electromagnetic waves that are involved in transmission mirrors. By using a technology that can penetrate non-metal materials like as wood or clothing, human physiological variables such as HB and respiration may be detected remotely. The frequency of its operation varies. Device sensitivity is closely connected to output power (e.g. microwave scale 24 GHz, UR: 40 kHz). One-channel and quadrature radar systems for measuring human physiological parameters have been tested. The single-channel system has a substantial drawback owing to the continual periodic change in the receiver end and the control signal. QRR has been suggested in order to choose between the two channels, I and Q, to deal with the subject's location sensitivty. Quadrature outputs have been suggested to be demodulated using two different ways, one linear (complex) and the other non-linear (arctangent). [23] The linear method spins the data to the optimal point, while the non-linear method overwhelms the sensitivity to the desired spot.

Due of Doppler radar's NC to the topic quality, it has received a lot of attention. It does not infringe on the privacy of the person in the same way as image sensors do. While the patient is on the bed, Higashi et al. [24] devised a technique that automatically extracts the information from the actual message. This method, which relies on ML, is able to be used in real time because of its usage of simple features. In a six-person sleep laboratory investigation, the

sleeping position was detected with 88.5 percent accuracy and the pulse rate estimated error ratios were reduced by 15.2 percent. Both signals (I and Q) from the radar may be separated into 10-second windows with a 5-s phase. Power spectrum analysis is used to estimate HR by moving messages through a bandpass filter (1–15 Hz), then transforming them to the frequency response using FFT. According to Figure 6, the MNR, also known also as SR, was used to predict the five possible postures for lying down when sleeping. Classifying multiclass issues using logistic regression, MNR is one way. Due to the fact that radar signals do not have the same high peak as ECG signals, other methods of measuring heart rate are required. Because of this, the FFT and WF were used for the measurement.

Fig.6: Block diagram of HR calculation for radar system

Continuous wave Doppler radar was used as a phase-locked loop in the process demodulator setup by Mercuri and colleagues [25] to monitor non-contact vital signs in vivo. Using the radar's architecture, it is possible to eliminate the heartbeat and breathing from the stage change caused by human cardiopulmonary activity concurrently [26]. To separate it from other systems, the single mixer's design is resistant to the blank point and does not need criteria for approximation of the tiny angle. In this way, the proposed radar may operate in a very linear manner, detecting phase modulations generated by any movement, independent of amplitude and velocity. Following calculations and technical testing, a realistic setup on healthy volunteers was presented to prove the proposed approach's functioning and safety. The baseband radar data was effectively analysed using independent element wavelet analysis to extract data on respiratory and heart rate. Tracking CW's Phase Non-contact surveillance of vital signs using Doppler radar has been shown to work on human subjects. Unlike other techniques, the architecture of the sole mixer is immune to the null point and does not need requirements of approximations of the tiny angle. It is possible to reliably monitor a person's breathing and heartbeat without touching them, according to a new hypothesis. CMOS output is also supported by the design. With the advent of WICA, HR calculation has been significantly improved. Additionally, WICA was introduced. The WICA strategy was shown to be superior to the alternative selecting features when it came to collecting that data. Although the HR RMSEs are less using the WICA technique, the HR relative errors are also smaller.

IV. Conclusion

Identifying the transmission of raw health information can help to save lives and diagnose anomalies in long-term heart conditions. Non-contact methods are reviewed in this study to understanding the non-contact approaches of heart rate measurement. Predicting heart disease in the medicine field is very difficult and important. However, the mortality rate can be significantly regulated if the illness is a investigator at an early stage and the preventive decision is made as soon as possible. This work acknowledged that the performance changes could be obtained through the utilization of non-contact methods.

References

[1] https://indianexpress.com/article/lifestyle/health/heart-disease-now-killing-more-people-than-ever-before-who-7099058

[2] Abdel-Basset, M., Gamal, A., Manogaran, G., and Long, H. V. (2019). A novel group decision making model based on neutrosophic sets for heart disease diagnosis. Multimedia Tools and Applications, 1-26.

[3] Abdul-Aziz, A. A., Desikan, P., Prabhakaran, D., and Schroeder, L. F. (2019). Tackling the burden of cardiovascular diseases in India: The essential diagnostics list. Circulation: Cardiovascular Quality and Outcomes, 12(4), e005195.

[4] Vijayavanan, M., Rathikarani, V., and Dhanalakshmi, P. (2014). Automatic classification of ECG signal for heart disease diagnosis using morphological features. International Journal of Computer Science and Engineering Technology, 5(4), 449-455.

[5] Yang, B., Li, X., Hou, Y., Meier, A., Cheng, X., Choi, J. H., ... and Li, A. (2020). Non-invasive (non-contact) measurements of human thermal physiology signals and thermal comfort/discomfort poses-A review. Energy and Buildings, 110261.

[6] Scalise, L. (2012). Non contact heart monitoring. Advances in Electrocardiograms-Methods and Analysis, 81-106.

[7] Hassan, M. A., Malik, A. S., Fofi, D., Karasfi, B., and Meriaudeau, F. (2020). Towards health monitoring using remote heart rate measurement using digital camera: A feasibility study. Measurement, 149, 106804.

[8] Kranjec, J., Beguš, S., Geršak, G., and Drnovšek, J. (2014). Non-contact heart rate and heart rate variability measurements: A review. Biomedical signal processing and control, 13, 102-112.

[9] Asthana, A., Zafeiriou, S., Cheng, S., and Pantic, M. (2013). Robust discriminative response map fitting with constrained local models. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3444-3451).

[10] Li, X., Chen, J., Zhao, G., and Pietikainen, M. (2014). Remote heart rate measurement from face videos under realistic situations. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4264-4271).

[11] Rouast, P. V., Adam, M. T., Chiong, R., Cornforth, D., and Lux, E. (2018). Remote heart rate measurement using low-cost RGB face video: a technical literature review. Frontiers of Computer Science, 12(5), 858-872.

[12] Hassan, M. A., Malik, A. S., Fofi, D., Saad, N., Karasfi, B., Ali, Y. S., and Meriaudeau, F. (2017). Heart rate estimation using facial video: A review. Biomedical Signal Processing and Control, 38, 346-360.

[13] Prakash, S. K. A., and Tucker, C. S. (2018). Bounded Kalman filter method for motion-robust, non-contact heart rate estimation. Biomedical optics express, 9(2), 873-897.

[14] Fernandes, S. L., Gurupur, V. P., Sunder, N. R., Arunkumar, N., and Kadry, S. (2017). A novel nonintrusive decision support approach for heart rate measurement. Pattern Recognition Letters.

[15] Alzahrani, A., and Whitehead, A. (2015, June). Preprocessing realistic video for contactless heart rate monitoring using video magnification. In 2015 12th Conference on Computer and Robot Vision (pp. 261-268). IEEE.

[16] Wadhwa, Neal, Hao-Yu Wu, Abe Davis, Michael Rubinstein, Eugene Shih, Gautham J. Mysore, Justin G. Chen et al. "Eulerian video magnification and analysis." Communications of the ACM 60, no. 1 (2016): 87-95.

[17] Moya-Albor, E., Brieva, J., Ponce, H., Rivas-Scott, O., and Gómez-Peña, C. (2018, July). Heart rate estimation using Hermite transform video magnification and deep learning. In 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp. 2595-2598). IEEE.

[18] Bennett, S. L., Goubran, R., and Knoefel, F. (2016, May). Adaptive eulerian video magnification methods to extract heart rate from thermal video. In 2016 IEEE International Symposium on Medical Measurements and Applications (MeMeA) (pp. 1-5). IEEE.

[19] Moya-Albor, E., Brieva, J., Ponce, H., and Martínez-Villaseñor, L. (2020). A non-contact heart rate estimation method using video magnification and neural networks. IEEE Instrumentation and Measurement Magazine, 23(4), 56-62.

[20] Sarkar, A., Abbott, A. L., Doerzaph, Z., and Sykes, K. (2016, January). Evaluation of video magnification for nonintrusive heart rate measurement. In 2016 IEEE First International Conference on Control, Measurement and Instrumentation (CMI) (pp. 494-498). IEEE.

[21] Huang, B. Y., and Lin, C. L. (2019, April). Improvement of Environment and Camera Setting on Extraction of Heart Rate Using Eulerian Video Magnification. In International Conference on Biomedical and Health Informatics (pp. 381-388). Springer, Cham.
[22] Kranjec, J., Beguš, S., Drnovšek, J., and Geršak, G. (2013). Novel methods for noncontact heart rate measurement: A feasibility study. IEEE transactions on instrumentation and measurement, 63(4), 838-847.
[23] Fletcher, R., and Han, J. (2009, June). Low-cost differential front-end for Doppler radar vital sign monitoring. In 2009 IEEE MTT-S International Microwave Symposium Digest (pp. 1325-1328). IEEE.
[24] Higashi, K., Sun, G., and Ishibashi, K. (2019, July). Precise Heart Rate Measurement Using Non-contact Doppler Radar Assisted by Machine-Learning-Based Sleep Posture Estimation. In 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp. 788-791). IEEE.
[25] Mercuri, M., Liu, Y. H., Lorato, I., Torfs, T., Wieringa, F., Bourdoux, A., and Van Hoof, C. (2018). A direct phase-tracking doppler radar using wavelet independent component analysis for non-contact respiratory and heart rate monitoring. IEEE transactions on biomedical circuits and systems, 12(3), 632-643.
[26] Singh, P., Kaur, A., Batth, R.S. *et al.* Multi-disease big data analysis using beetle swarm optimization and an adaptive neuro-fuzzy inference system. *Neural Comput & Applic* **33,** 10403–10414 (2021). https://doi.org/10.1007/s00521-021-05798-x

Paper 4:

Non-Contact Heart Rate Measurement using Video Magnification: A Review

Gaganjot Kaur[1*], Jeff Kilby[2]

[1,2] School of Engineering, Computer and Mathematical Sciences, Auckland University of Technology, Auckland, New Zealand
[*]gaganjot.kaur@aut.ac.nz

Abstract: This paper presents a comprehensive review of non-contact heart rate measurement techniques. From 2010 to 2020, over 90 publications from reputable international journals and conferences from 2010 and 2020 were reviewed related to the topic, with just 42 papers claiming to use non-contact approaches in their research. Video magnification can be characterised as either motion magnification or colour magnification, according to the study. The pyramid serves as the foundation for constructing the spatial decomposition. In this review, the following significant elements are considered: (a) the pyramid employed; (b) the filter used; and (c) the magnification factor. For spatiotemporal decomposition, all video magnification approaches for heart rate measurement used various image processing pyramids and filters.

1. Introduction

Heart Rate (HR) measurement plays an important role in the field of medical science and clinical studies. HR focuses on the number of heartbeats in a one-minute time period defined as beats-per-minute (bpm). Heart rate variability (HRV) is defined the time interval of an oscillation in the between consecutive heart beats in milliseconds, which is called 'RR' interval' or 'inter-beat interval (IBI)'.

In 1842 Italian physicist Carlo Matteucci, showed that an electric current is present with each heartbeat [1]. With the advancement in technology and biomedical engineering number of techniques are used to measure heart rate from electrocardiogram (ECG) signals, such as smart bands, smart jackets etc. The Dutch physiologist Willem Einthoven recorded the first ECG in humans [2].

ECG uses the detected R wave in the QRS complex and calculates the time between R waves the R-R interval [3] (see Fig. 1). The QRS the combination of three of the graphical deflections seen on a typical ECG output. The shape of the QRS depends upon different factors like age, physical activity, fitness, stress etc.

Photoplethysmography (PPG) measures heart pulses, by using a pulse oximeter which illuminates the skin and measures changes in light absorption and is usually placed on fingertips, toes and earlobe [4]. The PP interval is the time interval between two successive pulses, shown in Figure 1 (bottom).

ECG electrodes must be attached to the human body for HR detection and are not appropriate because the limitations of skin contact. These sensors attached with skin may discomfort when used for neonates, skin damaged for hyposensitive patients and cancer patients while radiation treatment as they required to be isolated in the treatment room. Using PPG for measuring HR through the variation of optical source. PPG needs enough light for pulse rate detection which obtains blood volume pulse (BVP) signals out from the data obtained. So non-contact methods for HR detection is convenient.

In the past decade, researchers have work on the HR and HRV parameters using a contactless or remote method with a video camera [5-30]. In 2012 a non-contact method Eulerian Video Magnification (EVM) was proposed and developed at Computer Science and Artificial Intelligence Lab (CSAIL) at Massachusetts Institute of Technology (MIT, USA). This enhances the visual effect of time varying blood flow of region of interest (ROI) by amplifying colour channels in video purpose is to amplify the subtle or low-amplitude variations in video which are often invisible to the naked eye [10].

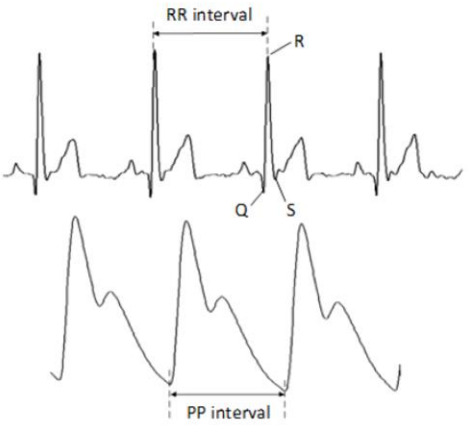


Fig 1 Electrocardiogram signal (top)
Photoplethysmography signal (bottom)

EVM magnify the subtle colour and invisible motion variations within the video. Figure 1 shows the framework of EVM. EVM combines spatial and temporal filtering which represents the subtle temporal changes in videos. Firstly, the recorded standard video sequence is spatially decomposed and after that temporal filtering performed and the output signal is then amplified with amplification factor which depends different applications. After amplification it shows the small motions in video which are not possible to see in original input video. The EVM is inspired Eulerian approach which deals with fluid motion properties such as pressure and velocity over time and space [31].

Figure 2 shows the overall picture of EVM. Firstly, the input video is decomposed into number of spatial frequency bands and for all bands same filter was used. All filtered spatial bands are amplified using a given factor, added back to the input signal and generate the output video. According to the applications different amplification factor and temporal filter is selected.

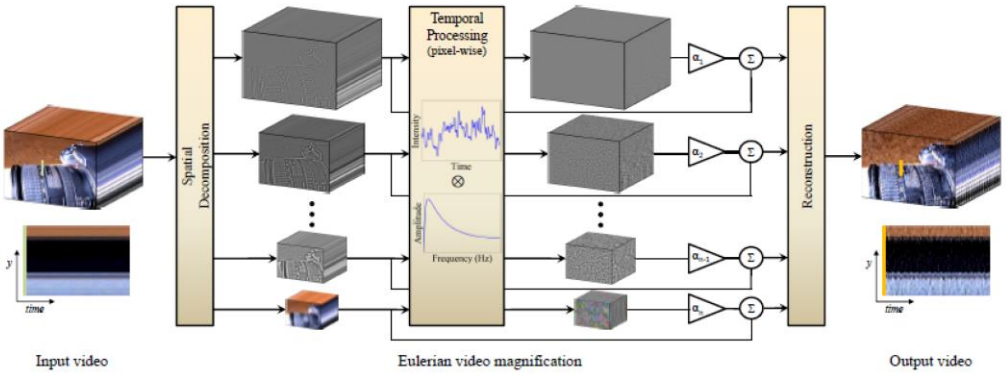


Fig. 2. *Overview of the Eulerian video magnification framework [10]*

2. Literature review of Video Magnification

A detailed literature review of video magnification was carried out, by examining review and research papers for research-related field of video magnification parameters and methods, from the year 2010 until 2020.

Over 90 papers were searched for the study and research related to non-contact measurement of heart rate. The main component of this research reveals temporal variation in videos that are unapparent to the naked eye. The two approaches of video magnification were mentioned in the literature review ***'motion magnification'*** and ***'colour magnification'.***

From the initial number of papers sourced from the journal and conference databases, only 42 of those papers met the criteria of relevance for the literature review of this research .

The literature review showed there were different parameters used like type of filters, amplification factor, different pyramids used for decomposition and various region of interest (ROI) consider like face, nose, wrist, neck etc.

The literature shows number of authors has used different kinds of pyramids and filters, named:

- Laplacian pyramid, Ideal bandpass [32, 33]
- Laplacian pyramid, Butterworth bandpass [34, 35]
- Laplacian pyramid, Second-order IIR bandpass [34]
- Gaussian pyramid [36]

And they use different combination for different examples and effects.

2.1 Pyramids

Pyramid is developed by the image and signal processing communities. Basically, pyramids are type of multiscale signal representation in which image is subject to repeated smoothing and subsampling [37]. Pyramids are built by using multiple copies of same image but different resolution. Two main operations of pyramids are "REDUCE" and "EXPAND". The lowest level of pyramid is the highest resolution and highest level having the lowest resolution. The two categories of pyramids are Low pass and Band pass.

A low pass pyramid is made by using the correct smoothing filter for smooth the image and then subsample the image. The resulting image is again following the same procedure and the process repeated several times. Each pattern of this procedure results in a modest small image with sampling density which results the drop of image resolution. The graphic representation of entire multi-scale will look like a pyramid as shown in figure 3, each cycle's following the modest picture stacked one a top the other.

A band pass pyramid created by calculate the difference between images at next levels in the pyramid and performing the image interpolation between adjacent levels of resolution, to enable computation of pixel wise differences.

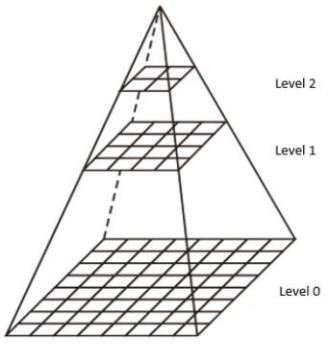


Fig. 3. *Representation of Image pyramid for three levels [38; Chapter 15]*

2.2 Gaussian Pyramid

Gaussian pyramid breaks down the image into successively smaller group of pixels, image is repeatedly filtered and subsampled to generate the sequence of reduced resolution for the purpose of blurring the image [36] .

Each pixel containing a local average that corresponds to a pixel neighbourhood on a lower level of the pyramid. Figure 4 shows the process to generate Gaussian pyramid which is one-dimensional graphic. Each row of dots represents nodes within a level of the pyramid [39]. The value of each node in the zero level is just the grey level of a corresponding image pixel. The value of each node in a high level is the weighted average of node values in the following lower level.

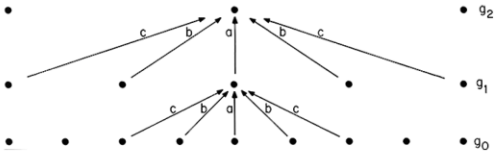


Fig. 4. *A one-dimensional graphic representation of the process which generates a Gaussian pyramid [39]*

REDUCE operation of Gaussian pyramid is basically simple convolution. Each level can be computed with a single convolution, given as:

$$g_L(i, j) = \sum_{m=-2}^{2} \sum_{n=-2}^{2} w\ (m, n)\ g_{L-1}\ (2i + m, 2j + n) \quad (1)$$

where g_{L-1} is level of pyramid and g_L gets the l^{th} level of Gaussian pyramid. In convolution there is (2i + m, 2j + n) which shows it reduced by half.

$$g_L = \text{REDUCE}\ [g_{L-1}] \quad (2)$$

The EXPAND operation of Gaussian pyramid is also used convolution, for this EXPAND applied to array g_L of the Gaussian pyramid would yield an array g_L,1 which is the same size as g_{L-1} .

Let $g_{L,n}$ be the result of expanding g_Ln times. Then

$$g_{L,n} = g_L \quad (3)$$

$$g_{L,n} = \text{EXPAND}\ (g_{L,n} - 1) \quad (4)$$

and

$$g_{L,n}(i, j) = \sum_{m=-2}^{2} \sum_{n=-2}^{2} w\ (m, n)\ g_{L,n-1}\left(\frac{i-m}{2}, \frac{j-n}{2}\right) \quad (5)$$

Only terms for which (i—m)/2 and (j—n)/2 are integers are included in this sum others will be rejected.

Figure 5 shows the Gaussian pyramid with six levels for the "Lady" image. The original image, level 0, 257 by 257 pixels, and each higher-level array is roughly half of its predecessor. So, level 5 measures just 9 by 9 pixels [40].

Fig. 5. *Six levels of the Gaussian pyramid for the "Lady" image [39]*

2.3 Laplacian Pyramid

A Laplacian pyramid is almost same as Gaussian pyramid but it saves the difference image of the blurred versions between each levels [10, 31, 37, 39, 40].

The Laplacian pyramid is a sequence of error images $L_0, L_1, \ldots, L_N$. Each pyramid level is the variance between two levels of the Gaussian pyramid.

$$L_1 = g_l - \text{EXPAND}\ [g_2] \quad (6)$$

$$L_2 = g_l - \text{EXPAND}\ [g_3] \quad (7)$$

To decode it can be shown that the input image can be retrieved completely by expanding and sum all the levels of the Laplacian pyramid. Compute the Gaussian pyramid from Laplacian pyramid.

$$g_N = L_N \quad (8)$$

$$g_3 = \text{EXPAND}\ [g_4] + L_3 \quad (9)$$

$$g_2 = \text{EXPAND}\ [g_3] + L_2 \quad (10)$$

$$g_1 = \text{EXPAND}\ [g_2] + L_1 \quad (11)$$

Last level of Gaussian g_1 is reconstructed image. Figure 5 shows the summary of the steps in Laplacian pyramid coding and decoding [39].

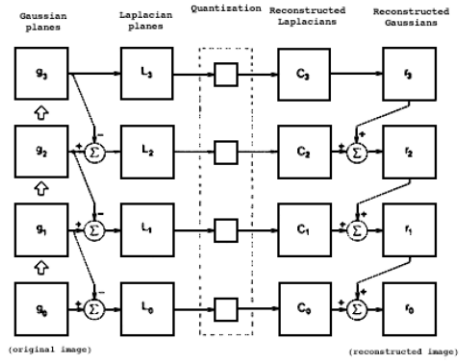


Fig. 6. *A summary of the steps in Laplacian pyramid coding and decoding [39]*

In figure 6 it shows first, the original image g0 (lower left) is used to create Gaussian pyramid levels g_1, g_2, … through the process of repeating the local averaging. Levels of the Laplacian pyramid L0, L_1, … are then computed as the differences between next adjacent Gaussian levels. Laplacian pyramid elements are quantized to yield the Laplacian pyramid code C_0 ,C_1…. Finally, a reconstructed or output image r_0 is generated by summing levels of the code pyramid.

2.4 Steerable Pyramid

A steerable pyramid is an implementation of a multi-scale, multi-orientation band pass filter bank. The basic function of steerable pyramid are K^{th} order directional derivative operators for any choice of K, which come in different sizes and K+1 orientation [41] [42].

Figure 7 illustrates a three level steerable pyramid decomposition of image, with k = 1. Shown are the bandpass images and the final lowpass image.

Fig. 7. A 3-level k = 1 (non-oriented) steerable pyramid. Shown are the bandpass images and the final lowpass image[42].

The advantage of steerable pyramid is self-inverting, and the errors introduce by quantization of the sub bands will not appear as low frequency distortions upon reconstruction.

Figure 8 illustrates a 3-level steerable pyramid decomposition with k = 3.

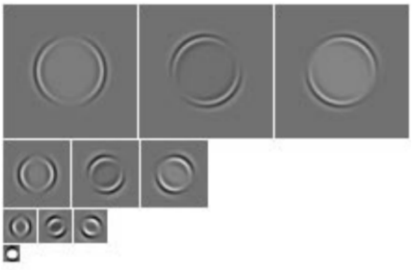


Fig. 8. *A 3-level k = 3 (second derivative) steerable pyramid. Shown are the three bandpass images at each scale and the final lowpass image [42]*

3. Eulerian Video Magnification Method

The essential methodology was to observe the series of colour values in time scale at any spatial location, say pixel, and then amplify the change in a certain temporal-frequency band [10]. Figure 9 shows the revised basic block diagram of EVM. The different steps followed to process video frame by frame and spatial and temporal processing are described as follows: [10, 12, 14, 25, 27, 43-45]

Step (1) Decompose Video: Performed spatial decomposition to identify the area of interest and apply full Laplacian pyramid.

Spatial Decomposition is basically the standard process of building Gaussian and Laplacian pyramid. During pyramid creation process of convolution some information might loss but at same time more generalized image received with less noise.

Step (2) Temporal Processing: Performed on each spatial band and the choice of band pass filter depends on the frequency band of interest. For motion magnification, a filter with a broad passband was preferred; for colour amplification a narrow passband produces a more noise-free result.

Step (3) Magnification/Amplification Factor α: The value of α depends on applications used and the boundary of α is decided by the video motion step δ and the image structure spatial wavelength λ. The equation is given by

$$(1+\alpha) \times \delta(t) < \lambda / 8 \qquad (12)$$

where, δ(t)= spatial displacement function
Spatial wavelength (λ)= $2\pi/\omega$
ω = spatial frequency

Step (4) Spatial Frequency: Select a spatial frequency cut-off (specified by spatial wavelength, λc) beyond which an attenuated version of α is used.

Step (5) Output: Amplified and reconstruct the signal. Reconstruction of final output video by adding the magnified signal to original signal. That process is first order approximation.

The pseudo code of EVM is shown in Table2. The limitation of this linear approach was, it amplifies the noise when the magnification factor was increased. That means it allows only limited or small magnification factors.

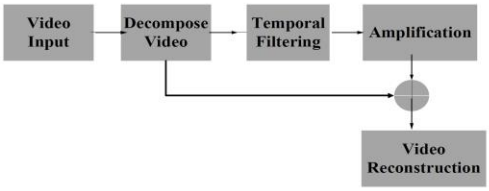


Fig. 9. *Revised block diagram showing EVM steps used to magnify the signals [12]*

4. Phase based Video Magnification

The phase-based method based on the linear Eulerian magnification method [5, 10, 44, 46, 47] and improves it in two important aspects: Firstly, large magnifications, and second has substantially better noise performance [46]. Method modifies phases instead of amplitudes which does not amplify the magnitude of spatial noise. Phase based method was following the Fourier shift theorem to make connection between the phase variations and motions in space.

Temujin G. et al. [5] described the three stages of phase-based approach. In first stage, the image sequence by spatially filtered using a set of quadrature pairs of Gabor filters, and compute the phase responses, yielding estimates of the velocity component in directions orthogonal to the filter pairs' orientations. In second step, if the corresponding filter pair's phase information is not linear over a given time span, a component velocity was rejected. In final step, the remaining component velocities at a single spatial location are combined to produce an estimate of full velocity, V, at that position.

4.1 Phase based video processing using complex Steerable Pyramid

Wadhwa et al. [46] proposed a new Eulerian method based on complex steerable pyramids, which was inspired by phase-based optical flow methods [5].

They used complex steerable pyramids to decompose the video and extract the amplitude of the local wavelets from their phase. Then, phases are passed through temporal bandpass filters to filter out the specific temporal frequencies and any temporal DC component. As these band passed phases correspond to motion, they multiply the band passed phases by some amplification factor (α) [43],and then amplified phase differences to magnify the motion for each frame. The limits of α for octave-bandwidth (for 4 orientation) steerable pyramid [42], used

$$\alpha\, \delta(t) < \lambda /4 \qquad (13)$$

and from Neal W. et al.[46] Limits of α for half-octave (for 8 orientation) pyramid

$$\alpha\, \delta(t) < \lambda /2 \qquad (14)$$

Figure 10 shows the revised block diagram of phase-based video magnification using complex steerable pyramid.

Their proposed method has advantages over EVM, such as, it supports larger magnification factors and has less noise. That method can also be used to attenuate and eliminate the low-amplitude motion during the process of colour magnification. After all, it was more complex, costly, over-complete and execution time is more as compared to EVM.

Fig. 10. *Phase based video Magnification using complex steerable pyramid[43]*

4.2 Riesz Pyramid for Fast Phase Based Video Processing

Neal W.[15] presented a new compact image Riesz pyramid which can be used for real time phase based motion magnification. Riesz pyramid was designed by divide the input image into non oriented sub bands by using an efficient invertible replacement for the Laplacian pyramid and then took approximate Riesz transform of each band. That all process was done in spatial domain.

Figure 11 shows the block diagram for motion magnification using Riesz pyramid. With this method the problem may occur at points where there is not a single dominant orientation.

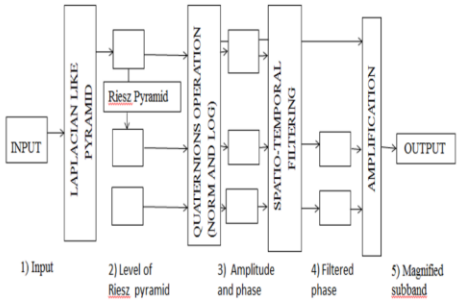


Fig. 11. *Block diagram of Motion magnification using Riesz Pyramid [43]*

5. Enhanced Eulerian Video Magnification

Le L. et.al [12, 15] introduced a post processing method to improve the EVM technique which was based on spatio temporal filtering .Authors used EVM as a video spatio temporal motion analyser to achieve pixel level motion mapping.

To amplify the motion input video pixels were wrapped based on pixel level motion mapping. Table 1 and Table 2 shows the pseudo codes of EVM and enhanced eulerian video magnification (E2VM) respectively.

Table 1 The Pseudo code of EVM [12]

Step	Processing
Input	Video $I(\mathbf{x},t)$, $t=0$
1	Read current frame I
2	Computing Gaussian and Laplacian pyramid: $\{G(I)\}_{l=0}^{n}$ $\{L(I)\}_{l=0}^{n}$
3	Applying band-pass filter: $y_L(\mathbf{x},t)=\omega_L L_l(\mathbf{x},t)+(1-\omega_L)L_l$ $(\mathbf{x},t-1)\ y_H(\mathbf{x},t)=\omega_H L_l(\mathbf{x},t)+(1-\omega_H)L_l$ $(\mathbf{x},t-1)\ B_l(\mathbf{x},t)=y_H(\mathbf{x},t)-y_L(\mathbf{x},t)$
4	Space-time video processing: $\tilde{L}_l(\mathbf{x},t)=L(\mathbf{x},t)+(1+\alpha)B(\mathbf{x},t)$
5	Reconstruction: $\tilde{L}(I)\leftarrow\tilde{L}(I)+upSample(\tilde{L}\ (I))$ $\tilde{I}\leftarrow\tilde{L}(I)$
6	Rewrite current frame: $I\leftarrow\tilde{I}$
7	If $t<t_{\max}$, t ++ goto <1>

E2VM has advantages over EVM such as, it supports larger magnification factors and less affected by frame noise, it does not include modification of pixel values. On the other side, processing time is more and some magnification specifications may be lost during image warping. Moreover, as E2VM uses the EVM as part of its algorithm, any failure in the EVM effects the results [47].

Table 2 Pseudo code of E2VM [12]

Step	Processing
Input	Video $I(\mathbf{x},t)$, $t=0$
1	Read current frame I
2	Computing Gaussian and Laplacian pyramid $\{G(I)\}_{l=0}^{n}$ $\{L(I)\}_{l=0}^{n}$
3	Applying band-pass filter: $y_L(\mathbf{x},t)=\omega_L L_l(\mathbf{x},t)+(1-\omega_L)L_l$ $(\mathbf{x},t-1)\ y_H(\mathbf{x},t)=\omega_H L_l(\mathbf{x},t)+(1-\omega_H)L_l$ $(\mathbf{x},t-1)\ B_l(\mathbf{x},t)=y_H(\mathbf{x},t)-y_L(\mathbf{x},t)$
4	Space-time video processing $\tilde{L}_l(\mathbf{x},t)=L_l(\mathbf{x},t)+(1+\alpha_l)B_l(\mathbf{x},t)$
5	Reconstruction $\tilde{L}_l(I)\leftarrow\tilde{L}_l(I)+upSample(\tilde{L}_{l+1}(I)$ $\tilde{I}\leftarrow\tilde{L}_0(I)$
6	Motion Analysis $U(x,y)=\frac{\hat{I}(x,y)-I(x,y)}{\frac{\delta t(x,y)}{\delta x}}$ $V(x,y)=\frac{\hat{I}(x,y)-I(x,y)}{\frac{\delta t(x,y)}{\delta y}}$
7	Warping Frame $\hat{I}(x,y)=I(x+\beta(x,y)U(x,y),y+\beta(x,y)V(x,y))$
8	Rewrite current frame I ← $\hat{I}$

6. Efficient Motion Magnification System

Efficient motion magnification system (EMMS) better than other methods in terms of noise removal, output video quality, overall performance at large magnification factors and reduce the execution time. Ali A.N. et.al [47] proposed EMMS as shown in Fig 12.

The algorithm of EMMS considered three main steps as follows:

Step (1) Separate the intensity information from the colour information by converting Y Cb Cr colour space of RGB from sequence of the source video.

Step (2) To reduce the processing time the intensity channel (Y) resized down to 50% by using Lanczos resampling method.

Step (3) The Y channel is decomposed into number of spatial frequency bands using a multiresolution pyramid analysis.

In EMMS method [47] they used multiresolution pyramidal wavelet decomposition to decompose image into

sequence of image transformations with different spatial resolution to extract the information content in an image.

For reconstruct the wavelet pyramid the inverse discrete wavelet transform (DWT) was used. They used 2-D DWT based on Haar family, as this is simplest to implement among all other wavelet families. Temporal filtering applied on every single level of wavelet pyramid based on zero phase Chebyshev type 1 band pass filter. The Chebyshev filter provides a smaller root mean square error (RMSE) than Butterworth filter during the process of magnification.

EMMS shows better results as compared to other methods of video magnification. EMMS has low execution time and better video quality.

Table 3 shows the performance based comparison of video magnification methods. Different pyramids are used by individual authors and minimized the noise from the result video. As shown in table EMMS is faster as compared to EVM process also reduce the noise.

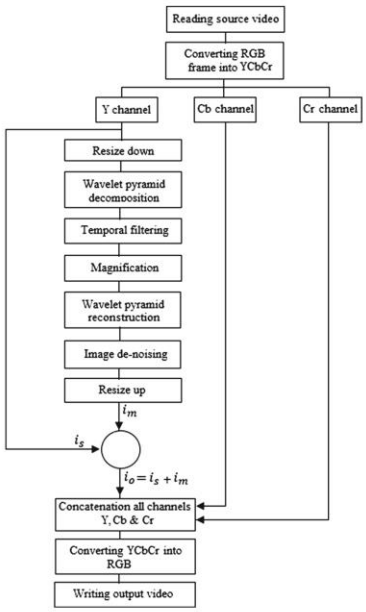


Fig. 12. *Block diagram of EMMS[48]*

Table 3 Comparison of video magnification method

Parameter	EVM	Phase Based	Fast Phase Based	E2VM	EMMS
Decomposition	Laplacian Pyramid	Complex Steerable Pyramid	Riesz Pyramid	Laplacian Pyramid	Wavelet Pyramid
Noise	Magnified	Translated	Translated	Minimized	Minimized
Amplification Factor (α)	Small Value of α	Large Value of α	Large Value of α	Medium Value of α	Large Value of α
Limits of α	$(1+\alpha)\delta(t) < \lambda/8$	$\alpha\delta(t) < \lambda n/4$	Similar as complex steerable pyramid	-	-
Over completeness	3/4 image	3-4 times slower than EVM with octave bandwidth ,2 orientations	~ 20-80% faster than phase based method	~15-20% slower than EVM	~60 -70% faster than EVM

7. Conclusion

Video magnification has a wide scope in the field of biomedical applications because it can extract the useful information from video and image sequences. The existing video magnification techniques having the limitations of decreased video quality and increased noise level when tried to increase the magnification factor. These current methods also take a long time for execution. To solve the problems, many improvements have been made by researchers based on the decomposition. The execution time is compared in table 3 for all video magnification methods, and table 5 shows the comparison between the different parameters of magnified videos for each technique.

From the literature review, it shows that the researchers mainly focused on motion magnification as compared to colour magnification.

8. References

[1] G. Moruzzi, "The electrophysiological work of Carlo Matteucci," *Brain Research Bulletin,* vol. 40, pp. 69-91, 1996. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/0361923096000366?via%3Dihub.

[2] L. S. Gettes and E. H. Chung, "Electrocardiography," in *Netter's Cardiology*, G. A. M. D. Stouffer, M. S. M. D. P. Runge, C. M. D. M. B. A. Patterson, and J. S. M. D. Rossi Eds., 2019, sec. 3-s2.0-B9780323547260000078, pp. 43-51.

[3] S. C. Mukhopadhyay, *Wearable Electronics Sensors: For Safe and Healthy Living*. 2015.

[4] S. C. Mukhopadhyay, *Wearable electronics sensors : for safe and healthy living* (Smart sensors, measurement and instrumentation: 15). Springer, 2015.

[5] T. Gautama and M. A. V. Hulle, "A phase-based approach to the estimation of the optical flow field using spatial filtering," *IEEE Transactions on Neural Networks,* vol. 13, no. 5, pp. 1127-1136, 2002, doi: 10.1109/TNN.2002.1031944.

[6] C. Liu, A. Torralba, W. T. Freeman, F. Durand, and E. H. Adelson, "Motion magnification," presented at the ACM SIGGRAPH 2005 Papers, Los Angeles, California, 2005. [Online]. Available: https://doi.org/10.1145/1186822.1073223 https://dl.acm.org/doi/pdf/10.1145/1073204.1073223.

[7] A. D. Droitcour *et al.*, "Non-contact respiratory rate measurement validation for hospitalized patients," in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 3-6 Sept. 2009 2009, pp. 4812-4815, doi: 10.1109/IEMBS.2009.5332635. [Online]. Available: https://ieeexplore.ieee.org/document/5332635/

[8] G. Lu, F. Yang, Y. Tian, X. Jing, and J. Wang, "Contact-free Measurement of Heart Rate Variability via a Microwave Sensor," (in eng), *Sensors (Basel, Switzerland),* vol. 9, no. 12, pp. 9572-9581, 2009, doi: 10.3390/s91209572.

[9] A. K. Abbas, K. Heimann, K. Jergus, T. Orlikowsky, and S. Leonhardt, "Neonatal non-contact respiratory monitoring based on real-time infrared thermography," *BioMedical Engineering OnLine,* vol. 10, no. 1, p. 93, 2011/10/20 2011, doi: 10.1186/1475-925X-10-93.

[10] H.-Y. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, and W. Freeman, "Eulerian Video Magnification for Revealing Subtle Changes in the World," *ACM Transactions on Graphics - TOG,* vol. 31, 07/01 2012, doi: 10.1145/2185520.2185561.

[11] N. Wadhwa, M. Rubinstein, F. Durand, and W. Freeman, "Phase-Based Video Motion Processing," *ACM Transactions on Graphics (TOG),* vol. 32, 07/01 2013, doi: 10.1145/2461912.2461966.

[12] X. He, R. A. Goubran, and X. P. Liu, "Using Eulerian video magnification framework to measure pulse transit time," in *2014 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 11-12 June 2014 2014, pp. 1-4, doi: 10.1109/MeMeA.2014.6860029. [Online]. Available: https://ieeexplore.ieee.org/document/6860029/

[13] J. Kranjec, S. Beguš, G. Geršak, and J. Drnovšek, "Non-contact heart rate and heart rate variability measurements: A review," *Biomedical Signal Processing and Control,* vol. 13, pp. 102–112, 09/01 2014, doi: 10.1016/j.bspc.2014.03.004.

[14] N. Miljković and D. Trifunović, "Pulse rate assessment: Eulerian Video Magnification vs. electrocardiography recordings," in *12th Symposium on Neural Network Applications in Electrical Engineering (NEUREL)*, 25-27 Nov. 2014 2014, pp. 17-20, doi: 10.1109/NEUREL.2014.7011447. [Online]. Available: https://ieeexplore.ieee.org/document/7011447/

[15] N. Wadhwa, M. Rubinstein, F. Durand, and W. T. Freeman, "Riesz pyramids for fast phase-based video magnification," in *2014 IEEE International Conference on Computational Photography (ICCP)*, 2-4 May 2014 2014, pp. 1-10, doi: 10.1109/ICCPHOT.2014.6831820. [Online]. Available: https://ieeexplore.ieee.org/document/6831820/

[16] K. Murakami, M. Yoshioka, and J. Ozawa, "Non-contact pulse transit time measurement using imaging camera, and its relation to blood pressure," in *2015 14th IAPR International Conference on Machine Vision Applications (MVA)*, 18-22 May 2015 2015, pp. 414-417, doi: 10.1109/MVA.2015.7153099. [Online]. Available: https://ieeexplore.ieee.org/document/7153099/

[17] S. L. Bennett, R. Goubran, and F. Knoefel, "Adaptive eulerian video magnification methods to extract heart rate from thermal video," in *2016 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 15-18 May 2016 2016, pp. 1-5, doi: 10.1109/MeMeA.2016.7533818. [Online]. Available: https://ieeexplore.ieee.org/document/7533818/

[18] E. Blackford, A. Piasecki, and J. Estepp, *Measuring pulse rate variability using long-range, non-contact imaging photoplethysmography*. 2016, pp. 3930-3936.

[19] X. He, R. A. Goubran, and X. P. Liu, "Wrist pulse measurement and analysis using Eulerian video magnification," in *2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, 24-27 Feb. 2016 2016, pp. 41-44, doi: 10.1109/BHI.2016.7455830. [Online]. Available: https://ieeexplore.ieee.org/document/7455830/

[20] R.-Y. Huang and L.-R. Dung, "Measurement of heart rate variability using off-the-shelf smart phones," *BioMedical Engineering OnLine,* vol. 15, 01/29 2016, doi: 10.1186/s12938-016-0127-8.

[21] J. Lai, S. Tsai, R. Lin, and S. Huang, "Using video magnification approach directly to extract heart-beat," in *2016 International Conference on Applied System Innovation (ICASI)*, 26-30 May 2016 2016,

pp. 1-4, doi: 10.1109/ICASI.2016.7539815. [Online]. Available: https://ieeexplore.ieee.org/document/7539815/

[22] A. Lam and Y. Kuno, *Towards Estimating Heart Rates from Video Under Low Light*. 2016, pp. 494-503.

[23] N. Wadhwa *et al.*, "Eulerian video magnification and analysis," *Commun. ACM,* vol. 60, no. 1, pp. 87–95, 2016, doi: 10.1145/3015573.

[24] K. Alghoul, S. Alharthi, H. Al Osman, and A. El Saddik, "Heart Rate Variability Extraction from Videos Signals: ICA vs. EVM Comparison," *IEEE Access,* vol. PP, pp. 1-1, 03/20 2017, doi: 10.1109/ACCESS.2017.2678521.

[25] M. A. Hassan *et al.*, "Heart rate estimation using facial video: A review," *Biomedical Signal Processing and Control,* vol. 38, pp. 346-360, 2017.

[26] Y.-C. Lin, N.-K. Chou, G.-Y. Lin, M.-H. Li, and Y.-H. Lin, "A Real-Time Contactless Pulse Rate and Motion Status Monitoring System Based on Complexion Tracking," (in eng), *Sensors (Basel, Switzerland),* vol. 17, no. 7, p. 1490, 2017, doi: 10.3390/s17071490.

[27] Y. S. Dosso, A. Bekele, and J. R. Green, "Eulerian Magnification of Multi-Modal RGB-D Video for Heart Rate Estimation," in *2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 11-13 June 2018 2018, pp. 1-6, doi: 10.1109/MeMeA.2018.8438741. [Online]. Available: https://ieeexplore.ieee.org/document/8438741/

[28] X. He *et al.*, "Video-Based Analysis of Heart Rate Applied to Falls," in *2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 11-13 June 2018 2018, pp. 1-5, doi: 10.1109/MeMeA.2018.8438773. [Online]. Available: https://ieeexplore.ieee.org/document/8438773/

[29] A. Melchor Rodríguez and J. Ramos-Castro, "Video pulse rate variability analysis in stationary and motion conditions," (in eng), *Biomed Eng Online,* vol. 17, no. 1, p. 11, Jan 29 2018, doi: 10.1186/s12938-018-0437-0.

[30] C. Ordóñez, C. Cabo, A. Menéndez, and A. Bello, "Detection of human vital signs in hazardous environments by means of video magnification," (in eng), *PloS one,* vol. 13, no. 4, pp. e0195290-e0195290, 2018, doi: 10.1371/journal.pone.0195290.

[31] K. Kamble, N. P. Jagtap, R. Patil, and A. A. Bhurane, "A Review: Eulerian Video Motion Magnification," *International Journal of Innovative Research in Computer and Communication Engineering,* vol. 3, pp. 2384-2390, 2015.

[32] L. Liu, L. Lu, J. Luo, J. Zhang, and X. Chen, "Enhanced Eulerian video magnification," in *2014 7th International Congress on Image and Signal Processing*, 14-16 Oct. 2014 2014, pp. 50-54, doi: 10.1109/CISP.2014.7003748. [Online]. Available: https://ieeexplore.ieee.org/document/7003748/

[33] Y. S. Dosso, A. Bekele, and J. R. Green, "Eulerian Magnification of Multi-Modal RGB-D Video for Heart Rate Estimation," in *IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 11-13 June 2018 2018, pp. 1-6, doi: 10.1109/MeMeA.2018.8438741. [Online]. Available: https://ieeexplore.ieee.org/document/8438741/

[34] W. H. Y, R. M, S. E, G. J, D. F, and F. W, "Eulerian video magnification for revealing subtle changes in the world," 2012.

[35] S. L. Bennett, R. Goubran, and F. Knoefel, "Adaptive eulerian video magnification methods to extract heart rate from thermal video," in *IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 15-18 May 2016 2016, pp. 1-5, doi: 10.1109/MeMeA.2016.7533818. [Online]. Available: https://ieeexplore.ieee.org/document/7533818/

[36] F. D. V. R. Oliveira, J. G. R. C. Gomes, J. Fernández-Berni, R. Carmona-Galán, R. d. Río, and R.-V. Á, "Gaussian Pyramid: Comparative Analysis of Hardware Architectures," *IEEE Transactions on Circuits and Systems I: Regular Papers,* vol. 64, no. 9, pp. 2308-2321, 2017, doi: 10.1109/TCSI.2017.2709280.

[37] E. H. Adelson, P. J. Burt, C. H. Anderson, J. M. Ogden, and J. R. Bergen, "PYRAMID METHODS IN IMAGE PROCESSING," 1984.

[38] J.Beyerer, F.P.Leon, and C.Frese, *Macine Vision*. Springer, 2016, pp. 721-749.

[39] P. Burt and E. Adelson, "The Laplacian Pyramid as a Compact Image Code," *IEEE Transactions on Communications,* vol. 31, no. 4, pp. 532-540, 1983, doi: 10.1109/TCOM.1983.1095851.

[40] E. H. Adelson, C. H. Anderson, J. R. Bergen, P. J. Burt, and J. M. Ogden, "Pyramid Methods in Image Processing," 1984.

[41] E. P. Simoncelli and W. T. Freeman, "The steerable pyramid: a flexible architecture for multi-scale derivative computation," *Proceedings., International Conference on Image Processing,* vol. 3, pp. 444-447 vol.3, 1995.

[42] E. P. Simoncelli and W. T. Freeman, "The steerable pyramid: a flexible architecture for multi-scale derivative computation," in *Proceedings., International Conference on Image Processing*, 23-26 Oct. 1995 1995, vol. 3, pp. 444-447 vol.3, doi: 10.1109/ICIP.1995.537667. [Online]. Available: https://ieeexplore.ieee.org/document/537667/

[43] K. S. Kamble, N. Jagtap, R. A. Patil, and A. A. Bhurane, "A Review Eulerian video motion magnification," *International Journal of Innovative Research in Computer and Communication Engineering* 2015.

[44] N. Wadhwa *et al.*, "Eulerian video magnification and analysis," *Communications of the ACM,* vol. 60, no. 1, pp. 87-95, 2016, doi: 10.1145/3015573.

[45] C. Ordonez, C. Cabo, A. Menendez, and A. Bello, "Detection of human vital signs in hazardous environments by means of video magnification," *PLoS One,* vol. 13, no. 4, p. e0195290, 2018, doi: 10.1371/journal.pone.0195290.

[46] W. N., R. M, D. F, and F. W. T., "Phase-based video motion processing," *ACM Transactions on Graphics,* vol. 32, no. 4, p. 1, 2013, doi: 10.1145/2461912.2461966.
[47] A. Al-Naji, S.-H. Lee, and J. Chahl, "An efficient motion magnification system for real-time applications," *Machine Vision and Applications,* vol. 29, no. 4, pp. 585-600, 2018, doi: 10.1007/s00138-018-0916-0.
[48] A. A. Al-Naji, S.-H. Lee, and J. Chahl, "An efficient motion magnification system for real-time applications," *Machine Vision and Applications,* vol. 29, pp. 585-600, 05/01 2018, doi: 10.1007/s00138-018-0916-0.

Paper 5:

International Journal of Advanced Science and Technology
Vol. x, No. x, (20xx), pp. xx-xx

An Optimized Artificial Bee Colony Algorithm for the Improvement of Heart Rate Detection from Video Frames

[1]Gaganjot Kaur [0000-0002-8379-4429], [2]Jeff Kilby [0000-0001-5010-7170]

School of Engineering, Computer and Mathematical Sciences, Auckland University of Technology, Auckland, New Zealand
[1]er.gaganaulakh@gmail.com

Abstract

Heart rate Variability (HRV) analysis plays a paramount role in monitoring the heart rate of the patient. It is termed as fluctuations occurred between the consecutive intervals of heartbeats. In this paper, a novel non -contact HRV measurement system has been developed to monitor the cardiac activity of the patient. The proposed algorithm is divided into two frameworks in which Framework 1 is used to extract the features from the video frames using the Euler Video Magnification (EVM) technique. The face detection and selection of forehead has been done using the optimization technique such as Pixel –Driven - Artificial Bee Colony (PD-ABC). The second framework is based on the training and classification using the Feed forward Back propagation Neural Network (FFBPNN) for the extraction of features without any noise. For robustness, the proposed algorithm has been further compared with the existing techniques such as Convolution Neural Network (CNN), FFBPNN only, and CNN with ABC. The simulation results in terms of TPR, FPR, and classification Accuracy for three different algorithms has been compared with different ratios. The evaluated results show that proposed algorithm performs better in comparison to existing algorithms and improved by 0.14% for TPR, 0.05% for FPR and 0.06% for Accuracy for 90:10 ratio analysis.

Keywords: *EVM, CNN, FFBPNN, Non-Contact Measurement of Heart Rate Based on Video.*

1. Introduction

Heart attacks have become common in working or office going human due to excessive amount of workload and the mind-set of the human to defeat everyone in all the possible races that can earn money and dignity for him/her [Charles and Nixon 2019]. The over workloads, day and night shifts, extra working times in the same environment put a lot of mental pressure that boosts the blood pressure of the human. Extreme of such case, could result into a mild or a serious heart attack, sometimes could be fatal too. The measurement of human physiological characteristics on a frequent basis is an essential component in health care, with implications for healthcare policies and economics [Patel et al. 2018]. For the genome level in the last few years, opening up potential for early identification of symptoms and to revamp the healthcare practices itself. Numerous studies have discovered a link between the autonomic nervous system (ANS) and the mortality rate of the heart system, among other things [Egger 2022]. The cardiac diseases have been identified by determining the perturbations and imbalances of the ANS that may lead to certain death due to cardiac imbalance. The variation in Heart Rate (HR) may lead to serious problems such as attacks that were measured using the fourier transform. The cardiac data measured during the Heart Rate Variability (HRV) analysis has been measured wearable devices, Electrocardiogram (ECG) and photoplethysmography (PPG) [Hinde et al. 2021]. Although both these techniques provide appropriate measurement and accurate results, but PPG is limited due to the handling of massive equipment and noise in an operative mode. ECG is widely regarded as the best method for determining Heart Rate (HR), Heart Rate Variability (HRV) or Inter Beat Intervals (IBI), or the time between two consecutive heart beats [Attaoui et al. 2020]. HR is calculated by measure the beats per minute (bpm) and HRV is time in milliseconds between successive heart beats also referred as RR interval.

ISSN: XXXX-XXXX IJXX

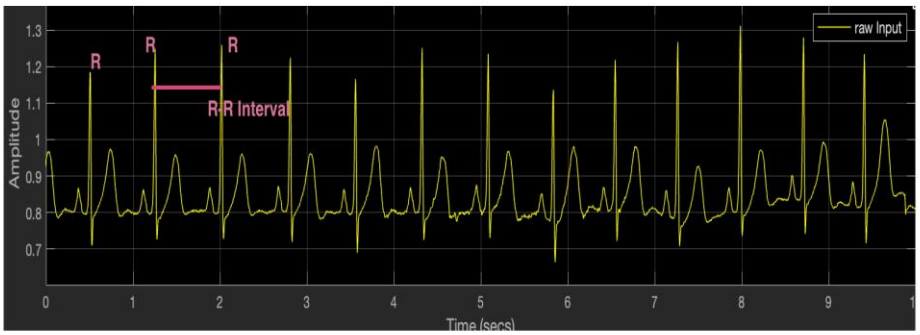


Figure 1. R-R interval: an interval between adjacent

The rate of heart related issues has risen due to the following aspects.

a) Stress economical
b) Stress mental
c) Food intake behaviour

HR has been measured by the oscillations between adjacent intervals R-R as shown in Figure 1 and consecutive intervals for the measurement of instantaneous HR. R-R interval is the duration spent between two consecutive R-waves of the QRS signal on an ECG is a result of both inherent sinus node features and autonomic factors and measured in milliseconds. QRS complex shows the ventricular depolarization, whereas the ST segment and T wave reflects ventricular repolarization and produced by relaxation of Ventricles. In the conventional studies, many automated HR measurement systems has been developed and popularised due to their extensive use for health care applications **[Hussein et al. 2018].** The analysis of the measured signal is very complex and incorrect decisions may lead to extrapolations. However, contact less measurement of the HR activity characteristics based on video could prove to be a useful tool in both clinical and non-clinical settings **[Chen et al. 2018].** In an ideal measuring system, the subject would be unaware of the measurement process, resulting in a lower psychological component to the measurement. Such measures would also result in more objective readings, in addition to removing various limitations of touch sensor-based approaches. For the monitoring of HR, researchers used various algorithms and methods to calculate bpm from different data sets like images, videos, audio and Iris. These all focused on the non-invasive method for the measurement of heart rate [**Nisar et al. 2016; Nogami et al. 2021**]. Several researchers have worked on HR monitoring using contactless devices or remote methods such as using a video camera **[Kranjec et al. 2014; Ni et al. 2021**; **Rouast et al. 2017**].

1.1 Challenges in measuring the HRV and HRV through Video monitoring

Despite many experimental studies has been presented on the measurement of HRV and vast amount of literature focused on providing the clinical tools. But, research is limited as several factors affecting the performance of the proposed tool among other methods. There are several reasons affecting the performance due to this fact. The different factors such as lack of standard methodology, parameters variability, consensus limitation about the accurate measurement of the Heart Rate affect the performance for the health care applications. As mentioned in Figure 1, the HRV analysis measurement is generally based on the R-R intervals, described as period between the adjacent intervals. Due to instantaneous changes in the dynamics of HR measurement system, it is difficult to detect the healthy subjects due to complexity level and subjects having heart –related disorders.

ISSN: XXXX-XXXX IJXX

The R-R intervals measurement usually taken place in the clinics using the ECG or PPG method. There are several factors that limit the use of these methods within a clinical environment. These methods are limited due to the direct contact of the subject with the sensors physically. Thus, such a measurement process is unpleasant and inappropriate for the subjects. For using the HRV, it is significant to determine that the variance is time dependent and increase in recordings length also affect the parameter. Therefore, the statistical variables and parameters obtained during the recording have been compared ensuring that these are of same duration. For this reason, practitioners defined 5 min and 24 h nominal recording as an appropriate data analysis option which is standardized. In comparison to bio-electrical signals, it is easy to obtain the electrical activity of the heart. The amplitude value of the recorded signal is comparatively small and prone to various interferences and noises. Such noises may dominate the original signal and may cause contaminate the actual signal and masking the morphology of the original signal. For the accurate realization of the R –waves, practitioners develop the robust R-wave algorithm to obtain the signal without any noise.

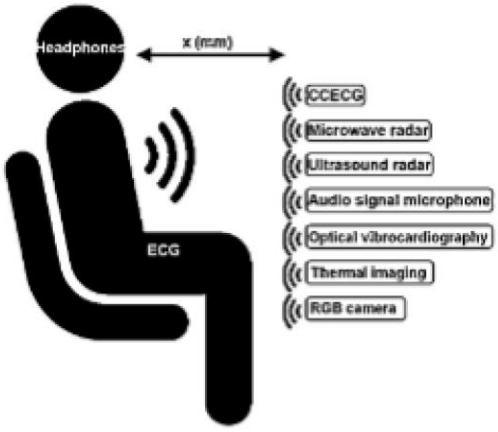


Figure 2. Non-contact method [Kranjec 2014]

Based on the non-contact measurement methods that has been opted in the past, video monitoring has been one of the methods that has been aimed to be opted in the past and has been studied in the literature as well. The general architecture of the processing that is gathered through input source, in case of the proposed work, the input set would be the video frames that has been collected as processed. In order to use the frames, the videos must be magnified as the changes in the face frames are so minute that they cannot be visualized using naked eyes. For the magnification of the video, the proposed algorithm architecture has utilized Euler Video Magnification (EVM) and is illustrated in the related work section as well. It is not necessary that the measurement of the heart can be done using forehead only, there are evidence that illustrates that, the detection can be even done using iris movements and is cited in the related work section as well. The related work section also incorporates the training and classification architectures that has been presented earlier for heart rate prediction using some source of input. The process of classification remains the same regardless of the input source.

This paper aimed to develop the robust HRV measurement system which is non-contact and based on the novel deep learning algorithm. The proposed detector is realized using the Feed Forward Back Propagation Neural Network (FFBPNN) in conjunction with the meta-heuristic algorithm Artificial Bee Colony (ABC). The process of data acquisition is followed by the measurement of cardiac activity signal by eliminating the interference and noisy signal for the analysis of HRV.

ISSN: XXXX-XXXX IJXX

The remained of this paper is structured as section 2 illustrates the conventional methods for video magnification and limitations of the technique. Section 3 describes the research methodology in which working of FFBPNN and optimization technique such as ABC has been illustrated well. Section 4 illustrates the discussion on results and finally conclusion is presented in section 5.

2. Literature Review

The measurement of cardiac activity has been done by fixing the electrodes on the body surface, which is reliable and capture the good signal quality with higher amplitude level. The conventional techniques are capable of measuring the physiological signals, vascular monitoring, and assessment of autonomic function. The related work section incorporates the utilization of HRV, Euler Video Magnification (EVM), training and classification mechanism to support the proposed work methodology.

Monkaresi et al. 2016 implement the tree different techniques to extract the features from the video frames. The methods were face tracking engine to track the face in the videos, determination of local binary pattern, and extraction of HR features using cascade classifier. Further, RELIEF-F feature selection method had been used to select the features from the 7 classes. The authors further used the Oversampling technique to avoid the data imbalance during the training process and classification via K-means clustering to train the data. The validation had been done using the Area under the Region of Convergence (ROC) curve and the simulation results show that 84 features were extracted, and decision level model developed for the different classifiers. The drawback of the study is that some of the features were lost from the video segments, and it was difficult to estimate the HR in a real scenario.

Wadhwa et al. 2016, the authors presented the EVM technique to visualize the motion patterns and subtle color in the recorded videos. It is a computational technique for visualizing the minor color and motion differences in conventional movies by enlarging them. The feature extraction had been done and video magnification to detect the small changes that are difficult or impossible to notice on our own. Furthermore, these minor variations may be quantified and used to retrieve sounds from distant objects, define material qualities, and assess a person's pulse remotely. The observations had been done considering the phase difference in an image sequence. The outcomes were robust and useful to reveal the tiny motions.

Bennett et al. 2017, the goal of this research is to build a robust model considering the video magnification system. The thermal and optical videos had been used to develop a simulation model and hot water was pushed through the apparatus. The captured video frames have been divided to extract the spatial features and each pixel had been processed to extract the temporal features. The experimental trials for facial perfusion using the five people had been performed while video data was collected using both cameras. The mean intensity was computed within a ROI considering the magnification factor of 200. The processing had been done with filter and amplification factor of low value, then temporal bandpass filter was used to extract the facial features. The modelling trials revealed that using thermal video to process the EVM helped to characterize the noise level.

Abnousi et al. 2019, the authors in this study used the EVM to determine the accuracy of a new noninvasive technique. The study collected the data of 50 patients to record the jugular venous pulsations. The Euler motion magnification was further applied and compared considering the three different categories such as examinations of the bedside by the physicians, (2) in-office examinations by the physicians and (3) direct invasive recording of the heart rate. The statistical analysis shows that accuracy had been improved for the clinical examinations of the recorded signal.

Petrovic et al. 2019 provides a unique approach for estimating HRV features using a 24-GHz Doppler radar. For high-accuracy estimation of beat-to-beat intervals, the proposed approach integrates frequency and temporal domain analysis. The authors initially used the in-phase (I) and quadrature (Q) radar components to detect heartbeat information. For coarse HR prediction, the

ISSN: XXXX-XXXX IJXX

authors employed frequency domain analysis. A filter bank having narrowband Band Pass Filter (BPF) processes the combined signal at the same time. Based on the coarse HR prediction, one of the BPF outputs is chosen as the correct output. The extracted error estimated in the range of 1.02-2.07%.

Bent et al. 2020, the authors used an optical method for sensing changes using the PPG to calculate HR from video frames using the wearable devices. PPG inaccuracies can be attributed to three key factors such as different skin types, artifacts, and signal crossover. There has been no systematic investigation of the accuracy of devices over the whole range of skin tones to yet. The authors had been investigated HR monitoring and PPG data from consumer and research-grade wearables. The authors found no statistically significant changes across skin tones but find significant disparities between devices and different activities with absolute inaccuracy during exercise being 30% higher on average than at rest.

Demirezen 2021 proposed the nonlinear model using the remote PPG method to estimate the HR using the facial videos. The extraction of PPG signal had been done by measuring the number of rotations and variations in skin color subjected to heart pulses. The estimation of HR had been done using the nonlinear Mode Decomposition (NMD) and Viola Jones detector used to locate the face region from the first video frame. The authors consider the full face and forehead as ROI using the detector and consistency had been performed to avoid the erroneous estimation of HR. The authors consider the PURE dataset [22 PURE] having 60 face videos and UBFC dataset [Bobbia 23] containing 42 videos and estimate the root mean square error (RMSE) and mean absolute error (MAE). The experimental results states that smallest RMSE and MAE was observed of about 2.68 and 1.91 respectively. The study is limited to work in a real time and does not provide any information for the first 15s of the video frames.

Wang et al. 2021 proposed a millimeter wave based HRV device to monitor the activity of heart. The authors introduced mm-HRV, the HRV monitoring system based on commercial mm wave radio. There are two main components to the design of mm-HRV. The authors created a detector to determine the location of each user. Second, extractor is developed that can maximize the decomposition modulated by movement of the body and, as a result, determine the heartbeat signal. The determination of the peak location can be used to estimate the exact time of heartbeats, and the inter-beat intervals that can be used to evaluate the HRV metrics for each specific target.

Zheng et al. 2022 proposed a novel method to measure the HR remotely by integrating the EVM with chirp model and quality assessment had been done to decompose the video in different scenarios. The video frames had been selected using the EVM and PPG signal used to extract the ground truth information. The temporal changes to determine the color in the facial video had been detected using the EVM algorithm. The feature extraction using the Convolutional Neural Network (CNN) to determine the facial points especially forehead area as a ROI. The Red, Green and Blue (RGB) channels had been used to represent the cropped facial video and green channel considered for average intensities. The outcomes show that the proposed method efficiently recognizes the facial videos having a frame rate of less than and equals to 15 frames per second.

3. Research Methodology

The research methodology is designed into two frameworks: Framework 1 is used to extract the video using HRV measurement system from the video frames. The face detection, extraction of morphological features, and application of the Swarm Intelligence technique has been applied. The system includes the optimization technique such as PD-ABC applied to the extracted signal. The second framework is based on the training and classification using the FFBPNN for the construction of original signal recorded during the measurement system.

3.1 Face Detection Framework

In order to process the generated video, MATLAB 2016 has been utilized which uses audio and video reader libraries. The first step is to extract the frames out of the video. To be on the approximated

ISSN: XXXX-XXXX IJXX

size, from a sample of length 12secs, a total count of 388 video frames has been generated. Each extracted frame is passed to Euler magnification. Each extracted frame is passed to forehead selection. The process can be illustrated using Figure 4 which explains the work architecture and used utilities from MATLAB. The main purpose is to enhance the forehead segment in order to train the system for its most efficient possible classification accuracy. To identify the forehead, face has to be detected from each frame that is going to be processed. The face detection process is not a complicated process in the modern world of technology. There are several ways through which a face detection framework can be designed and defined.

Figure 3. Magnified video through Euler magnification

There are several types of detection that are popular in algorithmic implementation designs and are generally termed as Upper Body Classification Model as described in Figure 4. The extracted forehead is further passed to the optimization process in order to select the forehead more precisely.

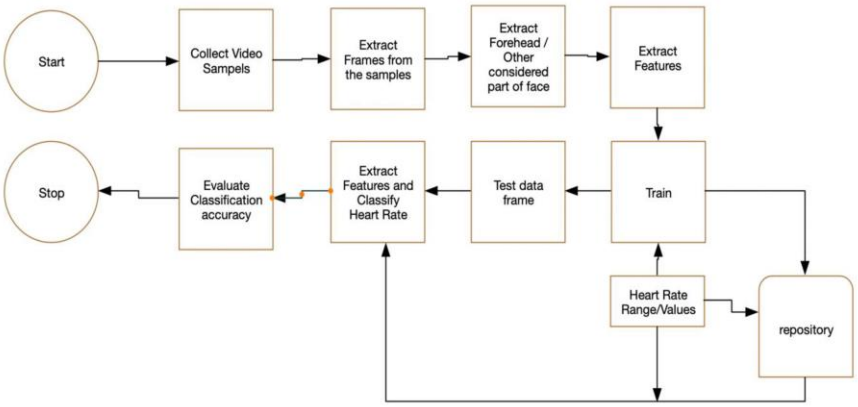


Figure 4: Forehead selection process along with training and classification

The proposed algorithm uses Cascaded Object Detection (COD) due to its accurate positioning of the face pixels. The COD method crops the outer segments and selects the inner segments for the processing. The implementation of the work has been done on MATLAB simulation tool which contains the Neural Network toolbox for the processing of Neural oriented engines whether it is Convolution Neural Network (CNN) or FFBPNN is integrated in the toolbox. The process of detection is suppressed in four subsequent steps as illustrated in Figure 5. For the face magnification, the proposed algorithm is utilizing the histogram equalization method for the selection of the forehead

ISSN: XXXX-XXXX IJXX

once it is magnified by Euler magnification. In order to perform the histogram equalization, following set of implementation steps have been designed and implemented. To perform histogram equalization, the red pixel components have been extracted.

Figure 5: Forehead detection

1.1.1 Histogram Equalization

Histogram equalisation is a common method for improving picture contrast (HE). Because of its simplicity and comparably greater performance on practically all sorts of photos, it is the most often used approach. HE operates by remapping the image's grey levels according to the probability distribution of the input grey levels. Many studies have been conducted on histogram equalisation, and several approaches have been presented. The authors used the Global Histogram Equalization (GHE) technique as shown in Figure 6. GHE improves picture contrast significantly, but it comes with various artefacts and unpleasant side effects, such as a washed-out look in the flower's grey levels. This is the primary cause of such adverse effects and visual detail reduction.

Assume that the input picture f(x, y) is made up of discrete grey levels in the range [0, L-1]. C(rk) is a transformation function defined as

$$s_k = C(r_k) = \sum_{i=0}^{k} P(r_i) = \sum_{i=0}^{k} \frac{n_i}{n} \quad (1)$$

Where, $0 \le s_k \le 1 and\ k = 0,1,2,.....,L = 1.$

ISSN: XXXX-XXXX IJXX

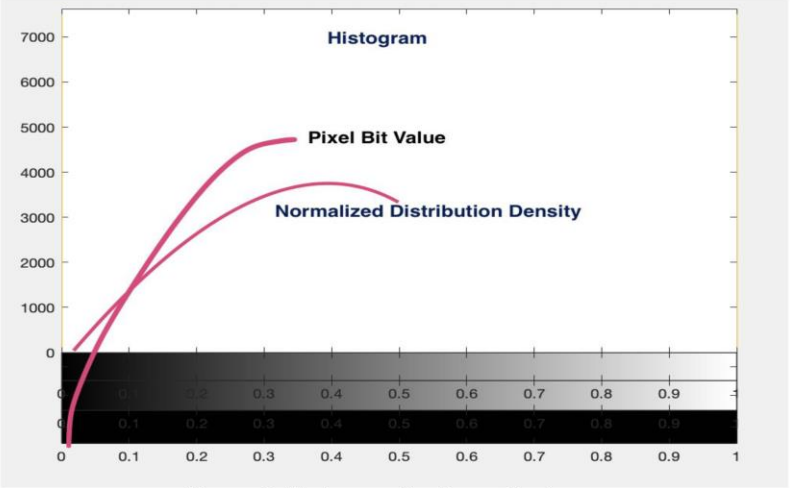


Figure 6: Histogram for the entire frame

Swarm Intelligence Optimization algorithm

There are three stages of the proposed ABC algorithm. Each stage has its set of significance in order to provide reward to the employed bee. The proposed algorithm generates a reward mechanism in the following order. For each flight one bee will be awarded with a maximum score of .10 and hence the total reward of 1 can be attained by one employed bee. If the employed bee gains more than .60 values as reward out of 10 Levy flights, the pixel is considered to be a good pixel for forehead. In order to compose the employed bee, pixel value of each distribution is considered. Each employed bee is formed by applying the arithmetic means of collected other bee sample. In order to create sub hives in bee list, k-means is applied to a population of 70% of total distribution. 65% of the population from the same hive and 35% distribution is taken from other hive group. The distribution is made in such a way that each employed bee gets a chance to co-ordinate with other bees working on different pixel distribution as well.

The ordinal architecture of ABC algorithm is illustrated using Figure 7(a) and the work process is illustrated in Figure 7(b).

ISSN: XXXX-XXXX IJXX

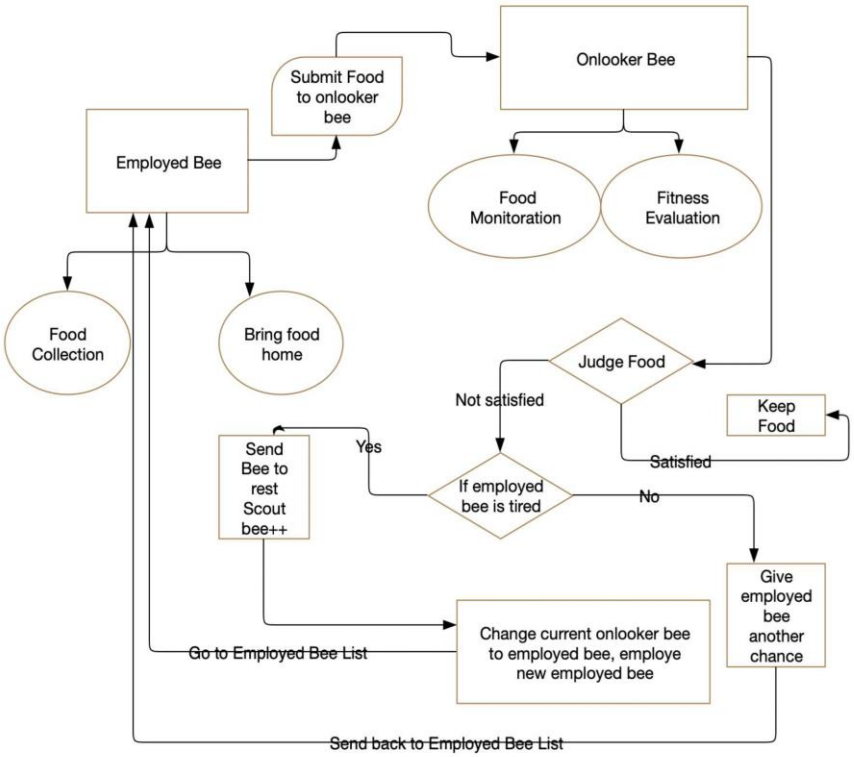


Figure 7(a) General Flow architecture of ABC algorithm

As illustrated in 7(a), there are three bees in the architecture of ABC algorithm. The first bee is the employed bee that has two responsibilities. The responsibilities are marked with circles as that in an E-R diagram. The employed bee goes for the search of food and submits the food to the onlooker bee. The onlooker bee holds two responsibilities namely monitoring of the gathered or aggregated food and judgement of the food that is brought by the employed bee. In order to judge the food of the employed bee, the onlooker bee applies a fitness function. If the food brought by the employed bee satisfies the food threshold of the onlooker bee, it keeps the food in the repository. If the fitness function is not satisfied, the onlooker bee again evaluates whether the employed is too tired to bring the quality food or the employed bee should be provided with another chance. If the onlooker bee finds the current employed bee to be tired enough, it sends the employed bee to another bee structure that is termed as scout bee. Scout bee is the bee group that takes rest if they are tired. The onlooker bee, as per the rulebook, employs herself as the current employed bee as the previous bee has gone to take rest. The motive behind the self-employment is the sense to get better food. The current employed bee, that was the onlooker bee a moment before, has the experience of previous food and hence rather than sending some other bee in search of food, is employs herself and brings a new bee on the position of the empty onlooker bee. The general Pseudo code of the ABC algorithm is mentioned in the algorithm.

ISSN: XXXX-XXXX IJXX

Algorithm: ABC

Input: N_D ← Normalization of the extracted features from the video frames
Output: S_D ← Consider the fitness value for the selection of features

```
Compute the size of the pixel in a face, [Row, Col] = Size (N_D)
FR=[] // Final record
Count = 1
For I in range ( (N_D, Col)
   Current Feature Col = N_D (All Row in a frame, I)
   All Bee Record = []
   For J in range (5)
      Ebee = [Current extracted Feature Col (1), Other five Current Feature Col (Randomly)]
      Obee = (Σ_1^J Ebee(J)) / (Number of Ebee)
      Define fitness function of ABC for the extracted frame (feature)
            All Fit Record = []
            Fit Status = 0
            For M in range(Ebee)
               If Ebee (M) > Obee
                  Fit Status = 1
               Else
                  Fit Status = 0
               End – If
            All Fit Record (M) = Fit Status
            End – For
      End – For
      All Fit = fitness function(Ebee, Obee)
      If count of non-zeros in All Fit > 1
         Bee Status = 1
      Else
         Bee Status = 0
      End – If
      All Bee Record (J) = Bee Status
      End – For
      If count of non-zeros in All Bee Record > Average (All Bee Record)
         Final Record (count) = I
         Count = Count + 1
      End – If
End – For
```

The illustration of proposed ABC algorithm is based on pixel distribution and hence, the proposed algorithm is named as pixel distributed ABC.
The employed bee and the onlooker bee can be defined using equation (2) and (3).

$$Employed_{bee} = \int_{i=1}^{n} P_{vlue_i}.BL.ohb \tag{2}$$

Where BL is the Bee List which is ultimately the pixel distribution of the extracted image. Pvalue is the pixel value of the image, ohb is the other hive group which is formed by 65-35 % distribution of the pixels and n is total number of pixels in the extracted image. The onlooker bee is the arithmetic mean of the collected employed bees.

$$Onlooker_{bee} = \int_{i=1}^{n}\int_{j=1}^{10} Ar(Employed_{bee_i})dn \tag{3}$$

The onlooker bee forms the bee food for 10 levy flights and hence the iterative value of j starts from 1 and ends at 10.
The fitness function of the proposed PD-ABC is expressed by equation (4)

ISSN: XXXX-XXXX IJXX

$$Fitness\ ABC \quad f_x = 1 \ if\ Eb.cr.cg > Eb.cr.mg \quad 0\ otherwise \tag{4}$$

Where Eb is the employed bee, cr is the co-relation, mg is the major group, cg is the current group. The ABC algorithm decides the keeping threshold based on the evaluated co-relation among the group members and the co-relation of the bees with their centroid of the major group that belongs to the employed bee and the steps can be expressed as given in algorithm.

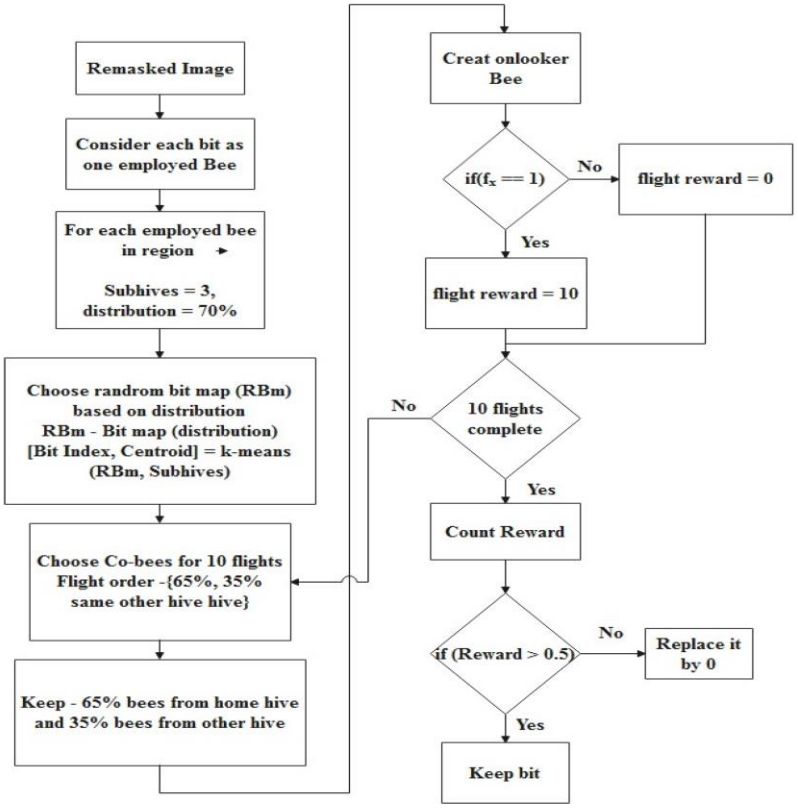


Figure 7 (b): Workflow of proposed ABC

Algorithm Apply ABC

Inputs :PD Output:SPD **where PD is the pixel distribution and SPD is the selected pixel distribution.**

1. for i in PD
   a. Eb = i;
   b. Sp = R.sample(PD) // Select a random sample space by 70% distribution law
   c. hv = kmean(Sp,3) // create hv as three hives from the sample space
   d. Ei = Find(hv.Eb.index) // Employed bee index, extract from hive indexes

ISSN: XXXX-XXXX IJXX

e. Initiate Reward = 0 // Initiate reward for Levy flights

for j = 1:10 // create 10 Levy flights

1. Create Ob_j // Create Onlooker bee for jth flight
2. Create Ei_j // Place employed bee into random groups and create employed bee group for jth flight
3. fx = $bee_{fitness}$(Obj,Eij) // Pass to bee fitness
   a. If fx == 1
      i. Reward = Reward + .10
   b. Else
      i. Reward = Reward + 0
   c. End If

f. End for_j

2. If Reward ≥ .60
   a. Accept bit value as forehead
   b. Append to SPD
3. **Else**
   a. Do nothing
4. **End If**
5. **End for_i**

To perform training and classification, Feed-Forward Back Propagation Network (FFBPNN) has been applied. The FFBPNN is a multi-layered architecture in which data passes through at least one hidden/middle layer on its way from the input to the output. Each layer has neurons that are linked to all of the neurons in the layers around it. The connections are paired with numerical values (weights) that will be modified throughout the training phase. The basic purpose of a (FFBPN) is to learn and map input-output correlations. The FFBPN learning rule is also used to alter a system's weight values and threshold values in order to obtain the lowest possible error. It's also known as a complicated connection between a network set's input and output values. The input received from other network system units determines the value of each node or neuron. Each input signal is multiplied by the weight value of the associated input line.

For training a model that performs two-way iterations, the FFBPN approach was utilised. The first method involves computing input weights in a forward step, whereas the second method involves updating weights and calculating mistakes in a backward step. The training data was standardised to a value between 0 and 1. Seventy percent of the data was utilised to train the model, while the remaining 30 percent was shared evenly between testing and validation. The model was then trained until the conditions has been fulfilled.

$$x_k = \sum_i^n W_{ki} x_i \tag{5}$$

ISSN: XXXX-XXXX IJXX

where xk is the new value of the variable, xi is the starting value of the variable, and wki is the weight link value of the neuron/variable. Equation shows that the activation function between the input and the buried layer was 'logsig' (6).

$$f(x) = 1/(1+e^{-x}) \quad (6)$$

As illustrated earlier, Neural Network is a three layer architecture that contains the input layer, the hidden layer and the propagation output layer. The hidden layer is optimized using the learning functions the rotates the data to generate weights. For the proposed work case, the input layer contains as many values that has been extracted as HOG features from each frame. If there are 30 frames and if each frame contains 12000 HOG values, there would be 12000 features at the input end as shown in Figure 8.

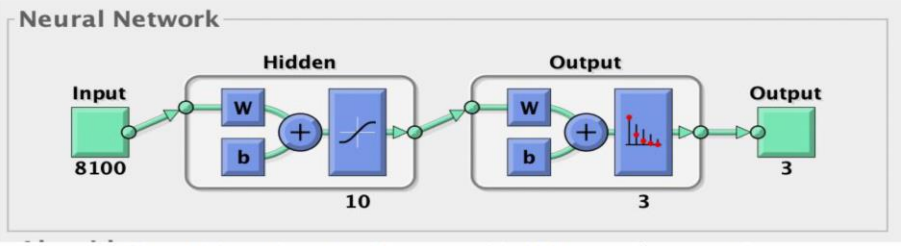


Figure 8 Input layer architecture with 10 layers of propagation

The training pattern propagates through provided layers with weight generation through the supplied function of propagation. In case of proposed work, the training stops the validation when the entropy of the training set is minimum. The entropy of the propagation is bundled in different bins of propagation as shown in Figure 9.

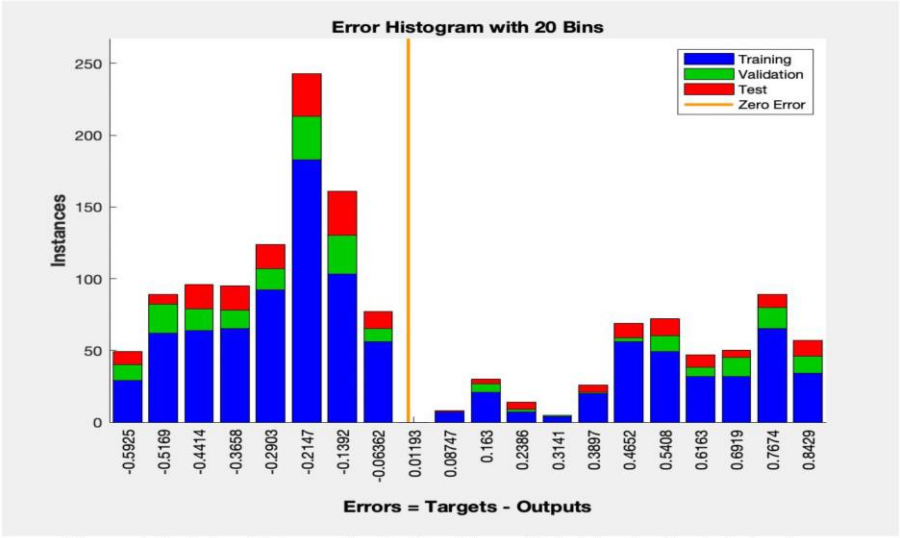


Figure 9 Training Entropy Analysis with multiple bins in the training layer

ISSN: XXXX-XXXX IJXX

The training through the propagation algorithm is attained in three steps of outlining namely the training section, the validation section and the output section. The total data, that is supplied for the training is separated into three subsequent sections that contains 70% as training data and rest 15-15% for validation and testing. The learning algorithm shuffles the input set in the propagation layer. For each layer of propagation, the training algorithm generates MSE for all the validations that are applied.

The FFBPN-based model is trained utilising historical data during the first training step. If the model fails to satisfy expectations, the Levenberg-Marquardt (LM) backpropagation method will enable the process to be re-propagated until it meets the optimal requirement. At the validation step, 15% of the data collected after the model training was utilised to verify the trained model. Once the model's correctness was determined, it was put through its paces and tested with the remaining 15% of the data sets.

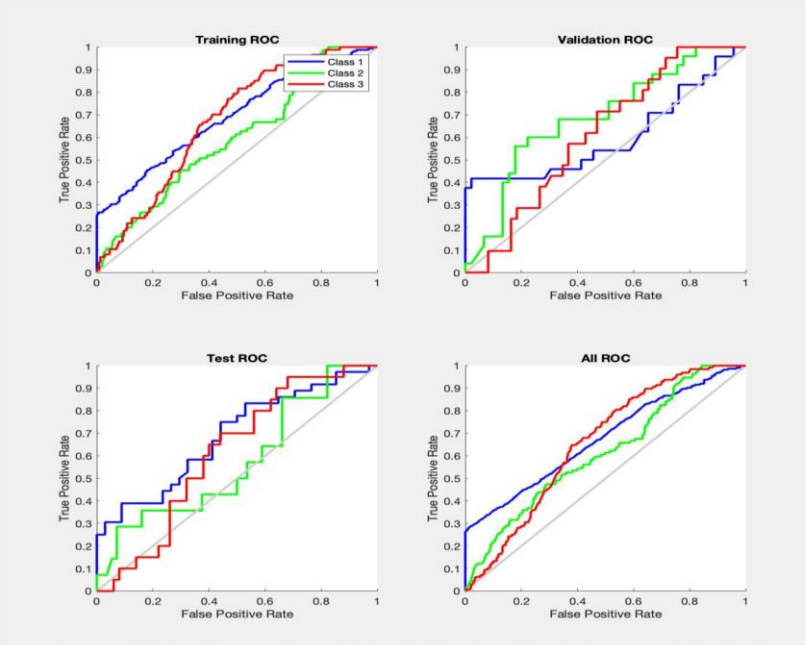


Figure 10 Validation using ROC architecture for all the three classes

The findings were found to be accurate, with an R2 value near to 1.0. As illustrated in Fig. 8, the network is made up of three layers: input, hidden, and output. The suggested network was employed in the hidden layer with a sigmoid function (tansig) and the output layer with a linear function (Purelin).Fig. 10 is used to display the TPR and FPR in which validation and training using ROC.

4. Results and Discussion

The performance metrics has been computed to test the robustness of the proposed technique. The simulation outcomes in terms of True Positive Rate (TPR), False Positive Rate (FPR), and Accuracy have been computed.

- TPR - It is also called as Recall. It is the ratio of TPR to sum of TPR and False Negative Ratio (FNR).

$$TPR = \frac{TPR}{TPR + FNR} \quad (1)$$

ISSN: XXXX-XXXX IJXX

- FPR – It is the analysis of false negatives and false positives.

$$FPR = \frac{FPR}{FPR + TNR} \quad (2)$$

- Accuracy- It is defined as accurate measurement of the original signal during the HRV measurement system.

$$Acc = (100 - 100\frac{|Monitor_{HR} - Video_{HR}|}{Monitor_{HR}}$$

$Monitor_{HR}$ is the monitoring of the HR and $Video_{HR}$ is the magnification from the RGB video.

Table 1. Ordinals of Simulation Analysis

Parameter	Description
Number of Frames	500 to 1000
Classifiers	CNN, FFBPNN
Number of Layers in each neural architecture	5 to 20
Evaluation Parameters	TRP, FPR, Accuracy

Table 2 displays the TPR analysis of CNN, ABC with CNN, FFBPNN, and ABC with FFBPNN using 90:10, 80:20, and 70:30 ratio analysis. The frames are separated into ten categories: 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, and 1000.The TPR have gradually increased with the number of frames. For 90:10 ratio, considering the 500 frames, the TPR of CNN is at 0.79 and ABC with CNN is 0.81 while TPR using the proposed technique is 0.91 and without ABC, with FFBPNN only is 0.90. The average TPR of CNN is 0.80 and ABC with CNN is 0.838. The frames are separated, and the TPR of the method is calculated. For 80:20 ratio, CNN has a TPR of 0.78 for 600 frames, whereas ABC has a true positive rate of 0.79, while TPR using the proposed technique is 0.91 and without ABC, with FFBPNN only is 0.90.

Table 2. Comparative Analysis to compute the TPR analysis

Number of Frames	using 90:10 distribution				using 80:20 distribution				using 70:30 distribution			
	TPR using CNN	TPR using ABC with CNN	TPR using FFBPNN	TPR using ABC with FFBPNN	TPR using CNN	TPR using ABC with CNN	TPR using FFBPNN	TPR using ABC with FFBPNN	TPR using CNN	TPR using ABC with CNN	TPR using FFBPNN	TPR using ABC with FFBPNN
500	0.798	0.812	0.907	0.920	0.782	0.795	0.888	0.901	0.767	0.779	0.871	0.884
550	0.799	0.812	0.914	0.926	0.783	0.796	0.896	0.903	0.767	0.780	0.878	0.889
600	0.799	0.815	0.921	0.933	0.783	0.799	0.903	0.914	0.767	0.783	0.885	0.900
650	0.799	0.819	0.928	0.939	0.784	0.803	0.911	0.922	0.768	0.787	0.892	0.903

ISSN: XXXX-XXXX IJXX

700	0.803	0.832	0.936	0.946	0.788	0.817	0.919	0.929	0.772	0.801	0.900	0.910
750	0.806	0.836	0.943	0.952	0.792	0.821	0.926	0.937	0.776	0.805	0.908	0.917
800	0.809	0.839	0.950	0.959	0.795	0.825	0.934	0.941	0.779	0.808	0.915	0.925
850	0.812	0.842	0.957	0.969	0.798	0.828	0.941	0.952	0.782	0.812	0.922	0.933
900	0.819	0.846	0.965	0.975	0.806	0.832	0.948	0.959	0.790	0.815	0.929	0.940
950	0.827	0.855	0.967	0.982	0.813	0.841	0.951	0.967	0.797	0.825	0.932	0.946
1000	0.834	0.875	0.968	0.986	0.820	0.861	0.952	0.970	0.804	0.848	0.933	0.952

Similarly, for 850 frames using 90:10 ratio, the CNN TPR algorithm is 0.81, whereas ABC with CNN is 0.845, while TPR using the proposed technique is 0.96 and without ABC, with FFBPNN only is 0.95. The algorithm CNN has a TPR of 0.83 for 1000 frames, whereas the algorithm ABC with CNN has a TPR of 0.8747.

The proposed technique has a TPR of about 0.98 while with FFBPNN only is 0.96. As a result, as compared to the CNN algorithm, ABC with CNN, and FFBPNN only, TPR has been improved by 0.17%, 0.14%, and 0.013%.

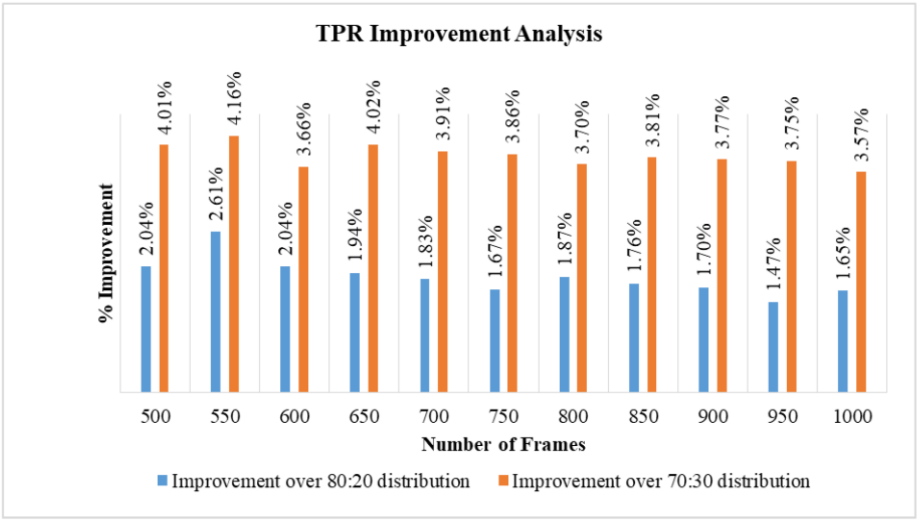


Figure 11 TPR improvement analyses for different ratios

Fig. 11 shows that TPR has been improved for 80:20 ratio in comparison to 70:30 ratio. The simulation results show that average TPR using 80:20 ratio is 93.6% while using 70:30 ratio, it is 91%. Thus, proposed technique shows better results with 80:20 ratio analyses.

ISSN: XXXX-XXXX IJXX

Table 3. Comparative Analysis of the proposed technique for FPR

Number of Frames	using 90:10 distribution				using 80:20 distribution				using 70:30 distribution			
	FPR using CNN	FPR using ABC with CNN	FPR using FFBPNN	FPR using ABC with FFBPNN	FPR using CNN	FPR using ABC with CNN	FPR using FFBPNN	FPR using ABC with FFBPNN	FPR using CNN	FPR using ABC with CNN	FPR using FFBPNN	FPR using ABC with FFBPNN
500	0.117	0.106	0.114	0.100	0.1188	0.1082	0.1159	0.1024	0.1210	0.1101	0.1180	0.1032
550	0.120	0.112	0.119	0.106	0.1223	0.1139	0.1216	0.1081	0.1244	0.1158	0.1237	0.1090
600	0.129	0.113	0.121	0.110	0.1314	0.1150	0.1229	0.1129	0.1336	0.1170	0.1250	0.1129
650	0.134	0.113	0.121	0.111	0.1370	0.1156	0.1235	0.1126	0.1392	0.1175	0.1255	0.1138
700	0.140	0.118	0.127	0.118	0.1429	0.1202	0.1290	0.1197	0.1452	0.1221	0.1311	0.1217
750	0.149	0.125	0.130	0.121	0.1517	0.1268	0.1325	0.1241	0.1542	0.1288	0.1347	0.1243
800	0.151	0.126	0.132	0.122	0.1536	0.1284	0.1344	0.1245	0.1562	0.1305	0.1367	0.1264
850	0.156	0.134	0.140	0.125	0.1591	0.1362	0.1422	0.1269	0.1618	0.1384	0.1445	0.1278
900	0.158	0.144	0.147	0.129	0.1611	0.1463	0.1498	0.1307	0.1637	0.1487	0.1523	0.1325
950	0.161	0.146	0.150	0.131	0.1637	0.1488	0.1524	0.1343	0.1664	0.1512	0.1549	0.1354
1000	0.163	0.149	0.157	0.134	0.1659	0.1516	0.1596	0.1361	0.1685	0.1540	0.1622	0.1375

Table 3 shows the FPR analysis of CNN, ABC with CNN, FFBPNN, and ABC with FFBPNN using 90:10, 80:20, and 70:30 ratio analysis. The FPR have gradually reduced with the number of frames. For 90:10 ratio using 500 frames, the FPR of CNN is at 0.11 and ABC with CNN is 0.10 while FPR using the proposed technique is 0.10, and with FFBPNN only is 0.11. The frames are separated, and the FPR of the proposed method is calculated. Further, CNN has a FPR of 0.14 for 700 frames, whereas ABC has a false positive rate of 0.1180, while FPR using the proposed technique is 0.11 and, with FFBPNN only is 0.117. Similarly, for 850 frames, the CNN FPR algorithm is 0.15, whereas ABC with CNN is 0.133, while FPR using the proposed technique is 0.125 and, with FFBPNN only is 0.139. The algorithm CNN has a FPR of 0.163 for 1000 frames, whereas the algorithm ABC with CNN has a 0.149. The proposed technique has a FPR of about 0.133 while with FFBPNN only is 0.1570. The average FPR of CNN and ABC with CNN is 0.1435 and 0.1260 respectively. The average FPR of the proposed technique and with FFBPNN only is 0.13 and 0.15 respectively. However, the average FPR using the 80:20 ratio and 70:30 is 0.12 and 0.122 respectively. As a result, as compared to the CNN algorithm, ABC with CNN, and FFBPNN only, FPR of the proposed technique has been improved by 0.17%, 0.05%, and 0.10% for 90:10 ratio analysis.

ISSN: XXXX-XXXX IJXX

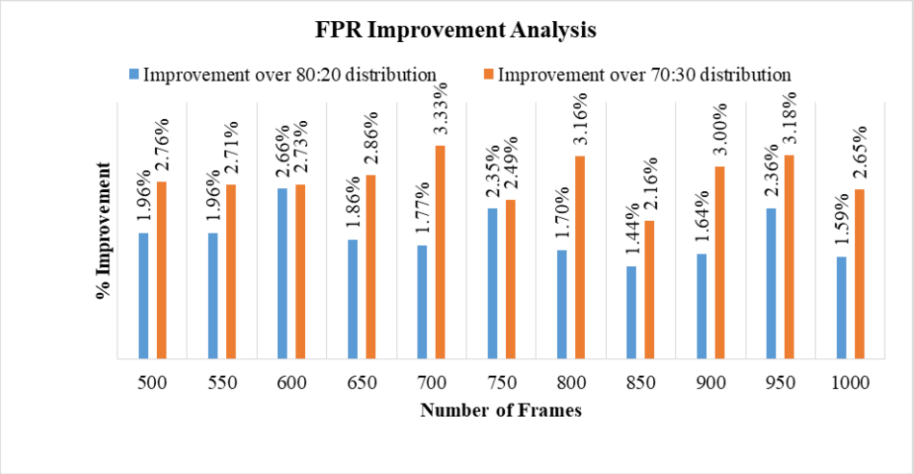


Figure 12 FPR Improvement Analysis

Fig. 12 shows that FPR has been improved for 80:20 ratios in comparison to 70:30 ratio. The simulation results show that average FPR using 80:20 ratios is 0.121 while using 70:30 ratio, it is 0.122. The average value using 70:30 and 80:20 is 0.146 and 0.149 for CNN respectively and 0.135 and 0.137 for FFBPNN only. Thus, proposed technique shows better results with 80:20 ratios analyses. Table 4 depicts the accuracy analysis of the proposed technique (ABC with FFBPNN), CNN, ABC with CNN, and FFBPNN using 90:10, 80:20, and 70:30 ratio analyses. There is a gradual increase in the accuracy with the number of frames. For 90:10 ratio using 550 frames, the accuracy of CNN is at 85 and ABC with CNN is 0.86 while accuracy using the proposed technique is 0.93 and without ABC, with FFBPNN only is 0.92. The frames are separated, and the accuracy of the method increases with increase in frames.

Table 4. Comparative Analysis of the proposed technique for Accuracy

Number of Frames	using 90:10 distribution				using 80:20 distribution				using 70:30 distribution			
	Accuracy using CNN	Accuracy using ABC with CNN	Accuracy using FFBPNN	Accuracy using ABC with FFBPNN	Accuracy using CNN	Accuracy using ABC with CNN	Accuracy using FFBPNN	Accuracy using ABC with FFBPNN	Accuracy using CNN	Accuracy using ABC with CNN	Accuracy using FFBPNN	Accuracy using ABC with FFBPNN
500	0.853	0.869	0.922	0.936	0.836	0.851	0.904	0.917	0.820	0.836	0.888	0.901
550	0.853	0.869	0.922	0.936	0.836	0.852	0.904	0.920	0.822	0.838	0.888	0.902
600	0.857	0.874	0.930	0.937	0.840	0.857	0.911	0.918	0.826	0.842	0.896	0.902
650	0.863	0.881	0.936	0.943	0.847	0.864	0.918	0.925	0.833	0.850	0.903	0.910

ISSN: XXXX-XXXX IJXX

700	0.874	0.899	0.940	0.958	0.858	0.883	0.923	0.946	0.844	0.868	0.908	0.925
750	0.882	0.906	0.941	0.966	0.866	0.891	0.924	0.948	0.852	0.876	0.909	0.933
800	0.889	0.914	0.945	0.971	0.873	0.898	0.929	0.954	0.859	0.883	0.913	0.938
850	0.895	0.921	0.952	0.977	0.880	0.905	0.936	0.961	0.866	0.890	0.920	0.945
900	0.907	0.930	0.958	0.989	0.891	0.914	0.942	0.972	0.877	0.899	0.927	0.953
950	0.908	0.931	0.959	0.984	0.893	0.915	0.943	0.962	0.878	0.901	0.928	0.951
1000	0.908	0.939	0.960	0.987	0.893	0.923	0.945	0.971	0.879	0.909	0.930	0.959

Further, CNN has an accuracy of 0.85 for 600 frames, whereas ABC has an accuracy of 0.87, while accuracy using the proposed technique is 0.93 and without ABC, with FFBPNN only is 0.92. Similarly, for 850 frames, the CNN accuracy is 0.89 for 90:10 and 0.88 for 80:20, whereas ABC with CNN is 0.92 for 90:10, while accuracy using the proposed technique is 0.97 for 90:10 and 0.945 for 70:30 and with FFBPNN only is 0.95 for 90:10. The algorithm CNN has average accuracy of 0.8807, whereas the algorithm ABC with CNN is 0.90. The proposed technique has average accuracy of about 0.96 for 90: 10 and 0.94 for 80:20 while with FFBPNN only is 0.94 for 90:10 and 0.92 for 80:20. As a result, as compared to the CNN algorithm, ABC with CNN, and FFBPNN only, accuracy of the proposed technique has been improved by 0.093%, 0.066%, and 0.022%.

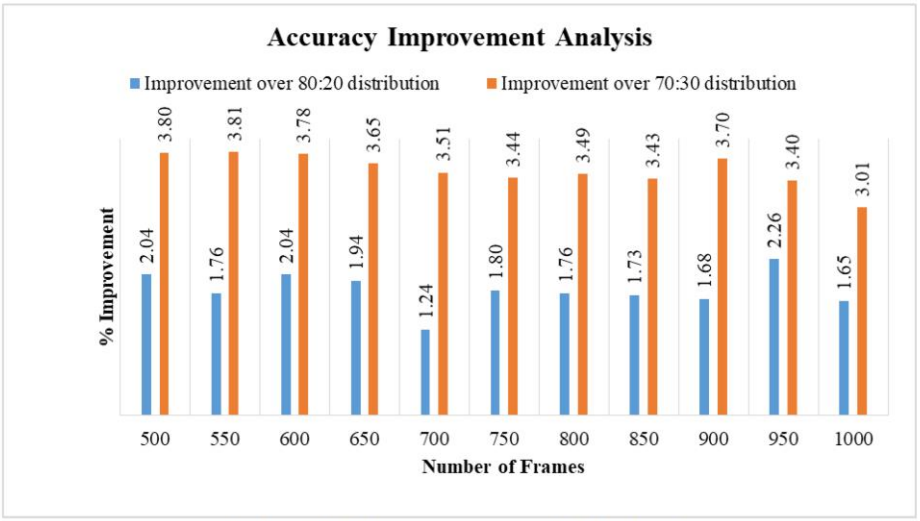


Figure 13 Accuracy Improvement Analysis

ISSN: XXXX-XXXX IJXX

Fig. 13 depict the improvement analysis that accuracy of the proposed technique using 80:20 shows better results in comparison to 70:30 ratio. The average accuracy using the proposed technique for 70:30 ratio is 0.92 while for 80:20, it is 0.94. Thus, proposed technique using 80:20 ratio shows better results in comparison to 70:30 ratio.

5. Conclusion

In this paper, we presented the non-contact heart rate monitoring system. The video frames have been processed to extract the features such as face and selection of forehead. The optimization of the video frames has been done using the Swarm Intelligence (SI) technique i.e. PD-ABC and then classification and training has been done using the Feed forward Back propagation Neural Network (FFBPNN). The proposed algorithm further compared with the existing techniques considering the ration analysis 70:30, 80:20, and 90:10. The simulation results for 90:10 shows better results as FPR of the proposed technique has been improved by 0.17% in comparison to the CNN algorithm, 0.05% from ABC with CNN, and 0.10% from FFBPNN only. The proposed algorithm has a TPR of about 0.98 while with FFBPNN only; it is 0.96 for 90:10. Thus, there is 0.02% improvement in the measurement system for heart rate.

References

[1] J. Kranjec, S. Beguš, G. Geršak, and J. Drnovšek. "Non-contact heart rate and heart rate variability measurements: A review." *Biomedical signal processing and control* 13 (2014): 102-112..

[2] H. Nisar, M. B. Khan, W. T. Yi, Y. K. Ho, and L. K. Chun. "A Non Invasive Heart Rate Measurement System for Multiple People in the Presence of Motion and Varying Illumination." *International Journal of Disease Control and Containment for Sustainability (IJDCCS)* 1, no. 1 (2016): 1-11.

[3] H. Monkaresi, N. Bosch, R. A. Calvo, and S. K. D'Mello. "Automated detection of engagement using video-based estimation of facial expressions and heart rate." *IEEE Transactions on Affective Computing* 8, no. 1 (2016): 15-28.

[4] S. Bobbia, Y. Benezeth, and J. Dubois. "Remote photoplethysmography based on implicit living skin tissue segmentation." In *2016 23rd International Conference on Pattern Recognition (ICPR)*, pp. 361-365. IEEE, 2016..

[5] N. Wadhwa, H. Y. Wu, A. Davis, M. Rubinstein, E. Shih, G. J. Mysore, J. G. Chen, O. Buyukozturk, J. V. Guttag, W. T. Freeman, F. Durand. "Eulerian video magnification and analysis." *Communications of the ACM* 60, no. 1 (2016): 87-95.

[6] S. Bennett, T. N. El Harake, R. Goubran, and F. Knoefel. "Adaptive eulerian video processing of thermal video: An experimental analysis." *IEEE Transactions on Instrumentation and Measurement* 66, no. 10 (2017): 2516-2524.

[7] P. V. Rouast, M. T. P. Adam, D. J. Cornforth, E. Lux, and C. Weinhardt. "Using contactless heart rate measurements for real-time assessment of affective states." In *Information Systems and Neuroscience*, pp. 157-163. Springer, Cham, 2017.

[8] S. C. Matta, Z. Sankari, and S. Rihana. "Heart rate variability analysis using neural network models for automatic detection of lifestyle activities." *Biomedical Signal Processing and Control* 42 (2018): 145-157.

[9] N. Patel, P. Patel, and N. Patel. "Heart attack detection and heart rate monitoring using IoT." *International Journal of Innovations and Advancements in Computer Science, IJIACS* 7, no. 4 (2018): 612-615.

[10] X. Chen, J. Cheng, R. Song, Y. Liu, R. Ward, and Z. J. Wang. "Video-based heart rate measurement: Recent advances and future prospects." *IEEE Transactions on Instrumentation and Measurement* 68, no. 10 (2018): 3600-3615.

[11] A. F. Hussein, N. A. Kumar, M. Burbano-Fernandez, G. Ramirez-Gonzalez, E. Abdulhay, and V. H. C. De Albuquerque. "An automated remote cloud-based heart rate variability monitoring system." *IEEE access* 6 (2018): 77055-77064.

ISSN: XXXX-XXXX IJXX

[12] J. Malik, Y.-L. Lo, and H.-T. Wu. "Sleep-wake classification via quantifying heart rate variability by convolutional neural network." *Physiological measurement* 39, no. 8 (2018): 085004.
[13] V. L. Petrovic, M. M. Jankovic, A. V. Lupsic, V. R. Mihajlovic, and J. S. Popovic-Bozovic. "High-accuracy real-time monitoring of heart rate variability using 24 GHz continuous-wave Doppler radar." *IEEE Access* 7 (2019): 74721-74733.
[14] R. L. Charles and J. Nixon. "Measuring mental workload using physiological measures: A systematic review." *Applied ergonomics* 74 (2019): 221-232.
[15] F. Abnousi, G. Kang, J. Giacomini, A. Yeung, S. Zarafshar, N. Vesom, E. Ashley, R. Harrington, and C. Yong. "A novel noninvasive method for remote heart failure monitoring: the EuleriAn video Magnification apPLications In heart Failure studY (AMPLIFY)." *NPJ digital medicine* 2, no. 1 (2019): 1-6.
[16] B. Bent, B. A. Goldstein, W. A. Kibbe, and J. P. Dunn. "Investigating sources of inaccuracy in wearable optical heart rate sensors." *NPJ digital medicine* 3, no. 1 (2020): 1-9.
[17] PURE dataset (2020). https://www.tu-ilmenau.de/en/neurob/datasets-code/pulse/
[18] A. El Attaoui, M. Hazmi, A. Jilbab, and A. Bourouhou. "Wearable wireless sensors network for ECG telemonitoring using neural network for features extraction." *Wireless Personal Communications* 111, no. 3 (2020): 1955-1976.
[19] H. Nogami, S. Ohgata, A. Saito, K. Ban, T. Akiyama, T. Hiejima, R. Takigawa, and T. Hosoya. "A non-invasive heart rate measurement system using laser Doppler blood flowmetry with husbandry training of the masked palm civet (Parguma larvata)." *Japanese Journal of Applied Physics* 60, no. SC (2021): SCCL13.
[20] A. Ni, A. Azarang, and N. Kehtarnavaz. "A review of deep learning-based contactless heart rate measurement methods." *Sensors* 21, no. 11 (2021): 3719.
[21] F. Wang, X. Zeng, C. Wu, B. Wang, and K. J. R. Liu. "mmHRV: Contactless heart rate variability monitoring using millimeter-wave radio." *IEEE Internet of Things Journal* 8, no. 22 (2021): 16623-16636.
[22] K. Hinde, G. White, and N. Armstrong. "Wearable devices suitable for monitoring twenty four hour heart rate variability in military populations." *Sensors* 21, no. 4 (2021): 1061.
[23] H. Demirezen and C. Eroglu Erdem. "Heart rate estimation from facial videos using nonlinear mode decomposition and improved consistency check." *Signal, Image and Video Processing* 15, no. 7 (2021): 1415-1423.
[24] C. Egger. "The autonomic nervous system." *Manual of Equine Anesthesia and Analgesia* (2022): 110-118.

Gaganjot Kaur received the M.Tech degree in Electronics and communication engineering from the Punjabi University Patiala, India. She is currently doing PhD from Auckland University of Technology, New Zealand. Main research topic is in the field of image processing.

Jeff Kilby is currently senior lecturer with the Department of Electronics and Computing, School of Engineering Auckland University of Technology, New Zealand. His research interests are biomedical signal processing, LABVIEW and wireless sensor network applications.

ISSN: XXXX-XXXX IJXX

References

[1] H. F. Jelinek, D. J. Cornforth, and A. H. Khandoker, *ECG time Series Variability Analysis: Engineering and Medicine*. CRC Press 2017, pp. 1-480.

[2] U. Ravens, M. C. Sanguinetti, and O. N. Tripathi, *Heart Rate and Rhythm : Molecular Basis Pharmacological Modulation and Clinical Implications*, Heidelberg: Springer, 2011.

[3] G. Ernst, *Heart Rate Variability*, London: Springer, 2014.

[4] M. AlGhatrif and J. Lindsay, "A brief review: history to understand fundamentals of electrocardiography," *Journal of community Hospital Internal Medicine Perspectives,* vol. 2, no. 1, 2012.

[5] J.-J. Goy, *Electrocardiography (ECG),* Sharjah: Bentham Science Publishers, 2013.

[6] J. Kranjec, S. Beguš, G. Geršak, and J. Drnovšek, "Non-contact heart rate and heart rate variability measurements: A review," *Biomedical Signal Processing and Control,* vol. 13, pp. 102-112, 2014.

[7] S. C. Mukhopadhyay, *Wearable Electronics Sensors: For Safe and Healthy Living*. Springer 2015.

[8] P. Shi, S. Hu, and Y. Zhu, "A Preliminary Attempt to Understand Compatibility of Photoplethysmographic Pulse Rate Variability with Electrocardiogramic Heart Rate Variability," *Journal of Medical and Biological Engineering,* vol. 28, pp. 173-180, 2008.

[9] C. A. G. Martinez, *Heart Rate Variability Analysis with the R package RHRV*, Cham, Switzerland: Springer, 2017.

[10] J. Burnett, "The Origins of the Electrocardiograph as a Clinical Instrument," *Medical History,* Article vol. 29, no. S5, pp. 53-76, 1985.

[11] J. R. Henson, "Descartes and the ECG Lettering Series," *Journal of the History of Medicine and Allied Sciences* vol. 26, no. 2, pp. 181-6, Apr 1971.

[12] N. J. Holter, "New Method for Heart Studies, Continuous electrocardiography of active subjects over long periods is now practical," *Science,* vol. 134, no. 3486, pp. 1214-1220, 1961.

[13] A. A. Alian and K. H. Shelley, "Photoplethysmography," *Best Practice & Research Clinical Anaesthesiology,* vol. 28, no. 4, pp. 395-406, 2014.

[14] A. B. Hertzman and C. R. Spealman, "Observations on the Finger Volume Pulse Recorded Photoelectrically," *American Journal of Physiology,* vol. 119, pp. 334-335, 1937.

[15] A. B. Hertzman, "The Blood Supply of Various Skin Areas as Estimated by the Photoelectric Plethysmograph," *American Journal of Physiology,* vol. 124, no. 2, pp. 328-340, 1938.

[16] A. B. Hertzman and J. B. Dillon, "Distinction between Arterial, Venous, and Flow Components in Photoelectric Plethysmography in Man," *American Journal of Physiology,* vol. 130, no. 1, pp. 177-185, 1940.

[17] A. B. Hertzman and J. B. Dillon, "Applications of Photoelectric Plethysmography in Peripheral Vascular Disease," *American Heart Journal,* vol. 20, no. 6, pp. 750-761, 1940.

[18] J. M. Schmitt, "Simple Photon Diffusion Analysis of the Effects of Multiple Scattering on Pulse Oximetry," *IEEE Transactions on Biomedical Engineering,* vol. 38, no. 12, pp. 1194-1203, 1991.

[19] W. B. Baker, A. B. Parthasarathy, D. R. Busch, R. C. Mesquita, J. H. Greenberg, and A. Yodh, "Modified Beer-Lambert law for blood flow," *Biomedical optics express,* vol. 5, no. 11, pp. 4053-4075, 2014.

[20] A. V. J. Challoner and C. A. Ramsay, "A photoelectric plethysmograph for the measurement of cutaneous blood flow," *Physics in Medicine and Biology*, Article vol. 19, no. 3, pp. 317-328, 1974.

[21] T. Aoyagi, "Improvement of an earpiece oximeter," *Abstracts, 13th Annual Meeting of the Japan Society of Medical Electronics and Biological Engineering, Osaka, Japan*, 1974.

[22] I. Yoshiya, Y. Shimada, and K. Tanaka, "Spectrophotometric monitoring of arterial oxygen saturation in the fingertip," *Medical and Biological Engineering and Computing*, vol. 18, no. 1, pp. 27-32, 1980.

[23] T. Valencell, "Optical heart rate monitoring: what you need to know," ed: Valencell. Diakses dari http://valencell. com/blog/10/optica l-heart …, 2015.

[24] K. Watanabe, T. Watanabe, H. Watanabe, H. Ando, T. Ishikawa, and K. Kobayashi, "Noninvasive Measurement of Heartbeat, Respiration, Snoring and Body Movements of a Subject in Bed via a Pneumatic Method," *IEEE Transactions on Biomedical Engineering* vol. 52, no. 12, pp. 2100-7, 2005.

[25] S. Li and Y. Minghui, "Video-based Heart Rate Measurement Using Head Motion Tracking and ICA," vol. 01, ed: IEEE International Congress on Image and Signal Processing (CISP), 2013, pp. 160-164.

[26] Y. Yong-Poh, K. Ban-Hoe, L. Chern-Loon, W. Siaw-Lang, and P. Raveendran, "Video-based Heart Rate Measurement using Short-time Fourier transform," ed: IEEE International Symposium on Intelligent Signal Processing and Communication Systems, 2013, pp. 704-707.

[27] L. Liu, L. Lu, J. Luo, J. Zhang, and X. Chen, "Enhanced Eulerian Video Magnification," in *7th International Congress on Image and Signal Processing*, 2014, pp. 50-54.

[28] K. Kamble, N. P. Jagtap, R. Patil, and A. A. Bhurane, "A Review: Eulerian Video Motion Magnification," *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 3, pp. 2384-2390, 2015.

[29] C. Liu, A. Torralba, W. T. Freeman, F. Durand, and E. H. Adelson, "Motion magnification," presented at the ACM Transactions on Graphics, Los Angeles, California, 2005.

[30] Y. Wu Hao, M. Rubinstein, E. Shih, J. Guttag, F. Durand, and W. Freeman, "Eulerian video magnification for revealing subtle changes in the world," *ACM Transactions on Graphics*, 2012.

[31] A. Davletcharova, S. Sugathan, B. Abraham, and A. P. James, "Detection and Analysis of Emotion from Speech Signals," *Procedia Computer Science*, vol. 58, pp. 91-96, 2015.

[32] A. P. James, "Heart rate monitoring using human speech spectral features," *Human-centric Computing and Information Sciences*, vol. 5, no. 1, p. Article 52, 2015.

[33] M. Usman, Z. Ahmad, and M. Wajid, "Dataset of Raw and Pre-processed Speech Signals, Mel Frequency Cepstral Coefficients of Speech and Heart Rate Measurements," ed: IEEE 5th International Conference on Signal Processing, Computing and Control (ISPCC), 2019, pp. 376-379.

[34] Y. Jin *et al.*, "Effect of Changing Heart Rate on the Ocular Pulse and Dynamic Biomechanical Behavior of the Optic Nerve Head," *Investigative Ophthalmology and Visual Science*, Article vol. 61, no. 4, 2020.

[35] S. Park, M. J. Won, D. W. Lee, and M. Whang, "Non-contact measurement of heart response reflected in human eye," *International Journal of Psychophysiology*, Article vol. 123, pp. 179-198, 2018.

[36] A. Parnandi and R. Gutierrez-Osuna, "Contactless Measurement of Heart Rate Variability from Pupillary Fluctuations," in *Humaine Association Conference on Affective Computing and Intelligent Interaction*, 2013, pp. 191-196.

[37] G. Balakrishnan, F. Durand, and J. Guttag, "Detecting Pulse from Head Motions in Video," ed: IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 3430-3437.

[38] Y. Cheng, C. Gene, and V. Stankovic, "Estimating Heart Rate and Rhythm via 3D Motion Tracking in Depth Video," *IEEE Transactions on Multimedia,* Periodical vol. 19, no. 7, pp. 1625-1636, 2017.

[39] J. P. Lomaliza and H. Park, *Detecting Pulse from Head Motions Using Smartphone Camera* (Lecture Notes in Electrical Engineering). Springer Verlag, 2017, pp. 243-251.

[40] S. Pare, A. Bhandari, A. Kumar, G. Singh, and S. Khare, *Satellite image segmentation based on different objective functions using genetic algorithm: A comparative study*. 2015, pp. 730-734.

[41] M. Besnassi, N. Neggaz, and A. Benyettou, "Face detection based on evolutionary Haar filter," *Pattern Analysis and Applications,* vol. 23, no. 1, pp. 309-330, 2020.

[42] X. He, R. A. Goubran, and X. P. Liu, "Wrist Pulse Measurement and Analysis using Eulerian Video Magnification," in *IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, 2016, pp. 41-44.

[43] Y. Zhang, S. L. Pintea, and J. C. V. Gemert, "Video Acceleration Magnification," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 502-510.

[44] H. Yu and B. M. Wilamowski, "Levenberg–marquardt training," in *Intelligent systems*: CRC Press, 2018, pp. 12-1-12-16.

[45] M. Kebe, R. Gadhafi, B. Mohammad, M. Sanduleanu, H. Saleh, and M. Al-Qutayri, "Human Vital Signs Detection Methods and Potential Using Radars: A Review," *Sensors (Basel)*, vol. 20, no. 5, p. 1454, 2020.

[46] C. Lockwood, T. Conroy-Hiller, and T. Page, "Vital signs," *JBI Library of Systematic Reviews,* vol. 2, no. 6, pp. 1-38, 2004.

[47] Barrett, K. E., Ganong, and W. F., *Ganong's review of medical physiology*, 24th ed. ed. New York : London: McGraw-Hill Medical, 2012.

[48] J. E. Hall, *Guyton and Hall Textbook of Medical Physiology*, Twelfth edition. ed. Philadelphia, Pa.: Saunders/Elsevier, 2011.

[49] R. Rajaganeshan, C. L. Ludlam, D. P. Francis, S. V. Parasramka, and R. Sutton, "Accuracy in ECG lead placement among technicians, nurses, general physicians and cardiologists," *International Journal of Clinical Practice*, vol. 62, no. 1, pp. 65-70, Jan 2008.

[50] T. Tabassum and M. Islam, "An approach of cardiac disease prediction by analyzing ECG signal," ed: IEEE 3rd International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), 2016, pp. 1-5.

[51] R. Gade *et al.*, "Accuracy of Electro Cardiogram (ECG) Interpretation and need for ECG Audit in Emergency Department, A prospective observational study from a tertiary care teaching hospital," *Asian Pacific Journal of Health Sciences*, vol. 2, pp. 173-178, 2015.

[52] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiological Measurement*, vol. 28, no. 3, pp. R1-39, 2007.

[53] D. Castaneda, A. Esparza, M. Ghamari, C. Soltanpur, and H. Nazeran, "A review on wearable photoplethysmography sensors and their potential future applications in health care," *International Journal of Biosensors & Bioelectronics*, vol. 4, no. 4, pp. 195-202, 2018.

[54] T. Tamura, Y. Maeda, M. Sekine, and M. Yoshida, "Wearable Photoplethysmographic Sensors—Past and Present," *Electronics*, vol. 3, no. 2, pp. 282-302, 2014.

[55] J. Přibil, A. Přibilová, and I. Frollo, "Comparative Measurement of the PPG Signal on Different Human Body Positions by Sensors Working in Reflexive and Transmission Modes," *Engineering Proceedings*, vol. 2, p. 69, 2020.

[56] K. M. van der Kooij and M. Naber, "An open-source remote heart rate imaging method with practical apparatus and algorithms," *Behavior Research Methods,* Article vol. 51, no. 5, pp. 2106-2119, 2019.

[57] L. Scalise, N. Bernacchia, I. Ercoli, and P. Marchionni, "Heart rate measurement in neonatal patients using a webcamera," in *IEEE International Symposium on Medical Measurements and Applications Proceedings*, 2012, pp. 1-4.

[58] M. A. Hassan *et al.*, "Heart rate estimation using facial video: A review," *Biomedical Signal Processing and Control,* vol. 38, pp. 346-360, 2017.

[59] R. McCluney, *Introduction to Radiometry and Photometry*. Artech House, 2014.

[60] S. S. Chowdhury, R. Hyder, M. S. B. Hafiz, and M. A. Haque, "Real-Time Robust Heart Rate Estimation from Wrist-Type PPG Signals Using Multiple Reference Adaptive Noise Cancellation," *IEEE Journal of Biomedical and Health Informatics,* Article vol. 22, no. 2, pp. 450-459, 2018.

[61] A. Temko, "Accurate Heart Rate Monitoring During Physical Exercises Using PPG," *IEEE Transactions on Biomedical Engineering,* Periodical vol. 64, no. 9, pp. 2016-2024, 2017.

[62] D. Jarchi and A. J. Casson, "Towards Photoplethysmography-Based Estimation of Instantaneous Heart Rate During Physical Activity," *IEEE Transactions on Biomedical Engineering,* Periodical vol. 64, no. 9, pp. 2042-2053, 2017.

[63] M. Kumar, A. Veeraraghavan, and A. Sabharwal, "DistancePPG: Robust non-contact vital signs monitoring using a camera," *Biomedical Optics Express,* Article vol. 6, no. 5, pp. 1565-1588, 2015.

[64] G. de Haan and V. Jeanne, "Robust Pulse Rate From Chrominance-Based rPPG," *IEEE Transactions on Biomedical Engineering,* Periodical vol. 60, no. 10, pp. 2878-2886, 2013.

[65] M. van Gastel, S. Stuijk, and G. Haan, "Motion Robust Remote-PPG in Infrared," *IEEE Transactions on Biomedical Engineering,* 2015.

[66] D. McDuff, S. Gontarek, and R. W. Picard, "Improvements in Remote Cardiopulmonary Measurement Using a Five Band Digital Camera," *IEEE Transactions on Biomedical Engineering*, Periodical vol. 61, no. 10, pp. 2593-2601, 2014.

[67] Y. Yan, X. Ma, L. Yao, and J. Ouyang, "Noncontact measurement of heart rate using facial video illuminated under natural light and signal weighted analysis," *Biomedical Materials and Engineering*, vol. 26 Suppl 1, pp. S903-9, 2015.

[68] E. Garfield, "Journal impact factor: a brief review," *Canadian Medical Association Journal*, vol. 161, no. 8, pp. 979-980, 1999.

[69] E. R. Smith, "The journal impact factor," *The Canadian Journal of Cardiology*, vol. 22, no. 9, p. 787, 2006.

[70] L. A. Vucovich, J. B. Baker, and J. T. Smith, "Analyzing the impact of an author's publications," *J Med Libr Assoc*, vol. 96, no. 1, pp. 63-66, 2008.

[71] V. Ronca *et al.*, "A Video-Based Technique for Heart Rate and Eye Blinks Rate Estimation: A Potential Solution for Telemonitoring and Remote Healthcare," *SENSORS*, vol. 21, no. 5, p. 1607, 2021.

[72] M. S. Mohsen, A. B. Fakhri, N. M. Ahmed, M. F. Mahmood, and S. L. Mohammed, "Video Magnification Techniques: Medical Applications and Comparison of Methods," in *IOP Conference Series: Materials Science and Engineering*, 2021, vol. 1105, no. 1: IOP Publishing, p. 012074.

[73] R. Song, S. Zhang, C. Li, Y. Zhang, J. Cheng, and X. Chen, "Heart Rate Estimation From Facial Videos Using a Spatiotemporal Representation With Convolutional Neural Networks," *IEEE Transactions on Instrumentation and Measurement*, Periodical vol. 69, no. 10, pp. 7411-7421, 2020.

[74] H. Shahadi, J. Zaid, Z. Jabbar Al-allaq, H. Albatat, I. Haider, and Shahadi, "Efficient denoising approach based Eulerian video magnification for colour and motion variations," *International Journal of Electrical and Computer Engineering*, vol. 10, pp. 4701-4711, 2020.

[75] H. Shahadi, Z. Jabbar Al-allaq, and H. Albatat, "Developed approach for phase-based Eulerian video magnification," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 18, pp. 2391-2400, 2020.

[76] H. Shahadi, H. Albatat, Z. Jabbar Al-allaq, A. Thahab, I. Haider, and Shahadi, "Eulerian video magnification: a review," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 18, pp. 799-811, 2020.

[77] A. d. F. G. Rosa and R. C. Betini, "Noncontact SpO2 Measurement Using Eulerian Video Magnification," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 5, pp. 2120-2130, 2020.

[78] P. Gupta, B. Bhowmick, and A. Pal, "MOMBAT: Heart rate monitoring from face video using pulse modeling and Bayesian tracking," *Computers in Biology and Medicine*, Article vol. 121, 2020.

[79] S. L. Fernandes, V. P. Gurupur, N. R. Sunder, N. Arunkumar, and S. Kadry, "A novel nonintrusive decision support approach for heart rate measurement," *Pattern Recognition Letters*, vol. 139, pp. 148-156, 2020.

[80] Z. Yu, W. Peng, X. Li, X. Hong, and G. Zhao, "Remote Heart Rate Measurement From Highly Compressed Facial Videos: An End-to-End Deep Learning Solution With Video Enhancement," ed: IEEE, 2019, pp. 151-160.

[81] V. Perrot, S. Salles, D. Vray, and H. Liebgott, "Video Magnification Applied in Ultrasound," *IEEE Transactions on Biomedical Engineering*, Periodical vol. 66, no. 1, pp. 283-288, 2019.

[82] X. Chen, J. Cheng, R. Song, Y. Liu, R. Ward, and Z. J. Wang, "Video-Based Heart Rate Measurement: Recent Advances and Future Prospects," *IEEE Transactions on Instrumentation and Measurement*, Periodical vol. 68, no. 10, pp. 3600-3615, 2019.

[83] S. Shourjya and N. Koushik Kumar, "Algorithms for Monitoring Heart Rate and Respiratory Rate From the Video of a User's Face," *IEEE Journal of*

Translational Engineering in Health and Medicine, article vol. 6, pp. 1-11, 2018.

[84] E. Moya-Albor, J. Brieva, H. Ponce, O. Rivas-Scott, and C. Gomez-Pena, "Heart Rate Estimation using Hermite Transform Video Magnification and Deep Learning," ed: 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2018, pp. 2595-2598.

[85] H. Ghanadian, M. Ghodratigohar, and H. A. Osman, "A Machine Learning Method to Improve Non-Contact Heart Rate Monitoring Using an RGB Camera," *IEEE Access*, vol. 6, pp. 57085-57094, 2018.

[86] A. Al-Naji, S.-H. Lee, and J. Chahl, "An efficient motion magnification system for real-time applications," *Machine Vision and Applications*, vol. 29, no. 4, pp. 585-600, 2018.

[87] Y. Yang *et al.*, "Blind identification of full-field vibration modes from video measurements with phase-based video motion magnification," *Mechanical Systems and Signal Processing*, vol. 85, pp. 567-590, 2017.

[88] H. Qi, Z. Guo, X. Chen, Z. Shen, and Z. Jane Wang, "Video-based human heart rate measurement using joint blind source separation," *Biomedical Signal Processing & Control*, vol. 31, pp. 309-320, 2017.

[89] S. S. Lokhande and B. K. Nair, "Heart rate and respiratory rate measurement using image processing," *International Journal of Innovative Research in Computer and Communication Engineering*, 2017.

[90] J. Kranjec, S. Beguš, G. Geršak, M. Šinkovec, J. Drnovšek, and D. Hudoklin, "Design and Clinical Evaluation of a Non-Contact Heart Rate Variability Measuring Device," *Sensors*, vol. 17, no. 11, p. 2637, 2017.

[91] Y.-P. Yu, P. Raveendran, and C.-L. Lim, "Dynamic heart rate measurements from video sequences," *Biomedical Optics Express*, vol. 6, no. 7, pp. 2466-2480, 2015.

[92] C. H. Antink, H. Gao, C. Brüser, and S. Leonhardt, "Beat-to-beat heart rate estimation fusing multimodal video and sensor data," *Biomedical Optics Express*, vol. 6, no. 8, pp. 2895-2907, 2015.

[93] A. Alzahrani and A. Whitehead, "Preprocessing Realistic Video for Contactless Heart Rate Monitoring Using Video Magnification," in *12th Conference on Computer and Robot Vision*, 2015, pp. 261-268.

[94] L. Xiaobai, C. Jie, Z. Guoying, and M. Pietikainen, "Remote Heart Rate Measurement from Face Videos under Realistic Situations," ed: IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 4264-4271.

[95] N. Wadhwa, M. Rubinstein, Durand, W. T., and Freeman, "Phase-based video motion processing," *ACM Transactions on Graphics*, vol. 32, no. 4, pp. 1-10, 2013.

[96] L. Shan and M. Yu, "Video-based heart rate measurement using head motion tracking and ICA," in *6th International Congress on Image and Signal Processing (CISP)*, 2013, vol. 01, pp. 160-164.

[97] K. Sungjun, K. Hyunseok, and P. Kwang Suk, "Validation of heart rate extraction using video imaging on a built-in camera system of a smartphone," ed: Annual International Conference of the IEEE Engineering in Medicine and Biology Society Engineering in Medicine and Biology Society (EMBC), 2012, pp. 2174-2177.

[98] C. G. Scully *et al.*, "Physiological Parameter Monitoring from Optical Recordings With a Mobile Phone," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 2, pp. 303-306, 2012.

[99] M. Poh, D. J. McDuff, and R. W. Picard, "Advancements in Noncontact, Multiparameter Physiological Measurements Using a Webcam," *IEEE Transactions on Biomedical Engineering*, Periodical vol. 58, no. 1, pp. 7-11, 2011.

[100] H. Monkaresi, R. A. Calvo, and Y. Hong, "A Machine Learning Approach to Improve Contactless Heart Rate Monitoring Using a Webcam," *IEEE Journal of Biomedical and Health Informatics*, Periodical vol. 18, no. 4, pp. 1153-1160, 2014.

[101] Y. Liu *et al.*, "Motion-Robust Multimodal Heart Rate Estimation Using BCG Fused Remote-PPG With Deep Facial ROI Tracker and Pose Constrained Kalman Filter," *IEEE Transactions on Instrumentation and Measurement*, Periodical vol. 70, pp. 1-15, 2021.

[102] A. Zolfi, S. Avidan, Y. Elovici, and A. Shabtai, *Adversarial Mask: Real-World Adversarial Attack Against Face Recognition Models*. 2021.

[103] M. D. Firdaus and H. Nugroho, "Classification of stamps and handmade batik based on pattern recognition," in *Journal of Physics: Conference Series*, 2019, vol. 1402, no. 6: IOP Publishing, p. 066053.

[104] D. Das *et al.*, "A comparative analysis of four classification algorithms for university students performance detection," 2020: Springer, International Conference on Electrical, Control & Computer Engineering, pp. 415-424.

[105] N. Sabri *et al.*, "A Comparison of Face Detection Classifier using Facial Geometry Distance Measure," in *9th IEEE Control and System Graduate Research Colloquium (ICSGRC)*, 2018, pp. 116-120.

[106] P. Musa, F. A. Rafi, and M. Lamsani, "A Review: Contrast-Limited Adaptive Histogram Equalization (CLAHE) methods to help the application of face recognition," in *Third International Conference on Informatics and Computing (ICIC)*, 2018, pp. 1-6.

[107] I. Febin, K. Jayasree, and P. T. Joy, "Violence detection in videos for an intelligent surveillance system using MoBSIFT and movement filtering algorithm," *Pattern Analysis and Applications*, vol. 23, no. 2, pp. 611-623, 2020.

[108] I. Gitman and B. Ginsburg, "Comparison of batch normalization and weight normalization algorithms for the large-scale image classification," 2017.

[109] Y. Wang, X. Luo, L. Ding, and J. Wu, "Object tracking via dense SIFT features and low-rank representation," *Soft Computing*, vol. 23, no. 20, pp. 10173-10186, 2019.

[110] H. Bindu and K. Manjunathachari, "Hybrid feature descriptor and probabilistic neuro-fuzzy system for face recognition," *Sensor Review*, 2018.

[111] S. Nazir, M. H. Yousaf, and S. A. Velastin, "Evaluating a bag-of-visual features approach using spatio-temporal features for action recognition," *Computers & Electrical Engineering*, vol. 72, pp. 660-669, 2018.

[112] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Computer vision and image understanding*, vol. 110, no. 3, pp. 346-359, 2008.

[113] Y. Kortli, M. Jridi, A. Al Falou, and M. Atri, "A comparative study of CFs, LBP, HOG, SIFT, SURF, and BRIEF techniques for face recognition," in *Pattern recognition and tracking XXIX*, 2018, vol. 10649: International Society for Optics and Photonics, p. 106490M.

[114] C. Shu, X. Ding, and C. Fang, "Histogram of the oriented gradient for face recognition," *Tsinghua Science and Technology*, vol. 16, no. 2, pp. 216-224, 2011.

[115] V. Nasteski, "An overview of the supervised machine learning methods," *Horizons. b*, vol. 4, pp. 51-62, 2017.

[116] M. R. Rejeesh, "Interest point based face recognition using adaptive neuro fuzzy inference system," *Multimedia Tools and Applications*, vol. 78, no. 16, pp. 22691-22710, 2019.

[117] G. Xie *et al.*, "SRSC: Selective, Robust, and Supervised Constrained Feature Representation for Image Classification," *IEEE Transactions on Neural Networks and Learning Systems*, Periodical vol. 31, no. 10, pp. 4290-4302, 2020.

[118] J. E. van Engelen and H. H. Hoos, "A survey on semi-supervised learning," *Machine Learning*, vol. 109, no. 2, pp. 373-440, 2020.

[119] R. Meyes, M. Lu, C. Waubert de Puiseau, and T. Meisen, *Ablation Studies in Artificial Neural Networks*. 2019.

[120] S. Li, *Global Face Pose Detection Based on an Improved PSO-SVM Method*. International Conference on Aviation Safety and Information Technology, 2020, pp. 549-553.

[121] Z. Kallenborn and P. C. Bleek, "Swarming destruction: drone swarms and chemical, biological, radiological, and nuclear weapons," *The Nonproliferation Review*, vol. 25, no. 5-6, pp. 523-543, 2018.

[122] J. van den IJssel, J. Encarnação, E. Doornbos, and P. Visser, "Precise science orbits for the Swarm satellite constellation," *Advances in Space Research*, vol. 56, no. 6, pp. 1042-1055, 2015.

[123] T. Tashtoush *et al.*, "Design of a Swarm Search Algorithm: DustySWARM Spiral Epicycloidal Wave (SEW) Code for NASA Swarmathon," *International Journal of Research Studies in Computer Science and Engineering (IJRSCSE)*, vol. 7, no. 1, pp. 28-36, 2020.

[124] M. M. al-Rifaie and A. Aber, "Identifying metastasis in bone scans with stochastic diffusion search," in *International Symposium on Information Technologies in Medicine and Education*, 2012, vol. 1: IEEE, pp. 519-523.

[125] M. M. Al-Rifaie, A. Aber, and D. J. Hemanth, "Deploying Swarm Intelligence in Medical Imaging Identifying Metastasis, Micro-Calcifications and Brain Image Segmentation," *IET Systems Biology*, vol. 9, no. 6, pp. 234-244, 2015.

[126] D. Karaboga, "An Idea Based on Honey Bee Swarm for Numerical Optimization, Technical Report - TR06," *Technical Report, Erciyes University*, 2005.

[127] D. Karaboga and B. Akay, "A comparative study of Artificial Bee Colony algorithm," *Applied Mathematics and Computation*, vol. 214, no. 1, pp. 108-132, 2009.

[128] S. Arslan and C. Ozturk, "Multi hive artificial bee colony programming for high dimensional symbolic regression with feature selection," *Applied Soft Computing*, vol. 78, pp. 515-527, 2019.

[129] X. Wu, X. Yang, J. Jin, and Z. Yang, "Amplitude-Based Filtering for Video Magnification in Presence of Large Motion," *Sensors (Basel)*, vol. 18, no. 7, 2018.

[130] R. Ullah *et al.*, "A Real-Time Framework for Human Face Detection and Recognition in CCTV Images," *Mathematical Problems in Engineering*, vol. 2022, p. 3276704, 2022.

[131] A. Dirin, N. Delbiaggio, and J. Kauttonen, "Comparisons of Facial Recognition Algorithms Through a Case Study Application," *International Journal of Interactive Mobile Technologies* vol. 14, pp. pp. 121-133, 2020.

[132] S. Videla and A. K. P M, "Modified Feature Extraction Using Viola Jones Algorithm," *Journal of Advanced Research in Dynamical and Control Systems*, vol. 10, pp. 528-538, 2018.

[133] O. Starostenko, C. Cruz-Perez, V. Alarcon-Aquino, and R. Rosas-Romero, "Real-time facial expression recognition using local appearance-based descriptors," *Journal of Intelligent & Fuzzy Systems*, vol. 36, pp. 5037-5049, 2019.

[134] T. Abd El-Hafeez, A. Ali, and Y. Mohany, "An Accurate System for Face Detection and Recognition," vol. 33, pp. 1-19, 2019.

[135] M. Collins. "The naive bayes model, maximum-likelihood estimation, and the em algorithm." http://www.cs.columbia.edu/~mcollins/em.pdf. (accessed June 2020).

[136] A. Zarkasi *et al.*, "Face Movement Detection Using Template Matching," in *International Conference on Electrical Engineering and Computer Science (ICECOS)*, 2018, pp. 333-338.

[137] G. Yang and T. S. Huang, "Human face detection in a complex background," *Pattern Recognition,* Article vol. 27, no. 1, pp. 53-63, 1994.

[138] J. Zhang, X. Wu, S. C. H. Hoi, and J. Zhu, "Feature agglomeration networks for single stage face detection," *Neurocomputing,* vol. 380, pp. 180-189, 2020.

[139] T. Soon *et al.*, "The Utilization of Feature based Viola-Jones Method for Face Detection in Invariant Rotation," *International Journal of Advanced Computer Science and Applications,* vol. 9, 2018.

[140] H. Md Khaled, A. Md. Shamim, M. Abdullah Al, S. H. S. Newaz, and L. Gyu Myoung, "Human Face Detection Techniques: A Comprehensive Review and Future Research Directions," *Electronics,* article vol. 10, no. 2354, pp. 2354-2354, 2021.

[141] S. Shivappriya, M. J. P. Priyadarsini, A. Stateczny, C. Puttamadappa, and B. Parameshachari, "Cascade Object Detection and Remote Sensing Object Detection Method Based on Trainable Activation Function," *Remote Sensing,* vol. 13, no. 2, p. 200, 2021.

[142] J. Cai, S. Gu, and L. Zhang, "Learning a Deep Single Image Contrast Enhancer from Multi-Exposure Images," *IEEE Transactions on Image Processing,* vol. 27, no. 4, pp. 2049-2062, 2018.

[143] F. Yang, S. He, S. Sadanand, A. Yusuf, and M. Bolic, "Contactless Measurement of Vital Signs Using Thermal and RGB Cameras: A Study of COVID 19-Related Health Monitoring," *Sensors,* vol. 22, p. 627, 2022.

[144] W. Verkruysse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light," *Opt Express,* vol. 16, no. 26, pp. 21434-21445, 2008.

[145] V. Nair, P. Ram, and S. Sundararaman, "Shadow detection and removal from images using machine learning and morphological operations," *The Journal of Engineering*, 2018.

[146] V. Bogdan, C. Bonchiş, and C. Orhei, "Custom Dilated Edge Detection Filters," *Journal of Winter School of Computer Graphics* vol. 28, pp. 161-168, 2020.

[147] A. I. Salhi, M. Kardouchi, and N. Belacel, "Histograms of fuzzy oriented gradients for face recognition," ed: IEEE International Conference on Computer Applications Technology (ICCAT), 2013, pp. 1-5.

[148] F. Suard, A. Rakotomamonjy, A. Bensrhair, and A. Broggi, "Pedestrian Detection using Infrared images and Histograms of Oriented Gradients," in *IEEE Intelligent Vehicles Symposium*, 13-15 June 2006 2006, pp. 206-212.

[149] B. Wang, W. Zhao, P. Gao, Y. Zhang, and Z. Wang, "Crack damage detection method via multiple visual features and efficient multi-task learning model," *Sensors*, vol. 18, no. 6, p. 1796, 2018.

[150] X. Teng and Y. Gong, "Research on Application of Machine Learning in Data Mining," *IOP Conference Series: Materials Science and Engineering*, vol. 392, p. 062202, 08/03 2018.

[151] M. Leon, S. Islam, J. Akter, N. Sakib, and M. Islam, "Analysis of EEG Signal Classification for Application in SSVEP-Based BCI Using Convolutional Neural Network," in *Proceedings of the International Conference on Big Data, IoT, and Machine Learning*, 2022: Springer, pp. 593-606.

[152] L. Bote-Curiel, S. Munoz-Romero, A. Gerrero-Curieses, and J. L. Rojo-Álvarez, "Deep learning and big data in healthcare: A double review for critical beginners," *Applied Sciences*, vol. 9, no. 11, p. 2331, 2019.

[153] M. F. Burg *et al.*, "Learning divisive normalization in primary visual cortex," *PLoS computational biology*, vol. 17, no. 6, p. e1009028, 2021.

[154] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *2017 International Conference on Engineering and Technology (ICET)*, 2017, pp. 1-6.

[155] R. Venkatesan, *Convolutional neural networks in visual computing : a concise guide*, Boca Raton, Florida ; London, England: CRC Press, 2018.

[156] N. Aloysius and M. Geetha, "A review on deep convolutional neural networks," in *2017 International Conference on Communication and Signal Processing (ICCSP)*, 6-8 April 2017 2017, pp. 0588-0592.

[157] *Convolutional neural networks for medical image processing applications*, Ş. Öztürk, ed., First edition. ed. Boca Raton, FL: CRC Press, 2022.

[158] A. Laith *et al.*, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *Journal of Big Data*, article vol. 8, no. 1, pp. 1-74, 2021.

[159] E. Siregar, H. Mawengkang, E. B. Nababan, and A. Wanto, "Analysis of Backpropagation Method with Sigmoid Bipolar and Linear Function in Prediction of Population Growth," in *Journal of Physics: Conference Series*, 2019, vol. 1255, no. 1: IOP Publishing, p. 012023.

[160] A. I. Abbas, O. Y. M. Alhamadani, and M. U. Mohammed, "The application of an artificial neural network for 2D coordinate transformation," *Journal of Intelligent Systems*, vol. 31, no. 1, pp. 739-752, 2022.

[161] Vandana and N. Kaur, "Fingerprint and Face-Based Secure Biometric Authentication System Using Optimized Robust Features," 2020.

[162] A. Ben-David, "Comparison of classification accuracy using Cohen's Weighted Kappa," *Expert Systems with Applications*, vol. 34, no. 2, pp. 825-832, 2008.

[163] S. Vieira, U. Kaymak, and J. Sousa, *Cohen's kappa coefficient as a performance measure for feature selection*. 2010, pp. 1-8.

[164] Z. Somogyi, *The application of artificial intelligence : step-by-step guide from beginner to expert*, Cham: Springer, 2021.