Preventing Falls & Accidents of Babies in a Crib using Machine Learning-Based Intelligent Alert System

Neda Khan

A thesis submitted to

Auckland University of Technology

in fulfilment of the requirements for the degree of

Doctor of Philosophy (PhD)



School of Engineering, Computer and Mathematical Sciences

New Zealand

2026

Abstract

Babies play a very important role in a healthcare setting. Babies aged 6-24 months are at a crucial stage of growth. They need to be supervised because of their growing age, developmental stage, and motor skills. These are the reasons why babies are in dangerous postures and unsafe scenarios, leading to falls, which sometimes lead to fatal injuries, severe head injuries, bone fractures, etc. Approximately 52% of hospitalizations for falls involve babies aged 0-4 years. Around 31% of fall-related hospitalizations happen at home. Parents who work can't always keep an eye on their newborns, so they use nannies, daycare services, or smart technology to monitor their babies 24/7 and keep them safe. But concerns about caregivers' trustworthiness and the safety of childcare arrangements cause stress and underscore the importance of affordable, reliable e-health solutions that let you always keep an eye on your infant. In modern times, Artificial Intelligence (AI) and Machine Learning (ML) play a very important role in daily life in healthcare applications, serving as a vital companion for babies at home in the crib, providing continuous vigilance of babies' activities, early fall risk predictions, and resolving the issues of negligence.

The focus of this thesis is to identify fall risk scenarios among babies within the crib setting by analyzing behaviour patterns using AI and Machine Learning techniques. The proposed approach, termed the ADL Conditioned Cascaded AI model for potential baby fall risk prediction, employs Pose features (MediaPipe), a Gated Recurrent Unit (GRU), a Soft Spatial Attention mechanism, and a Dual-Stage Trigger mechanism to predict high-risk ADLs that lead to future baby fall risk. In this research, all the models use Activities of Daily Living (ADLs) within the crib, such as sitting, playing, lying down, leaning over the fence, standing near the fence, and climbing over the fence as input parameters. This thesis presents three significant contributions:

- A detailed background analysis of baby fall detection, baby fall prediction, and baby safety monitoring was done in terms of behaviour, identifying scenarios within or outside the crib.
- The system architecture mainly consists of the AI-3 model, leveraging the techniques and overcoming the flaws of the AI-1 and AI-2 models to predict the high-risk scenarios leading to falls and generate a more reliable outcome.
- Data is collected from online platforms for all the models, and a few videos are generated specifically by an AI tool to test the AI-2 and AI-3 models on an unseen dataset. By testing the AI-3 model in the time domain, it effectively predicts high-risk ADLs that lead to future baby fall risk.

This thesis summarizes the research findings and model advancements that underscore the considerable potential of sophisticated deep learning techniques to enhance future predictions of babies' fall risk, establish an evolving warning system, and ultimately, forecast and mitigate fall incidents. This cutting-

edge technology offers a revolutionary solution that enables prompt interventions, making babies in the crib safer and healthier while also relieving parents of the stress of constant vigilance.

Contents

Abstract ii

List of Figures ix

List of Tables xi

Attestation of Authorship xii

Co-Authored Works xiii

Artificial Intelligence Declaration xiv

Acknowledgments xvi

Ethics Approval xviii

Chapter 1. Research Motivation, Focus, and Thesis Structure 1

1.1. Introduction 1

1.2. Advances in AI/ML for Healthcare Applications 1

1.3. Rationale and Significance of Baby Safety in-Crib and Fall Prediction System 2

1.4. Motivation for Advancing AI-driven Fall Prediction 4

1.5. Research Objectives and Contributions 6

1.5.1 Research Objectives 6

1.5.2 Research Contributions 7

1.7. Thesis Structure 9

1.8. Chapter Summary 11

Chapter 2. Literature Review 12

2.1 Introduction 12

2.2 Baby Healthcare Monitoring Systems: Contributing Factors 12

2.3 Overview of Baby Monitoring for Health & Safety 14

2.3.1.1 *Using sensors, hardware, and IoT* 15

2.3.1.2 Using AI and ML techniques 15

2.3.1.3 *Using integration of Sensors and AI/ML* 16

2.3.2 Without Crib 18

2.3.2.1 Using Sensors, hardware and IoT 18

2.3.2.2 Using AI and ML techniques 18
2.3.2.3 Using integration of Sensors and AI/ML 20
2.4 Overview of Baby Monitoring for Fall Detection 21
2.4.1 Within Crib 21
2.4.1.1 *Using sensors, hardware, and IoT* 21
2.4.2 Without Crib 22
2.4.2.1 Using Sensors, hardware and IoT 22
2.4.2.2 *Using AI/ML techniques* 22
2.4.2.3 *Using integration of Sensors and AI/ML* 22
2.5 Technologies Used for Baby Fall Prediction Inside the Crib 23
2.5.1 Sensors/IoT 23
2.5.2 Using AI and ML Techniques 23
2.6 Analysis of Smart Devices Powered by AI 26
2.7 Comparison of AI-Based Fall Prediction within Baby Crib 29
2.8 Conclusions 31
Chapter 3. Research Methodology 33
3.1 Introduction 33
3.2 Methodology and Model Development Phase 34
3.3 AI/ML Modeling and Development Simulation Tools 37
3.4 Python-Based AI Modeling and Jupyter Notebook IDE 38
3.5 Conclusion 41
Chapter 4. Proposed AI-1: Pose Estimation-Based Baby Unsafe Scenarios Prediction System 43
4.1 Introduction 43
4.2 Significance of Human Pose Estimation (HPE) 43
4.2.1 Deep Learning Models in HPE 44
4.3 Role of Supervised Machine Learning 45
4.3.1 Applied ML Algorithms/Classifiers 46
4.4 Architectural Overview of Pose-Estimation-Based Baby Unsafe Scenarios Prediction Model 47

4.4.1 MediaPipe Pose Estimation 49
4.4.2 Classifiers Used for Prediction 49
4.5 Algorithm for Pose-Estimation-Based Baby Unsafe Scenarios Prediction Model 50
4.6 Data Collection and Ground Truth Selection 52
4.6.1 Evaluation Metrics 53
4.7 Assessment of the Developed PE-ML Model Based on AI 53
4.8 Conclusions 55
Chapter 5. Proposed AI-2: CNN-Attention-Based Baby Unsafe Scenarios Prediction System 57
5.1 Introduction 57
5.2 CNN-based Monitoring 59
5.3 Importance of ADLs and Parameter Selection 59
5.4 Convolutional Neural Network (CNN) 61
5.4.1 Standard Techniques for Reducing Overfitting 61
5.4.1.1 Data Augmentation 62
5.4.1.2 Increasing Training Data 62
5.4.1.3 Simplify the model 63
5.4.1.4 Batch Normalization 63
5.4.1.5 Regularization 63
5.4.1.6 Dropout 63
5.4.1.7 Cross-validation 63
5.4.1.8 Early Stopping 63
5.4.1.9 Ensemble Methods 64
5.4.1.10 Transfer Learning 64
5.4.2 Benchmarking CNN 64
5.4.2.1 The Proposed Architecture 65
5.4.2.2 Experimentation and Evaluation Results 66
5.4.3 Soft Spatial Attention (SSA) Architecture 70
5.4.4 CNN-SSA 70
5.5 Algorithm for CNN-Attention-Based Baby Unsafe Scenarios Prediction System 72

5.6 Dataset Preparation and Ground Truth Selection 73
5.6.1 Dataset Analysis and Bias 74
5.6.2 Data Preprocessing 74
5.6.3 Implementation Setup and Evaluation Metrics 74
5.7 Evaluation of the AI-2 Model 76
5.7.1 Frame-based (Time-Independent) Evaluation of the model 78
5.7.2 Statistical (Time-Dependent) Parameters 78
5.7.3 Results of Time-Independent vs Time-Dependent Datasets 79
5.8 Comparison of the AI-2 model with State-of-the-Art methods 81
5.9 Conclusions 82
Chapter 6. Novel ADL-Conditioned Cascaded Transfer Learning-based Baby Fall Risk Prediction Model 84
6.1 Introduction 84
6.2 Brief Overview of the approaches used in the model 85
6.2.1 Pose Estimation Approach 85
6.2.2 Temporal Model 86
6.2.3 Attention Mechanism 87
6.2.4 Rule-based Decision System 87
6.3 Novel ADL-Conditioned Cascaded Transfer Learning-Based Baby Fall Risk Prediction (BFRP) Model Architectural Summary 88
6.4 ADL-Conditioned Cascaded Transfer Learning -Based Hybrid BFRP Model Algorithm .. 90
6.5 Data Collection and Ground Truth Selection 92
6.6 Assessment of the Created Dual-Trigger Hybrid AI-3 Model 95
6.7 Comparison of Previous Studies and the Proposed Method 101
6.8 Conclusions 102
Chapter 7. Conclusions and Future Directions 103
7.1 Introduction 103
7.2 Conclusions 103
7.3 Future Directions 105
References 109

List of Abbreviations 116

Appendices: 121

A. Model Development Analysis 121

A.1. AI-2 Model – Unsafe Scenarios Prediction using Temporal Coefficients 121

A.2 Predict Video Segments using Model Probabilities 123

B. Results and Evaluation 124

B.1. AI-2 Model: Attention Map Visualization over Test Datasets 124

List of Figures

Figure 1.1: AI and ML advancement in baby healthcare applications....2
Figure.1.2: Rationale for motivating the baby-safety monitoring system....3
Figure 1.3 AI: Transforming the crib into high-tech without human intervention....6
Figure 2.1: Classification of Baby Monitoring System emphasizing our research focus. 13
Figure 2.2: Risk Factors of Fall 14
Figure 2.3: Classification of AI-Based Smart Devices 27
Figure 2.4: Categorization of Cognitive Devices in Baby Monitoring Systems....28
Figure 2.5: Smart Device Analytics....28
Figure 3.1: Research Methodology Procedure....35
Figure 3.2: Context of Research 36
Figure 3.3: Framework of the Proposed Research Approach 37
Figure 3.4: The AI-3 Model's simulation technique using Python code. 39
Figure 3.5: Python code output in a terminal window 40
Figure 3.6: Graphical analysis of unsafe frames using Python Code. 41
Figure 3.7: Graphical analysis of Risk Timeline using Python Code. 41
Figure 4.1. Hierarchical overview of Supervised Learning 46
Figure 4.2: Architectural overview of the AI-1 model 47
Figure 4.3: Employed system methodology 48
Figure 4.4. (a). The original frame was extracted from the climbing the fence video clip. (b). Extracted frame with 33 key points. (c). The frame shows 33 key points with the background subtracted. 49
Figure 5.1 ADLs Parameter Selection 60
Figure 5.2 Traditional CNN (TCNN) architectural diagram 65
Figure 5.3: Architectural diagram of Benchmark CNN with BN, Global Average Mapping, Regularization, Dropouts, and SGD optimizer (BCNN-BGRDOS)....65
Figure 5.4 (a) Training and Validation Loss of the TCNN Model (b) Training and Validation Accuracy of the TCNN Model....67
Figure 5.5: (a)BCNN-BGRDOS Training and Validation Accuracy (b) BCNN-BGRDOS Training and Validation Loss....67
Figure 5.6: General structure of CNN-Attention-based BUSPS....68
Figure 5.7: Methodology of SSCNN outlining the entire system....69
Figure 5.8: A detailed illustration of the SSA mechanism....70
Figure 5.9: Proposed model SSCNN architecture....71
Figure 5.10. Activity Daily Living (ADLs): (a) Unsafe (b) Safe....73

Figure 5.11: Graph showing the improvement in accuracy and loss against the number of Epochs (a) Accuracy Vs Epochs (b). Loss Vs Epochs 77
Figure 5.12: Confusion matrix 77
Figure 5.13: Depicting the scenarios frame-by-frame and alert accordingly 78
Figure 6.1 Definition of keypoints in MediaPipe Pose [122] 85
Figure 6.2: Working of GRU 86
Figure 6.4: Flowchart showing the working of the PE-GRU-SSA 92
Figure 6.5: Flowchart showing the mechanism of Dual-Trigger Alert Logic 93
Figure 6.6: Confusion matrix 97
Figure 6.7 (a) Confusion matrix for the Safe class: One vs. Rest 97
Figure 6.7 (b) Confusion matrix for the Near-Fence class: One vs. Rest 98
Figure 6.7 (c) Confusion matrix for the High-Risk class: One vs. Rest 98
Figure 6.8: Graph showing Trigger-1 to Trigger-2 escalation 100
Figure 7.1: Integration of AI-3 Model with Sensor Technology and Smart App. 106
Figure 7.2: Combined Integration Structure of Future Work – Stage 2 107
Figure 7.3: Future Directions - Baby Fall Risk Prediction Model using Transfer Learning 107

List of Tables

Table 2.1: Comparison of the previous research work based on novelty and techniques 29
Table 4.1: Dataset for Activity Distribution 52
Table 4.2: Signs Of Dangerous Activities with Labels While the Baby is in the Crib 52
Table 4.3: Specifications of the Baby Pose Dataset 52
Table 4.4: Comparative analysis of F1-score of keypoint-based classifiers in FFCV under ADLs 54
Table 4.5: Precision Comparison of Keypoint-Based Classifiers in FFCV Under ADLs 54
Table 4.6: Recall Comparison of Keypoint-Based Classifiers in Five-Fold Cross-Validation Under.. 55
Table 4.7: Comparative Study for the Average Performance Of 6 ADLS 55
Table 4.8: Comparison of F1-Score for the ADL-Based Outcome 55
Table 4.9: Comparison of Average Accuracy for the ADL-Based Outcome 55
Table 5.1: Parameter Configuration for the Implementation of the Model 66
Table 5.2: Comparative Analysis of Training and Validation of the Models 67
Table 5.3: Model performance: A comparative look 68
Table 5.4: Hyperparameter settings for the proposed model 75
Table 5.5: Fall Risk Prediction Metrics 76
Table 5.6: Classification of SSCNN Vs BCNN in the prediction of Unsafe Scenarios 77
Table 5.7: Comparison of SSCNN Vs BCNN in the prediction of Unsafe Scenarios 80
Table 5.8: Comparison of SSCNN with state-of-the-art methods for fall prediction 81
Table 6.1: Summary of the Statistics of the customized benchmark dataset 94
Table 6.2: Optimized Hyperparameter Settings 95
Table 6.3: Model Convergence with epochs 96
Table 6.4: Classification Performance of the Spatial-Temporal GRU Model on the Testing Data 96
Table 6.5: Comparison of metrics of our current proposed method with AI-1 and AI-2 99
Table 6.6: Comparison of unsafe event detection using SSCNN and the proposed model 99
Table 6.7: Risk Density and Segment-level ADL Statistics 100
Table 6.8: Summary of the comparison of AI-based studies for baby fall prediction in the crib 101

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 has it used artificial intelligence tools or generative artificial intelligence tools (unless it is clearly stated, and referenced, along with the purpose of use), 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.

27/10/2026

_____ _____

Signature Date

Co-Authored Works

Peer-Reviewed Journals

1. **N. Khan**, D. Z. Al-Hamid, P. H. J. Chong, F. Mirza, and A. Kumar, "Systematic Review of Sleep Monitoring Systems for Babies: Research Issues, Current Status, and Future Challenges," *Health & Social Care in the Community*, vol. 2024, Article ID 5510164, 11 pages, 2024, doi: 10.1155/2024/5510164.
2. **N. Khan**, P. H. J. Chong, and F. Mirza, "Baby Unsafe Scenarios Prediction Using Soft Spatial CNN and Statistical Parameters", Intelligent Sensors, MDPI, *Under revision*.
3. **N. Khan,** P. H. J. Chong, F. Mirza, and A. Kumar, "Pose-Based Spatio-Temporal Dual-Trigger Model for Baby Crib Fall-Risk Monitoring and False Alarm Reduction", IEEE Transactions on AI, *Under review.*

Peer-Reviewed Conferences

1. N. Khan, "Baby in a Crib: Fall or Prevention," presented at the Auckland Univ. of Technology. Postgraduate Research Symp., Auckland, New Zealand, Sep. 8, 2023.
2. **N. Khan**, D. Z. Al-Hamid, P. H. J. Chong, and F. Mirza, "Fall Prediction and Prevention: Smart Onesie," in *Proc. International Conference on Future Communication and Networks (FCN)*, IEEE, Dec. 2023, pp. 1–5.
3. **N. Khan**, P. H. J. Chong, and F. Mirza, "Comparison of Four Classifiers Using Keypoints Technique for Babypose in-Crib," in *Proc. IEEE 4th Int. Conf. Micro/Nano Sensors for AI, Healthcare, and Robotics (NSENS)*, Kuala Lumpur, Malaysia, 2025, pp. 9–14, doi: 10.1109/NSENS65807.2025.11428140.
4. **N. Khan**, P. H. J. Chong, F. Mirza, and D. Z. Al-Hamid, "Baby Safety Monitoring: Improving the CNN Model Using Batch Normalization and SGD," in *Proc. 10th Int. Conf. Commun., Image, Signal Process. (CCISP)*, IEEE, Nov. 2025, pp. 40–45, doi: 10.1109/CCISP67522.2025.11282124.
5. **N. Khan,** P. H. J. Chong, and F. Mirza, "Baby Fall Prediction System using GRU-SSA", International Conference on Artificial Intelligence and Technology (ICAIT 2026), Auckland, New Zealand. *Accepted for publication*.

Artificial Intelligence Declaration

Artificial Intelligence Guidelines:

Artificial Intelligence (AI) tools, including generative AI tools approved for use by AUT (AI Tool Register), can be used as learning tools when preparing to write your research proposal or research component. AI tools also have capability to assist with data modelling, data analysis and data visualisation. However, the research component you submit for examination must be substantively your own work. AI tools cannot solely be used to generate content when writing or creating an artwork/artefact, as this constitutes plagiarism. The formal guidelines for appropriate use of AI on the AI Hub must be followed. Also refer to the Using Artificial Intelligence in your Research: Self-Assessments, Checklists, and Sample Text, for examples and sample text.

If you do plan to use AI tools, this declaration should be submitted with the initial research proposal, Confirmation of Candidature, and when you submit your research component for examination. You must complete the template below for each AI output used detailing the following:

- **Chapters where AI was used in any way**
- **Purpose of AI Use**
- **AI Tool(s) Used**
- **Prompts Used**
- **Post-AI Processing Methods**

Specify the chapter number(s) in your research proposal/research component where AI was used.
Chapter number(s): **1,2, 4,5,6**
Briefly explain why the AI tool was used (e.g., idea generation, summarization, data analysis, image creation, etc.).
Purpose of AI Use: Paraphrase the statements to reduce plagiarism and to generate a few test videos (synthetic data)
Clearly state the AI tools used.
AI Tool(s) Used: **Quillbot, Gemini**
Prompts or queries entered into the AI tool (only the instructive portion of the prompt is required). Include all relevant variations.
Prompts Used: a combination of ADLs
A summary or description of the AI-generated output. (Optional: attach output in an appendix.)
Output Received: ---
Describe how the AI output was used, edited, or transformed.
Post-AI Processing Methods: **Sentences are rewritten.** **The generated video was treated as synthetic data and used to test the model as an unseen test dataset.**

You may use this free text box to explain the intent for the use of the AI generated content. *For example, rather than stating what prompt(s) were used, you may articulate the purpose for which the prompt was used (supported by examples) here:*

STUDENT DECLARATION			
By signing you are confirming that the AI use stated in the table(s) above are accurate and follow AUT's recommended guidelines detailed in the AI Hub and Postgraduate Handbook.			
Student Name:	**Neda Khan**	**Signature:**	**Date:** **25/04/2026**

Acknowledgments

"A PhD is not a measure of intelligence; it's a measure of perseverance." – Jordan Peterson. "Everything is theoretically impossible, until it is done" – Robert A. Heinlein. Endurance and perseverance are the keys to a PhD candidate completing the journey. In the meantime, motivation, inspiration, guidance, and support from the supervisors, families, and friends help make this journey easier.

I would like to express my sincere gratitude and deep appreciation to Professor Peter Chong, Associate Head of School (Research) at the School of Engineering, Computer, and Mathematical Sciences, Auckland University of Technology, for his indispensable supervision, encouragement, and insightful advice throughout my PhD journey. I am extremely thankful for the remarkable and exceptional advice and experiences he has provided throughout this journey. I greatly appreciate his profound understanding, scientific knowledge, and passion for engineering and research, which have been a great source of inspiration for me, enriching my research skills and elevating my enthusiasm to learn new skills and gain practical experience. Additionally, I started my PhD as an overseas student during COVID-19, and many struggles and failures entered my life, preventing me from continuing my PhD. Since then, his continual, unwavering support has been invaluable; it has not let me down and has motivated me to continue my PhD. It has been a great privilege to be accompanied on this journey by such a great supervisor.

I would also like to express my thanks to my secondary supervisor, Dr. Farhaan Mirza, from the School of Engineering, Computer, and Mathematical Sciences, Auckland University of Technology, for his insightful comments, support, and cooperation. I would like to extend my thanks to Duaa Zuhair Al-Hamid for mentoring, supporting, and providing valuable feedback throughout various stages of my research. I am deeply grateful to Deepika Mohan for her commendable support and help.

I want to express my gratitude to the Postgraduate Research Programs office staff for their exceptional assistance and support. The AUT Doctoral Scholarship, the Summer Research Award from the School of ECMS at AUT, and the Postgraduate Research Assistantship Award from the Faculty of Design and Creative Technologies at AUT all helped pay for this research and provided resources and staff.

I genuinely thank my family and friends for their consistent support, unconditional love, and faith in me, all of which have been a continual source of strength during this path. A big thanks to my father, brother, sister, and my in-laws who stood by me in all ups and downs and continually motivated me towards my success. A special thanks to my husband, who stood by me like a pillar since the beginning of this journey and made many inescapable compromises and sacrifices to let me achieve this goal, and to my loving daughter, Elaina Khan, whose sensible and mature behaviour greatly contributed to completing this research. Finally, I would like to extend my thanks to myself, Neda Khan, for completing this research, balancing family and work, being a good researcher, and being a sincere mother. I thank everyone who helped me finish this thesis. Thank you for your indispensable assistance.

Ethics Approval

In this research, all the data were generated from videos collected from online video-sharing platforms, such as YouTube, Dailymotion, and TikTok. All the models use only the data provided by these platforms. AI-2 and AI-3 have been tested using videos specifically generated using the AI tool Gemini.

Chapter 1. Research Motivation, Focus, and Thesis Structure

1.1. Introduction

This chapter provides a comprehensive overview of hazardous crib scenarios and infant falls. It includes hospitalization statistics associated with these events, along with the primary strategies currently employed to mitigate them. In this chapter, the integration of Artificial Intelligence (AI) into fall prediction is presented as a key focus, outlining how advanced algorithms learn from in-crib baby datasets. This enables proactive interventions and underpins customized fall-prevention strategies. The research gap, research questions, and objectives are then articulated to provide a structured roadmap. Thus, ensuring that the study's aims lead to meaningful outcomes. In addition, the scope and boundaries of the research are clearly defined. To orient the reader, the chapter also outlines the remaining thesis structure and explains how each subsequent chapter contributes to the central research problem. The chapter concludes by summarizing the key issues discussed and introducing the next chapter, which provides an in-depth review of the relevant literature.

1.2. Advances in AI/ML for Healthcare Applications

Artificial intelligence (AI) and machine learning (ML) are rapidly advancing across various domains, particularly in healthcare [1], including radiology, cardiology, oncology, public health, and hospital operations. In these domains, they support early diagnosis, risk prediction, and personalized treatment planning. In healthcare, deep learning and computer vision models are increasingly utilized for safety-critical monitoring tasks, including human fall detection using wearable sensors [2], vision-based gait analysis [3], and real-time video surveillance [4]. A new AI tool is being rolled out across the NHS that can predict a patient's risk of falling with 97% accuracy. This may prevent as many as 2000 falls and hospital admissions each day in England [5].

Recent work has begun specifically to help infants and toddlers explore AI-driven systems. This research work utilizes object detection, pose estimation, and spatial reasoning to monitor children in home or crib environments. They detect hazardous postures and trigger alarms when a fall or high-risk configuration is detected. These developments provide a strong foundation and motivation for AI-based frameworks for baby fall detection and prediction. These frameworks leverage keypoint-based pose estimation and temporal modelling to anticipate unsafe states before a fall occurs. One promising approach to predicting and preventing baby fall risk based on daily activities is the use of AI and the Internet of Things (IoT). These techniques are well-suited to both Pose Estimation (PE) and Healthcare Monitoring (HM) for infants in crib environments. In these techniques, they identify the baby's posture and compare the current configuration with predefined high-risk Activities of Daily Living (HR-ADLs) over a given time window to infer future outcomes. In this context, prediction refers to the system's ability to collect data and adapt to its environment. By leveraging machine learning methods, the system

automatically improves its performance through experience rather than explicit rule-based programming [6]. Nowadays, the availability of data from online and sensor-based sources is increasing rapidly. AI systems that integrate image processing with machine learning algorithms have become increasingly powerful across domains, particularly in healthcare. In baby monitoring, such systems can continuously observe a baby's activities and capture and store multimodal data. Further, process these signals to infer and predict aspects of the child's health status and safety [7].

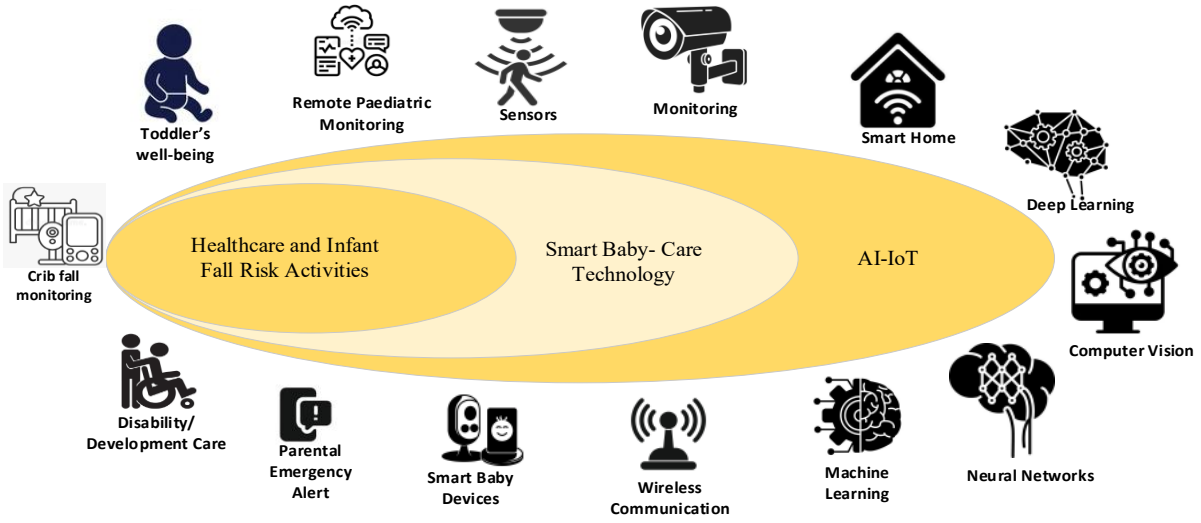


Figure 1.1: AI and ML advancement in baby healthcare applications

1.3. Rationale and Significance of Baby Safety in-Crib and Fall Prediction System

Working mothers have made up approximately one-third (32%) of all employed women in recent decades. The World Bank reports that in 2020, the employment-to-population ratio for women aged 15 and over was 45.92% globally. It was 65.25% in New Zealand. Higher ratios for this metric are found in several emerging nations, including Iceland (68.06%), Paraguay (56.93%), and Nepal (75.70%). [8]. Mothers with children under the age of 18 made up 72.3% of the labour force in 2019. According to the Bureau of Labour Statistics, the rate was 77.6% for single, widowed, or divorced mothers [9]. Two-thirds of the 23.5 million working women with children under the age of 18 were full-time employees, according to the American Community Survey conducted by the U.S. Census Bureau. [10]. Since 1990, significant progress has been made worldwide in lowering childhood mortality, according to the World Health Organization (WHO). In 1990, there were 12.6 million under-5 fatalities globally; by 2020, that figure had dropped to 5 million. From 93 fatalities per 1000 live births in 1990 to 37 in 2020, the global under-5 mortality rate has decreased by 60%. This translates to 1 in 11 children dying before turning 5 in 1990 as compared to 1 in 27 in 2020. [11]. The New Zealand Ministry of Health research states that the rise in COVID-19 cases, which has killed 497 people in the past two years and affected nearly the whole world, has increased the threats to the safety of babies. [12]. It would not be a smart idea to build

childcare centers because of this. Furthermore, if the parents rely on a nanny who does not maintain proper hygiene for the infant, the risk increases. Additionally, grandparents should not be caregivers. Figure 1.2 shows the different motivations for developing a baby safety monitoring system. Additionally, children between the ages of 0 and 4 are especially vulnerable to falls, including male and Māori children. They are more likely to be hospitalized for falls than for other injury hazards. Around 40% of falls at the specified location occurred at home and remained the primary reason for non-fatal unintentional injuries until 2012 [13].



Figure. 1.2: Rationale for motivating the baby-safety monitoring system

Recently, the workforce has increased and is expected to continue growing [14], making it necessary for parents to fulfill their children's needs and provide a good life for them. In turn, an unsafe environment is created with no proper monitoring and surveillance of children at home. Caregivers, such as nannies, may be unable to provide full attention, and the accident can occur at any time. Occupancy in early childhood education is down by approximately 15% to 20% from its pre-COVID levels, thus affecting the affordability of these centres for parents [15]. On the other hand, the growing stage of children involves a range of activities and motor skills. Baby Safety has become a major concern in current research. Safety concerns span various age groups, including infants (0-11 months), babies (0-24 months), and toddlers (1-3years). It is most prevalent when babies aged 12-24 months have developed motor and cognitive skills, such as climbing, leaning, and jumping, which can lead to risky behaviours, including potential falls, in the absence of consistent supervision [16]. These activities, however, may turn dangerous, such as falls. In New Zealand, falls are the most common cause of injuries, particularly among children aged 0 to 4, who have varying developmental skills. Thus, AI and Deep Learning (DL) have emerged as leading technologies in the research domain of baby safety monitoring [17], particularly for preventing baby falls.

Recent research on monitoring baby safety [18] and predicting accidents [19] has advanced significantly through deep learning-based video analysis. For example, early accident prediction frameworks use spatiotemporal feature extraction with recurrent models and custom loss functions to predict the occurrence of dangerous events. Smart home monitoring systems also use action recognition models to sort baby activities and send alerts. Even with these improvements, there are still some big problems.

- First, most of the methods that are already out there use implicit feature learning from raw RGB frames or optical flow instead of explicit human position information. As a result, these models don't understand body posture, joint dynamics, or biomechanical markers very well. These are all things needed to accurately identify early signs of imbalance or fall risk.
- Second, most approaches lack spatial attention mechanisms; thus, the entire frame is processed in the same manner. This means that background noise and clutter impair the models, reducing their accuracy in real-world settings where irrelevant objects or motions can interfere with decision-making.
- Third, LSTM or ConvLSTM architectures are often used for temporal modelling. However, these models can be very complicated and have many parameters, which can make them hard to use in real time and to generalize, especially when trained on small datasets. There hasn't been enough research on GRU-based options in this area.
- Fourth, most modern systems use a one-stage alert system, which means that an alarm goes off right away when a prediction comes true. This method doesn't distinguish between gradual increases in risk and immediate threats, which can lead to more false alarms and reduced user trust.
- Finally, much of the work already done focuses on reactive detection rather than proactive prediction. This makes it harder for them to allow enough time for preventive action.

In the domain of baby safety, many studies focus on fall detection; however, fall-risk prediction is less prevalent in the research literature. In addition, activity-based fall-risk prediction in the crib setting remains limited. Thus, to overcome these challenges, a baby fall-risk prediction model should be developed with higher accuracy, a lower false-alarm rate, and high interpretability.

1.4. Motivation for Advancing AI-driven Fall Prediction

The risk of fall-related injury is present from birth, emerging as soon as infants acquire sufficient mobility to roll or shift off elevated surfaces. As children progress to climbing and jumping, their exploratory behaviour, combined with immature balance and motor coordination, substantially increases the risk of falls. This study focuses on infants aged 12–24 months in the crib environment,

where typical hazardous behaviours include leaning out of the crib on the rails, attempting to climb over the fences, or reaching toward objects outside the crib perimeter. Under such conditions, infants may slip, trip, lose their footing, or topple over the crib boundary.

Although minor falls are common and often benign during early development, a subset results in clinically significant outcomes such as fractures, skull injury, lacerations, or other significant injuries. In severe cases, falls can lead to traumatic brain injury or spinal damage with potentially lifelong functional consequences. 49% of traumatic brain injuries in children are caused by falls. According to the Starship Foundation in New Zealand, more than 3 children are hospitalised every day from falls, including 110 children who fall from beds, cots, or cribs. Although the rate of hospitalisations for fall injuries has declined by about 19% over the past 10 years, the survey reveals that the hospitalisation rate for babies aged 12-24 months from 2013 to 2017 was the second-highest among fall injuries across age groups, at 1,282. Additionally, 31% of fall-related hospitalisations occur at home [20]. Continuous monitoring of infants' activities in the crib is therefore critical for mitigating the risk of falls. The central aim of this research is to enhance protection for babies, including male and Māori children, who are at elevated risk of falls linked to developmental stage, growth patterns, and motor skill maturation. Most prior work has focused on fall risk management, detection, and prediction for older adults, often complemented by balance or strength training programs to mitigate risk. In contrast, fall prevention for babies, especially within the crib environment, requires technology that operates continuously, with minimal human intervention, and that can distinguish normal exploratory movements from genuinely dangerous behaviours. Artificial Intelligence and Machine Learning are essential in achieving this goal in the following ways. Large video datasets of baby activities can be used to train deep learning models to learn intricate spatiotemporal patterns. This allows more accurate prediction of unsafe scenarios and more reliable early warnings. Existing approaches typically rely on cameras and environmental sensors to capture movement patterns, transmitting these data to a processing unit where algorithms analyse posture, motion trajectories, and sudden changes to infer potential fall events. Once a high-risk event is detected, an alert can be automatically issued to caregivers, enabling rapid intervention and reducing response time [4]. Taken together, these studies demonstrate that existing work has primarily focused on behavioural monitoring, fall detection, and fall prediction. At the same time, neural-network-based ensemble methods and sensors for human action recognition (HAR) in elderly monitoring have demonstrated promising accuracy for safety-related applications [21]. Consequently, there is a clear need to emphasize activity-based fall-risk prediction and prevention in crib baby monitoring, using reliable, low-intervention technologies (as illustrated in Figure 1.2) that can reassure parents during both daytime and nighttime care.



Figure 1.3 AI: Transforming the crib into high-tech without human intervention

Fall-risk prevention research adopts a proactive approach by forecasting potential crib falls through continuous monitoring of high-risk activities of daily living (ADLs) in babies and issuing early warnings to caregivers or parents when elevated fall risk is detected. AI is playing a pivotal role in improving people's lives, and healthcare is a major area [22].

1.5. Research Objectives and Contributions

In this section, the objectives of our research are explained, along with the clarification of research gaps, research questions, and research contributions.

1.5.1 Research Objectives

In many similar research studies based on baby monitoring, baby fall detection, and prediction discussed in Section 1.4 and Chapter 2, it has been observed that the existing work has the following research gaps and fails to satisfy at least one of the criteria:

- Many of the developed models in the existing work are not suitable for crib scenarios.
- Some of the pretrained models used physical sensors to estimate the pose, and due to dislocation or misalignment of their deployment, are unable to accurately estimate the postures.
- Numerous studies have been conducted on the detection and prediction of falls among the elderly. Few people are paying attention to babies, especially when they are in the crib. Most of the current crib research focuses on health, particularly breathing and sleep issues in infants aged 0-11 months. Rare research targets babies aged 12-24 months within the crib setting and predicts fall risk based on behaviour.

The preceding background on babies' falls or safety from critical or dangerous scenarios, and crib-related falls in babies, together with existing work on fall detection and prediction, leads to the central research question that underpins this study:

- Due to the development of motor skills and the growth of babies, multiple activities develop at this age. Thus, because babies in the crib engage in different Activities of Daily Living, such as sitting, sleeping, standing, and climbing, it would be difficult to identify dangerous or unsafe scenarios. Therefore, how can we identify baby fall risk scenarios in the crib?

- How can we predict unsafe scenarios and ADLs using pose estimation and deep learning models?

- Due to continuous monitoring of behaviour or activities, how can we study and predict the risk of a baby falling out of the crib and give early warning?

Thus, this research aims to identify key fall-risk factors critical to improving prediction performance. This study introduces a Novel ADL Conditioned Transfer Learning-based Hybrid Network model for baby-in-crib fall prediction, designed to forecast future falls by continuously analyzing spatial features for specific areas, scenarios, and postures, and time-domain features for continuously monitoring ADLs, changes in activities, and movements around the crib using AI.

The following objectives are formulated to achieve this research aim:

- To develop a deep understanding of how AI technology can be used to enhance and predict the risk of falls in babies, particularly for identifying unsafe baby-fall scenarios.

- To investigate deep learning models for in-crib baby safety across diverse scenarios and risk-related activities.

- To build a robust model that surpasses existing approaches, this study exploits joint coordinate values, spatial, and time domain features from videos while jointly leveraging a two-stage alert mechanism.

1.5.2 Research Contributions

To address the research gaps outlined above, this study introduces an innovative framework for predicting the risk of baby falls. This framework incorporates pose-based representations, attention mechanisms, efficient temporal modelling, and an intelligent alert system. The proposed method uses explicit pose estimation and keypoint coordinates to get detailed information about the skeleton and joints. This allows the model to learn biomechanically significant features, making it easier to detect

and understand subtle posture changes associated with dangerous situations. Second, a spatial attention mechanism is added to help the model focus on key areas, such as the baby's body and the surrounding risk zones. This reduces the impact of unimportant background information, making the system more stable and accurate. Third, the framework uses a Gated Recurrent Unit (GRU)-based temporal model to quickly capture motion dynamics and temporal dependencies. Compared to traditional LSTM-based methods, GRUs are easier to compute and generalize better. Thus, making them better suited for real-time applications with limited data. Fourth, a two-stage alert system has been implemented to make it easier to use in real-world settings. The first stage issues an early warning when it detects potentially dangerous posture patterns, and the second stage issues a critical alert when it detects a high-risk or imminent fall. This hierarchical alert system reduces false positives and provides caregivers with useful information at different levels of risk. Lastly, the proposed system changes the focus from passive monitoring to proactive risk prediction. This enables earlier intervention and improved safety outcomes. Hence, improves baby fall risk prediction systems by combining pose features, attention-driven learning, and efficient temporal modelling.

1.6. Scope of the Study

The significance of this research lies in its ability to predict unsafe scenarios involving babies in cribs. Additionally, preventing risk of falls before they occur through real-time monitoring by intelligent devices. This research focuses on the 6- to 24-month age range, with particular emphasis on the 12- to 24-month period. The data consist of videos of babies performing ADLs in their cribs, collected from online platforms including YouTube Shorts, YouTube, TikTok, and DailyMotions. These ADLs are classified as safe or unsafe, particularly those that pose a fall risk, as identified by two different AI Models. These two models employ distinct classification approaches to estimate the probability of future falls. The first model, AI-1, employs pose estimation techniques that use joint coordinates as keypoints. With these estimated poses, the model can identify dangerous or unsafe poses and categorize the defined ADLs into safe or unsafe outcomes. Key points enable the model to accurately identify postures by providing joint coordinate values, which are highly beneficial for addressing baby safety.

However, the second model, AI-2, employs a deep neural network (DNN), the Soft Spatial Convolutional Neural Network (SSCNN), which combines deep learning models. It comprises a 2D Convolutional Neural Network (CNN) built by stacking multiple convolutional, pooling, and fully connected layers to learn hierarchical feature representations from data. Then, a Soft Spatial Attention mechanism is used to help the CNN reweight its features by learning a continuous weight map over spatial locations in each feature map. During pre-training, real-time data augmentation is applied to training images. This augmentation process is conducted dynamically during training, ensuring the model is exposed to diverse variations of the input data at each epoch. This thereby enhances generalization and reduces the risk of overfitting. The fine-tuning stage further adapts the pretrained

model via supervised learning to maximize classification accuracy. Leveraging both pretraining and fine-tuning enables AI-2 to capture complex nonlinear structure in the data. This makes it an effective model for discriminating unsafe outcomes. The model was tested on long-duration unseen data using temporal features extracted from statistical parameters. Further, these features, combined with spatial features, demonstrate their efficiency and effectiveness.

Both frameworks are proving effective on their respective datasets, indicating the flaws to address for the next upgraded model. AI-1 generates outcomes by classifying posture features estimated using the PE technique with four different classifiers. However, AI-2 generates outcomes by classifying posture features extracted from the SSCNN model using the Spatial Attention layer.

Thus, the third model is developed by leveraging the best of AI-1 and AI-2. Adding temporal modelling and a dual-trigger alert mechanism to predict baby falls early. Reducing false alarms by providing progressive alert mechanisms with warning and high-risk alerts. The proposed method acquires accurate skeleton and joint data from explicit posture estimates and keypoint coordinates. This helps the model learn biomechanically important qualities, making it easier to spot and understand subtle posture variations associated with dangerous situations. A spatial attention mechanism is included to help the model focus on important regions, such as the baby's body and surrounding dangerous areas. This makes the system more stable and accurate by reducing the effect of irrelevant background information. The framework uses a Gated Recurrent Unit (GRU)-based temporal model to quickly capture how motion evolves over time and how different events are related. GRUs are easier to compute and apply to new situations than regular LSTM-based algorithms, making them preferable for real-time use. A two-stage alert mechanism has been implemented to make it easier to utilize in real life. The first stage issues an early warning when it detects potentially risky posture patterns. The second stage issues a critical alert when it detects a high-risk posture or an impending fall. This hierarchical alarm system avoids false positives and provides carers with meaningful information at varying levels of risk.

1.7. Thesis Structure

Chapter **1** presents an overview of the fall-risk problem in babies aged 6–24 months and motivates the development of an AI-driven fall-risk prediction system for crib environments. It surveys the roles of image processing, artificial intelligence, and deep learning in healthcare. The specific focus is on existing methods for detecting and predicting the risk of baby falls in cribs, as well as on engineering solutions to mitigate such events. Furthermore, the chapter articulates the research gap, motivation, problem statements, research objectives, and research contributions, thereby framing the scientific novelty and contribution of this work.

Chapter 2 presents a critical review of the literature relevant to the key dimensions of this research. It first examines the factors contributing to baby health monitoring for babies aged 6–24 months. This is

followed by a detailed analysis of the subcategories and prior work across two categories: with a crib and without a crib. This is further subcategorized into baby health & safety monitoring, baby fall detection, and baby fall prediction using various techniques. Later, the subsections include only sensors or IoT, only AI/ML, and sensor-AI fusion. The chapter then evaluates AI-enabled smart devices for safety and healthcare monitoring, outlining their advantages and limitations. Then, concludes with a comparison of the AI-based Fall Prediction System with the baby crib. At last, a discussion of how the proposed approach addresses their shortcomings through the novel contributions of this study.

Chapter **3** outlines the research methodology, presenting a step-by-step framework that encompasses data processing, model development, and the tools used for system design and experimental validation. It examines the capabilities of various software and computational resources for exploring and implementing an AI-based system to predict the risk of baby falls. In particular, the use of the Python programming environment and the Jupyter Notebook IDE is discussed comprehensively.

Chapter **4** highlights the significance of joint coordinate parameters, i.e., keypoints, in the Pose Estimation–based AI-1 model, which operates solely on video frames. It provides a comprehensive description of the proposed Pose Estimation–based Baby Fall Prediction and Prevention System (AI-1), detailing its conceptual foundation and core design components. This includes the selected age group, crib environment, joint coordinate extraction, keypoint-based fall prediction mechanism, and safe/unsafe activity classification. The chapter further elaborates on data collection procedures and model evaluation strategies. Activities of Daily Living (ADLs) are annotated as safe or unsafe to serve as ground truth for validating the model's predictive capability, and performance is compared across four distinct classifiers. From a system performance standpoint, the chapter emphasizes the design, implementation, and assessment of the AI-1 model for unsafe scenarios (potential fall) classification in babies within crib settings.

Chapter **5** introduces the significance of the spatial parameters used in developing the AI-2 model. The proposed model, a CNN + Attention mechanism-based Baby Unsafe Scenarios Prediction System, is described in detail. This chapter discusses the rationale behind the proposed approach's core components. This chapter includes improving the CNN model, the SSA mechanism, designing an Unsafe Scenarios prediction model, and testing the model using statistical parameters that leverage temporal features. This chapter also elaborates on data collection procedures and model evaluation strategies. Activities of Daily Living (ADLs), annotated as either safe or unsafe, serve as ground truth to validate the model's predictive capability. Thus, the findings are compared with existing works, particularly attention-based vision systems for detecting falls or monitoring babies. From a system performance standpoint, the chapter emphasises the design, implementation, and assessment of the AI-1 model, using statistical parameters to classify unsafe scenarios, including fall-risk behaviour in babies in crib settings.

Chapter **6** presents the network modelling of the novel mutual AI-based baby fall-risk prediction model. The technique of keypoint utilization in the temporal domain, dual-trigger logic for alerting, and a detailed summary of the final developed system are discussed. This chapter also discusses the data collection mechanism, how the final model is evaluated against ground truth, and how the dual-trigger logic helped reduce false alarms. A comparison between the current and existing work is highlighted.

Chapter **7** presents a summary of the research findings and discusses the key limitations encountered during the model development phase. It further outlines the strategies and solutions implemented to address these challenges. The chapter concludes with an overview of potential advancements enabled by sensor fusion techniques, providing a direction for future research and system enhancements.

1.8. Chapter Summary

This chapter provides a comprehensive introduction to the risk of crib falls among babies aged 6–24 months, with particular emphasis on the 12–24-month group. Also, focused on fall-related injury as a multifactorial problem intensified by systemic constraints in healthcare. It highlights current challenges in fall prevention and underscores the need for more coordinated, proactive strategies. This can include using fall-risk prediction based on actions to support timely preventive interventions. Building on this rationale, the chapter motivates the development of a predictive model to identify potential crib falls. Additionally, using dual-trigger logic to capture different stages of fall risk through an advanced, data-driven approach tailored to babies is appreciated. The expanding role of AI and ML in healthcare is discussed as an enabling technology for such predictive systems, providing the methodological foundation for this work. Furthermore, the chapter delineates the research gap, research questions, and research objectives that guide the design of an accurate fall-risk prediction model. Moreover, articulates the study's novelty by clarifying how it differs from existing fall prediction methods. The chapter concludes by specifying the research scope and outlining the overall thesis structure.

Chapter 2. Literature Review

2.1 Introduction

The lack of real-time posture tracking, continuous monitoring, and timely data transmission exacerbates the difficulties of managing fall risk for babies in crib environments [23]. Across the baby monitoring, fall detection, and prediction studies reviewed in Section 1.5 and Chapter 2, the existing literature consistently fails to satisfy at least one key requirement,

- Many of the existing models reported in the literature are not adapted to, or validated within, constrained crib environments.
- In several approaches, pose is inferred from external sensors, where displacement or misalignment of the devices degrades posture estimation accuracy.
- The detection and prediction of falls in the elderly have been the subject of much research. Few people are paying attention to babies, especially when they are in the crib. Most current research focuses on health, specifically respiratory and sleep problems in infants aged 0-11 months. A rare study found that babies aged 12 to 24 months are at risk of falling out of a crib.

These challenges necessitate innovative designs for healthcare monitoring systems that can operate under constrained computational resources while dynamically responding to fall events. Consequently, this chapter undertakes a comprehensive review of the literature to elucidate the critical requirements for developing a flexible, minimally supervised framework for fall-risk prediction and prevention. It examines state-of-the-art technologies, such as AI-based, sensor-driven, and pose estimation (PE)-based predictive models, IoT-enabled devices, wearable and non-wearable platforms, data mining methods, and real-time analytics, within rapidly evolving smart-care environments. The literature review primarily concentrates on methodological approaches to fall-risk prediction. Particularly, the emphasis is on frameworks grounded in motion detection, behavioural modelling, action recognition, pose estimation, machine learning, deep learning, and related attention-based architectures. Different scenarios that lead to distinct approaches to monitoring health and predicting baby safety are highlighted, along with the challenges they pose for the baby. Thus, the literature indicates that machine learning and AI-based approaches can detect and predict baby fall risk based on activities with high accuracy, specificity, and sensitivity.

2.2 Baby Healthcare Monitoring Systems: Contributing Factors

Babies are sometimes left without parents but kept under surveillance either through nannies or monitoring systems. Babies are under surveillance for two main reasons: to address health issues and

to ensure safety. Thus, the baby monitoring system (BMS) is further classified into three categories: a baby health & safety monitoring system (BHSMS), a baby fall detection system (BFDS), and a baby fall prediction system (BFPS). Figure 2.1 illustrates the classification of baby monitoring systems and their subcategories based on the crib scenario. Thus, this classification diagram focuses on the different research, highlighting our research domain in green.

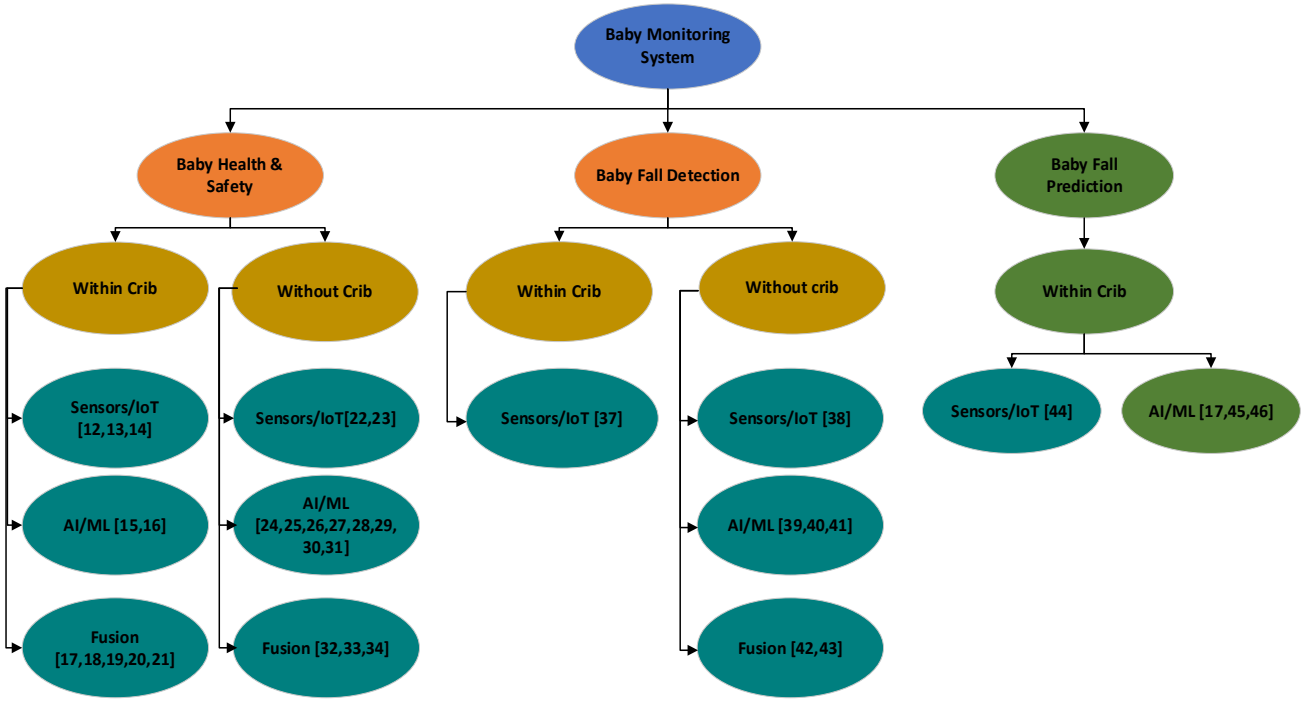


Figure 2.1: Classification of Baby Monitoring System emphasizing our research focus.

Different categories of falls result in various types of injuries in babies, such as falling out of the crib; for instance, an accidental crib fall can lead to skull fracture or traumatic brain injury (TBI) [24]. Various factors that lead to a baby's falls are separated into two groups: intrinsic and extrinsic. The intrinsic factors include age band, developmental stage, and overall mobility. Overall mobility can sometimes lead to typical movement patterns, such as a tendency to climb or grasp rails. Extrinsic factors include device factors (crib design, mattress height, camera viewpoints), environment layout (lighting, hard flooring, presence of sharp objects or furniture), and caregiver practices and supervision. Figure 2.2 elaborates on the two major types of risk factors that lead to babies falling out of their cribs. The most frequent cause of injuries in children who need hospitalization is falls. Falls account for half of all injuries in children under five. [25]. Babies can fall when they roll over and move around for the first time [26]. As they start to walk, children take more chances. They get inquisitive and eager to explore and climb. An important aspect of their growth is climbing. This may result in injuries from falls. [27]. They can move around, fall, and get into objects at various points over the first two years. Toddlers may attempt to climb but lack the coordination to respond appropriately to certain dangers.

So, the possibility of a dangerous fall or a tumble into a sharp edge might happen in almost any section of the home. [28].

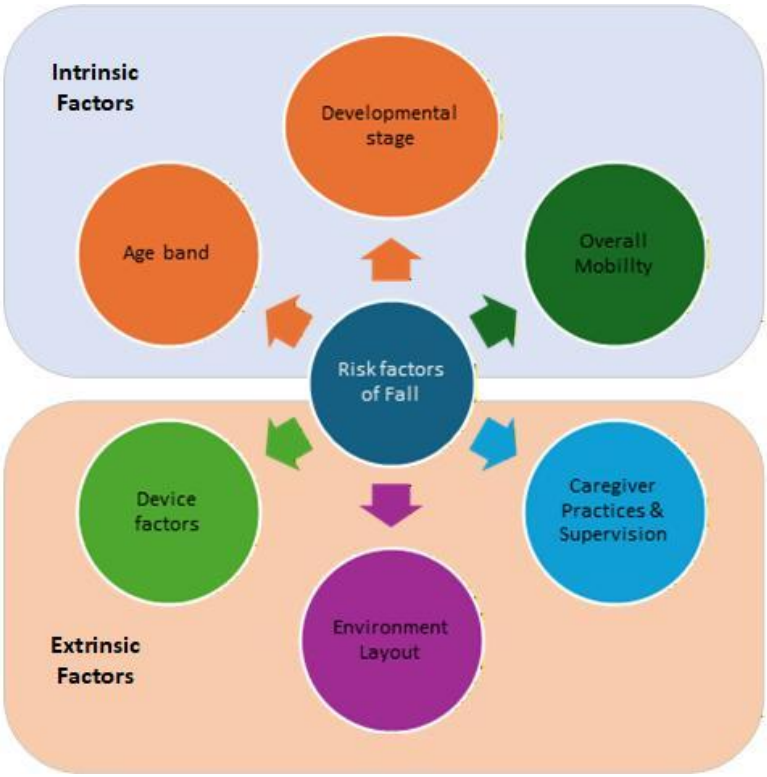


Figure 2.2: Risk Factors of Fall

2.3 Overview of Baby Monitoring for Health & Safety

This research primarily focuses on baby safety monitoring, including fall prevention and other safety measures. Baby safety is considered when the infant is sleeping, rolling over, crawling, or engaging in other activities. As a result, a lot of accidents involve babies and kids who are susceptible to health risks like suffocation, high fever, sharp objects, fear, staircases, open water zones, toxic substances, or any other hazardous objects. The literature study is categorized broadly into two scenarios: within the crib and without the crib, under sections 2.3.1 and 2.3.2, respectively. They are further classified into three broad technology areas: Sensors/IoT, AI/ML, and the Fusion of Sensors/IoT and AI/ML. The fall detection category is explained in Section 2.4, whereas fall prediction is covered in Section 2.5 for detailed discussion.

2.3.1 Within Crib

There are numerous studies examining different scenarios; however, this section primarily focuses on studies conducted in the crib setting. According to the New Zealand Consumer Protection [29], to prevent the baby from climbing out and falling, the crib must be deeper than 60 cm, measured from the

top of the mattress to the top rail. According to the U.S. Consumer Product Safety Commission, the minimum height of a crib side rail is 26 inches [30]. Although cribs are the safest places for kids to sleep, the risk varies by age. Babies between the ages of 0 and 11 months are at a high risk of SIDS [31] [32] because of suffocation, rollover, and an abundance of toys and bedding. Babies between the ages of 12 and 30 months are at great risk of falling because they are climbing, leaning over, or jumping close to the fence. Given that a previous study revealed that kids can climb out of cribs, raising the minimum rail height may be necessary. [33]. These dangers result in serious injuries and occasionally fatalities.

2.3.1.1 *Using sensors, hardware, and IoT*

An IoT-enabled smart crib was developed [34]to monitor a baby's physiological and environmental factors, such as temperature, humidity, wetness, and crying. The proposed system monitors and controls baby care by combining microcontrollers, Internet of Things sensors, and a mobile application. The system's essential component for data collection and communication with the mobile application is a Node MCU ESP32. The servo motor, DC fan, sound sensor, and DHT11 sensor all cooperate to monitor health indicators, control sleep patterns, and create the best possible sleeping environment. When a baby begins to cry while asleep, the swing mechanism helps quiet them. Therefore, the purpose of this study was to improve the comfort and well-being of infants in rural Somalia by utilising smart technologies that enable real-time monitoring. However, the system is purely sensor-based, which does not provide accurate results when sensors are misaligned, and sensors cannot perform advanced data analysis to provide a baby's sleep patterns and health trends, enabling more accurate monitoring and alerting parents or caregivers. The Tenshi Baby Crib [35], is an advanced crib system designed to replace existing baby monitors. The crib monitors a baby's surroundings and sleeping posture and can promptly alert parents when their child may be in danger, enabling them to take immediate action. Furthermore, parents simply need to place their infant in the crib to begin monitoring, as the sensor technology is effortlessly integrated into the cot. When paired with a mobile application, parents can easily assess their baby's sleep health and the baby's surroundings. All things considered, this crib can help newborns grow healthily and lower the risk of SIDS. In this study [36], an IoT-based smart baby cradle was designed to let parents watch and keep an eye on their children whether they are at home or at work, and it can recognise each infant's behaviour from any remote location in the world. The designed framework contains a camera, a temperature sensor, a gas sensor, and a sound sensor to gather necessary data and always ensure the babies' safety and well-being within the confines of the cradle.

2.3.1.2 Using AI and ML techniques

In place of sensors, a deep learning algorithm approach was suggested [37] to assess the infant's position and condition. The proposed approach reduced model weights by using MobileNetv2, a simpler CNN model architecture. Performance was enhanced by accelerating the deep learning method's processing

time to 20 frames per second using a Video Graphic Array (VGA)-equipped development board. The approach achieved 94% accuracy across a variety of scenarios, including infants in safe positions and in potentially dangerous positions (e.g., prone, encased in a barrier, or standing close to a fence). Thus, the infant's position and well-being can be accurately determined by the model with a misclassification rate of 6%. However, TinyML's present restrictions prevent it from processing images quickly. The integration of TinyML devices with high-FPS image processing may be made easier by advances in chip technology or specialized hardware such as VGA. CribNet, a novel vision-based infant-centric hazard detection system, was proposed in [38] to evaluate risks to in-crib safety posed by dangerous toys and blanket occlusions. This was done to address the wide-ranging and crucial issue of infant sleep safety. When estimating risks, CribNet considers the features and the vicinity of the items it detects around the babies. The first crib hazard detection (CribHD) dataset, comprising 1,620 images from infant-focused settings, was made publicly available to assess the system. Clutter, occlusion, different lighting conditions, and the presence or absence of infants are just a few of the many real-world issues that these images present. An unprecedented breakthrough in infant safety was made when the framework successfully segmented toys and blankets and accurately assessed dangers, achieving a mean average precision (mAP) of over 80%. CribNet's overall performance, including high accuracy in toy detection and occlusion prediction. This highlights the dataset's generalizability and utility for proactive infant hazard detection. These findings validate both the dataset and the approach as trustworthy resources for improving child safety by demonstrating CribNet's capacity to clearly identify in-crib vulnerabilities.

2.3.1.3 *Using integration of Sensors and AI/ML*

An AI- and IoT-based hazard prediction system was designed [39] to identify potential dangers to babies in the crib. An IoT hardware component and AI multimodal, multidimensional hazard prediction algorithms made up most of the system. The system detected the baby's real-time status and predicted potential dangers. Kicking the quilt, peeing the bed, having a fever, weeping, climbing over the crib, and rolling over are all ways to empower conventional baby cribs and turn them into intelligent ones. The basic idea behind this system is that infants' activity levels can be tracked in real time via mobile devices. The system utilizes the Real-time State Monitoring Algorithm in conjunction with ESP32-CAM and OV2640 cameras. Second, it can avoid suffocation by determining whether the baby has rolled over by using the Turn-over Detection Algorithm and an AONI high-definition camera (the brand "AONI" is special). The Turnover Detection algorithm used a Haar cascade classifier. It achieved face/nonface classification using integrated images, the AdaBoost algorithm, and cascade classifiers. Then, using an Arduino UNO and a variety of sensors, the Environmental Detection Algorithm identifies indicators such as distance, sound, temperature, and humidity. The three algorithms were combined to create a multimodal Hazard Prediction Algorithm that can both predict impending threats and thoroughly assess an infant's condition from multiple angles. A genius nest was designed to ensure

comprehensive infant safety during sleep [40] using a Convolutional Recurrent Neural Network (CRNN) with an attention mechanism and MediaPipe. Two experimentally verified modules for posture detection and audio analysis were combined with a proof-of-concept biomedical module for vital sign monitoring to create a low-cost edge-deployed AI-driven multimodal monitoring system. With a false-positive rate of 2.8%, the CRNN-based cry categorisation module attained 96.0% accuracy. SIDS prevention needs were directly addressed by the posture detection module, which used MediaPipe landmarks with a voting ensemble classifier, achieving 96.0% precision while maintaining a critically low 3.1% false-negative rate for hazardous poses. Additionally, the biomedical module used artificial data that simulated clinically significant defects to show architectural viability. The cry categorisation module ran asynchronously with an average response time of 3–4 seconds, while the posture and biomedical modules showed an average end-to-end latency of 231 ms and a power usage of 4.5 W when installed on a Raspberry Pi 4. Overall, the technology enables autonomous intervention capabilities while achieving a 60% cost savings when compared to market systems. Genius Nest is a major improvement over disjointed single-modality systems because it combines verified behavioural and physiological monitoring frameworks, providing a dependable, accessible, and real-time platform to improve child safety. AI and image processing were used to create an intelligent baby monitor that automatically identifies sleeping posture and sends notifications [41]. The monitoring system automatically identified the following: (a) whether the baby was sleeping on their stomach, which covered their face; (b) if the baby threw the blanket off; (c) if the baby was moving about a lot; and (d) whether the baby had woken up and had opened their eyes. Every time one or more of these circumstances occurred, the device generated warnings and sent notifications to the caregiver's smartphone. As a result, it was not necessary for the carers to check on the baby periodically. When their attention was needed, they were informed. The Jetson Nano microcontroller from NVIDIA was used in the device's development. Wi-Fi connectivity and a night vision camera were interfaced. Deep learning models for pose, face, and landmark detection were created using the microcontroller. Both the smartphone app and the monitoring gadget prototype were created and successfully tested in various settings. An automated infant monitoring system was developed [42] to ensure continuous, real-time observation of an infant’s behaviour using deep learning and IoT technologies. The system processes sensor data and detects critical conditions such as apnoea, seizures, and unfavourable environmental factors using advanced neural network architectures, including ResNet-152, MobileNet, VGG19, and AlexNet. It successfully identified respiratory abnormalities by examining breathing patterns and SpO2 values. Because of its 99.5% accuracy and lightweight architecture, MobileNet was selected for implementation. Real-time predictions on the edge device were made possible by converting the training model to TensorFlow Lite and embedding it on the ESP8266 microcontroller. When excessive vibrations were detected, the ESP8266 gathered sensor data in real time and either moved the cradle with a servo motor or sent an email, and Blynk alerted when a baby cried. In addition to improving infant monitoring reliability, this system's integration of AI, edge computing, and IoT provided parents

and carers with timely insights and alarms, ultimately promoting safer, more responsive childcare settings. This study proposed a smart baby-monitoring system that uses ML to identify emotions [43]. Using various sensors, the required monitoring features, such as room temperature and humidity, cry detection, and face detection, were monitored. The sensor data is sent to the Blynk server by controllers with an Internet connection. Using a machine learning model, the device can also identify emotions on the faces of registered newborns.

2.3.2 Without Crib

This section explains how the baby's safety can be considered if he/she is around at home. There are numerous studies that focused on babies' safety when they were not in the crib. There are so many other risks at home, such as electric shock, sharp objects, heated appliances, and falling on the floor or down the stairs [29]. Children may wander off while at home and suffer injuries like burns or falls from heights.

2.3.2.1 Using Sensors, hardware and IoT

A wearable child-safety device was developed [44] to make it easier for parents to find their kids. An SMS-enabled communication channel was created between a child's wearable and a parent operating on the GSM mobile network. The wearable device incorporates a GPS location sensor, a temperature sensor, and a UV sensor. When the parent texts the wearable IoT device with specific words such as "LOCATION", "TEMPERATURE", "UV", "SOS", "BUZZ", and so on, the wearable will respond with a text containing accurate measurements of the parameters, allowing the parent to receive immediate updates and assurance of their child's safety. This study [45] presented a multimodal baby monitoring system designed to bridge the accessibility gap for deaf parents. The system combined auditory analysis (300-600Hz frequency range) with visual (OLED) and haptic (vibration motor) notifications, attaining 98.1% cry identification accuracy, 92.4% specificity in rejecting non-baby noises, and <= 7.6% false alarms in controlled testing. A dual-stage architecture used an Arduino-based sound sensor (MAX9814 microphone) and Bluetooth 4.0 (HC-05) wireless transmission to provide real-time notification within a 10-meter range. Experimental validation across varied noise levels (30-70dB SNR) revealed consistent performance, with spectral analysis effectively distinguishing baby vocalisations from adult speech (80-300Hz).

2.3.2.2 Using AI and ML techniques

A new system, BabyPose [46], was developed that used an LSTM network to decode a baby's body language and learn skeletal keypoints to identify activities in films. 156 video clips from internet video-sharing platforms were used to construct a dataset of baby poses. Three distinct posture estimators—AlphaPose, KAPAO, and OpenPose—provided key points. Based on a posture-estimation model

trained on public videos, this neural network used the OpenPose pose estimator to predict baby behaviours with 99.83% accuracy and 0.0712 loss ratio. The DeepSORT algorithm was used to track babies in real time. As a result, baby safety awareness was raised, and real-time baby tracing was ensured. General Movement Analysis (GMA) was proposed [47] to conduct thorough movement analysis from video recordings using 2D and 3D pose estimation techniques for precise and effective analysis of baby's movement. Angle orientations at different projected joint positions, postural information, postural variability, movement velocity, movement variability, and left-right movement coordination are among the metrics that this approach computes. This served as a substitute for GMA in the early identification of cerebral palsy (CP). A collection of robust features designed to improve the model's interpretability was used to classify CP using a range of machine learning algorithms. The proposed method achieved classification accuracies of 92% and 97.37% on the MINI-RGBD and RVI-38 datasets, respectively. A deep learning approach was created [48], to categorize objects into two classes: child-safe and child-unsafe. This work provided a unique dataset, ChildSUn, comprising over 5350 photos across 10 categories of child-safe and child-unsafe items. The produced dataset was used to test several CNN architectures, including custom CNN, VGG16, VGG19, ResNet50, Inception-V3, EfficientNet-B0, and MobileNetV2. As a result, the EfficientNet-B0 model performed exceptionally well, obtaining an astounding 99.63% accuracy on the test dataset. The classification report confirmed its robustness, with near-perfect precision, recall, and F1 scores, and minimal false-positive rates. These results highlight EfficientNet-B0's promise as a practical solution for child-safety applications and encourage further research into deep learning methods to improve child protection worldwide. A computer-vision-based accident detection system was developed [49] for infants and toddlers to enhance their safety. The proposed system detects burns and cuts, two common types of incidents, using machine learning and closed-circuit television (CCTV). The proposed system tracks motions, records frames, detects children and hazardous objects, and calculates the risk. The technology reduces accidents, gives parents peace of mind, and provides a complete solution for child monitoring. The results show that detection accuracies are 87%, 84% and 75% for scissors, knife, and both items, respectively. The project's findings indicate that detecting accident risks achieves a satisfactory level of accuracy. However, when multiple dangerous objects are present in the field of view, accuracy declines slightly. A real-time evaluation system for fall-related brain injuries was developed. [50]. Every day, 200 RGB videos of toddlers aged 13-30 months playing near guardrails, furniture edges, and upside-down falls were used in this study. From them, 500 video clips were split into an 8:2 training and validation set. This research involved two phases to process the framework Phase I involves using Open Pose to process surveillance video to acquire joint coordinates. The spatial and temporal information of important places was then integrated using a 3D transform model and a long short-term memory (LSTM) network. Phase II involves developing a classification model to evaluate injuries, deriving the head acceleration, and incorporating it into the HIC calculation. An extra collection of 300 video clips depicting toddlers falling at home with their parents every day was produced in order to evaluate the

framework's efficacy. The classification accuracy was 96.67% in the experimental results. A system was proposed [51] that uses YOLO to detect objects and alert to child-safety hazards, continuously monitoring and detecting hazardous objects near a child in real time. The system achieved detection accuracies of 86-95% and response times of 18-30% across different hazard types, including sharp objects, toxic substances, open water zones, and staircases. However, there were certain problems, such as improving accuracy in complex or congested media, minimizing false warnings, and optimizing real-time processing to increase efficiency. An intelligent surveillance solution [52] leverages computer vision and deep learning to detect and mitigate potential threats to children in real time. Using the YOLOv8 object detection model, the system identified the children and hazardous objects around them, achieving 98.3% object recognition accuracy and 96.5% child recognition accuracy. Even though the system outperforms recent benchmarks in accuracy and successfully connects theoretical AI capabilities with real-world kid safety requirements, further research is needed on activity recognition and behaviour analysis to enable context-aware alarms. An AI-powered system was introduced [53] to detect fear in children by analyzing facial expressions and behavioural patterns. Using CNN and KNN, the model outperformed traditional methods, achieving 92.4% accuracy. Though the approach enabled faster intervention, reduced risk, and strengthened emergency response by adding a panic button that automatically notified emergency services, shared key details such as GPS location, and sent real-time images, the panic button is for the child to press, which is impractical for toddlers or babies. Real-world deployment is expected to demonstrate the system's robustness and reliability.

2.3.2.3 Using integration of Sensors and AI/ML

To improve the infant's welfare, this study used AI as an advanced technique in cry interpretation and monitoring [54]. This study explores the interpretation and monitoring of infant cries using artificial intelligence (AI) techniques based on sound recognition. The research aims to identify sounds indicative of distress and classify crying patterns from male and female infant screams. The study extracted features from MP3 and WAV audio data collected online using the Mel-Frequency Cepstral Coefficients (MFCC) approach. Using a variety of machine learning models, such as Random Forest and XGBoost, the method effectively captured the distinctive characteristics of each weeping sound. With accuracy scores of 94.5% and 94.2%, respectively, these models outperformed others. These results demonstrate how well these algorithms classify different infant screams. In this study, an AI-driven baby monitoring system was developed [55] to detect faces and cries using multiple models. To manage multi-threaded programs, the method was implemented on a Jetson Nano single-board computer. It then connected various other modules, including a camera, microphone, humidity and temperature sensors, and a Wi-Fi and Bluetooth module. TensorFlow and Haar Cascade face detection models were employed. A short-term Fourier Transformation (FT) called Mel-Frequency Cel Cepstrum (MFCC) transforms auditory information into a visual depiction of how people understand the audio. TensorFlow's conventional neural network and a CNN were used to evaluate the cry detection model's

accuracy. Additionally, a manual test using a real-time audio recording of a crying infant verified the test's accuracy. The CNN-created model outperformed its conventional neural network equivalent in terms of accuracy. This study [56] presented an intelligent infant monitoring system based on CNN analysis. Using YOLOv3-tiny to recognize the baby's body and face, this system assessed the baby's sleep quality. The infant monitoring system's performance was assessed through several experiments, which yielded 94.46% face detection and 86.35% body detection over 31000 images from the Electronics and Telecommunications Research Institute (ETRI) and a virtual dataset.

2.4 Overview of Baby Monitoring for Fall Detection

Baby falls are serious and a major concern, leading to a boom in research. Now, this baby fall may happen at home, around the park, at school, etc. This literature study focuses on baby falls at home, particularly. There should be a mechanism in place to identify falls and take appropriate action promptly to prevent significant injuries. If a fall occurs, it can happen on the floor, down the stairs, out of the crib, etc. AI has captured the attention of today's generation and is a leading technology in this era. This domain is playing a major role in healthcare applications [57]. The need for the smart healthcare system has grown recently due to significant advancements in contemporary technologies, particularly AI and ML. One of the main drivers of smart healthcare adoption is the rising birth rate in New Zealand. According to the New Zealand Statistics of births and deaths release, the rising fertility rate is 1.58 births per woman, up from 1.54 [58]. Numerous studies have demonstrated that AI benefits baby healthcare systems. AI and ML are integrated into baby healthcare systems for a range of applications, including fall detection, activity recognition, behaviour monitoring, vital monitoring, sleep monitoring, and safety monitoring. Although baby healthcare systems fall into several categories, this study focuses on how baby falls that can be divided into fall detection and fall prediction. Fall prediction is reviewed and detailed in Section 2.5; however, fall detection is reviewed under this section

2.4.1 Within Crib

2.4.1.1 *Using sensors, hardware, and IoT*

A smart monitoring system was developed [59], equipped with an ESP32 microcontroller and various sensors, including the APDS9960 for gesture recognition and the MPU6050 Multi-Axis Accelerometer for fall detection. The system was designed into three modules: Module 1 for infant movement detection, Module 2 for buzzer alerting, and Module 3 for notification via the Twilio App. The designed system aimed to constantly monitor the activities of a physically handicapped infant by detecting movement. The system has been implemented with alerts to keep a close watch and help relieve parents from the burden of physically supervising the infant. Thus, the system helps working parents to maintain a balance between home and their work life. However, this system lacks performance evaluation and cross-validation to support its real-world deployment.

2.4.2 Without Crib

2.4.2.1 Using Sensors, hardware and IoT

This study created an IoT-based child safety device [60] that comprises a GPS NEO-6M for precise position tracking and an MPU6050 sensor that improves environmental awareness by monitoring temperature changes and detecting falls. During testing, the device achieved 90% accuracy in identifying falls. This technology was created to continuously monitor a child's environment and identify potential hazards in real time.

2.4.2.2 *Using AI/ML techniques*

An OpenPifPaf-based camera-based method for estimating human pose [61] was employed to extract keypoints of toddlers; these keypoints are then passed to an LSTM model for classification. The LSTM-based neural network predicts four fall classifications: "Normal," "Fall Warning," "No Fall," and "Fall." This system trained the model using the UP-fall dataset. With a 93% F1-score, the model decreased false alarms and, consequently, the number of child fatalities. However, the designed system can be improved by predicting falls and avoiding fatal injuries rather than merely reducing them. A method was introduced [4] for extracting the toddler target and implementing a real-time falling alarm. This method used YOLOv8 to extract body features and further calculated head features using those keypoints. Later, the Generalized Hough Transform was used to determine the safe zone for a toddler, achieving an accuracy of 96.33%. However, the system cannot detect the features if the toddler is in bed and covered with a quilt. Error detection may still occur when toddlers confuse some toys and patterns. A technique comprising a background-subtraction module was presented [62] that detects the ROI, a tracking module that uses a Kalman filter to track the toddler's movement, and a decision module that uses a decision tree to determine the toddler's state. The model achieved 78.26% accuracy, 84.62% sensitivity, and 77.78% specificity, demonstrating the system’s strong ability to detect and track the toddler. Nevertheless, the method is limited to a single camera view and requires the addition of context-aware alerts for parents or caregivers.

2.4.2.3 *Using integration of Sensors and AI/ML*

To make research into the risk of fall-related injuries easier, a fall-motion database and a browser were created [63]. The authors' recognition method was used to identify falls from accumulated sensor data. Video clips of identified falls were then automatically retrieved from the recorded video streams. Fall motion analysis was then performed using the retrieved video segments. A computer vision (CV) method facilitated the collection of fall motion data in the database and automated fall motion analysis. By defining search conditions, such as child properties and specific fall circumstances, the corresponding database browser enables users to perform conditional searches of fall data. The creation of a children's real-world fall motion database, which might help reduce the risk of damage in everyday

situations, was new. To detect infant falls, a wearable sensor powered by a triboelectric nanogenerator (TENG) was proposed. [64]. This sensor has a copper foil electrode and a bridge-structured PDMS layer. By attaching the TENG sensor to the skin or joints of a human body or a baby model, fall status, frequency, impact intensity, and impact localization were determined. The sensor's minimum observable acceleration on human skin is 0.4 g, and its sensitivity is 2.6 V/g. The device predicts fall locations with over 94% accuracy when paired with artificial intelligence algorithms. Furthermore, the sensor exhibits outstanding stability and repeatability even following tens of thousands of cycles.

2.5 Technologies Used for Baby Fall Prediction Inside the Crib

Previous studies have shown that fall detection has been widely used for babies. Several studies propose various approaches to detecting when a baby falls out of the crib. Thus, various technologies are used in this domain, including AI/ML, sensors, wearable sensors, and integrated AI-and-sensor technologies. We have clearly stated each technology in the section below.

2.5.1 Sensors/IoT

This study used an innovative approach to remote infant and baby monitoring [65]. The subject's exact placement within the crib is determined by the technology. This method identifies the three activities the subject participates in and alerts the user if a hazardous location or activity is found. These alerts appeared on a mobile phone and on the display panel. A microprocessor-based device with an array of membrane switches installed on the top of the crib mattress, behind the mattress cover, was used in the procedure. The entire crib is intended to be covered by the array. The switches will determine the subject's location and generate a signal that a microprocessor-based device can use as input. The controller was chosen as a microprocessor-based integrated device for its reliability. Hardwiring an electronic circuit lowers the failure risk because it is not necessary. Furthermore, switch bounce prevention is a built-in feature. The Siemens Logo device was chosen for the task because it has a display panel (Logo TD) that can be connected to another part of the house. The four red switches at corners A, B, C, and D were chosen as the boundaries to illustrate the idea. When any of these switches were pressed, the subject reached a corner of the crib, and an alarm went off. The display turned red as a result. When the movement frequency was very low, the baby was either sleeping or resting; when it was medium, the baby was awake and moving. Though this work overcame the barrier to remotely forecasting falls with non-wearable sensor-based devices, software mechanisms are now preferred over hardware mechanisms, saving time and money by avoiding hardware design and sensor placement. Thus, AI dominates in this arena, offering a better, more effective method for predicting baby falls than sensor- or IoT-based systems.

2.5.2 Using AI and ML Techniques

The aim of this study was to prevent accidents [19], including occasional falls, falls from furniture, and bumping into objects, by predicting and preventing them. This study proposed a temporal Pyramid LSTM (TP-LSTM) to capture multi-scale temporal information and introduced a temporal attention mechanism to focus on important frames. The proposed network included a new Early Exponential Loss (EEL) to encourage early accident prediction and focused specifically on baby accident prediction, which is less explored than adult action recognition. This research involved a two-stream CNN, ResNet152 (pretrained on UCF101), that extracted action features from each frame and converted each frame into a feature vector (2048 to 1024). Thus, spatial information was compressed into a vector and no longer attended to spatial regions. After feature extraction, each frame yielded 1 feature vector, and the resulting sequence of vectors was fed into the TP-LSTM, where only temporal sequence information remained. The model then applied temporal attention, which decides which time steps (frames) are important without focusing on "where" in the image (say, head, torso, arms, legs, etc). The proposed network may select critical frames that have the greatest impact on the accident and capture significant video information at multiple temporal resolutions. Additionally, the proposed early exponential loss (EEL) function is included to improve prediction. According to the test data, the proposed network achieved an average precision of 61.13% and predicted accidents with 80% recall, 4.196 seconds before they occurred. The model outputted moderate precision of 53-61%, indicating that false alarms remain and are not ideal for real-world deployment. Additionally, this research study does not explicitly model pose using skeletons or keypoints, thereby limiting its interpretability. CNN-based features are extracted from single frames, but the spatial information is converted into feature vectors before being used to model temporal dependencies. Even though the model used a CNN, it was still considered temporal-only attention because the attention mechanism was applied only over time, not over spatial locations, such as pixels or body parts. The suggested attention mechanism operates only on temporal sequences and does not directly model the importance of spatial information within frames. So, the method can't account for space. Moreover, the model outputs a single probability for accident prediction, rather than implementing a progressive warning system. This resulted in no differentiation between prolonged standing in a safe state and a high-risk state that may lead to a fall. Users require graded alerts, as early, subtle risks are not always immediate dangers and may mislead parents and caregivers.

A study [18] was conducted to develop a smart baby-care system that predicts a baby's actions. A custom smart baby care dataset was developed for this system. A Conv2D LSTM layer and transfer learning were coupled to create an automated model. The pre-trained I3D model that was trained on the Kinetics dataset was used to extract features. This system required a smaller dataset and fewer computational resources. This study aimed to design a lightweight action recognition framework using transfer learning, combining I3D (Inflated 3D CNN) and ConvLSTM2D, and integrating it with an IoT-based smart home monitoring system. The developed framework focused on efficient action recognition for baby monitoring using pretrained features and ConvLSTM on a small dataset. Thus, the objective

was to monitor baby activities in a crib, detect unsafe actions, such as jumping and risky movements, and provide real-time alerts to caregivers. The model inferred posture and activity from spatiotemporal patterns in video frames and was trained with I3D + ConvLSTM. The I3D-ConVLSTM2D model achieved the best results, with 90% accuracy, 86% precision, 87% Recall, and 86% F1-score. In this study, video augmentation was implemented to enhance the model's performance in the task of smart baby care, which had dropped to 84-85% due to the task's realistic complexity. The model had a good balance between precision and recall and high accuracy because it used a controlled dataset with only a few action classes and did not perform early prediction evaluation. In this study, body orientation and fall dynamics are not analyzed. In the absence of spatial attention, the model failed to focus on critical regions, treating the entire frame equally, making it susceptible to background noise. Advanced temporal reasoning techniques, such as multi-scale or attention mechanisms, are limited by the use of ConvLSTM. Also involves an optimization gap, as ConvLSTM is computationally intensive and has more parameters. This model was a detection-only system with no intelligent alert strategy, allowing only a single alert mechanism, which increases false alarms.

A crib was created as an intelligent, multimodal fusion hazard prediction system [39] that leverages IoT and AI to anticipate potential threats before they occur and swiftly notify parents. This system involved multiple sensors and AI to implement the algorithm. The aim of this research was to protect the baby from multiple threats, including flipping or rolling over, climbing out of the crib, kicking the quilt, wetting the bed, fever, and crying excessively. The proposed method developed three algorithms for hazard prediction: a real-time state-monitoring algorithm, a turnover-detection algorithm, and an environmental-detection algorithm. The results of the experiment showed the intelligent multimodal fusion risk prediction system's real-time monitoring capabilities. Additionally, it demonstrated precise, efficient detection and warning capabilities across six hazardous situations. However, the prediction accuracy is not quantified. The experimental results showed that the system achieved high accuracy and sensitivity across different scenarios, with a quick reaction time. Using change-rate data for hypothermia, hyperthermia, bedwetting, crying, climbing over, and turning over, the accuracy was assessed, and wide fluctuations in the data due to sensor placement were observed, leading to increased uncertainty. Deviation risk may affect accuracy, leading to a higher False Positive Rate (FPR) or False Negative Rate (FNR) when the predicted value approaches the upper or lower limits of the range of variation. This might have affected the sensitivity. AI was utilized solely for image processing in this study, but it can also be used to train and learn from other multidimensional data, enabling a multidimensional risk prediction system. Thus, an AI-based system would be more reliable, avoiding hardware constraints.

Most current algorithms for detecting and predicting falls rely on implicit feature learning from raw RGB frames or optical flow, not on explicit human pose data. This limitation impedes their ability to

record vital body posture, joint dynamics, and biomechanical markers, which are crucial for the precise detection of early signs of imbalance or fall risk. Also, these approaches usually lack spatial attention mechanisms, meaning they treat the entire frame the same and are sensitive to background noise and clutter, which can make things harder in real life when there are many people around. They also typically utilize complex LSTM or ConvLSTM designs for temporal modeling. These architectures have too many parameters, making them challenging to use in real time and apply to new data, especially when the datasets are small. They don't often look for lighter GRU-based options. Most systems also have a one-stage alert system that sends alarms immediately upon forecast issuance. This means there is no difference between worsening threats and imminent hazards. This makes users less confident in the results, leading to more false positives. Finally, the current focus on reactive detection rather than proactive prediction reduces the time available for preventive interventions.

2.6 Analysis of Smart Devices Powered by AI

A thorough background investigation indicates a similar trend in healthcare monitoring, where smart devices are crucial components of initiatives, studies, and research. These smart devices use Machine learning/artificial intelligence and image processing techniques to enable IoT devices to automate processes, learn from data, and continuously improve performance. Figure 2.3 depicts three types of smart devices: wearable sensor-based, non-wearable sensor-based, and the fusion of AI/ML and Sensors/IoT.

The collected data is transferred after wearable sensors are affixed to body parts, such as the hands, ears, wrists, legs, thighs, and chest. According to some research, wearable sensor-equipped smart devices yield results comparable to those of medical devices. This category includes vibration and sound sensors, as well as other wearable devices, such as wristbands. The research cited in [45] uses a sound sensor to detect severe crying and alerts parents or caregivers via vibration, using vibration sensors specifically designed for deaf parents. When the child is not with an adult or parent, the wearable device must be a wristwatch that stays on the child's wrist. This study presented a wearable child safety device [44] that incorporated a GPS location sensor to track the child's location and ensure safety. Body-worn sensors were used in about 15% of the studies analyzed for data collection and communication, with most focusing on baby safety.

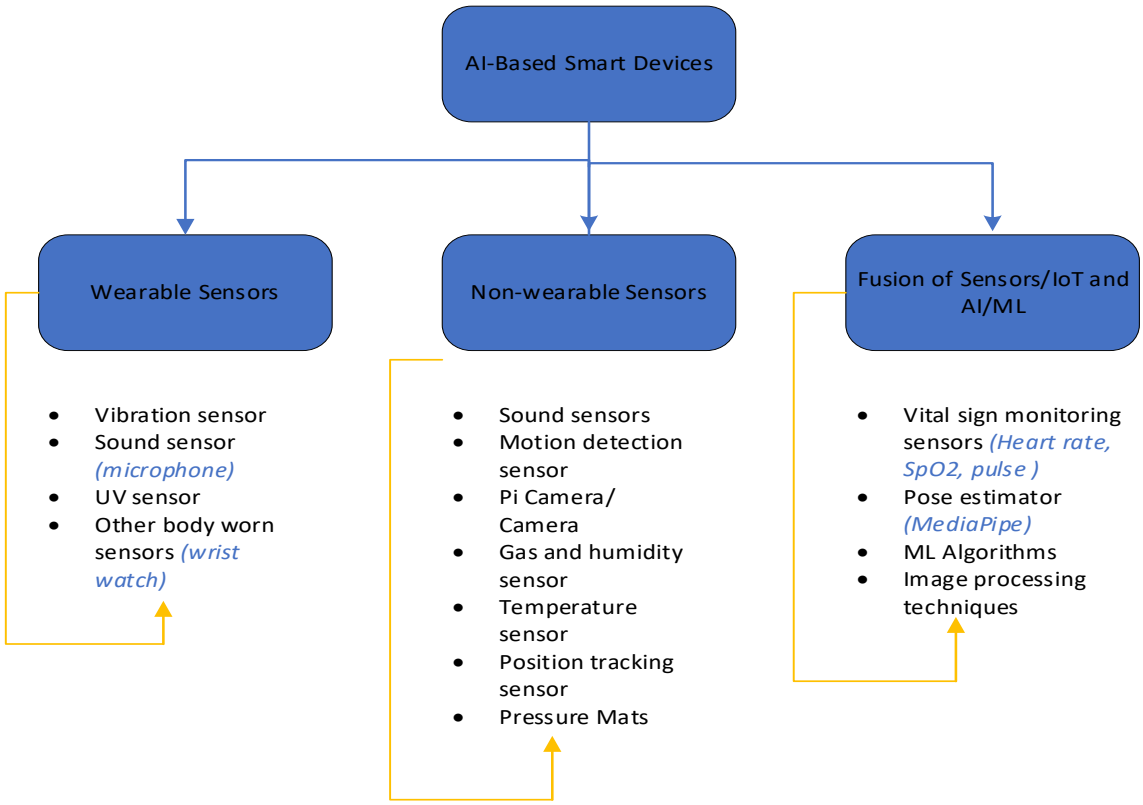


Figure 2.3: Classification of AI-Based Smart Devices

Non-wearable monitoring device-based studies, as cited in [60, 66, 67], use sound sensors, motion detectors, cameras, pressure mats, position-tracking sensors, moisture, and temperature sensors in various ways to enhance baby safety. These devices are employed in approximately 46% of the studies examined for data collection, with 83% addressing newborn safety and 17% covering baby fall detection. Research in [35, 36, 40, 64, 68, 69] investigates fall and non-fall detection and prevention utilizing integrated technologies such as sensors/IoT, AI/ML, or image processing. These integrated smart devices are used in 39% of the studies examined for data collection and communication, among which 80% focus on baby safety and 20% on fall detection. Interestingly, fusion technology outperforms its counterparts in both wearable and non-wearable monitoring devices in terms of accuracy. These smart gadgets are renowned for their exceptional intelligence and ease of use. They can respond to situations by instantly alerting parents and carers when irregularities are detected. Research on these smart gadgets is becoming increasingly popular, and this trend is expected to continue. However, one major downside of wearable sensors is that they must be worn continuously by the newborn, which can increase pain and misalignment due to the baby's movements. Furthermore, there are issues about data disconnection while charging. Non-wearable monitoring devices can be used to address these issues. Most smart devices are focused on baby safety, such as excessive crying, proper sleep, breathing concerns, and so on, but very few are for falls. Figure 2.4 clearly demonstrates the usefulness of smart gadgets in healthcare applications, particularly in baby health & safety and baby

fall detection. This analysis shows that the research is primarily focused on baby health & safety (85%), with 15% on baby fall detection and no baby fall prediction.

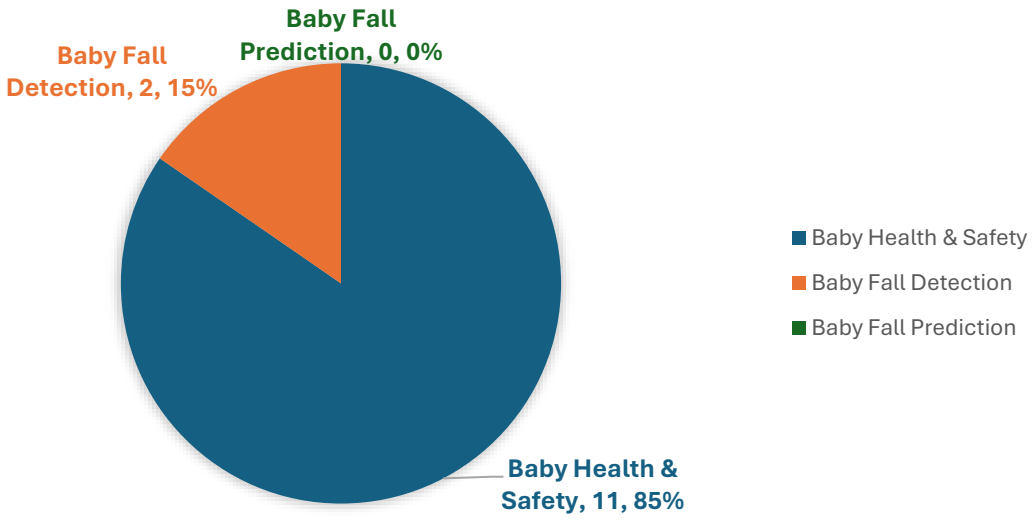


Figure 2.4: Categorization of Cognitive Devices in Baby Monitoring Systems

As shown in Figure 2.5, an outline of the metrics for using smart devices is shown. Non-wearable monitoring devices for ongoing data collection account for about 46% of research activities. With approximately 15% focussing on wearable sensors for health monitoring, particularly baby safety. However, baby activities, health, posture, and motion detection research employ 39% of sensor fusion and AI/ML techniques for baby safety and fall detection. In general, the research is being led by these three kinds of AI-based smart gadgets. These devices play an important role in home monitoring, both within and outside crib settings, to identify fall hazards and anticipate falls in babies.

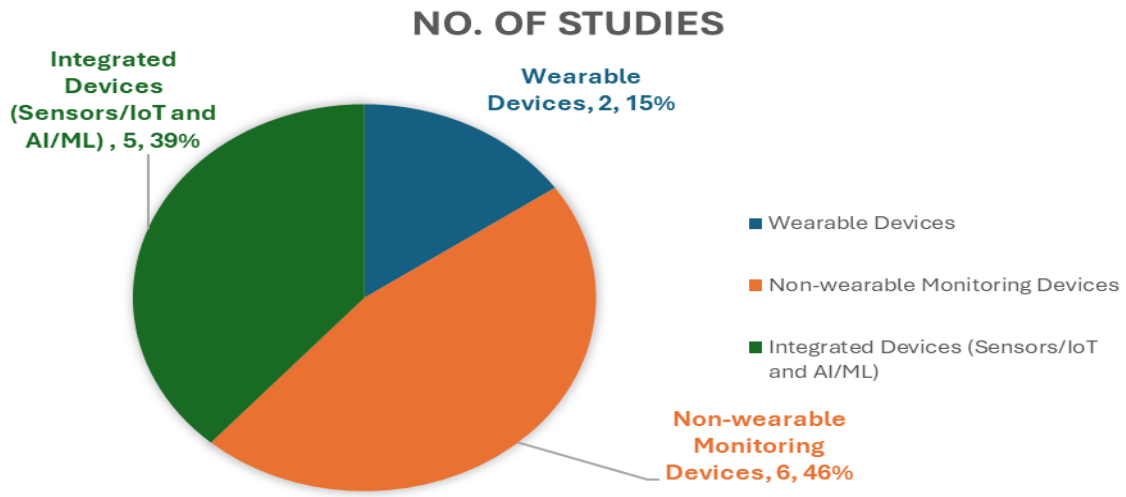


Figure 2.5: Smart Device Analytics

2.7 Comparison of AI-Based Fall Prediction within Baby Crib

We have reviewed numerous studies and found that research on baby fall detection without a crib has been conducted; however, research on baby fall risk prediction remains very limited. Table 2.1 provides a detailed overview of the comparisons among the existing work. The details of these studies are already given in Section 2.5. We offer cutting-edge techniques, pinpoint the pertinent gap that spurred our exploration of this field, and make a novel contribution. Most existing fall detection and prediction systems rely on implicit feature learning from raw RGB frames or optical flow rather than explicit human pose data. This constraint limits their capacity to record keypoint-based posture, joint dynamics, and biomechanical markers, all of which are critical for accurately detecting early signs of imbalance or fall risk. Furthermore, these systems typically lack spatial attention mechanisms, so they treat the entire frame equally and are sensitive to background noise and clutter, making real-world scenarios with many people difficult to handle. They also commonly use complicated LSTM or ConvLSTM designs for temporal modelling. These architectures have too many parameters, making it difficult to deploy them in real-world settings. Finally, existing alerting frameworks typically utilize single-threshold mechanisms that fail to categorize risk severity, leading to frequent false alarms and reduced user reliance. Concurrently, the prevailing focus on reactive detection rather than anticipatory prediction severely restricts the lead time available for proactive intervention.

Thus, we introduced AI-1, which uses the keypoint technique to extract joint coordinates and identify accurate postures, thereby predicting fall risk. However, key points alone are insufficient for posture identification; therefore, AI-2 was implemented using a CNN+Soft Spatial Attention mechanism that provides detailed spatial cues, edge textures, and other spatial features to accurately identify postures. Nevertheless, the AI-2 model, which uses only spatial features, poses challenges for accurately identifying postures and predicting risky scenarios that lead to falls. Consequently, the AI-3 model uses AI-1 pose features, AI-2 spatial features, and additional temporal features, analyzing video as input, yielding the best result among all three models. Besides this, a two-stage alert mechanism reduces the frequency of false alarms in AI-3.

Table 2.1: Comparison of the previous research work based on novelty and techniques

Research Work	**Novelty**	**Participation Age (months)**	**AI/ML Techniques**			**Prediction**	**Alert System**
			Keypoints-based	**Attention**	**Temporal Modelling**		

[39]	Using AI and IoT to detect dangers among the children within the crib	Not explicitly mentioned (addressing both infant and baby)	No	No	No (Haar Cascade Classifier)	Detection only	Single stage
[19]	Temporal Modelling + customized loss function -TP-LSTM + attention + EEL, focus mainly on Loss	Not explicitly mentioned	No (implicit motion + posture patterns via deep features using CNN)	Temporal (extracts implicit motion features from temporal attention)	LSTM, Proposed TP-LSTM (multiscale temporal learning), temporal attention mechanism	Prediction	Single stage
[18]	ConvLSTM +I3D (transfer learning), light weight model for small datasets and integration with IoT smart monitoring systems, focuses mainly on	Not explicitly mentioned	No	No attention	ConvLSTM	Recognition /Detection only (Only detects current action)	Single stage

	the architecture						
Our proposed work	Pose-driven approach for explicit spatial structure, spatial attention to focus more on body parts, efficient temporal modelling and intelligent alert strategy	12-24	Explicit	Spatial	GRU	Early + Real-time	Two-stage

2.8 Conclusions

Current research on baby safety monitoring either focuses on temporal prediction using LSTM-based architectures with tailored loss functions or on action recognition using CNN-LSTM frameworks.

However, they lack clear pose-based modelling, spatial attention mechanisms, efficient temporal architectures, and progressive alert systems. This study addresses these shortcomings by introducing a pose-driven, attention-based temporal GRU model, coupled with a two-stage alert system for precise, timely fall prediction.

When a baby is in a crib, early fall-risk prediction utilising AI technology can trigger an alarm and call for emergency help on time. Families or carers may feel more at ease and be able to keep an eye on their baby without being physically present by monitoring their baby's everyday activities in the crib. Babies' postures can be precisely assessed by analysing poses generated using AI-based Pose Estimation (PE) approaches, without the need for wearable sensors or sensor implants near the infant. Recurrent neural networks and deep learning models can predict more accurate and dependable postures. Nevertheless, the AI-1 PE-based and AI-2 Deep Neural network-based models, described in depth in Chapters 4 and 5, respectively, each contributed to shortcomings in baby fall prediction. Therefore, AI-Transfer Learning-based monitoring can be seen as a new era in the healthcare industry for identifying, preventing, and forecasting fall risk, as well as for triggering emergency alerts, especially for fall prediction in babies. However, more work and real-time studies are needed to develop a comprehensible and reliable AI system. Therefore, the elements mentioned above serve as the foundation for developing a novel ADL-Conditioned Transfer Learning-based Hybrid AI model that uses the MediaPipe Pose model, the Spatial Attention mechanism, the Gated Recurrent Unit (GRU) temporal model, i.e. Recurrent Neural Network (RNN), and the Dual-Trigger Alert Mechanism to anticipate falls in babies in cribs. Using a cascaded transfer learning-based AI model and predictive modelling, the requirements for robust baby fall prediction and the minimization of baby falls can be met. Thus, this model provides posture transitions and improves interpretability and early detection using the Pose model. The spatial attention mechanism focuses on body parts to localize fall events, and the temporal GRU maintains performance with lower complexity, offering faster training and inference and better suitability for edge devices, enabling real-time alert systems. Moreover, the two-stage alert system provides a warning for prolonged standing without risky postures and a critical alert for high-risk activities such as climbing over or jumping near the fence, improving usability and trust while reducing false panic among parents or caregivers.

Chapter 3. Research Methodology

3.1 Introduction

Artificial Intelligence has the potential to revolutionize healthcare systems. The quality of life has significantly improved over the last few decades because of AI, ML, and image processing, especially in healthcare applications. The methods and tools employed must be genuinely consistent with the research's main ideas. At the end of Chapter 1, we talked about the two main purposes of this research. First, the need for a Baby Fall risk prediction system, and second, to make it easier for parents and caregivers to keep an eye on their children at night. This research introduces a novel Semantic-Conditioned Hybrid AI model for future fall prediction that leverages analysis of Activities of Daily Living. The proposed AI-based fall risk prediction system combines a dual-trigger alert mechanism with several deep learning algorithms to provide parents or caregivers with early alerts about the baby's fall risk.

This research uses three AI frameworks to predict the risk of baby falls in the crib. The first model, AI-1, consists of a Pose model that uses keypoint extraction to analyze babies' postures in the crib and predict fall risk levels (Safe and Unsafe) based on ADLs. This model was validated using a customized BabyFall dataset and classified as Safe or Unsafe using four classifiers across six ADLs. The AI-1 model outperformed the other classifiers, such as Support Vector Machine (SVM), Multi-Layer Perceptron (MLP), and Extra Gradient Boosting (XGB), and Random Forest (RF). The second model, AI-2, employs a Deep Neural Network, specifically a 2D-CNN with spatial attention, to monitor ADLs such as sitting, lying down, playing, climbing, leaning, and standing near the crib's fence. The model detects postures by analyzing those ADLs into two categories: Safe and Unsafe. A three-layer 2D CNN and a spatial attention mechanism extract spatial cues from the images and feed those pooled features to the dense layers. Using the sigmoid activation function, the classification block outputs 0 as Safe and 1 as Unsafe. Both AI-1 and AI-2 models will independently assess fall risk using Safe and Unsafe outcomes. However, because these models are assessed frame-by-frame or image-by-image, they struggle to anticipate dangerous situations and generate a high number of false warnings or false positives. Even so, the AI-2 model outperforms AI-1 when tested using statistical parameters on an untested dataset. Nevertheless, these models are inefficient and yield unpromising results.

Conversely, the final model, AI-3, is developed by analyzing videos, extracting keypoints, and engineering pose features, including Center of Mass, torso angle, hand-to-rail distance, etc., feeding those sequences of feature vectors to a temporal model for classification, forming an ADL-Conditioned Transfer Learning-based Hybrid AI model. The approach allows for prompt risk identification by adjusting posture in response to changes over time through continuous learning. The model then applies temporal smoothing and classifies fall risk levels into *Safe, Near-Fence, and High-Risk*. The final

system decides based on ADL's conditions and duration and provides an alert in three levels: *Safe, Warning: Prolonged Standing. High-Risk.* Prediction accuracy improves with the AI-based cascaded transfer learning model. Additionally, by providing high sensitivity and specificity, it lowers false alarms. To demonstrate the model's efficacy and efficiency, it is tested on an unknown dataset created by an AI tool. The suggested approach demonstrates its reliability in identifying infants at risk of future falls, as evidenced by strong performance metrics. This study advances AI-driven healthcare solutions by offering a proactive, cost-effective approach to predicting the risk of baby falls. This chapter provides a detailed methodology for evaluating the model development process. The framework and research background are also covered. This chapter is followed by an overview of current simulation tools, a thorough summary of pertinent research on general AI simulation tools used in current studies, and an explanation of why Python was used in this study.

3.2 Methodology and Model Development Phase

In this research, a quantitative method is utilized for data collection, analysis, and model development. Behaviour is the primary input for our model development. During data collection, the baby's behavioural parameters are analyzed in either the spatial or time-series domain. The behaviour data relating to daily activities, such as ADLs, performed by babies in the crib are analyzed. The behaviour parameters are analyzed in the spatial domain for the AI-1 and AI-2 models using the Pose-ML pipeline and deep learning algorithm, respectively. However, the behaviour parameters are analyzed in the spatio-temporal domain for the AI-3 model. Thus, ADLs are required in phases 1 and 2 of the AI development process for reliable and accurate baby fall risk estimation. The individual AI models, AI-1, AI-2, and AI-3, use the same parameters in both phases. Thus, in phase 4, data collection is conducted based on the parameter selection from phase 3. Once the model is developed following a specific algorithm for each case, the system can predict future falls. The system then sends alerts to parents or caregivers for high-risk activities. Figure 3.1 depicts the research methodology's procedure. The fall risk prediction model is created via a methodical, interconnected process.

Observation Phase: Literature review and data analysis are completed.
Parameter Selection Phase: Suitable parameters are determined for AI-1, AI-2, and AI-3 models based on behaviour.
Data Collection Phase: The sources are an online platform and an AI tool.
Development Phase: The AI-1 model is developed using the Pose-ML pipeline; the AI-2 model is a deep neural network (DNN) algorithm; and the AI-3 model is a hybrid transfer learning algorithm.
Experimentation Phase: Training and testing are done on the developed models. The AI-2 model is tested on an unseen dataset leveraging temporal coefficients. Then, the final hybrid model, AI-3, is tested on an unseen dataset using the Dual-Stage Trigger mechanism, and a fall-risk prediction system is created.

Although these phases appear independent, they are intrinsically interrelated and contribute to the overall development of the future baby fall-risk prediction model in this research.

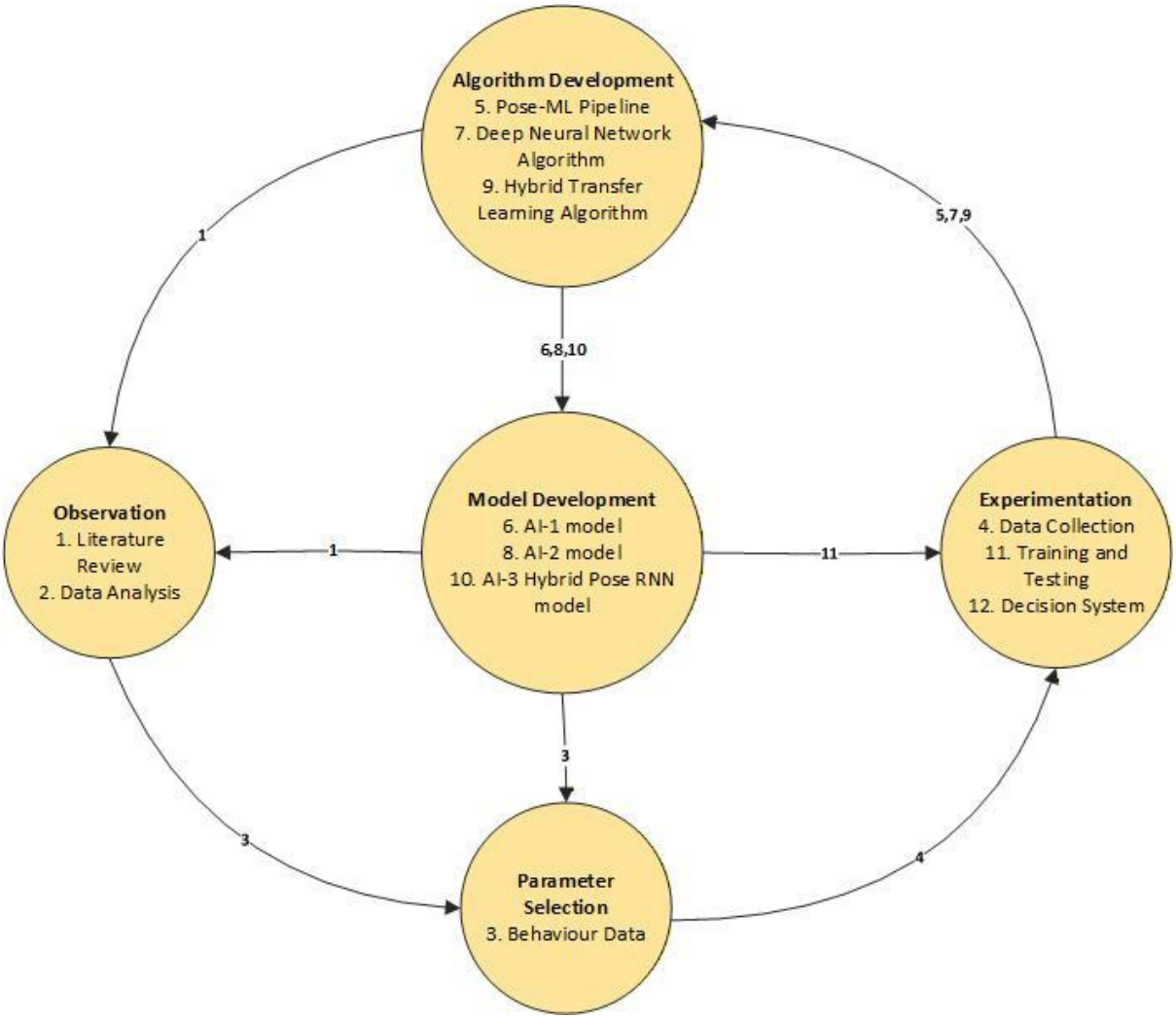


Figure 3.1: Research Methodology Procedure

In Chapter 2, a literature analysis of the baby monitoring system was undertaken, organized into categories that highlight the diversity of the safety domain and its specific areas. Based on the study, behaviour is the primary characteristic examined. This characteristic has had a significant impact on the incidence of baby falls. The data assessment was done based on the selected parameters, i.e.,

- *Behaviour Parameters*: Activities of Daily Living (Sitting, Lying Down, Playing, Standing, Leaning, and Climbing).

Data are collected for selected ADLs, and deep learning algorithms are developed. In the model development, the AI-1 model uses the Pose-ML pipeline, the AI-2 model uses DNN, and the final AI-3 model employs a hybrid pose RNN to detect and classify abnormalities based on behavioural inputs. The training and development process is performed for each individual model to assess the accuracy of

predictions of unsafe or fall-risk behaviour, thereby supporting the successful development of the system. The literature review is conducted at every phase of the development process to verify that the baby fall prediction system is constrained by prior research. Additionally, the shortcomings of the AI-1 and AI-2 models, along with the selection of key methods from each, spurred the development of the AI-3 model, as illustrated in Figure 3.2. This figure illustrates the context of this research and its applications. This research is carried out in three phases: *Data Collection, Design and Development, and Training, Validation, and Decision System*.

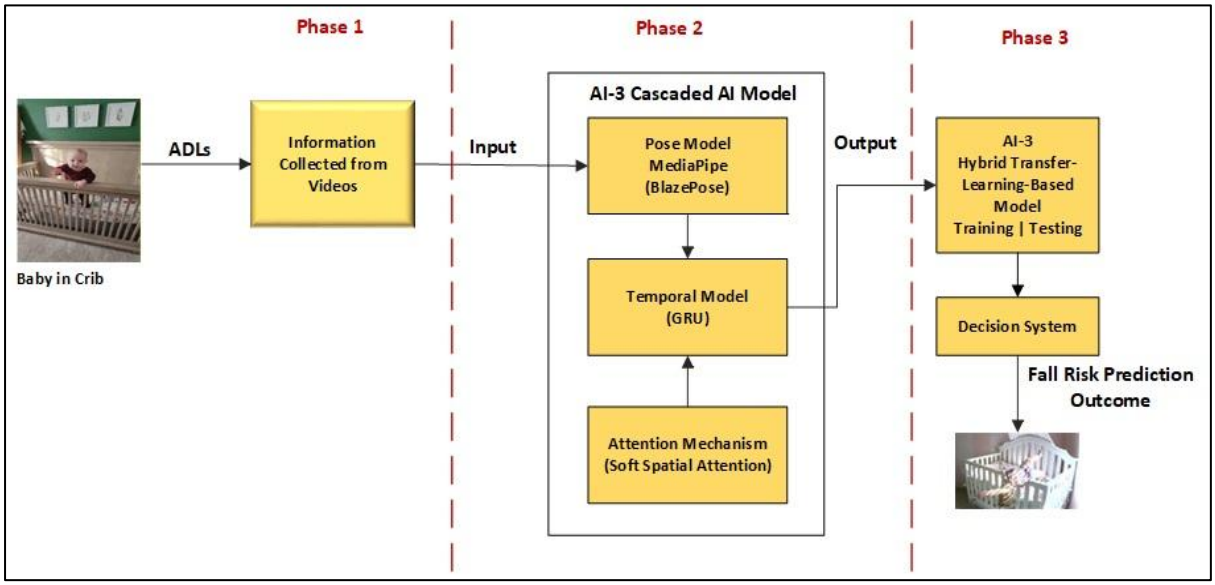


Figure 3.2: Context of Research

Phase 1 - Data Collection: Baby behaviour data is collected from online platforms such as YouTube, Dailymotion, TikTok, etc. A few datasets are generated using an AI tool for testing purposes.

Phase 2 – Design and Development of an AI model: The AI-3 model was developed using behavioural parameters extracted from babies' videos sourced from online platforms, and its results were tested on unseen videos generated using an AI tool. A hybrid Pose-RNN-based algorithm for baby fall risk prediction is formulated.

Phase 3 – Training, Validation, and Decision System: In this stage, the AI-3 model used fusion techniques during training. Pose features are extracted based on the BlazePose pretrained model. Engineered pose features and spatial features from the attention mechanism, whose outputs are aggregated together. A Gated Recurrent Unit (GRU), an RNN framework, is trained on aggregated and temporal features via transfer learning, and a two-stage alert mechanism is incorporated to estimate future baby fall risk. This thesis presents an innovative method, a "Novel ADL-Conditioned Transfer Learning-Based Hybrid AI Model for Future Baby Fall Prediction".

Figure 3.3 presents the research framework, which focuses on the creation and evaluation of data models for deep learning algorithms. The data is obtained from the given parameters (shown in Figure 3.2), and the collected data is filtered and pre-processed to eliminate extraneous data and keep only relevant information. For instance, when gathering data for the behaviour parameter using videos found online, the videos may show parents or siblings, pets or toys in subsequent frames, or additional information, such as a video log showing the door opening and then the camera being directed to the baby. So, all this

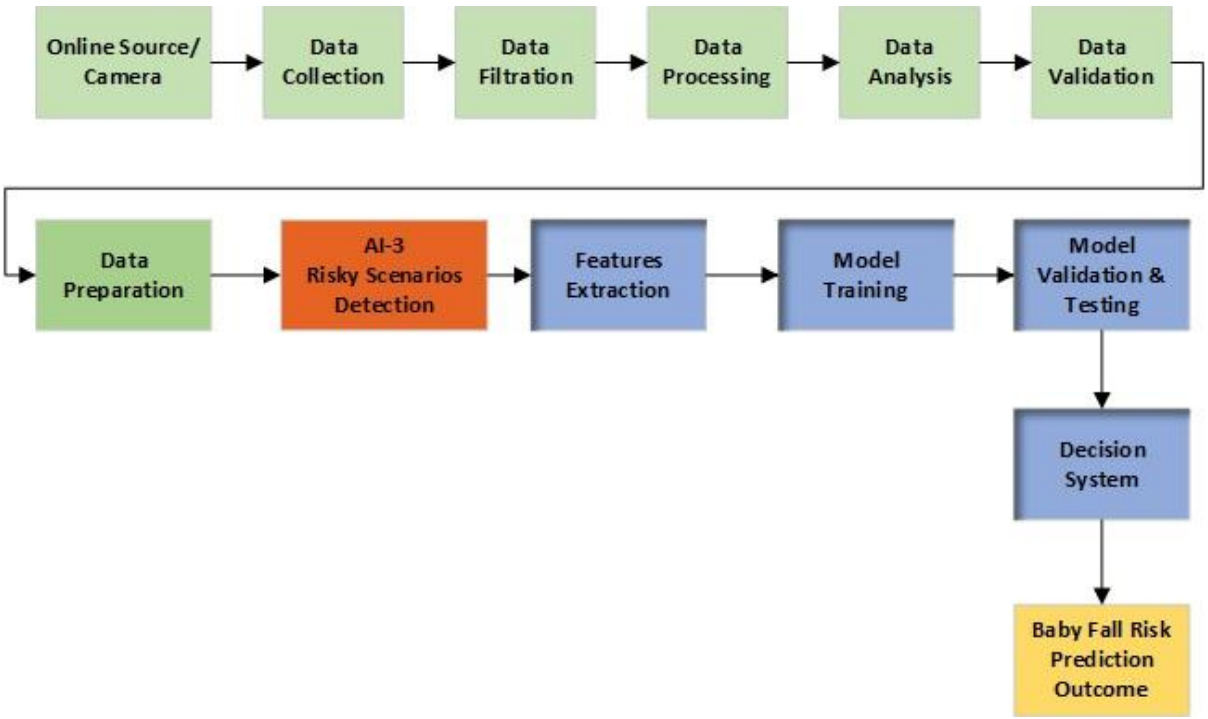


Figure 3.3: Framework of the Proposed Research Approach

information was not useful when we were collecting data, so it was left out when we made the datasets. This procedure filters data. To ensure data quality, preprocessing is performed before proceeding to the next phase. Data analysis and validation follow the same process to ensure data accuracy. After that, the acquired data module is converted into the correct format for data preparation and feeding into the AI-3 model. A Hybrid Transfer Learning Algorithm is employed in the AI-3 model. The model is trained and validated based on the provided data. The decision logic then allows the model to categorize the severity of a fall based on the behavioural time-series data, enabling accurate prediction of baby fall risk. The decision system generates an alert and determines the baby's fall-risk outcome.

3.3 AI/ML Modeling and Development Simulation Tools

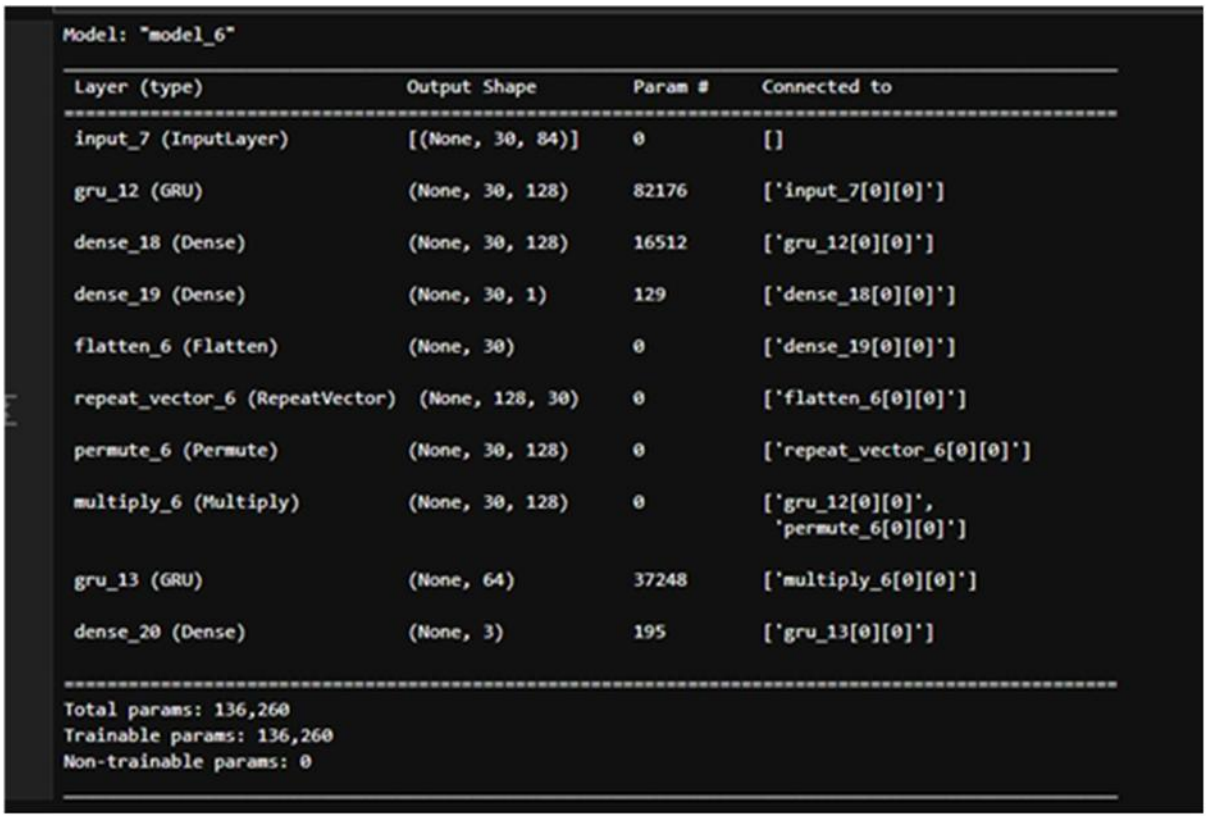
Artificial Intelligence (AI) is revolutionizing healthcare applications through the transformative power of simulation tools and coding software. The integration of AI-based simulation tools in healthcare

enables the provision of learning and training on crucial data and reduces the shortage of effective systems. These technologies support medical education, advance research, and improve baby care.

Various tools have been used in previous research on fall detection. Other visual languages, such as Scratch and C/C++, are also used as coding software. Scratch makes it easy to get started with programming concepts like variables, loops, and method calls, but it's not good for building professional applications. C++ is also an incredibly powerful language for expert developers. Most of the web browsers used today are written in C or C++, or even in a strategy video game; however, they are completely unfriendly to beginners seeking a basic introduction. Instead, to support rich applications, novices must immerse themselves in the language's full complexity.[70]. Python code is also simpler to write and read. Python code is used in many commercial applications, such as Instagram, YouTube, Reddit, and Spotify, and it has all the features of other languages. [71]. Python is utilized in several sectors, including healthcare, aviation, industry, finance, GPS technology, and consumer products. Governments employ it for governance, national security, traffic management, public safety, and other purposes. Python is utilized to manage firmware updates in network programming. Python is essential in computer graphics, data mining, embedded systems, and other domains. Python is perfect for scientific computation, artificial intelligence, and data analysis. [71].

3.4 Python-Based AI Modeling and Jupyter Notebook IDE

In this research work, the dominant programming language is Python, with libraries such as TensorFlow/PyTorch/Keras, Scikit-learn, and numpy. Jupyter Notebook is used as the integrated development environment (IDE). Python is a full-spectrum language that has become easy to learn for beginners [70, 72]. In addition to being a fun and approachable language, Python underpins the technology used by numerous top-tier businesses. Python is an open-source, high-level, general-purpose programming language. [71]. Python is used for its wide range of ready-made libraries that can be freely used in any program and integrated with other languages. Python is compatible with all major operating systems, including Windows, Linux/Unix, OS/2, Mac, and Amiga. [73]. Much of the research has been conducted using Python code for data science tasks, such as baby activity recognition [46] and fall predictive modeling. Python is supported by Jupyter Notebook [74] as a programming environment, allowing for interactive data exploration, graphical presentation, and visualization—all of which are essential for medical research. The retrieval of Activities of Daily Living (ADLs) for all models was performed using Python code. This simulation tool can extract key features and provide an explanation of each step of the simulation process. For example, the simulation analysis of the AI-3 model is depicted in Figure 3.4, where each stage of the simulation process was clearly explained, and the Hybrid Pose RNN, including the input layer, GRU layer, and attention mechanism layer, was examined separately.



```
Model: "model_6"
_________________________________________________________________________________________________
Layer (type)                 Output Shape              Param #     Connected to
=================================================================================================
input_7 (InputLayer)         [(None, 30, 84)]          0           []

gru_12 (GRU)                 (None, 30, 128)           82176       ['input_7[0][0]']

dense_18 (Dense)             (None, 30, 128)           16512       ['gru_12[0][0]']

dense_19 (Dense)             (None, 30, 1)             129         ['dense_18[0][0]']

flatten_6 (Flatten)          (None, 30)                0           ['dense_19[0][0]']

repeat_vector_6 (RepeatVector)  (None, 128, 30)        0           ['flatten_6[0][0]']

permute_6 (Permute)          (None, 30, 128)           0           ['repeat_vector_6[0][0]']

multiply_6 (Multiply)        (None, 30, 128)           0           ['gru_12[0][0]',
                                                                   'permute_6[0][0]']

gru_13 (GRU)                 (None, 64)                37248       ['multiply_6[0][0]']

dense_20 (Dense)             (None, 3)                 195         ['gru_13[0][0]']

=================================================================================================
Total params: 136,260
Trainable params: 136,260
Non-trainable params: 0
_________________________________________________________________________________________________
```

Figure 3.4: The AI-3 Model’s simulation technique using Python code.

The Jupyter Notebook terminal effectively presents epoch-wise analysis and the duration of each process utilizing Python code. Python code efficiently delivers a clear, well-organized output of findings for each epoch used in the AI-3 model as time advances. The model's enhancements during training are readily comprehensible, facilitating understanding of how improvements are realized over multiple iterations. Figure 3.5 illustrates how the model’s accuracy and efficiency improve at each successive epoch, as shown in the visualization of the training process.

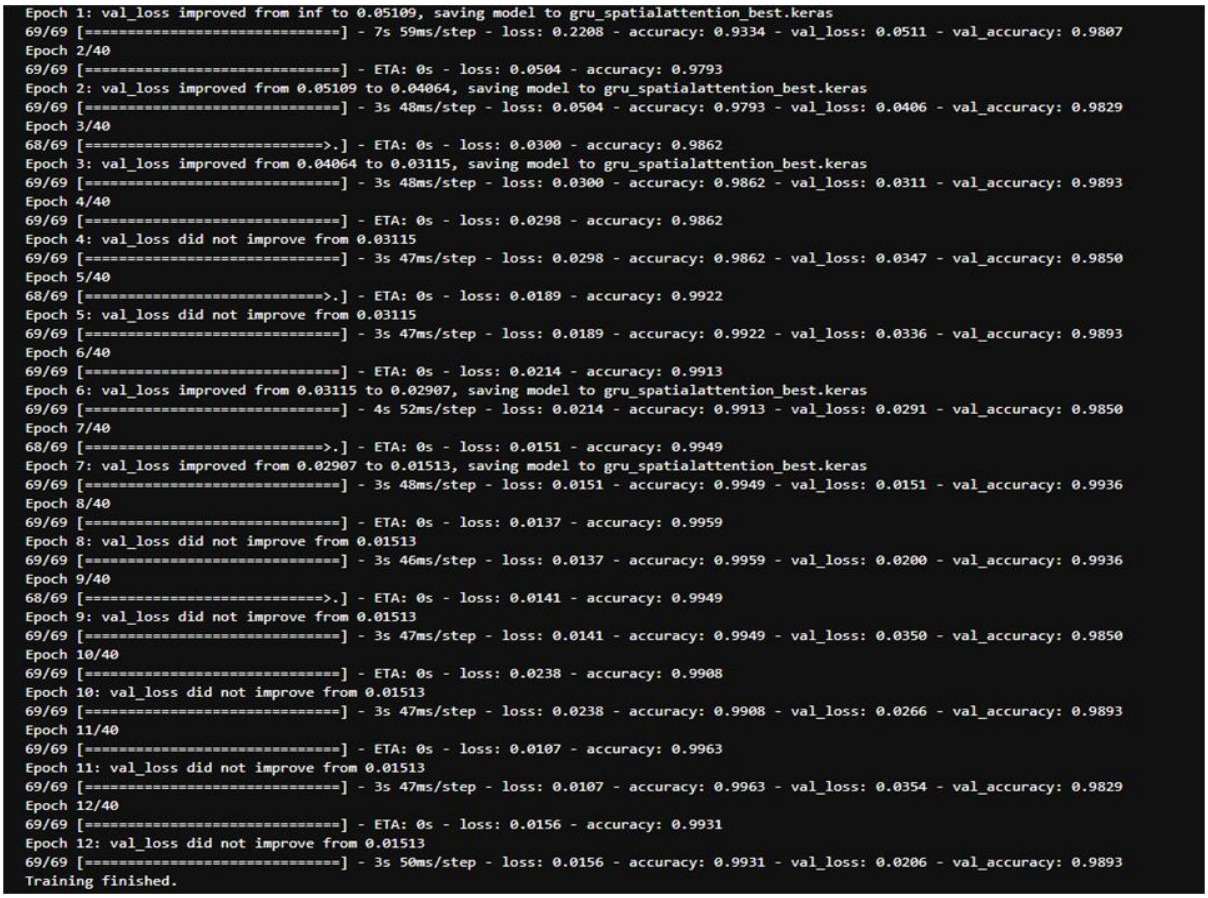


Figure 3.5: Python code output in a terminal window

The graphical representation of Python Code is clearer, more comprehensive, and more readable, thanks to its simple, efficient visualization of data points. In addition to its simplicity, it effectively interprets complex data and provides clear visualizations. Figure 3.6 shows the graphical analysis of the AI-2 model, illustrating the correlation between the probability of unsafe frames vs. time (in seconds) for a given threshold of 0.5, using Python Code. Figure 3.7 shows the graphical analysis of the AI-3 model, depicting the probability of events such as safe, near fence, high risk, or hand near rail over a 10-second time interval, using Python Code.

Because of these factors, as well as the successful application of the research findings by other researchers, Python was chosen as the simulation tool for this study. The simulation findings using Python Code are discussed in Chapters 4, 5, and 6. For this research, we created a virtual environment with TensorFlow and ran the code in Python (version 3.10.0) using the Jupyter Notebook IDE on a machine equipped with a first-generation Intel Core Ultra 7-155U (1.7 GHz) with 32 GB of RAM, running Windows 11 Enterprise. TensorFlow 2.12.0, a deep learning framework, was used to construct the final model in these experiments.

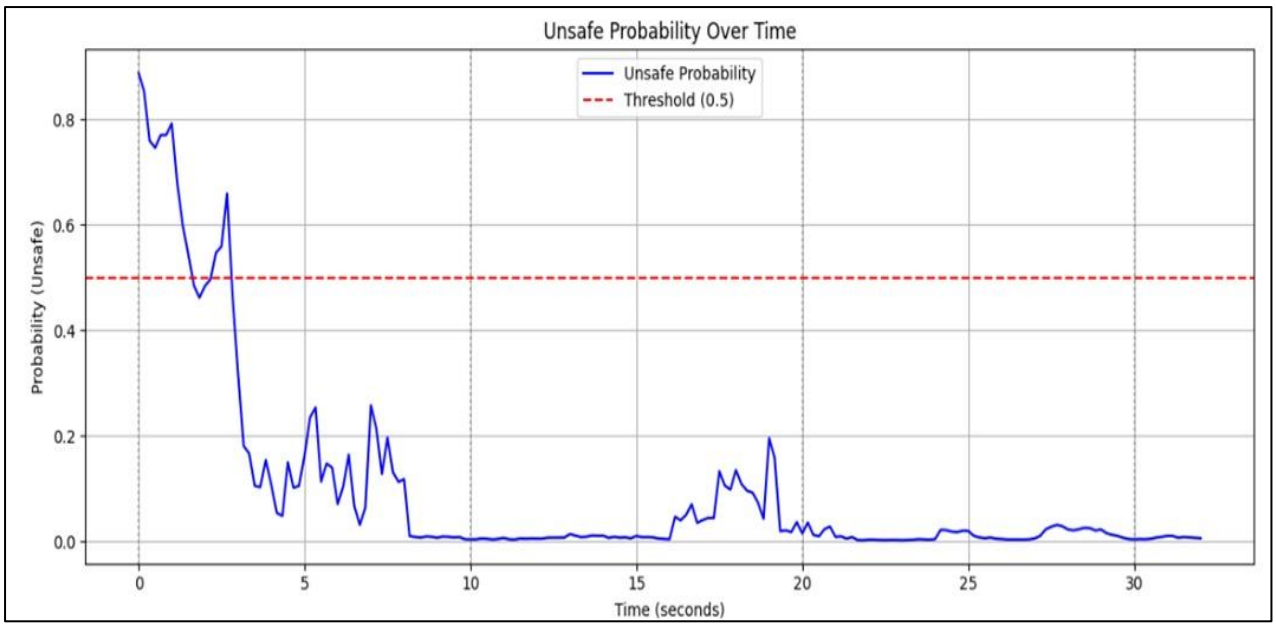


Figure 3.6: Graphical analysis of unsafe frames using Python Code.

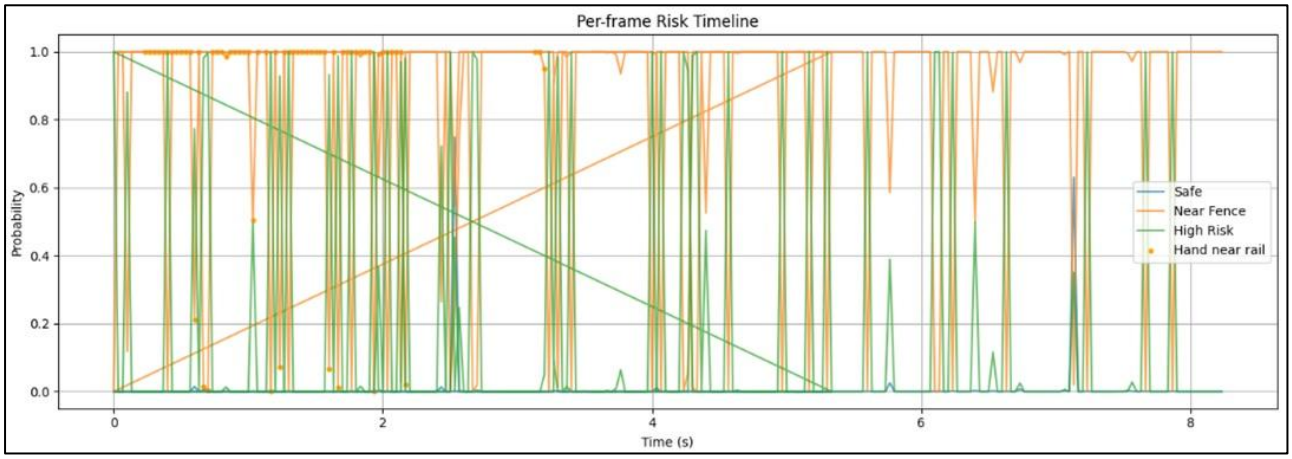


Figure 3.7: Graphical analysis of Risk Timeline using Python Code.

3.5 Conclusion

The developed AI prediction models are conceptualized individually within either the Pose ML pipeline or the Deep Neural Network to create a hybrid system that can be regulated based on each model's strengths and weaknesses. This is achieved by envisioning data within a virtualized software environment, enabling the effective simulation and testing of predictive models and thereby optimizing them prior to actual implementation. The concept of the final model relates to the methods of the earlier two models, AI-1 and AI-2, ensuring that the system is sensitive to spatial and temporal features and improving the effectiveness and efficiency of fall risk prediction.

The approach and framework employed in this study are explained in detail in this chapter. The identification and selection of suitable simulation tools that enable model construction, testing, and

validation in a controlled virtual environment are key features of the research. Here, Python code is utilized in a Jupyter Notebook, a popular programming environment. To make the language simple for anyone to understand, it uses a few prompts and a clear syntax. It has important features, including virtualization, a scalable network model, and excellent reliability. These qualities make it a perfect place to run AI-based simulations, as they allow testing across varied settings, improving algorithm effectiveness, and checking the model all at once. This research ensures that AI prediction models are thoroughly tested in a controlled, virtual environment before deployment in the real world. This method makes the fall risk prediction system more reliable, flexible, and responsive, ensuring it performs well in real-world healthcare settings.

Chapter 4. Proposed AI-1: Pose Estimation-Based Baby Unsafe Scenarios Prediction System

4.1 Introduction

In this research, the fall-prone, unsafe-scenario prediction methodology can tailor interventions to individual needs. It can enable timely preventive warnings before the baby falls. The focus of this chapter is to identify unsafe scenarios for babies in the crib by monitoring ADLs. Additionally, identifying the postures upon extracting the keypoints. Finally, notifying the risk of an unsafe scenario earlier (parents/caregivers) based on ML classification analysis. This chapter introduced a PE-ML-based methodology for predicting unsafe scenarios in babies in cribs by monitoring ADLs. A gap in the literature is filled by highlighting the importance of sensor data for estimating the baby's pose. Movement sensors incorporated into babies, such as accelerometers and gyroscopes, capture motion data. However, these sensors often provide limited information about complex body postures and joint positions, essential for precise pose estimation. Pose estimation techniques use image or video data to analyze skeletal key points. However, movement sensors do not inherently provide details about specific joint angles or spatial relationships between body parts. This limitation can restrict their ability to accurately identify and quantify complex postural changes. In turn, making pose estimation techniques more suitable for detailed movement analysis and identification of key points. The first model, AI-1, uses continuous tracking to detect dangerous situations involving infants in the crib. It is supported by an innovative Unsafe Scenarios Prediction Algorithm and an ML algorithm. This proposed prediction technique, with the Unsafe Scenarios Prediction Algorithm (USPA), uses multiple classifiers to learn and perform tasks. The model analyzes the level of scenario risk as *Safe or Unsafe* and was validated across the datasets using multiple classifiers. This technology anticipates unsafe scenarios for babies and may save lives in critical situations. Notably, it’s also an incremental contribution, as it is the sole research that depends on the specific crib scenario, is restricted to the age range of 6-24 months, and uses pose features as input for further classification. As previously indicated, the chapter additionally offers a summary of the algorithm's development and rigorous testing. A specially designed dataset has been used and verified against several classifiers to achieve optimal accuracy. The goal of this study is to create a more accurate model for predicting unsafe-scenario risk. AI-1 model was evaluated using the provided datasets and achieved successful predictions, as discussed at the end of this chapter. With these advancements in the proposed hybrid two-stage ML pipeline approach, the safety of babies can be predicted earlier, saving lives.

4.2 Significance of Human Pose Estimation (HPE)

HPE, a critical area of computer vision, involves identifying the positions and orientations of human body parts from sensor data such as images and videos. In this study, HPE is integrated with machine learning algorithms to monitor a baby's posture, movements, and potentially hazardous behaviours. The

system collects data on key points related to various activities of daily living (ADLs) for babies in cribs. Human pose estimation (HPE) [75, 76] is a key area of computer vision that estimates the configuration of a human's body parts from sensor-captured data, such as images and videos. It provides geometric and motion insights for AR, VR, healthcare, and human-computer interaction applications. Recent advancements in deep learning have outperformed traditional methods across tasks such as image classification, semantic segmentation, and object detection. In response to the conditions of body reality, HPE is extensive and has many applications. HPE is used across a variety of fields, including animation, medical applications, activity recognition, and robot training. Therefore, our work falls broadly within the field of activity recognition and is related to baby healthcare applications.

In this chapter, HPE is integrated with machine learning algorithms to monitor a baby's posture, movements, and potentially hazardous behaviors. The system collects data on key points associated with various activities of daily living (ADLs) specific to babies in cribs.

4.2.1 Deep Learning Models in HPE

Various PE techniques [77] identified in the research are very effective for identifying keypoints or joint coordinates. These techniques are based on deep learning models, and thus, there are a few major deep learning models for human pose estimation. The main DL models are OpenPose, DeepCut, AlphaPose, Mask R-CNN, Iterative error feedback, and MediaPipe.

The first bottom-up, multi-person, real-time system to identify 135 human body keypoints in a single image is called OpenPose. It was proposed by Carnegie Mellon University researchers. [78]. Because OpenPose is built to run on GPUs and achieve high accuracy without compromising implementation quality, it is far more accurate. However, it is limited to low-resolution results. This restricts the quantity of data in keypoint forecasts and marginally reduces accuracy and speed. [77]. DeepCut is a bottom-up method for multi-person human pose estimation. In this study [46], a novel system was proposed to recognize infant activities in videos by decoding body language through 2D pose estimation. The approach employs an LSTM network to analyze skeletal key points, using a pose representation model trained on publicly available video datasets. Using the OpenPose [77] pose estimator, the system achieved an impressive 99.83% classification accuracy for infant activities. A limitation of this approach is the dependency on reliable ground truth, which requires collecting substantial amounts of raw data, making it a resource-intensive process. This study [79] developed a machine learning-based model to identify multiple infant activities simultaneously using a chest-mounted accelerometer with a low sampling rate, employing multilabel classification with XG-Boost, Random Forest, and Multilayer Perceptron. This method was presented for classifying 14 infant activity types, including scenarios with overlapping behaviors, targeting infants aged 6 to 24 months. Sensor data was collected using a chest-mounted, low-sampling-rate accelerometer commonly used in childcare settings. A multi-label classification model was then developed to enable the automated detection of multiple infant behaviors

with high accuracy. The proposed method achieved an F1 Score of 0.88 for classifying 14 distinct behaviors. A drawback of the current approach is its limited feature set, which needs to account for vertical movements, potentially impacting accuracy. The classification performance could also be enhanced by incorporating features that capture more extended observation periods, accelerometer peak counts, contextual data from preceding and subsequent actions, and time-based characteristics. An innovative early prediction method for cerebral palsy was introduced [47], leveraging MediaPipe's pose estimation. This method is used to observe and analyze detailed attributes of infant movements using classifiers such as SVM and XGBoost. The MediaPipe pose estimation technique was used to analyze detailed attributes of infant movement. This method provided an advanced alternative to the General Movement Assessment (GMA). It captured a range of key CP indicators. This includes postural dynamics, motion diversity, antigravity motion patterns, and left-right coordination. The suggested method achieves classification accuracies of 92% and 97.37% on the MINI-RGBD and RVI-38 datasets, respectively. These outcomes demonstrated the effectiveness of using pose estimation for early forecasting of young-onset cerebral palsy (CP). It emphasized the critical role of tracking temporal variations in joint angles for precise diagnosis and well-informed therapy strategy. A limitation of the current approach is its reliance on a limited dataset and feature set. This may restrict the model's accuracy and generalizability. Additionally, its focus on cerebral palsy (CP) limits its applicability. The model left unexplored potential insights into other conditions, such as Developmental Dysplasia of the Hip (DDH), Ataxia, and Torticollis. Thereby hindering broader advancements in pediatric motor development diagnostics.

Despite their potential, there is a shortfall in detecting falls by integrating ADL classification outcomes for babies in the crib with image-based outcomes. Most of the research also highlighted the importance of sensor data for estimating the baby's pose. Movement sensors embedded in baby wearables, such as accelerometers and gyroscopes, capture motion data. However, these sensors mostly provide limited information about the baby's joint positions and complex postures, which are essential for accurate pose estimation. Pose estimation techniques use image or video data to analyze skeletal key points. However, movement sensors do not inherently provide details about specific joint angles or spatial relationships between body parts. This limitation can restrict their ability to accurately identify and quantify complex postural changes. Thus, pose estimation techniques are becoming more suitable for detailed movement analysis and identification of key points. MediaPipe is the most effective at accurately identifying a baby's pose landmarks, or key points, among all HPE techniques.

4.3 Role of Supervised Machine Learning

In machine learning, Supervised learning is the most common learning method. It teaches the system what to do by assigning each input a label specifying its action. Linear regression, support vector machines (SVMs), random forests (RFs), Naïve Bayes, and other approaches are the most commonly

used supervised learning methods. There are two stages in supervised learning: training and testing. This method uses data and target outcomes as input to train a system. After training, the system is tested with input that lacks target labels. It is an important, quick, and precise method. Regression and classification are the two types of supervised learning [80]. Classifiers are pivotal in distinguishing between data classes by mapping input features (key points) to six predefined ADL categories. These classifiers enhance the system's ability to accurately recognize complex patterns. Thus, they enable highly reliable automated decision-making and activity recognition. To achieve robust performance and scalability, algorithms such as support vector machines and ensemble models (e.g., XGBoost and Random Forest) are used. Figure 4.1 shows a hierarchical overview of supervised learning techniques. The system's ability to accurately identify intricate patterns is enhanced by these classifiers.

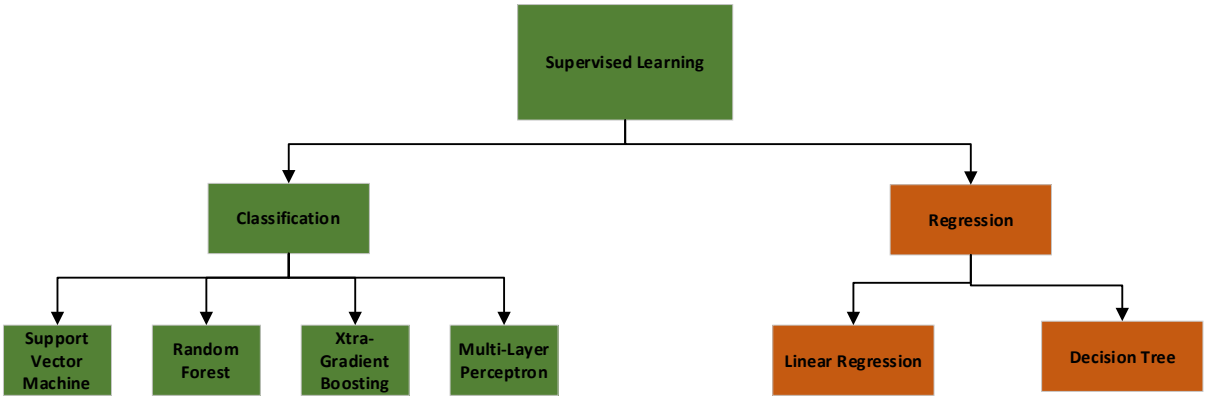


Figure 4.1. Hierarchical overview of Supervised Learning

4.3.1 Applied ML Algorithms/Classifiers

• Support Vector Machine (SVM): It is a model for supervised learning that is mainly applied to classification. It finds the best hyperplane to divide data points in an n-dimensional feature space. An n-dimensional space is reduced to an (n-1) dimensional space via SVM. Both linear and non-linear classifiers are supported by SVM. The linear classifier separates classes by maximizing the margin. The hyperplane serves as the decision boundary for classification [81].

• Random Forest (RF): Another type of algorithm based on supervised learning is RF. This is a widely used algorithm. It is a highly effective classification method for handling large and complex problems. It operates as an ensemble of decision trees. Each tree votes to categorize a specimen. The final data class is determined by selecting the category with the most votes. [82].

• Extreme Gradient Boosting (XGBoost): Extreme Gradient Boosting (XGBoost) is a machine learning method based on gradient-boosted decision trees (GBDT). It is highly scalable, parallelizable, and efficient, often outperforming other algorithms. XGBoost uses regularization to reduce overfitting and

employs gradient boosting with iterative model updates via gradient descent. Its fast-processing speed and multi-threaded CPU optimization make it a robust classifier [82].

• Multilayer Perceptron: One popular neural network model is the multi-layer perceptron (MLP). It is used for regression and classification tasks. Each input is treated as a class in this algorithm, making it suitable for solving binary and multi-class classification problems [82].

4.4 Architectural Overview of Pose-Estimation-Based Baby Unsafe Scenarios Prediction Model

The AI-1 model presents the architectural overview in Figure 4.2 and has been implemented in 5 phases, as shown in Figure 4.3. In the initial phase of data gathering, datasets are created for varying baby activities in a crib. Those datasets are processed before being fed to the pose estimation model in the second data preprocessing phase. These processed datasets extract key points as features during the third feature-extraction stage. Using selected features extracted from the HPE extraction, a few sets of blocks will be created. These blocks will then be fed to the machine learning model in the fourth stage to classify the pose as “Safe” or “Unsafe”. In the last stage, performance metrics are evaluated to further analyze the model's performance on the labelled video.

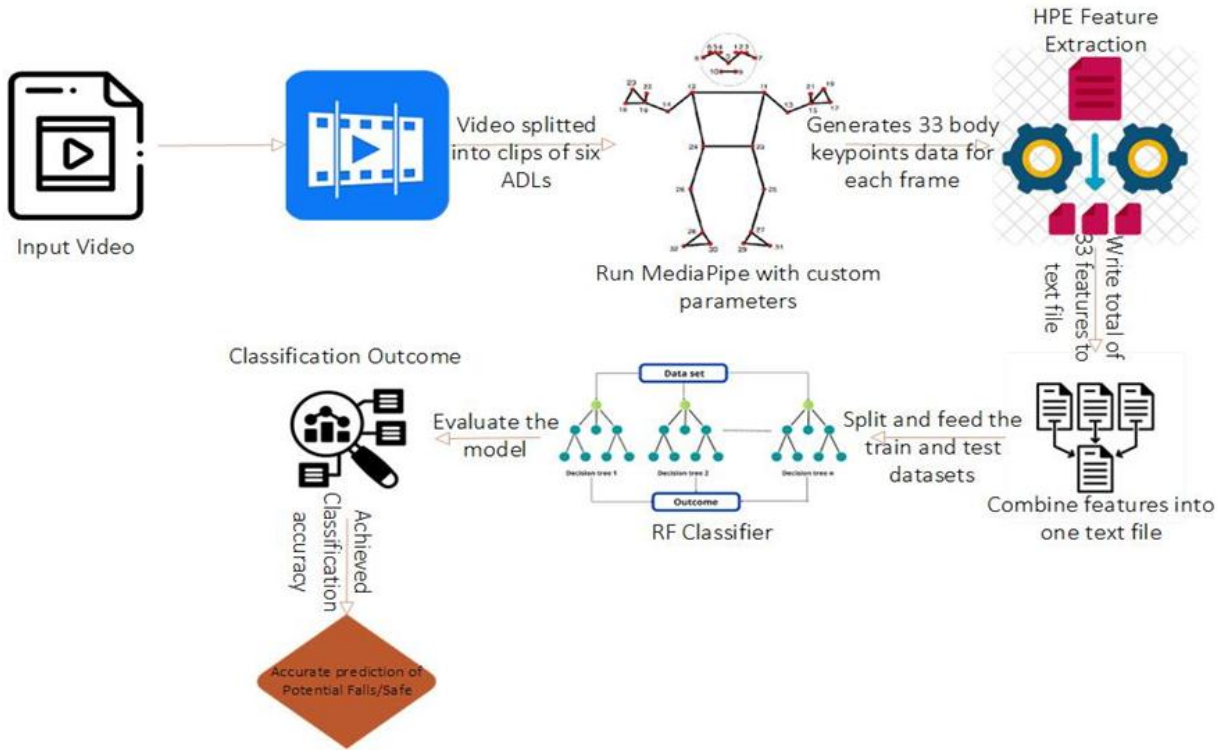


Figure 4.2: Architectural overview of the AI-1 model

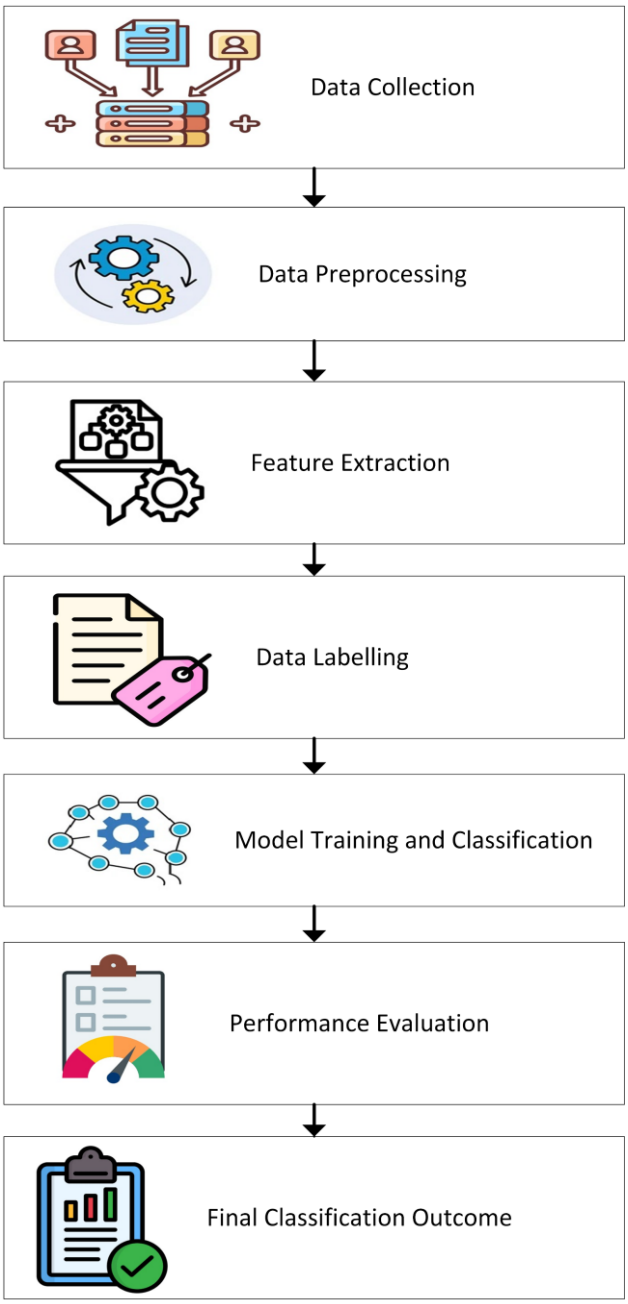
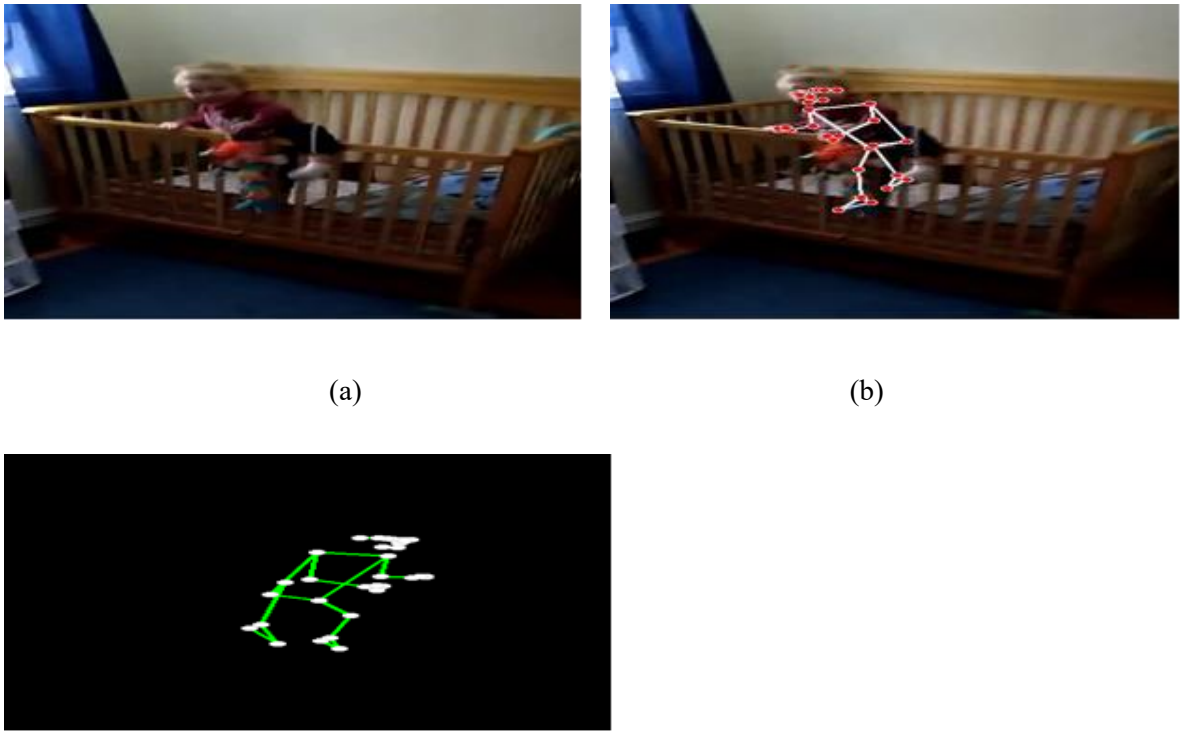


Figure 4.3: Employed system methodology

The employed system methodology, as shown in Figure 4.3, follows a sequential machine-learning classification framework. Initially, raw data are collected from relevant sources and pre-processed to improve data quality and consistency. Subsequently, relevant features are extracted from the pre-processed data (keypoint-based data), and class labels are assigned to the corresponding samples. The labelled dataset is then used to train and evaluate selected machine learning algorithms. Model performance is assessed using appropriate evaluation metrics, and the best-performing validated model is selected to generate the final classification outcomes.

4.4.1 MediaPipe Pose Estimation

One of the HPE methods is used to identify the key points. In this chapter, MediaPipe is used with custom parameters to detect 33 keypoints in each video frame. MediaPipe is a two-step machine learning system for bottom-up, multi-person posture estimation, involving detection and tracking [83, 84]. The system first identifies the frame's region of interest (ROI). Next, within the designated ROI, it identifies key regions and a segmentation mask, as shown in the figure. 4.4 (a), (b), and (c). These key points are considered features for creating training and testing datasets. Before separating them into training and test datasets, the key points for each ADL were saved in CSV (text) files. Later, all CSV files containing key points were combined into a single CSV file using Python code. 80% of the datasets are used for training, and 20% for testing.



(a) (b)

(c)

Figure 4.4. (a). The original frame was extracted from the climbing the fence video clip. (b). Extracted frame with 33 key points. (c). The frame shows 33 key points with the background subtracted.

4.4.2 Classifiers Used for Prediction

We analyzed our data collected using four standard algorithms for pose estimation and fall detection efficiency. (a). SVM [81] (b). RF [81, 82, 85] (c). XGB [82, 86] (d). MLP [87, 88]. We employed keypoints from different combinations of ADL activities to comprehend variations in correctness. This was carried out to comprehend the impact of changing baby poses on their accuracy. We also analyzed which ADLs and Potential Falls were incorrectly detected concerning the performance metrics.

4.5 Algorithm for Pose-Estimation-Based Baby Unsafe Scenarios Prediction Model

Algorithm 1: AI-1: Pose-Estimation-Based Baby Unsafe Scenarios Prediction Algorithm (BUSPA): Pseudocode of the Methodology of the Proposed Model (Figure 4.2)

Input: Video file V,

Split V into clips C = {C1, C2, ……, Cn},

ADL Class Labels: A= {Sitting =0,

Lying down = 1,

Playing = 2,

Standing = 3,

Leaning = 4,

Climbing over fence = 5},

Safety Labels: S = {Safe =0 for classes {0, 1, 2}

Unsafe = 1 for classes {3, 4, 5}}

Supervised Learning Model M

Input Dataset D0

Output: Prediction Outcome: O (Unsafe or Safe)

Evaluation/Performance Metrics: R (Accuracy, Precision, Recall, F1-score)

Phase 1: Initialization

1. Receive clips C one by one.
2. Initialize empty dataset D

Phase 2: Data Preprocessing

Phase 3: Pose Extraction and Label Generation

3. FOR each clip Ci in C D0
    Assign ADL label yi according to clip content
    Assign safety label si based on yi
    FOR each frame Fj in clip Ci D0
        Extract pose landmarks Lj using MediaPipe Pose
        IF landmarks are detected THEN
            Convert Lj into keypoint feature vector Xij
            Append (Xij, yi, si) to dataset D
        END IF
    END FOR
   END FOR
4. Save dataset D into a CSV file
5. Split dataset D into: Training set = 80%
    Validation/Test set = 20%

Phase 4: Training and Evaluation of the model

6. Select supervised learning model M
7. Train M using the training set
8. Validate M using the validation/test set
9. Compute evaluation metrics R
    Accuracy, Precision, Recall, F1-score

Phase 5: Validating/Testing the model

10. FOR each sample in validation/test set DO
    Predict ADL class y_pred
    Predict safety class s_pred
   END FOR

   IF y_pred in {0,1,2} THEN
    O: "Safe"
   ELSE
    O: "Unsafe"
   END IF

11. Return *O* as the Prediction Outcome and *R* as Performance Metrics.

4.6 Data Collection and Ground Truth Selection

Table 4.1 shows that 60 video clips are divided into six categories. Thus, it demonstrates six different activities of daily living (ADLs), further categorized into Safe and Unsafe. The predicted risk levels and the ADLs of the babies that encountered those levels are considered ground truth and categorized as shown in Table 4.2. The baby poses dataset, as specified in Table 4.3, is produced from 60 video segments obtained from websites. such as YouTube and Daily Motion. The long-duration videos of six different ADLs were split using online video split software. The split videos of 3-4 seconds were collected in the respective ADLs as the required videos. Then, Python programming was used to extract the video frames from those video clips., as shown in Fig. 4.4. (a).

Table 4.1: Dataset for Activity Distribution

ADLs	**Total number of videos**
Climbing the fence	15
Leaning	6
Standing holding the fence	15
Lying down	10
Playing inside the crib	5
Sitting	9

Table 4.2: Signs Of Dangerous Activities with Labels While the Baby is in the Crib

S.No.	**Activities**	**Clue Type**	**ADLs Classes**	**Outcomes**
1	Climbing the fence	Unsafe/Potential fall	0	0
2	Leaning	Unsafe/Potential fall	1	0
3	Standing holding the fence	Unsafe/Potential fall	2	0
4	Lying down	Safe	3	1
5	Playing	Safe	4	1
6	Sitting	Safe	5	1

Table 4.3: Specifications of the Baby Pose Dataset

Subject	Applied Deep Learning, Body posture understanding
Specific subject area	Comparison of model accuracy for ADLs and Safe/Unsafe scenario
Type of Data	Image, Video, Pre-processing Python code
Methods of data collection	From web video collections using search terms like standing, climbing, and leaning that indicate potential falls

Description of data collection	Raw data: Some keywords describing the baby's potential fall were climbing, standing, and leaning. Videos from video-sharing sites such as DailyMotion, YouTube, and others were crawled.

4.6.1 Evaluation Metrics

We used a set of 33 key points of 6 ADLs to understand the algorithm's behaviour and variations in accuracy, Precision, Recall, and F1-score. The models are specified using various classification parameters that determine their accuracy. These metrics consist of F1-score, recall (REC), accuracy (ACC), and precision (PRC), as listed in Equations (4.1- 4.5) below:

$$Accuracy(ACC) = \frac{TP+TN}{TP+TN+FP+FN} \tag{4.1}$$

$$Macro\ f1score = \frac{\sum_{i=1}^{n} F1\ score\ i}{n} \tag{4.2}$$

$$f1score = 2 \times \frac{PRC \times REC}{PRC + REC} \tag{4.3}$$

$$Precision(PRC) = \frac{TP}{TP + FP} \tag{4.4}$$

$$Recall(REC) = \frac{TP}{TP + FN} \tag{4.5}$$

True positives, true negatives, false positives, and false negatives are denoted by the letters TP, TN, FP, and FN, respectively.

4.7 Assessment of the Developed PE-ML Model Based on AI

In this work, we assess each classifier's learning performance using five-fold cross-validation (FFCV). The training set is arbitrarily divided into five portions. Four components are used for training, while one is reserved for validation. We initially analyzed the effect of changing the poses of 6 ADLs on ML algorithms' performance, before examining the impact of ADL-based outcomes on that performance. We have employed the common machine learning models (a) SVM, (b) RF, (c) XGB, and (d) MLP, as previously discussed.

The F1-score, Precision, and Recall for 6 ADLs of babies in the crib are shown in Tables 4.4, 4.5 and 4.6, respectively. Analysis shows that among the 6 ADLs, standing achieved the best F1-score, and leaning the worst, as shown in Table 4.4. However, standing leads to being the best and leaning to

being the worst under an Unsafe scenario, whereas sitting leads to being the best and playing to being the worst under a safe scenario when categorizing the ADLs by outcome. Additionally, RF performs best overall as a classifier, achieving the highest F1 Score of 60% for ADL identification. Tables 4.5 and 4.6 compare precision and recall metrics, respectively, for all 6 ADLs categorized as safe or unsafe. It was found that standing (ADL) achieved the best precision of 57% with RF under unsafe outcomes. However, playing and sitting achieved competitive performance under safe outcomes, with 50% and 53%, respectively, using the RF classifier, as shown in Table 4.5. Table 4.6 shows that the recall values are best for standing under unsafe outcomes, whereas for sitting under safe outcomes. Based on the Recall metric analysis, SVM performs better, achieving 72% for standing, and MLP performs well, achieving 73%. A performance comparison across all 6 ADLs has been conducted for our datasets, as shown in Table 4.7. The highest accuracy was obtained with key points as features in the RF classifier, at 51%. As mentioned, all 6 ADLs are categorized into two outcomes: safe and unsafe/potential fall. RF and MLP achieved the best F1-score of 79% and 68% for unsafe/potential falls and safe outcomes, respectively, as shown in Table 4.8. Moreover, the model achieved 75% accuracy under the RF and MLP classifiers, as shown in Table 4.9.

Table 4.4: Comparative analysis of F1-score of keypoint-based classifiers in FFCV under ADLs

Models	**Unsafe/Potential Fall**			**Safe**		
	ADL_0	**ADL_1**	**ADL_2**	**ADL_3**	**ADL_4**	**ADL_5**
	Climbing	**Leaning**	**Standing**	**Lying Down**	**Playing**	**Sitting**
SVM	0.35	0.00	**0.49**	0.22	0.13	**0.49**
Random Forest	**0.49**	**0.35**	**0.60**	**0.48**	**0.37**	**0.56**
XGBoost	0.45	0.35	**0.60**	0.42	0.37	**0.56**
MLP	0.42	0.06	**0.51**	0.39	0.35	**0.56**

Table 4.5: Precision Comparison of Keypoint-Based Classifiers in FFCV Under ADLs

Models	**Unsafe/Potential Fall**			**Safe**		
	ADL_0	**ADL_1**	**ADL_2**	**ADL_3**	**ADL_4**	**ADL_5**
	Climbing	**Leaning**	**Standing**	**Lying Down**	**Playing**	**Sitting**
SVM	0.39	0.00	**0.37**	0.43	**0.43**	**0.38**
Random Forest	0.49	0.39	**0.57**	0.48	**0.50**	**0.53**
XGBoost	0.48	0.41	**0.55**	0.43	**0.45**	**0.51**
MLP	0.44	0.43	**0.43**	0.41	**0.50**	**0.49**

Table 4.6: Recall Comparison of Keypoint-Based Classifiers in Five-Fold Cross-Validation Under ADLs

Models	**Unsafe/Potential Fall**			**Safe**		
	ADL_0	**ADL_1**	**ADL_2**	**ADL_3**	**ADL_4**	**ADL_5**
	Climbing	**Leaning**	**Standing**	**Lying Down**	**Playing**	**Sitting**
SVM	0.31	0.00	**0.72**	0.15	0.08	**0.69**
Random Forest	0.49	0.31	**0.65**	0.48	0.30	**0.60**
XGBoost	0.44	0.30	**0.66**	0.41	0.31	**0.63**
MLP	0.39	0.03	**0.62**	0.38	0.27	**0.73**

Table 4.7: Comparative Study for the Average Performance Of 6 ADLS

Models	**Precision**	**Recall**	**Accuracy**	**F1-score**
SVM	0.34	0.32	**0.38**	0.28
RF	0.49	0.47	**0.51**	0.48
XGB	0.47	0.46	**0.49**	0.46
MLP	0.45	0.40	**0.45**	0.39

Table 4.8: Comparison of F1-Score for the ADL-Based Outcome

Models	**Outcome**	
	Unsafe/Potential Fall	**Safe**
SVM	0.78	0.65
RF	**0.79**	**0.68**
XGB	0.78	0.67
MLP	**0.79**	**0.68**

Table 4.9: Comparison of Average Accuracy for the ADL-Based Outcome

Model	**Accuracy**
SVM	0.73
RF	**0.75**
XGB	0.74
MLP	**0.75**

4.8 Conclusions

Although multiple HPE techniques [3] have been used in previous research, more evidence is needed to identify accurate models grounded in key points. In this chapter, we examined and contrasted how

the accuracy of several popular machine learning methods was affected by incorporating sensor or keypoint position data from the HPE posture estimator. The system is modelled with four classifiers, SVM, RF, XGB, and MLP, to achieve the best accuracy and retrieve 75% with RF, since Random Forest's decision trees perform effectively on large and small datasets [82]. These results are still not promising because of the size of the datasets. Hence, the dataset size should be increased to further improve the model's accuracy. In conclusion, this paper identifies the optimal human pose estimation (HPE) classifier for a baby fall-prevention and prediction system. In future work, the system aims to provide caregivers with real-time alerts to mitigate the risk of unsafe scenarios, such as babies falling out of cribs. As a result, we intend to create a Deep Neural Network model that uses the Soft spatial attention mechanism to leverage the spatial cues. This model will predict safe and unsafe scenarios using two approaches: frame-based (time-independent) and statistical parameter-based (time-dependent). We will be able to predict scenarios more accurately, resulting in a more efficient model. Furthermore, the system's predictive capabilities can reduce caregiver anxiety by providing reassurance through proactive monitoring and early warning.

Chapter 5. Proposed AI-2: CNN-Attention-Based Baby Unsafe Scenarios Prediction System

5.1 Introduction

In the current socio-economic climate, rising inflation and a declining employment rate of 67.4% have imposed significant financial pressures on households [89, 90]. As a result, both parents are increasingly engaged in professional and entrepreneurial pursuits to meet financial demands. This change has created considerable challenges in balancing work and family responsibilities, particularly impacting the care and supervision of babies. Babies aged between 6 and 24 months are especially vulnerable, as they require continuous attention, affection, and care during this critical phase of rapid growth and development. At this age, babies exhibit increased curiosity and motor activity, actively exploring their surroundings, including the crib environment. Common behaviours, including jumping, leaning over, and climbing on crib fences, pose safety hazards when supervision is inconsistent [16]. The reduced presence of caregivers during critical periods of physical and cognitive growth poses serious implications for child safety. Without consistent monitoring, infants are more likely to engage in behaviours that may lead to falls, entrapments, or other injury-causing events, especially in home environments, such as cribs, which are often assumed to be safe. The inability to provide timely intervention in these scenarios substantially increases the danger of falling.

The most common cause of serious injuries and fatalities among adults 60 years of age and older and children less than four is fall. According to the Starship Foundation in New Zealand [20], 31% of falls occurred at home and required hospitalization. The most affected age group is 1-4 years, due to falls from slips, trips, and stumbles, as well as from falling out of bed. The study says that mostly 0-4 years accounted for around 51% of falls that needed hospitalization, in which boys are 1.5 times more likely to be injured than girls, and urban children. Though the rate of hospitalizations for fall injuries has declined about 19% over the past 10 years, more than 3 children are hospitalized every day from falls, with 110 children hospitalized due to falls from beds, and 49% of traumatic brain injuries are caused by falls.

Among elderly people, extensive research has been conducted on fall prediction, detection, and prevention [91]. However, our research focuses on fall prediction and prevention of babies aged 12-24 months. Research on babies' safety around the home has recently led to a breakthrough, with most research undertaken using sensors, AI, or a fusion of technologies, including accelerometer-based [87] and wearable-sensor-based [92] approaches. Accelerometer-based systems use motion sensors to measure an individual's motion. While mitigating potential privacy risks and LOS effects, this method faces challenges, including false positives and false negatives, data noise and interference, and limited

context awareness. However, wearable sensor-based approaches capture detailed motion data and kinematic parameters by integrating accelerometers, gyroscopes, and magnetometers within body-worn devices. These systems enable mobility and support continuous, real-time monitoring, making them highly suitable for fall detection applications. However, body-worn devices face challenges in terms of cost and maintenance, data transmission and connectivity issues, and concerns about data reliability due to signal noise and motion artifacts. Limited battery life and continuous sensor wear can cause user discomfort, skin irritation, or inconvenience.

In recent years, conventional machine learning (ML) techniques, including random forests, multilayer perceptrons, extreme gradient boosting, and support vector machines, have been widely used in fall detection applications. Subsequently, these approaches have been integrated with pose estimation techniques to facilitate the identification of individual postures. Despite their merits, such methods often rely on expert knowledge, exhibit domain-specific limitations, and are frequently time-intensive due to the need for expert-driven feature extraction. To overcome these issues, research has focused on the direction of DL models. DL models, such as CNNs, are effective at capturing local features and learning hierarchical representations. Although a conventional CNN achieves translation invariance through pooling, important fine-grained spatial details can be lost during pooling and convolution, thereby reducing sensitivity to subtle yet significant cues. If all spatial regions have uniform focus across the image, potentially overlooking critical areas relevant to specific tasks, such as fall detection or pose estimation, and reducing interpretability, fall prediction accuracy will be compromised. To address this concern, this study [93] proposed a CNN-based self-attention approach to extract importance weights (weights representing the degree of association between the applied signals and the occurrence of disequilibrium) required, along with the balance score calculated from IMU sensors, to predict fall events. However, the target was adult and integrated with sensors. Thus, AI-2 is motivated by an approach that aims to overcome these limitations and provide a reliable and efficient Baby Unsafe Scenarios Prediction System (BUSPS).

In this chapter, the motivation behind developing an attention-based deep neural network architecture is to acknowledge that Pose Estimation (PE) or PE+ML methods address only part of the baby's safety concerns, both inside and outside the home, while we focus on ensuring crib safety for babies and supporting parental peace of mind. Thus, this chapter proposes a deep learning-based approach to enhance in-crib baby safety by classifying behaviour using a customized 2D Convolutional Neural Network (CNN) architecture. We presented a unique dataset and proposed a model, the SSCNN (Soft Spatial Convolutional Neural Network), a 2D CNN architecture with a soft self-attention mechanism for extracting and classifying safety-related features.

5.2 CNN-based Monitoring

Several studies have been conducted in the field of baby safety monitoring using deep learning techniques. Deep learning approaches are the most prevalent in the AI world, encompassing methods such as machine learning, pose estimation, and more. Recent deep learning-based systems for monitoring baby and toddler safety primarily employ advanced vision models, pose analysis, and IoT-enabled frameworks to detect falls and facilitate real-time intervention. CNNs have particularly come into the picture as leading technologies, with transfer learning, pose estimation, and additional LSTM layers or transformer models to classify dangerous scenarios involving babies. Yang [4] presents a study that integrates an improved YOLOv8 model with the GELAN architecture and GHT-based safe-zone estimation to robustly detect toddler falls, achieving 96.33% accuracy and high real-world applicability despite challenges such as occlusions and variable lighting conditions. Another introduced approach [61], combines human pose estimation using OpenPifPaf with an LSTM classifier to identify different classes of falls in real-time, achieving a 93% F1-score, but noting performance drops with ambiguous postures or lost skeleton points. [18] presented the "Smart Baby Care" monitoring system for smart homes and childcare centers. AR task, dataset, and a ConvLSTM-I3D model was proposed, achieving 85-87% test accuracy in action recognition. Madhu [94] further extended this approach by blending deep action recognition, IoT/edge deployment, and multimodal learning. For instance, ConvLSTM-I3D networks for event recognition and CNN-BiLSTM or YOLOv8-OpenPose-LSTM pipelines for monitoring cry, movement, and sleep. Thus, achieved above 97% accuracy across several subtasks and providing instant alerts to caregivers. In the domain of environmental injury prevention, a curated dataset was introduced [48] , namely ChildSUn. They used deep CNNs, such as EfficientNet-B0, to classify objects as hazardous or safe. Thus, achieved nearly 99.6% accuracy in supporting proactive robotic responses, though without direct event detection. In this study, Amisha [95] modified YOLOv8 for fall detection, utilizing a multi-feature approach to IoT/edge/ML-based daycare safety.

Across these methods, most systems emphasize real-time performance, strong generalization, and comprehensive scenario coverage. However, the key research gap lies in addressing common limitations, such as hardware requirements, sensor calibration. In addition, these studies focus on elderly people, generic HAR tasks such as drinking, cooking, and walking, and rely on pose estimation and fall detection. Most of them do not address crib falls or unsafe crib scenarios, either, with a *soft spatial CNN* for 12–24-month-old babies. This motivated our contribution to develop a novel approach, BUSPS, which leverages the strengths of soft spatial attention (SSA) mechanisms with a 2D CNN.

5.3 Importance of ADLs and Parameter Selection

A baby aged 6-24 months is in a period of major developmental growth and increasing motor skills, which can contribute to hyperactivity. When left unattended, this can lead to sudden falls out of the crib and cause serious injuries. This leads to sometimes fatal injuries among babies [96]. Many studies have

focused on continuously monitoring babies' ADLs in recent decades [65-67, 97]. By improving safety and quality of life and enabling early identification of key events, such as falls, that prompt appropriate action, ADL monitoring can significantly enhance infant care. Long-term monitoring also provides parents and other carers with important information. This enables them to respond quickly and seriously to updates on the baby's activity. A strong grasp of the significance of ADLs in dangerous situations, such as fall-risk behaviour prediction, is informed by the background studies conducted in [4] and [13]. These studies explored monitoring safe and risky behaviours among newborns. As a result, this study focuses on six essential ADLs that help predict unsafe scenarios in babies in the crib. Figure 5.1 depicts the ADL parameters chosen for this proposed research. These six ADLs are fundamental to a baby's basic functioning.

- Sitting
- Lying down
- Playing inside the crib
- Leaning over the fence
- Climbing over the fence
- Standing near the fence

This chapter focuses on predicting dangerous situations for infants during their normal crib activities. This proposed work excludes activities such as jumping, as they may result in the same ADL parameter as standing. Section 5.6 and its parts provide detailed information on how the dataset was assembled and collected. Then, the SSCNN model uses these ADLs as input to predict which dangerous situations might lead to falls.

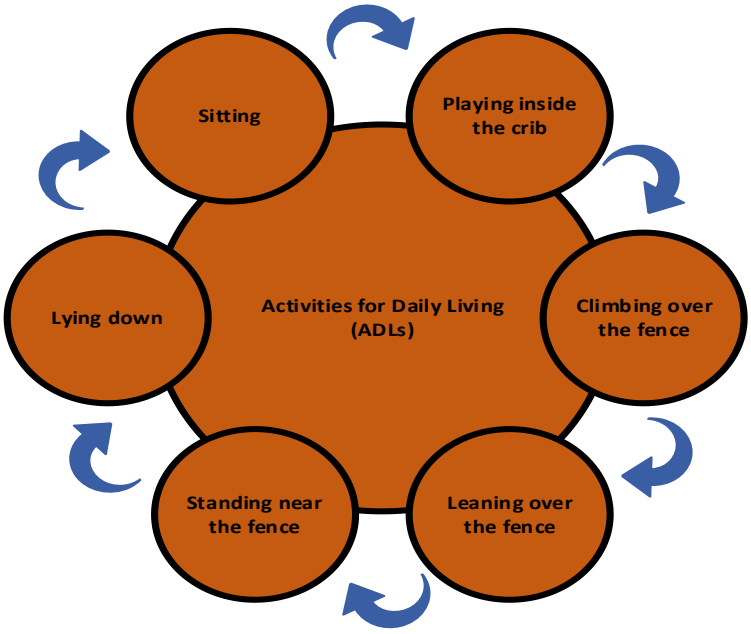


Figure 5.1 ADLs Parameter Selection

5.4 Convolutional Neural Network (CNN)

Deep Neural Networks (DNNs), Convolutional Neural Networks (CNNs), and Recurrent Neural Networks (RNNs) are some of the supervised learning techniques for Deep Learning (DL) [98]. One family of NNs that is commonly used for processing digital data images is the Convolutional Neural Network (CNN or ConvNet). CNN is a computer vision deep learning network that can identify and categorise features in images [99]. CNN architecture was impacted by the composition and operation of the visual cortex. It is meant to resemble how neurons in the brain connect to each other. The CNN architecture has a very significant layer called the convolutional layer. As mentioned in Sections 5.5.2 and 5.6, it takes an image as input and uses a 3x3 or 5x5 filter. A CNN has several parts that make it up. The first layer is the input image, followed by the convolutional layer, the pooling layer, the activation function, the fully connected layer, and finally the output layer [99]. Every training image has filters applied at different resolutions. Tens or even hundreds of layers can make up a CNN, each of which learns to recognise different characteristics of a frame. Each convolved image's result serves as the input for the layer after it [100]. Thus, CNN architectures have attained significant outcomes on diverse computer vision applications, including image classification [101] and object detection. However, the model often suffers from overfitting, in which it overfits to meaningful patterns and spurious noise in the training data, learning them excessively well yet generalizing poorly to unseen or validation datasets [102]. In the supervised learning paradigm, labelled datasets made up of input-output pairs are used to train models. If a model is too complex or trained for too long, it may memorize the training samples, learn noise or irrelevant information, and fail to provide an accurate inference on unseen data. Common causes of overfitting include overly complex models (e.g., deep neural networks with many layers), limited training data, excessive feature extraction, or prolonged training. Overfitting can be identified by comparing performance metrics; a gap indicates the model is not generalizing well. Thus, a lack of generalization [103] leads to unreliable predictions in real-world scenarios. If the training data is not representative, it may reinforce biases and undermine transparency in decision-making. If models are not properly validated and monitored against evaluation metrics, such as validation accuracy and loss, it can perpetuate systematic inequalities. Deployment failures can occur, especially in healthcare, finance, and autonomous systems. While measures such as accuracy, F1-score, precision, sensitivity, and specificity are frequently used to evaluate AI/ML models, these metrics can sometimes be misleading and lead to socially harmful or unfair outcomes. Poor performance in minority classes may be masked by high overall accuracy, particularly in unbalanced datasets. Thus, several standard techniques for reducing overfitting are explained in Section 5.5.1.

5.4.1 Standard Techniques for Reducing Overfitting

Although AI and ML play a critical role in real-world applications, models, particularly CNN architectures, remain subject to criticism due to overfitting. Overfitting poses a significant obstacle to

practical model training. DL's significance was discussed [98] and outlined various kinds of networks and DL approaches. This study described the development and key features of the commonly used CNN architecture. This research also highlighted problems and solutions to aid researchers in comprehending the current research gap, with a focus on overfitting. Various studies have explored different methods to reduce overfitting. [102] addressed the overfitting issue in CNNs in practical applications. To reduce overfitting, the author suggested an approach that combines maximum pooling, dropout, and weight attenuation. The proposed method effectively avoided overfitting and reduced the average error rate in classifying the dataset by more than 10%. Compared to alternative approaches, the performance of deep learning architectures are enhanced. They were especially designed for next-generation networks' data processing and management. [103] has demonstrated that the proposed noise injection method may provide marginally better protection against overfitting than the conventional early stopping approach. However, the author noted that implementing this method in real-world applications remains challenging. This study was conducted in the Annapolis River catchment, characterized by significant seasonal environmental variations. Thanapol [104] applied various combinations of techniques, including data augmentation alone, data augmentation with dropout, data augmentation with batch normalization (BN), and a combination of all methods The optimal performance occurred at epoch 30, by employing noise injection in conjunction with spatial data augmentation (height and width shifts), and dropout, effectively mitigating the effects of overfitting. [105] proposed a method for improving CNNs using Adam optimizer, drop out, and Batch Normalization (BN). Using CIFAR-10 as the benchmark, the proposed model outperforms the traditional CNN. The author demonstrated improved model performance by incorporating BN and dropout layers, achieving training and testing accuracies of 0.6904 and 0.6861, respectively, with corresponding training and testing losses of 0.8910 and 0.9136. In supervised learning, overfitting is mitigated through a combination of model design, training strategies, and data handling techniques. There are multiple effective methods to reduce overfitting.

5.4.1.1 Data Augmentation

This technique [104] is utilized to artificially expand the training dataset's size and variety. Common data augmentation methods include rotating, flipping, cropping, and injecting noise into images; temporal jittering for time-series data; and paraphrasing text. These techniques enhance model robustness and reduce the likelihood of overfitting to specific patterns. In this study, we applied rotation, flipping, height and width shifting, image resizing, and zoom range adjustments.

5.4.1.2 Increasing Training Data

The model will get more effective with more training data, enabling it to learn the underlying patterns rather than memorize noise. We have increased the dataset from 5000 to approximately. 18000 images.

5.4.1.3 Simplify the model

Network compression reduces the likelihood that the model will retain training data by reducing the number of layers or parameters in a neural network. Also, applying feature selection techniques that use fewer features will be an additional step toward simplifying the model. We reduced the number of layers from four to three, adjusted the learning rate, and replaced the Adam optimizer with Stochastic Gradient Descent (SGD).

5.4.1.4 Batch Normalization

It normalizes inputs to each layer in a neural network and helps stabilize and speed up the model's training. It can have a slight regularizing effect. We have introduced batch normalization when building the model [105].

5.4.1.5 Regularization

This technique introduces a penalty into the loss function to prevent over-parametrization and improve generalization. In neural networks, this is often implemented as a weight decay term [106].

- L1 Regularization (Lasso): promotes sparsity, which causes a few weights to drop to zero.
- L2 Regularization (Ridge): makes the model simpler by penalizing higher weights.

We have used L2 = 0.001 for improvement.

5.4.1.6 Dropout

This technique randomly drops the neurons during training, preventing co-adaptation and encouraging the network to generalize [104, 105]. In our case, we have reduced the dropouts from 0.5 to 0.3.

5.4.1.7 Cross-validation

During training, validation loss is monitored by this trigger. It stops when it increases to prevent the model from unintentionally modeling the noise inherent in the training dataset [107]. We have used this technique with patience=5 and verbose=1 to wait for five epochs for improvement before stopping, and print a message when stopping, respectively.

5.4.1.8 Early Stopping

During training, this trigger monitors validation loss and stops when it increases, keeping the noise present in the training dataset from inadvertently being modelled by the model [7]. We have used this technique with patience=5 and verbose=1 to wait for five epochs for improvement before stopping, and print a message when stopping, respectively.

5.4.1.9 Ensemble Methods

This is helpful in the case of using multiple models, combining predictions, either bagging, boosting, or stacking, to reduce variance and improve generalization [108].

5.4.1.10 Transfer Learning

To reduce the risk of overfitting, especially with small datasets, a pretrained model is used and fine-tuned on the given dataset [16].

5.4.2 Benchmarking CNN

In supervised learning, overfitting is mitigated through model design, training strategies, and data handling. One effective way to improve model performance is to prevent or reduce overfitting. Several techniques can be used, such as early stopping [7, 8], cross-validation, regularization (L1, L2) [9], dropout in neural networks, data augmentation [10], and simplifying the model architecture. Roseline [11] proposed an enhanced CNN method based on Batch Normalization (BN), Dropouts, and the optimizer (Adam) to address overfitting in Conventional CNNs, particularly when facing constraints in extracting convolutional features. Aliferis [109] has examined several common scenarios where overfitting played an important role in a model's performance metrics. Roth [110] presented an extensive literature survey to define specification metrics leading to overfitting. Pothuganti [111] describes overfitting as a critical issue that significantly undermines the model's performance. Previous studies have focused on reducing overfitting and improving the performance of CNN models. However, they often lack experimentation with large datasets and application-oriented model improvements. This study's main goal is to emphasize the significance of overfitting on model performance, identify its causes, present common mitigation techniques, and discuss the performance of our benchmark CNN (BCNN-BGRDOS) model [112]. This study validates the model by augmenting the training data and fine-tuning its parameters. We initially used the traditional CNN model shown in Figure 5.2 and later applied regularization and early stopping. Finally, the model was validated using a training dataset for RGB images.

Approximately 22,000 (224 x 224) RGB images are collected from the online platform. It is further divided into two classes, Safe and Unsafe, having 11000 and 10,500 frames, respectively. The dataset is divided into training and validation subsets, consisting of 18,000 and 3,500 images, respectively. We have used the Python Programming language in the Jupyter Notebook IDE to develop the eight-layer improved CNN model. The established model's performance is evaluated in a TensorFlow environment on a 64-bit Windows 11 machine, equipped with an Intel® Core™ i7-1185G7 (11th Gen) processor running at 3.00GHz.

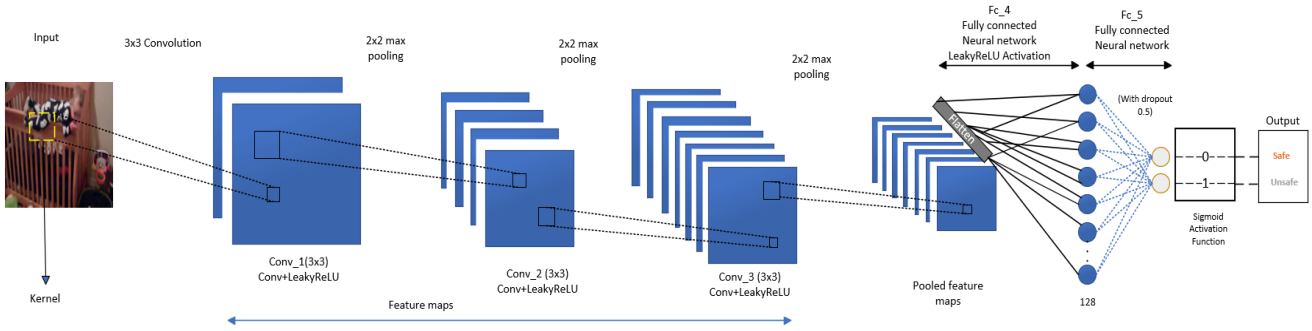


Figure 5.2 Traditional CNN (TCNN) architectural diagram

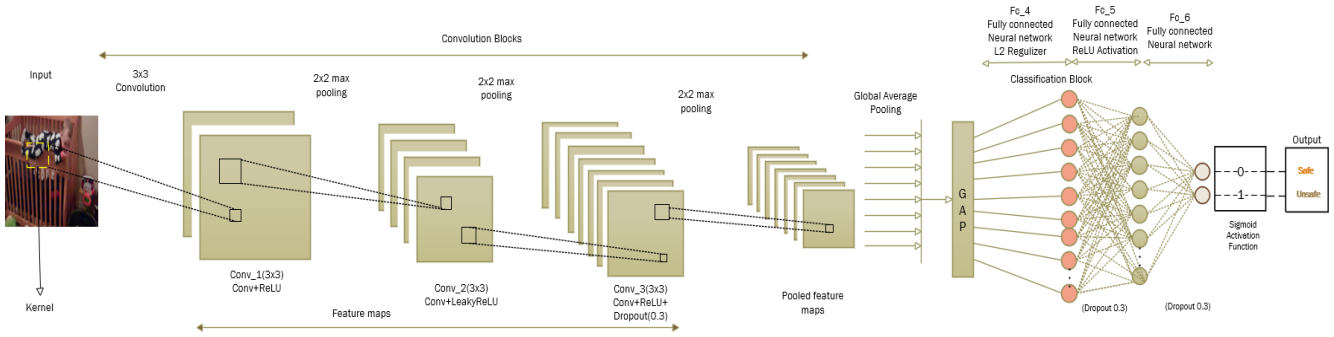


Figure 5.3: Architectural diagram of Benchmark CNN with BN, Global Average Mapping, Regularization, Dropouts, and SGD optimizer (BCNN-BGRDOS)

5.4.2.1 The Proposed Architecture

In this study, an eight-layer CNN model is developed, as shown in Figure 5.3, comprising,

- an input layer incorporating BN followed by a max-pooling layer,
- 3 hidden layers (blocks), two consisting of convolution and pooling layers with BN, and the third comprising BN with a DO layer (0.3),
- a global average pooling (GAP) layer,
- a fully connected block with two dense layers and a dropout layer after each (0.3),
- an output layer utilizing sigmoid activation.

The kernel in the first dense layer of the fully connected block is regularized using L2 regularization with a factor of 0.001. Apart from the second hidden layer (block), which uses LeakyReLU, all other layers use the rectified linear unit (ReLU) activation function (AF). The Conv2D layer applies two-dimensional convolutional operations to extract spatial characteristics from frame data. The BCNN-BGRDOS model incorporates Batch Normalization (B), Global Average Pooling, Regularization (R), Dropout (DO), and the Stochastic Gradient Descent (SGD) Optimizer. It is based on the TCNN architecture, the ReLU activation function, and an approach to reduce overfitting. approach to reducing

overfitting relies on BN, GAP, regularization, and SGD. After feature extraction, a 2x2 pooling window is applied to maintain consistent input and output dimensions. ReLU and LeakyReLU activation functions are used across different CNN blocks to activate neurons.

5.4.2.2 Experimentation and Evaluation Results

The proposed CNN model was trained using early stopping, ReduceLROnPlateau, and a fixed number of epochs. The model was run for 20 epochs, with validation loss used as the monitoring metric for both early stopping and ReduceLROnPlateau. When the learning rate plateaued, the minimum learning rate was set to 1e-6. Table 5.1 shows the parameters used to execute the improved model. By comparing the metrics of TCNN and BCNN-BGRDOS, we minimize the difference between the training and validation losses, as shown in Table 5.2. Figures 5.4 (a) and (b) illustrate the correlation between training and validation loss and accuracy of the TCNN model, respectively, and Figures 5.4 (a) and (b) show the correlation between training and validation loss and accuracy of the BCNN-BGRDOS model, respectively. The comparative look at models' performance on the validation dataset, as shown in Table 5.3, includes BN, GAP, R, DO, and SGD. As shown in the graphs below, the major spikes indicate significant fluctuations in TCNN's loss and accuracy, which the BCNN model mitigates. As shown in Figs. 5.5 (a) and (b), the accuracy and loss fluctuated only during the major spikes between epochs 2.5 and 7.5. Thus, we have achieved a 99.68% improvement in validation accuracy and reduced the gap between validation loss (1.93%) and training loss (2.71%), as shown in Table 5.2. In addition to that, BCNN-BGRDOS outperformed TCNN by achieving an accuracy of 99.68%, precision of 100%, sensitivity of 99.35%, specificity of 100%, and F1-score of 99.67%, as shown in Table 5.3.

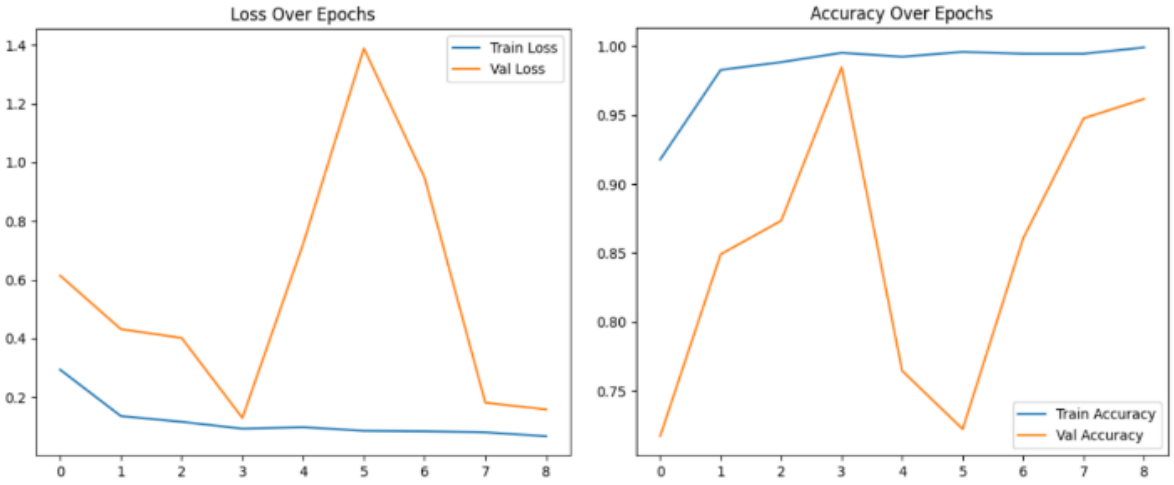
Thus, we developed a benchmark CNN based on a traditional CNN, a corrected version for predicting baby safety, as proposed and implemented in the article currently in press. Techniques like data augmentation, early stopping, and ReduceLROnPlateau are used to reduce overfitting. These strategies resulted in a substantial improvement in validation loss, from 0.4544 to 0.0193.

Table 5.1: Parameter Configuration for the Implementation of the Model

Specifications	**Values**
No. of Epochs	20
Batch Size (16-32)	32
Validation set	15
Training set	85
Rate of Optimizer Learning	0.01
Momentum	0.9
Loss_operation	Cross-entropy in binary

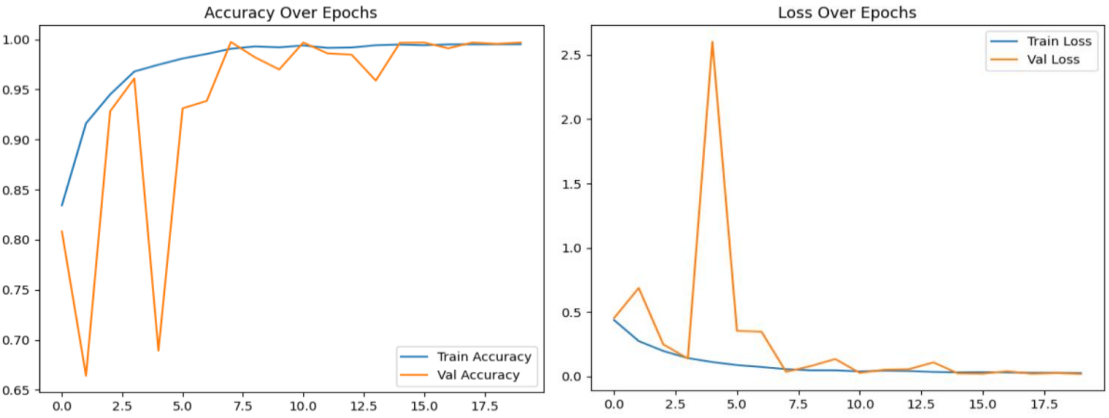
Table 5.2: Comparative Analysis of Training and Validation of the Models

Dataset	**TCNN**		**Proposed BCNN-BGRDOS**	
	Accuracy (%)	**Loss (%)**	**Accuracy (%)**	**Loss (%)**
Training	99.66	8.65	99.47	2.71
Validation/Test	88.67	74.25	99.68	1.93



(a) (b)

Figure 5.4 (a) Training and Validation Loss of the TCNN Model (b) Training and Validation Accuracy of the TCNN Model



(a) (b)

Figure 5.5: (a)BCNN-BGRDOS Training and Validation Accuracy (b) BCNN-BGRDOS Training and Validation Loss

Table 5.3: Model performance: A comparative look

Evaluated Metrics (%)	**TCNN (%)**	**Proposed BCNN-BGRDOS (%)**
Accuracy	92.45	99.68
Precision	98.37	100.00
Recall/Sensitivity	90.25	99.35
Specificity	96.94	100.00
F1-score	94.13	99.67

5.6 Proposed CNN-Attention-Based Baby Unsafe Scenarios Prediction System (BUSPS)

This section presents our unique dataset and introduces a new hybrid neural architecture—SSCNN (Soft Spatial Convolutional Neural Network)—that leverages a 2D CNN with a soft spatial attention mechanism to extract and classify safety-related cues and features from benchmarks. As shown in Figure 5.6, the general structure of the baby-unsafe-scenarios prediction system has been outlined. Our approach is illustrated in the Figure. 5.7, which outlines the entire system. The system is divided into three steps to implement the proposed methodology. First, we analyze the collected dataset, then examine the deep learning architecture, and assess overfitting to validate the model. Finally, we detailed the training and evaluation of the proposed model.

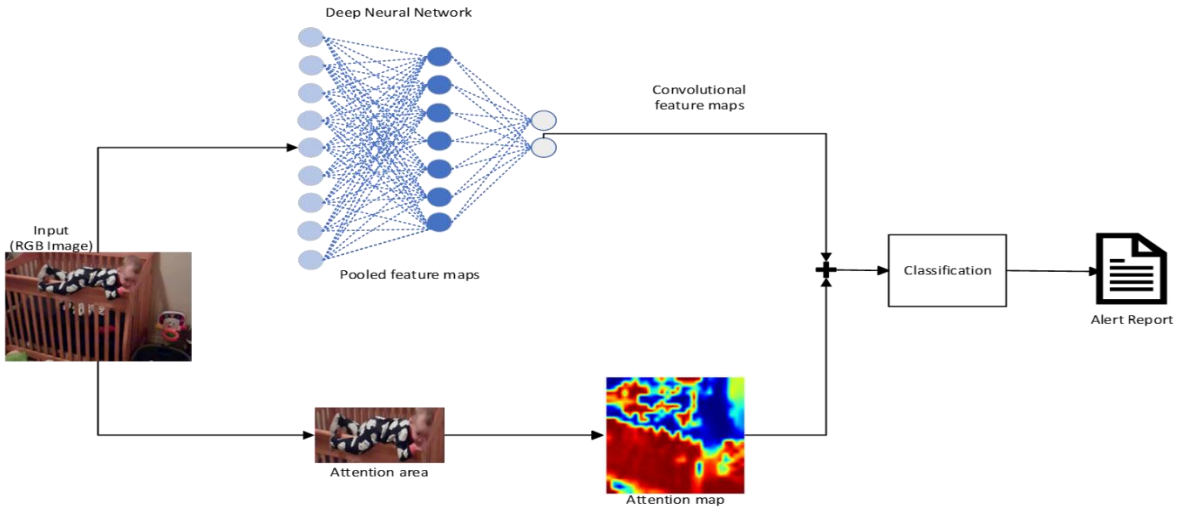


Figure 5.6: General structure of CNN-Attention-based BUSPS

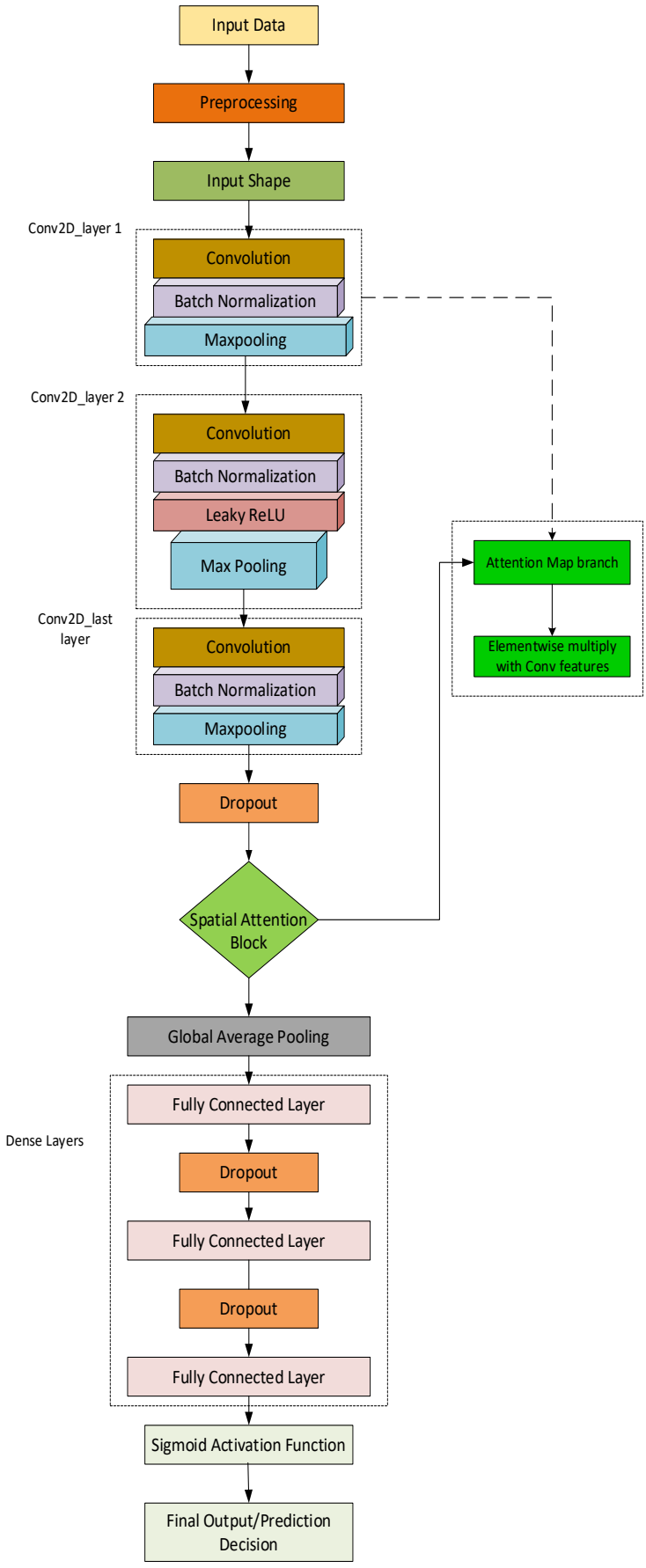


Figure 5.7: Methodology of SSCNN outlining the entire system

5.4.3 Soft Spatial Attention (SSA) Architecture

To enhance model performance, a spatial attention layer with a soft self-attention (SSA) mechanism is added to SSCNN after the CNN blocks. The architecture consists of a block designed as a single two-dimensional convolutional layer with a size of kernel 1 (1x1). The output of this block is then retrieved using a sigmoid activation function, shown in Figure. 5.8. In this architecture, input feature maps are fed to a single two-dimensional convolutional layer and a sigmoid activation function, extracting conv features. From those extracted features, an output attention map is created by multiplying the features elementwise, yielding fused features. Thus, attention mechanisms prefer soft masking during training to improve differentiability, especially when the output range is continuous.

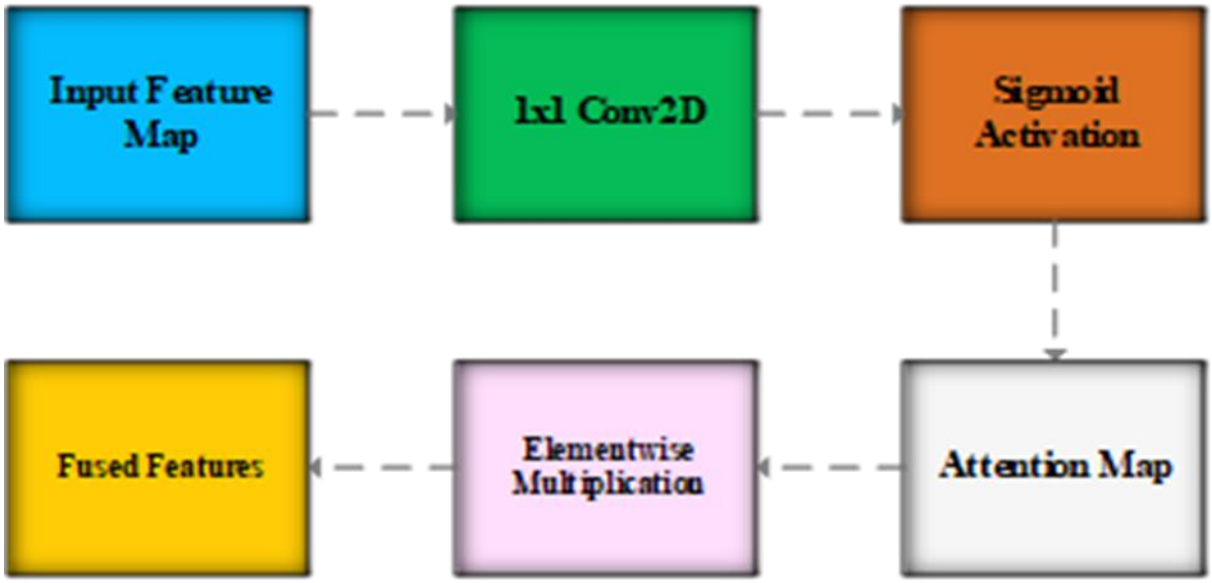


Figure 5.8: A detailed illustration of the SSA mechanism

5.4.4 CNN-SSA

The proposed model, illustrated in Figure. 5.9 comprises a customized CNN, BCNN (Benchmark CNN)-BGRDOS, a modified version of the traditional CNN [112]. It incorporates an SSA mechanism that emphasizes where to focus. CNN-SSA to extract CNN features, focusing on relevant local information, such as edge, texture, and colour context. To comprehensively represent the input data, we concatenate the CNN features with the spatial attention map to form combined_features. During the forward pass, the fused_features are processed through the meta-model's layers. At the last layer, the sigmoid activation function anticipates the input class by processing the normalized output from the preceding dense layer. The suggested design employs the Stochastic Gradient Descent (SGD) optimizer to reduce the training-related binary cross-entropy loss. Training proceeds through repeated cycles of forward and backward propagation, continuing until the model achieves the predefined accuracy, thereby ensuring effective learning of the task. Early stopping and ReduceLROnPlateau are employed

to avoid overfitting by monitoring the model metric 'accuracy'. The loss function for training the CNN-SSA model is defined in (5.1):

$$BCE = -\frac{1}{N}\sum_{i=1}^{N}[y_i \,.\, (\log p_i) + (1 - y_i) \,.\log(1 - p_i)] \quad (5.1)$$

Compared with lightweight or transformer-based attention models and generic spatial-attention CNNs, SSA performs well on medium-sized datasets that are neither too large nor too small. Additionally, SSA thus provides localized cues for small objects with limited data and computational budget, making it suitable for baby crib safety. Henceforth, rather than just a standard attention block, SSA placement, feature fusion, and training strategies collectively prove novelty in building up a BUSPS because BUSPS utilize customized scenario-based datasets, leverage the detailed spatial cues from the input images by fusing the conv features and attention map, evaluate the model, are trained with extensive strategies to overcome overfitting, and then predict the output by classifying as safe or unsafe.

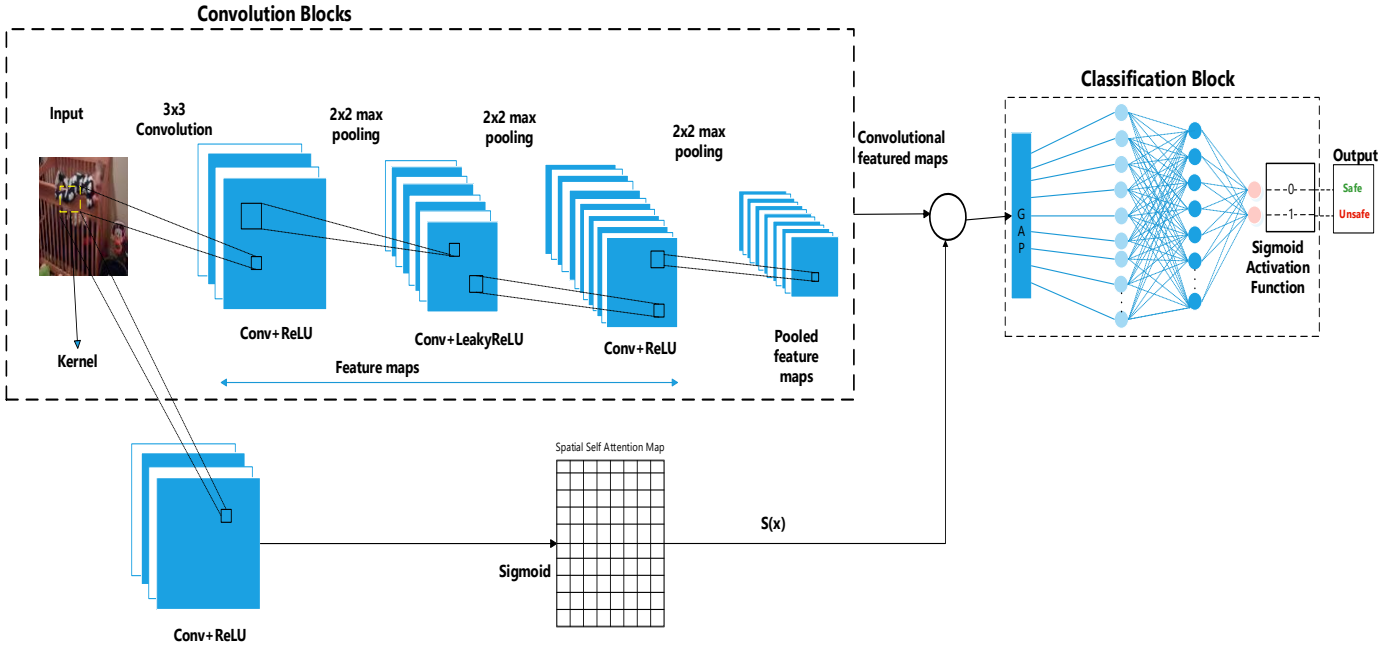


Figure 5.9: Proposed model SSCNN architecture

5.5 Algorithm for CNN-Attention-Based Baby Unsafe Scenarios Prediction System

Algorithm 1 Baby Unsafe Scenarios Prediction Algorithm (BUSPA): Pseudocode of the Methodology of the Proposed Model (Figure 5.8)

Input: Image data *I* from an online source platform

Output: Unsafe scenario prediction decision: *U* (Unsafe or Safe)

Phase 1: Initialization

1. Receive an input image $I \in R^{224X224X3}$
2. Initialize the CNN base model
   - A Convolutional Neural Network with Soft Spatial Attention (CNN-SSA) for extracting local spatial features.
3. Initialize meta-model with dense layers and sigmoid activation function at the output.

Phase 2: CNN Feature Extraction

4. Pass *I* through a 3-layer CNN block to extract features:

F_{CNN} = CNN

Phase 3: Soft Spatial Attention Mechanism (SSA)

5. Attention map generation:
   - Apply 1X1 Conv2D (kernel=1, Sigmoid activation) to F_{CNN} to produce a spatial attention map F_{SSA}

6. Elementwise Feature Fusion:
   - Multiply the F_{CNN} features and F_{SSA}

F_{fused} = $F_{CNN} \odot F_{SSA}$

Phase 4: Meta-Model Processing

7. Perform forward propagation:
   - Input F_{fused} into the meta-model
   - Compute class probabilities using sigmoid activation function:

   P = Sigmoid (Meta - model(F_{fused}))

8. Predict the class U with the highest probability:

$$U = \arg\max(P)$$

Phase 5: Training and Decision Output

9. Define the loss function:

BCE = $-\frac{1}{N}\sum_{i=1}^{N}[y_i \,.\, (\log p_i) + (1 - y_i) \,.\, \log(1 - p_i)]$

10. Train the meta-model using the SGD optimizer, ReduceLROnPlateau and early stopping to avoid overfitting.

11. Predict the class U as:

$$U = \arg\max(Sigmoid(meta - Model(F_{fused})))$$

12. Return U as the Unsafe Scenario prediction decision.

5.6 Dataset Preparation and Ground Truth Selection

We have collected 88 videos originally sourced from online platforms such as TikTok, Dailymotion, YouTube, and YouTube Shorts. Some were long, so we chopped them into 4- to 5-minute segments, totalling 97 videos and 25,284 frames. There are a few datasets comprising 3-4 minutes of video footage generated by an AI tool that serve as an unseen test set for evaluating the model's performance. A video was purposely generated by an AI tool to test the model in an unsafe zone. These datasets fall under the category of babies' ADLs in a crib. It can include leaning, lying down, jumping, climbing over the fence, standing near the fence, playing, or sitting inside the crib. We have categorized ADLs into six activities and classified each as either safe or unsafe, as demonstrated in Fig. 5.10 (a) and (b).



(a) (b)

Figure 5.10. Activity Daily Living (ADLs): (a) Unsafe (b) Safe

5.6.1 Dataset Analysis and Bias

The dataset comes from social media and includes AI-generated videos. Videos downloaded from online sources are used for training, validation, and testing; AI-generated videos are used as unseen data to test the model. The camera viewpoints were categorized into top-down, side, and oblique views on a stratified subset of 88 videos. Top-down views account for 11%, side views 73%, and oblique views 16%, indicating that the dataset is dominated by side views, which may limit generalisation to other camera setups. The videos are roughly categorized as bright vs dim, natural vs artificial, and stable vs flickering or variable lighting. Thus, we estimated lighting conditions by measuring frame-wise brightness and contrast, grouping videos into bright, medium, and low-light categories. Approximately 5% of videos are low-light, 22% medium, and 73% bright. The AI-generated videos for the unseen test dataset exhibit more uniform lighting than real videos. Crib variability is generally assessed by crib colour, bar spacing, the presence of additional objects such as toys and blankets, and crib orientation within the frame. In our annotated subset, 48% of cribs are light-coloured with narrow vertical bars, 38% have mesh or solid walls, and 14% have horizontal bars or wide spacing, suggesting limited representation of certain crib designs commonly found in real homes.

Despite this analysis, the dataset remains biased toward online platforms, side viewpoints, bright lighting, and narrow vertical crib spacing. We therefore caution that models trained on this dataset may underperform in underrepresented conditions and encourage future work to expand coverage of these scenarios.

5.6.2 Data Preprocessing

This section contains datasets with long-duration videos. The training, validation, and test sets comprised 70%, 15%, and 15% of the dataset, respectively. In this section, the Python code utilizes Keras's ImageDataGenerator to perform real-time data augmentation on training images. The generator is configured to rescale pixel values to a 1/255 range and apply a suite of random transformations. This includes rotations up to 10 degrees, zoom operations within a 20% range, horizontal and vertical translations up to 10% of the image dimensions, shearing by up to 0.1 radians, and horizontal flips. The fill_mode='nearest' argument specifies that any introduced empty pixels during these transformations are filled with the nearest pixel value. During training, this augmentation procedure is applied dynamically, ensuring the model is exposed to a variety of input data variations at each epoch. Thus, it enhances generalization and reduces the risk of overfitting.

5.6.3 Implementation Setup and Evaluation Metrics

In this section, a thorough explanation of the experimental settings and evaluation metrics used in this approach. The experiments were carried out using a Jupyter Notebook on a PC with 32GB of RAM and a 3.00 GHz Intel® CoreTM i7-1185G7 (11th Gen) CPU. TensorFlow 2.10.0, a deep learning

framework, was utilized in these experiments to construct the BUSPS model. The dataset is randomly split using the scikit-learn library's split function in Python, with allocations of 70%, 15%, and 15% for training, validation, and testing, respectively. Using a learning rate of 0.01 and a momentum of 0.9, the SGD (Stochastic Gradient Descent) optimiser is used to construct this model. The model will incur a loss, and to mitigate this, binary cross-entropy is incorporated. The model is trained using a batch size of 32 for 20 epochs. It mitigated overfitting with early stopping and reduced the learning rate by at least 1e-6 using the ReduceLROnPlateau callback. The SGD optimizer optimizes the neural network, and dropout layers are incorporated to enhance generalization. The final dense layer of the model computes and returns the probability of the final output using a sigmoid activation function. Table 5.4 presents the hyperparameter settings used to execute the proposed model.

We evaluated our model's performance using common metrics, such as Sensitivity, Specificity, precision, and accuracy, to assess its ability to predict safe and unsafe cases. The higher the values of these performance metrics, the better the classifier's performance.

Table 5.4: Hyperparameter settings for the proposed model

Specifications	Values
No. of Epochs	20
Batch Size	32
Validation set	15
Training set	70
Rate of Optimiser Learning	0.01
Momentum	0.9
Loss_function	Binary cross-entropy

Sensitivity (Recall) is the proportion of actual unsafe cases that the BUSPS model correctly predicts, presented in (5.2):

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP+FN}} \tag{5.2}$$

The specificity assesses the ability of the Baby Unsafe Scenarios Prediction System (BUSPS) to correctly identify safe scenarios from Activities of Daily Living (ADLs) without triggering false alarms, defined in (5.3):

$$\text{Specificity} = \frac{\text{TN}}{\text{TN+FP}} \tag{5.3}$$

Precision, shown in (5.4), measures the proportion of predicted unsafe cases that are truly unsafe:

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{5.4}$$

Accuracy is the metric that measures total correct labelling, providing the overall proportion of correct predictions (both safe and unsafe) among all cases, and can be represented in (5.5):

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \tag{5.5}$$

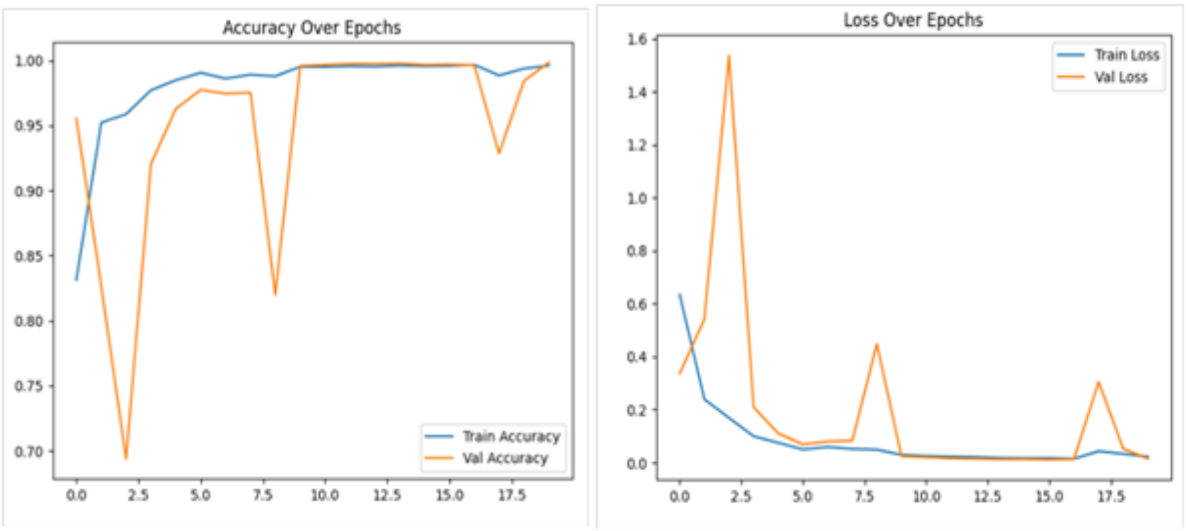
The unsafe scenarios prediction approach uses the metrics from Table 5.5. with a binary classification system to predict unsafe scenarios. In this evaluation, TP (True Positives) signifies correct unsafe identification, while FP (False Positives) shows incorrect classifications of safe as unsafe. TN (True Negatives) indicates correctly identified safe scenarios; however, FN (False Negatives) represents incorrectly identifying unsafe scenarios as safe.

Table 5.5: Fall Risk Prediction Metrics

Prediction Class	Binary Classifier	Description
True Positive (TP)	1	When correctly identified as Unsafe
True Negative (TN)	0	When correctly identified Safe scenarios (sitting/lying down)
False Positive (FP)	1	Incorrect classification of Safe as Unsafe
False Negative (FN)	0	Incorrectly identified Unsafe scenarios as Safe

5.7 Evaluation of the AI-2 Model

Fig. 5.11 (a) and (b) show the comparison of training and validation data's accuracy and loss over epochs of the SSCNN model using CNN-SSA architecture. In this case, we observed that the training accuracy and validation accuracy converged at the 1st epoch to 0.752 and 0.955, respectively, while the training and validation losses were 0.820 and 0.336, respectively. At the 20th epoch, the training and validation accuracies are 0.995 and 0.998, respectively; however, the training and validation losses are 0.024 and 0.15, respectively. This indicates that the SSCNN model improves validation loss and reduces overfitting. The proposed model is tested on 15% of the split unseen datasets, yielding the classification results shown in Table 5.6. This model achieved an accuracy of 99.79 (F1-Score: 99.78%; Sensitivity: 99.67%; Specificity: 99.90%), demonstrating high predictive reliability that aligns with the physical activity levels observed in various safety scenarios. The true and predicted labels clearly support the results, as shown in the confusion matrix in Figure. 5.12.



(a) (b)

Figure 5.11: Graph showing the improvement in accuracy and loss against the number of Epochs (a) Accuracy Vs Epochs (b). Loss Vs Epochs

Table 5.6: Classification of SSCNN Vs BCNN in the prediction of Unsafe Scenarios

Evaluated Metrics (%)	**Proposed Model (%)**
Accuracy	99.79
Precision	99.89
Recall/Sensitivity	99.67
Specificity	99.90
F1-score	99.78

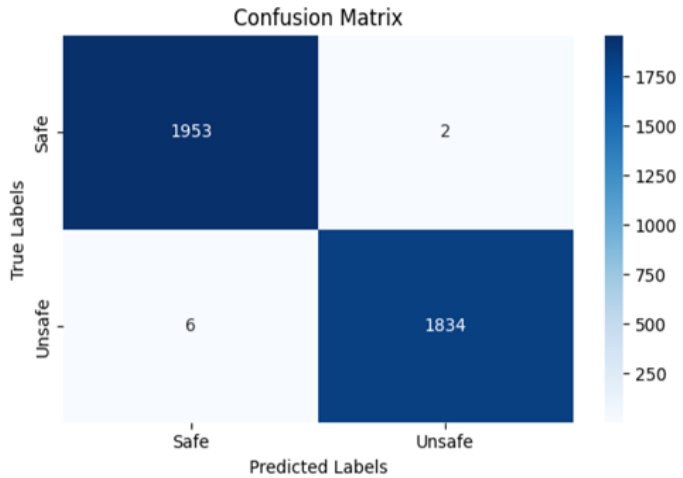


Figure 5.12: Confusion matrix

5.7.1 Frame-based (Time-Independent) Evaluation of the model

This method was used in the model to produce results every 10 seconds, based on frames. The model was evaluated on a long-duration video generated with an AI tool. The video is 199.70 seconds long and has 5991 frames. The interpretation of safe and unsafe scenarios for this model is as follows: if there are 300 frames in each 10-second interval, it is considered unsafe if a single unsafe frame occurs within that interval. Figure 5.13 depicts the outcomes of anticipating scenarios based on each frame and alerting to Unsafe and Safe correspondingly. Though it tallied accurately and sent the Unsafe and Safe signals for each period, it generates more false alerts. As a result, this strategy proved unpromising and did not enhance BUSPS's effectiveness. Henceforth, we added the statistical parameters detailed in Section 5.8.3 to interpret the temporal characteristics of the video, predict unsafe and safe scenarios, and improve the model's effectiveness and efficiency.

```
# ------------------------
video_path = "C:/Users/hcb1500/finalproject/Datasets/final_test_revised.mp4"
results = analyze_video_intervals(video_path, interval_sec=10, threshold=0.5)
```

```
WARNING:absl:Compiled the loaded model, but the compiled metrics have yet to be built. `model.compile_metrics` will be empty until you train or evaluat
e the model.
🎥 Video length: 199.70 sec, FPS: 30, Total Frames: 5991
Checking every 10-second interval...
Interval 1 (0.0s - 10.0s): UNSAFE  | Frames: 300, Unsafe Frames: 42
Interval 2 (10.0s - 20.0s): SAFE  | Frames: 300, Unsafe Frames: 0
Interval 3 (20.0s - 30.0s): UNSAFE  | Frames: 300, Unsafe Frames: 1
Interval 4 (30.0s - 40.0s): SAFE  | Frames: 300, Unsafe Frames: 0
Interval 5 (40.0s - 50.0s): SAFE  | Frames: 300, Unsafe Frames: 0
Interval 6 (50.0s - 60.0s): UNSAFE  | Frames: 300, Unsafe Frames: 1
Interval 7 (60.0s - 70.0s): UNSAFE  | Frames: 300, Unsafe Frames: 107
Interval 8 (70.0s - 80.0s): SAFE  | Frames: 300, Unsafe Frames: 0
Interval 9 (80.0s - 90.0s): SAFE  | Frames: 300, Unsafe Frames: 0
Interval 10 (90.0s - 100.0s): SAFE  | Frames: 300, Unsafe Frames: 0
Interval 11 (100.0s - 110.0s): SAFE  | Frames: 300, Unsafe Frames: 0
Interval 12 (110.0s - 120.0s): SAFE  | Frames: 300, Unsafe Frames: 0
Interval 13 (120.0s - 130.0s): UNSAFE  | Frames: 300, Unsafe Frames: 268
Interval 14 (130.0s - 140.0s): SAFE  | Frames: 300, Unsafe Frames: 0
Interval 15 (140.0s - 150.0s): UNSAFE  | Frames: 300, Unsafe Frames: 3
Interval 16 (150.0s - 160.0s): UNSAFE  | Frames: 300, Unsafe Frames: 169
Interval 17 (160.0s - 170.0s): UNSAFE  | Frames: 300, Unsafe Frames: 26
Interval 18 (170.0s - 180.0s): SAFE  | Frames: 300, Unsafe Frames: 0
Interval 19 (180.0s - 190.0s): SAFE  | Frames: 300, Unsafe Frames: 0
Interval 20 (190.0s - 199.7s): SAFE  | Frames: 291, Unsafe Frames: 0
```

Figure 5.13: Depicting the scenarios frame-by-frame and alert accordingly

5.7.2 Statistical (Time-Dependent) Parameters

The SSCNN model incorporated time-dependent coefficients, such as Standard Deviation [113], Entropy [114], and Autocorrelation [115], to predict unsafe scenarios. Assuming a dataset that produced a sequence of probabilities, P = [p_1, p_2, p_3, ..., p_N], where pi ∈ [0,1] represents the **predicted probability of "Unsafe"** at frame i, and $\bar{p} = \frac{1}{N}\sum_{i=1}^{N} p_i$.

- *Standard Deviation (Variance)*

This parameter captures variabilities in predicted probabilities. If σ is high, it leads to fluctuating probabilities (unstable over time), whereas if σ is low, it leads to consistent predictions (either mostly safe or mostly unsafe). The equation for the mean of the standard deviation is shown in (5.6),

$$std_mean = \sqrt{(\frac{1}{N}\sum_{i=1}^{N}(p_i - \bar{p})^2)} \quad (5.6)$$

- *Entropy*

This parameter measures uncertainty in model predictions. Shannon's binary entropy is calculated for ([p, 1-p]) for each probability p_i. If entropy is high, the model is uncertain (probabilities near 0.5); however, if entropy is low, confident predictions (probabilities near 0). Thus, the entropy equation holds for each p_i and across all samples. The entropy for each probability, p_i, and the mean of all entropies are expressed in (5.7) and (5.8), respectively:

$$entropy = -p\log_2(p) - (1-p)\log_2(1-p) \quad (5.7)$$

$$ent_{mean} = \frac{1}{N}\sum_{i=1}^{N} entropy(pi) \quad (5.8)$$

- *Autocorrelation*

This parameter computes the lag-1 autocorrelation coefficient, i.e. correlation between successive probabilities p_i and p_{i+1}. It measures temporal dependency, that is, how much the current prediction correlates with the next one. Thus, if autocorr is 1, there is a strong positive temporal dependency; if it is 0, the correlation is random (random predictions); if autocorr is -1, there is an alternating pattern (unstable predictions). The autocorrelation for p_i and p_{i+1} is expressed in (5.9),

$$autocorr_mean = \frac{1}{N}\sum_{i=1}^{N-1} p_i . p_{i+1} \quad (5.9)$$

5.7.3 Results of Time-Independent vs Time-Dependent Datasets

The model was tested on a 20-second unseen video (30 frames per second) and successfully generated up to 20 safety alerts. To avoid this constraint, the model was further tested on long-duration, unseen videos that generated safety alerts, even if a single frame was deemed unsafe, as detailed in Section 5.8.1. Later, the model was tested on unseen long-duration videos using both time-independent and time-dependent parameters. The model's testing using time-independent data was conducted within a frame-based approach, whereas the time-dependent analysis considered time-dependent statistical parameters, including autocorrelation, entropy, and standard deviation. When tested on a long-duration video purposely generated for this model using an AI tool (as mentioned in the Dataset section), the SSCNN model performs better at identifying unsafe scenarios. The thresholds were calculated from the datasets for each parameter: standard deviation (0.419), entropy (0.109), and autocorrelation (0.951). We compared the results collected from testing the BCNN-BGRDOS and proposed model (SSCNN) with the CNN-SSA architecture and found that the SSCNN model provided better predictions of 2

unsafe scenarios for the time-independent and 1 for the time-dependent approach, whereas the BCNN model did not give any improvement in both approaches, giving 4 as a count.

We discovered that the SSCNN model outperforms the BCNN model in terms of false alarms, regardless of whether the time-independent or time-dependent approach is used. However, BCNN shows consistent alerts in both cases. These results are compared by manually testing the video and analyzing the scenarios every 10 seconds. Table 5.7 compares the time-independent and time-dependent approaches for both models. Thus, the developed model, SSCNN, reduced false alarms compared to BCNN. The logic for interpreting the underlying unsafe-scenario prediction using statistical parameters is described in Appendix A.1. Moreover, a detailed explanation of how to predict whether a 10-second video segment is safe or unsafe is provided in Appendix A.2.

Table 5.7: Comparison of SSCNN Vs BCNN in the prediction of Unsafe Scenarios

Time duration (secs)	**Time-independent[1]**		**Time-dependent[2]**		**Manually**
	BCNN	**SSCNN**	**BCNN**	**SSCNN**	
0-10	SAFE	UNSAFE	SAFE	UNSAFE	UNSAFE
10-20	SAFE	UNSAFE	SAFE	UNSAFE	UNSAFE
20-30	UNSAFE	UNSAFE	UNSAFE	UNSAFE	UNSAFE
30-40	UNSAFE	UNSAFE	UNSAFE	UNSAFE	UNSAFE
40-50	UNSAFE	UNSAFE	UNSAFE	UNSAFE	UNSAFE
50-60	UNSAFE	UNSAFE	UNSAFE	UNSAFE	UNSAFE
60-70	UNSAFE	UNSAFE	UNSAFE	UNSAFE	SAFE
70-80	UNSAFE	UNSAFE	UNSAFE	UNSAFE	UNSAFE
80-90	UNSAFE	UNSAFE	UNSAFE	UNSAFE	UNSAFE
90-100	SAFE	UNSAFE	SAFE	UNSAFE	UNSAFE
100-110	UNSAFE	UNSAFE	UNSAFE	UNSAFE	UNSAFE
110-120	UNSAFE	SAFE	UNSAFE	UNSAFE	UNSAFE
120-130	UNSAFE	UNSAFE	UNSAFE	UNSAFE	UNSAFE
	4	2	4	1	

[1] Time-independent: frame-based

[2] Time-dependent: temporal features using statistical parameters

5.8 Comparison of the AI-2 model with State-of-the-Art methods

To contextualize execution of the suggested SSCNN, we compare it with recent attention-based vision systems developed for fall detection or baby safety monitoring. Table 5.8 compares SSCNN with recent attention-based methods for baby safety monitoring. El Marhraoui [93] utilized CNN self-attention with IMU sensors for fall-risk prediction in adults or older adults. Thus, achieving a 71% detection rate but required wearables and controlled settings, unlike SSCNN's RGB frames-only approach, which targets babies' safety. CribNet [38] utilized YOLOv8 for segmentation and pose estimation of crib hazards, achieving 86-88% mAP in object detection. It still lacks direct classification of safe or unsafe behaviours. [21] proposed a CNN-BiLSTM ensemble with channel/temporal attention for general fall detection of elderly people. This approach attained 97.93-98.99% accuracy on sensor datasets (KFall/SisFall), but relies on waist accelerometers/gyroscopes rather than vision and targets only elderly people. [116] apply Spatial-Temporal Graph Convolutional Network (ST-GCN) to estimate keypoints data from infrared videos. Thus, achieved 98.6% accuracy across fall datasets, but discarded the RGB texture/context essential for baby monitoring and understanding their crib behaviours.
SSCNN achieves superior 99.79% accuracy (with 99.89% precision and 99.90% specificity) on its custom 25,284-frame crib dataset. It outperformed these baselines while utilizing non-intrusive RGB frames and soft spatial attention for real-time, privacy-friendly deployment. The soft mask enables differentiable focus on safety-critical regions, reducing false alarms compared to traditional CNN. This approach addresses crib-specific behavioral nuances that are absent in sensor- or object-centric methods. SSCNN achieves top performance by using direct RGB frames, which provide richer cues than skeletons or sensors, for crib-specific safe/unsafe classification.

This comparative analysis focuses exclusively on fall prediction methodologies to justify the integration of Convolutional Neural Networks (CNNs) with attention mechanisms. Furthermore, it highlights a critical research gap: the lack of literature applying these combined architectures to predict unsafe scenarios for babies in crib environments. Most of these studies do not specifically target crib falls or unsafe crib conditions, and those that do use soft-spatial CNNs for babies aged 12-24 months, thereby motivating our contribution.

Table 5.8: Comparison of SSCNN with state-of-the-art methods for fall prediction

Reference	Approach	Methods	Dataset	Performance
El Marhraoui [93]	Self-Attention fall prediction	CNN with self-attention mechanism + IMU signals	IMU-based child balance/fall motion recordings	71% detection of fall risk events in a 1s window
Zhu [38]	CribNet: in-crib hazard detection	YOLov8 + Detectron2 segmentation + infant pose estimation for toy or	CribHD dataset (toys, blankets, crib hazards; multi-environment crib videos)	mAP 86.1% (toys), 88.3% (blankets) at IoU 0.5;

		blanket detection in cribs		
Sahni [21]	CNN-BiLSTM ensemble method for fall detection	CNN + channel attention + BiLSTM + temporal attention; ensemble for sensor fusion	KFall (5075 samples), SiSfall (7303 samples); accelerometer/gyroscope waist sensors	Accuracy 97.93% (KFall), 98.99% (SiSFall)
Keskes [21]	Vision-based fall detection using ST-GCN	Spatial-Temporal Graph Convolutional Network (ST-GCN) on skeleton data from infrared videos for fall event classification.	Multiple fall detection datasets (UR Fall, Multiple Cameras Fall Datasets etc); skeleton keypoints only	98.6% accuracy
Proposed Method - SSCNN	Soft Spatial CNN (SSCNN) for baby unsafe scenarios prediction	2D CNN + Soft Spatial Attention (SSA) via sigmoid mask; temporal statistical post-processing	Custom Crib dataset: 25284 RGB frames from 97 videos (safe/unsafe; 7:1.5:1.5 Dataset split)	Accuracy 99.79%, Precision 99.89, Recall 99.67%, Specificity 99.90%, F1 99.78%

5.9 Conclusions

The proposed Baby Unsafe Scenarios Prediction Model, based on the CNN-SSA architecture, predicts unsafe scenarios for babies in cribs at risk of falls. The primary focus in developing this model was to establish real-time in-crib infant monitoring to enhance safety during periods of limited parental supervision. Although not real-time, we developed an unsafe-scenario prediction model with reduced false alarms. In addition to that, it addressed the concern of sudden falls from a crib or serious head injuries in babies, which in-crib safety supports parental peace of mind. The significance of this research lies in its contribution to the limited body of research on predicting baby fall risk across scenarios. Additionally, to provide a safe environment within the crib for babies aged 12-24 months with developing motor skills and a tendency to fall. Continuous monitoring of babies' ADLs related to safe and unsafe scenarios, particularly sitting, lying down, and playing inside the crib as safe, whereas standing, leaning, and climbing are considered unsafe. These six activities are the critical ADLs used to predict unsafe scenarios for babies in the crib. The key benefit of our proposed model is its exclusive reliance on spatial features, leveraging SSA to extract more precise spatial features. In addition, using statistical parameters to extract temporal features. Thus, we predict risky and unsafe scenarios over the test datasets. The results analysis shows that our model delivers outstanding performance, achieving 99.79% accuracy, 99.67% sensitivity, and 99.90% specificity. Numerous studies focused on babies' fall detection and pose estimation suggest that fall risk among babies needs to be reduced. Thus, the issues encountered in these studies are considered resolved in this model.

However, the system is limited to a dataset of only 97 videos and biased towards specific crib and camera conditions. Despite augmentation, there is a risk of overfitting, and to avoid overestimating accuracy, we also need cross-environment validation. The model's real-time deployability still needs to be verified, as it was evaluated on datasets of previously produced videos. Additionally, further investigation is required to establish the exact location of the baby in the crib and the duration during which the baby remains near the fence.

In the future, we intend to incorporate temporal features, integrating pose-based temporal modelling, specifically the Gated Recurrent Unit (GRU), as the temporal model, and add dual-trigger logic to accurately predict unsafe or high-risk scenarios for babies in their cribs. We anticipate publishing our later findings based on the real-time input videos and the upgraded model. The long-term goal of our research is to evolve the application into a comprehensive monitoring system that integrates IoT devices and sensors (pressure and accelerometer) placed under a baby crib's mattress. This will enable the development of a smart baby monitor that safeguards the baby and features a user interface, such as an API or a mobile app. It would serve as a reliable, efficient device that provides safety alerts in threatening situations. Baby safety, specifically the prediction of fall risk, is a vital yet underexplored area in AI, making this application socially significant.

Chapter 6. Novel ADL-Conditioned Cascaded Transfer Learning-based Baby Fall Risk Prediction Model

6.1 Introduction

Falls are the leading cause of injury in early childhood, yet most monitoring technologies focus on vital signs rather than physical safety in the crib. Various factors that lead to a baby's falls are separated into intrinsic and extrinsic groups. The intrinsic factors include age band, developmental stage, and overall mobility. Overall mobility can sometimes lead to typical movement patterns, such as a tendency to climb or grasp rails. The most frequent cause of injuries in children who need hospitalization is falls. Falls account for half of all injuries in children under five. [25]. When babies turn over and move for the first time, they may fall. [26]. As they start to walk, children take more chances. They get inquisitive and eager to explore and climb. An important aspect of their growth is climbing. This may result in injuries from falls. [27]. Recently, significant progress has been made in baby health monitoring systems in various ways, including sensors, computer vision-based approaches, PE techniques, AI, and ML. Previous studies have focused on frequently detecting risky postures[117] [97], suffocation, respiration [118], or continuous crying [119] [120]. However, few have focused on baby falls from the crib. Wipulasundara [65] emphasized the Siemens Logo module's non-contact technology as a benefit of the microprocessor-based integrated solution for identifying potentially dangerous areas or behaviours in a baby's crib. Yin [39] developed a multidimensional danger prediction algorithm and a smart multimodal fusion danger forecasting system that makes use of IoT devices and artificial intelligence. These devices detect risky postures, but they use sensors to detect environmental issues and rely on AI only for image processing. They seldom simulate the spatiotemporal evolution of behaviour in the crib environment, which can lead to falls, and they often misalign accuracy and issue false alarms.

Therefore, the research model presented in this thesis is a novel ADL-Conditioned Cascaded Transfer Learning-based Hybrid Baby Fall Risk Prediction Model. This may generate false alarms infrequently when predicting future falls among infants in the crib. The proposed model is a vision-based fall-risk prediction system that leverages pose-based spatio-temporal features, a GRU [121] with soft spatial, and a dual-trigger alert mechanism to anticipate dangerous crib events while suppressing spurious alerts. This is the first scientific study to use a temporal model with a soft spatial attention layer and a pose estimation model, leveraging spatio-temporal attention features, to predict the baby's correct pose in the crib at the corresponding time. This methodology incorporates an additional decision-based event to accurately classify and notify infants in the crib of the risk of falls. An AI-generated video is used to calibrate this developed model. After validation, real-time footage of the babies will be added to the model in the next stage (future development). The alert system will be incorporated into a user's or caregiver's communication arrangement, creating a mobile application. Before integrating real-time

data into later phases, the technique was first verified using custom datasets to ensure stability. The suggested AI-powered system may enable early predictions, reduce the risk of fall-related accidents, reduce false alerts, and allow babies in the crib to live safely and comfortably. The basis for improved fall prediction, safety, and medical outcomes is strengthened by more recent advancements in AI and ML. In this chapter, the Dual-Trigger Decision-based event of the Semantic-conditioned AI-based baby fall risk prediction model is used as an example.

6.2 Brief Overview of the approaches used in the model

In this section, a quick synopsis of the techniques and approaches is explained in sections 6.2.1 and 6.2.2 and used to develop the model.

6.2.1 Pose Estimation Approach

Research on human pose estimation technology is ongoing in a variety of fields, including sports, surveillance, health monitoring, home care for the elderly, baby monitoring, gesture control, and even entertainment. As discussed in detail in Chapter 4, human pose estimation plays an important role in healthcare research. Three types of human pose estimation exist: 2D or 3D coordinate estimation techniques; single-image or video-based approaches, depending on the type of input; and single-person or multi-person approaches, depending on the no. of target subjects. In this study, Google's open-source, cross-platform each image frame's 2D human joint coordinates were estimated using the MediaPipe Pose (MPP) framework. MPP processes cognitive data from videos and builds machine-learning pipelines. As seen in Figure 6.1, MPP extracts 33 2D keypoints (KPs) from the human body using BlazePose. A lightweight machine learning architecture called BlazePose delivers real-time output on PCs and smartphones via CPU inference. As illustrated in Figure 6.1, we employed nine landmarks among the estimated MPP keypoints to calculate the arbitrary positions and movements, as well as to compute the torso angle, the Center of Mass, joint velocities, and joint angular velocities (left elbow angle).

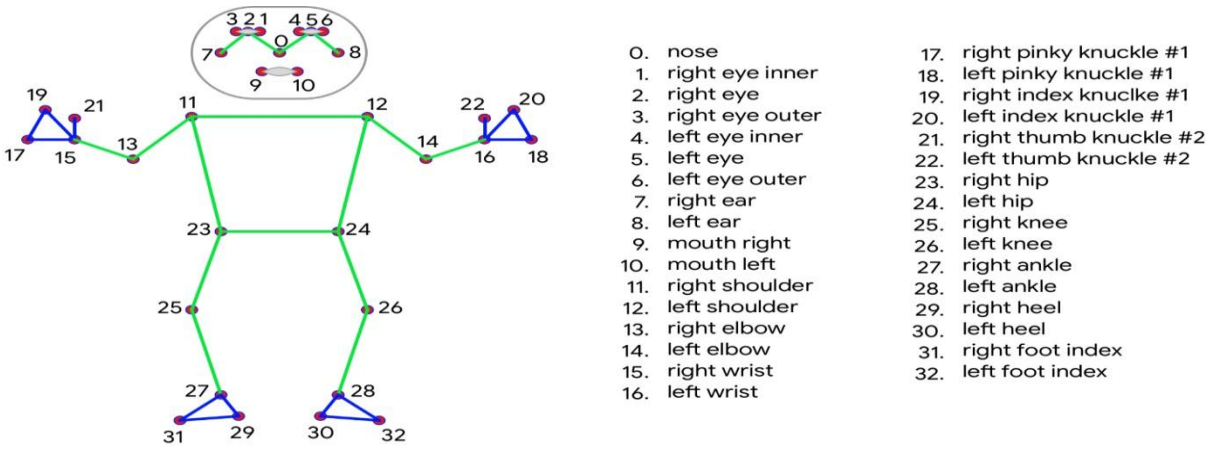


Figure 6.1 Definition of keypoints in MediaPipe Pose [122]

MPP is a novel approach that uses machine learning to identify a baby's posture in a crib by analyzing video frames. The steps listed below can be used to integrate MPP into a posture-detection system for detecting baby fall risk out of the crib:

- Extracting keypoints using BlazePose, a pretrained model.
- Save those keypoints into a CSV file.
- Perform feature engineering to calculate the important parameters depicting the posture and distance of the baby to the fence.
- Encode these features with state labels pointing to the seriousness of scenarios.

These features serve as inputs to the temporal model for further risk-level classification, as discussed in Section 6.6.

6.2.2 Temporal Model

Fall prediction uses a variety of temporal models due to their temporal dependencies. Gated Recurrent Units (GRUs) are a specific type of recurrent neural network (RNN) cell.[121, 123]. It is designed to simulate sequences, such as time series and videos. GRU is an advanced version of LSTM. It creates update and reset gates by combining several gates. Thus, using fewer parameters and faster training to learn long-term dependencies. Its two primary gates, the update gate and the reset gate, are used to update its hidden state, h_t, at each time step. [124].

- Update gate z_t :determines the amount of memory/long-term information from the preceding hidden state, h_{t-1}.
- Reset gate r_t :determines how much of the prior hidden state should be ignored while calculating the new candidate state.

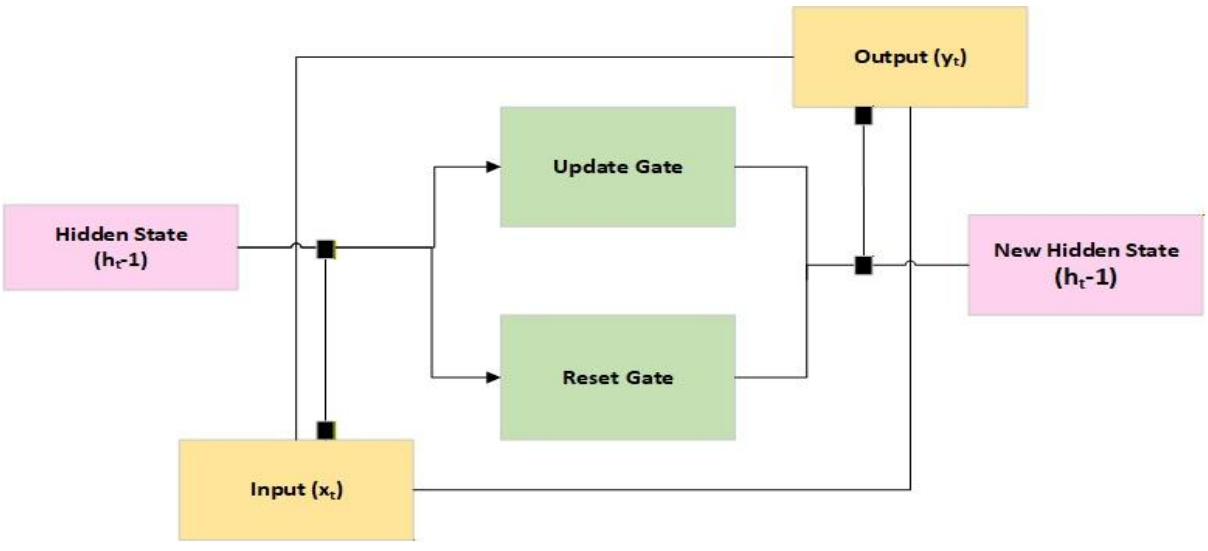


Figure 6.2: Working of GRU

These gates are essentially used by GRU to update the hidden state at each time step in a selected manner. This enables it to retain crucial details while eliminating unimportant specifics. By employing gates to control gradient flow during training, GRUs help alleviate the vanishing gradient problem by preventing gradients from shrinking too much over time and preserving crucial information. Figure 6.2 illustrates the working of GRU.

6.2.3 Attention Mechanism

This section highlights the significance of the attention mechanism. Attention mechanisms have been very helpful across many visual tasks, including image classification, object detection, semantic segmentation, video understanding, image generation, 3D vision, multi-modal tasks, and self-supervised learning. This study explored six different categories of attention from existing attention methods [125]. The four basic categories are: spatial attention (*where to pay attention*), temporal attention (*when to pay attention*), channel attention (*what to pay attention to*) [21] and branch channel attention (which to pay attention to), in addition to channel and spatial attention and spatial and temporal attention, two hybrid combination categories. [126, 127].

Spatial attention can be conceptualized as a mechanism for adaptation that selects regions of focus in space. There are four different types of spatial attention methods: the Recurrent Attention Model (RAM) [128], Spatial Transformer Networks (STN) [129], Gather-Excite Network (GENet) [130], and Non Local [131]. The spatial attention mechanism is further classified into two variants: hard attention and soft attention.

The AI-3 model is developed using a soft spatial attention (SSA) mechanism. Given a set of features vectors $a = \{a_1, \ldots, a_L\}$, where $a_i \in R^D$. It directly uses the context vector's expectation, z_t [125] and is expressed in (6.1):

$$z_t = \sum_{i=1}^{L} \alpha_{t,i} \, a_i \qquad (6.1)$$

By enabling the user to comprehend where the model is focusing, the application of a soft spatial attention mechanism enhances the interpretability of the detailed image information and cues. It also improves the network's representational capability.

6.2.4 Rule-based Decision System

This method applies threshold rules to the Dual-Trigger Alert Logic block, as shown in Figure 6.5, to determine the outcome based on the PE-GRU-SSA network's prediction. The three classes were formulated using a temporal model based on the state probabilities of babies in the crib, for example: 0: Safe, 1: Near the fence, 2: High risk. A temporal smoothing approach is employed for these formulated classes and fed to the rule-based decision system. The system provides the final output based on the two-level events. For example, if the posture is standing and the duration exceeds 4 seconds, it will trigger a first-level alert as “Trigger-1: Prolonged Standing.” If the first-level alert is activated,

ADL is checked to see if it is High-risk, and then the final alert is triggered as “Trigger-2: High-risk alert”.

6.3 Novel ADL-Conditioned Cascaded Transfer Learning-Based Baby Fall Risk Prediction (BFRP) Model Architectural Summary

The purpose of this research is to present a novel approach that integrates the Pose Estimation technique (MPP), GRU, an Attention network, and Dual-Trigger Alert logic to improve baby fall risk prediction while suppressing false alarms. The presented framework is the process of the Vision-based Baby Fall Risk Prediction System. The proposed method combines several building blocks to construct a pose-based 2-layer GRU network for fall risk prediction, as illustrated in Fig. 6.3. In this research, we used the MediaPipe Pose [83] using the BlazePose model to extract 2D human 33 keypoints from raw input frames. Thereafter, the keypoints are used to perform feature engineering and to compute relevant features for pose estimation. During feature engineering, the keypoints are used to calculate the torso angle the center of mass in the x and y directions, the elbow angle, the wrist velocity, and the ankle velocity. Further, the distance between the crib fence and the baby's hand is also calculated, and labeling is performed based on the required calculated features.

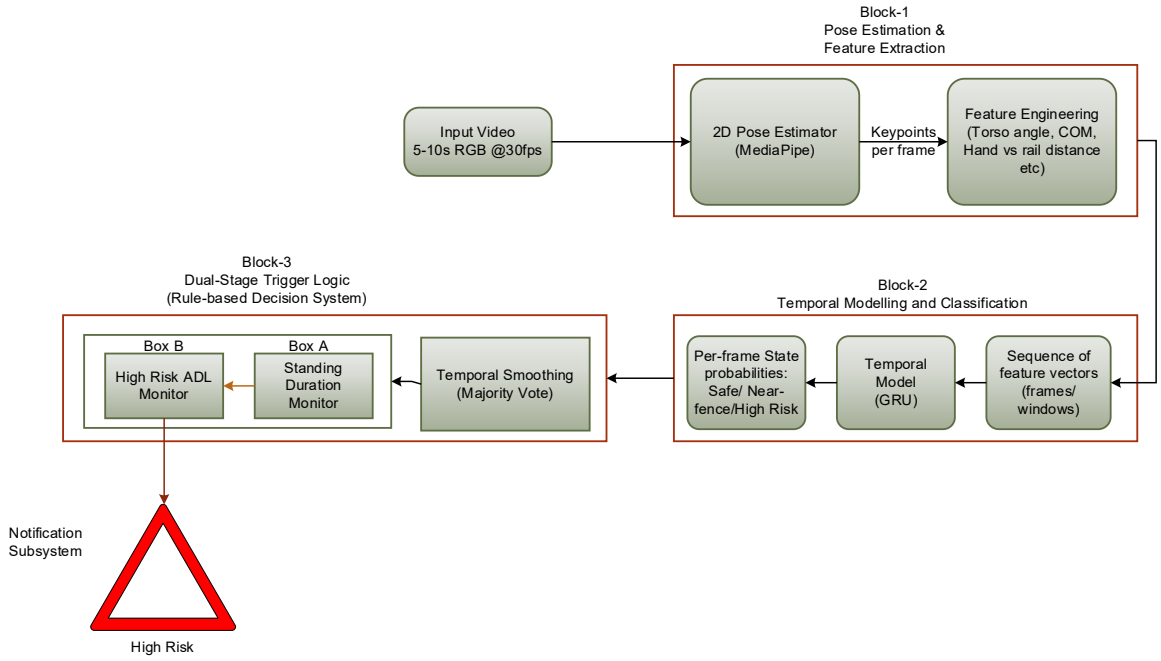


Figure 6.3: Architecture of an AI-based Semantic Conditioned Hybrid Neural Network for Baby Fall Risk Prediction

The second block of Fig. 6.3 presents a model architecture of a spatio-temporal GRU with SSA, as explained in Fig. 6.4. A single GRU layer is a gated recurrent unit, i.e., an RNN cell with update and reset gates designed for sequence data. This is a 2-stack GRU model with 128 neurons in the first layer and 64 in the second. The overall model wraps the spatial attention layer in a GRU-based sequence

classifier, given below:

- Each timestep in the input layer has a feature dimension of 84 features and a fixed sequence length of 30.
- A hidden state is produced for each time step by the first GRU layer, which has 128 units and processes the sequence: (batch=None, time=30, units=128). At each time step, this layer compresses raw characteristics into more complex representations and encodes temporal dependencies.
- First layer (None, 30, 128) and second layer (None, 30, 1) are two dense layers that are added to the first GRU layer.
- An attention layer generates attention weights over time steps using the GRU's sequence of hidden states. Before moving on to the next GRU layer, the attention weight for each time step is multiplied by its 128-dimensional vector, which emphasizes informative time steps and down-weights noisy or irrelevant ones.
- The second GRU layer just outputs the last hidden state (None, 64) because it has 64 units and does not return sequences. The attention-weighted sequence is condensed into a single vector that contains relevant temporal information.
- Lastly, the output layer converts the 64-dimensional summary vector into class probabilities across three classes.

To capture motion patterns and temporal context, this framework was designed to take a sequence of 30 pose feature vectors as input. Stacked GRU layers are used for temporal modeling. Prior to classification, a soft spatial attention layer is included, which learns weights across timesteps to highlight informative frames. As in the earlier SSCNN model, neither convolutions nor kernels are used in this model. Rather, sequence length controls the temporal receptive field. Figure 6.4 illustrates the approach of Block 1 and Block in the flowchart.

Block 3 receives the GRU output to provide state probability, as shown in Fig. 6.3. Safe, Near-Fence, and High-Risk are examples of state probabilities that provide a risk-oriented presentation. The ADLs Class, which offers comprehensible tasks including sitting, lying down, sleeping, playing, standing, climbing, jumping, and leaning, is mapped to these states. We added two triggers to the system to implement dual-trigger alert logic. Trigger-1 activates when standing posture persists for more than 4 seconds in a risky spatial context, such as near a fence or in a high-risk area, and it provides an early warning of "Prolonged Standing". Trigger-2 only activates when Trigger-1 is activated, and a long window of 300 frames of Dominant ADL is a high-risk ADL, like climbing or jumping, confirming a high-risk alert. Consequently, warnings are raised only when both triggers are satisfied, to reduce false positives and avoid continuous unsafe flags.

6.4 ADL-Conditioned Cascaded Transfer Learning -Based Hybrid BFRP Model Algorithm

The pseudocode for Algorithm I describes how the AI-based Semantic Conditioned Hybrid model's fall risk prediction algorithm is constructed using a spatio-temporal model and pose estimation. Features are extracted from keypoints and fed to the temporal model. The model is trained, weights are updated, and attention weights are added to provide more detailed spatial cues in time series. The labeled classes are defined. Later, the decision mechanism provides an alert based on majority voting and a twofold trigger mechanism, as shown in Figure 6.5. Figure 6.4 presents the workflow of the PE-GRU-SSA architecture as a flowchart, whereas Figure 6.5 illustrates the mechanism of a dual-trigger alert and is explained in Appendix A as Algorithm A.2.

ADL-Conditioned Cascaded Transfer Learning-Based Hybrid AI BFRP Model Algorithm I: The Proposed Model's Pseudocode (Figure 6.3)

Input:

Let:

- Video file V, Split V into clips C = {C1, C2, ……, Cn},
  State Label, S = {Safe =0, Near-Fence=1, High-Risk = 2}

Output:

- Prediction Outcome: O
- Evaluation of the Performance Metrics: R

1. **Define landmarks**:
   Shoulders, hips, wrists, elbows, neck, rail (fence) reference points
2. Create an empty list all_frame_features, D_0.
3. Feature Extraction – All features extracted and saved in a CSV file, D_0.
4. Label Encoding: S → specify labels using label encoding logic

$$S_{encoded} = \text{Label Encoder (S)}$$

5. Assign $S_{encoded}$ values to D_0.
6. Create a window sequence using the sample in $D_0 \rightarrow$ (X, Y)
7. Data Splitting → Split D_0 into training, validation, and testing sets:
   [(X_train, Y_train, X_val, Y_val, X_test, Y_test): (0.7, 0.15, 0.15)]
8. Transform train, val, and test into a scaler and save scaler features.
   Apply one-hot encoded labels

9. **Define AI-3 Architecture**: Initialize a Functional GRU-attention model with two dense layers corresponding to the attentional mechanism, SSA

    GRU = Model ([gru (128), dense (128), dense (1), flattening, repeat/permute, elementwise multiplication, gru (64), dense (3)])

    Activation function = SoftMax

    Optimizer = adam, loss = categorigal_crossentropy

10. Train the **Functional GRU-Attention Model:** Fine-tune the model using early stopping and model checkpointing.
11. **Validating the model** utilizing the dataset for testing → Assess the Measures of Performance, R.

    R = {Accuracy, Sensitivity, Specificity, Precision, and Confusion Matrix}

12. **Test the model** over an unseen dataset using the Dual-Trigger Mechanism:

    Outcome: $O \rightarrow$ Predicted State as Baby Fall Risk.

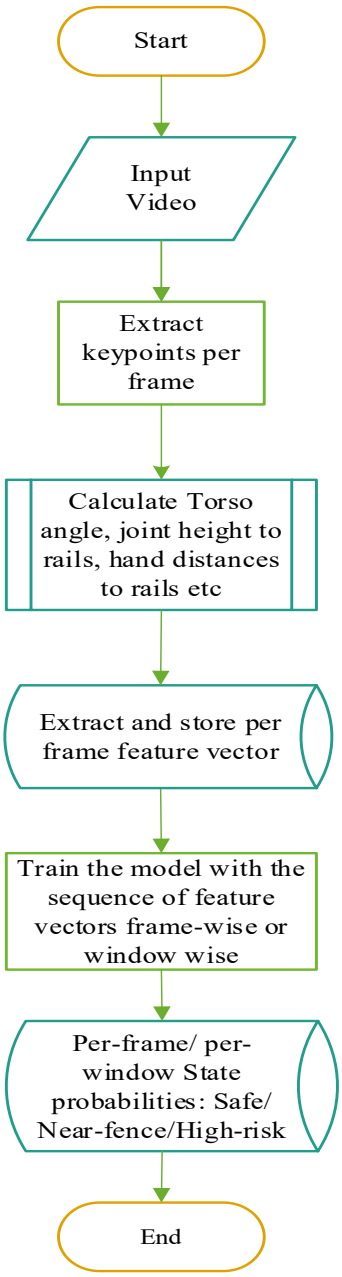


Figure 6.4: Flowchart showing the working of the PE-GRU-SSA

6.5 Data Collection and Ground Truth Selection

This section assesses the efficacy of the suggested ADL Conditioned Cascaded Transfer Learning-based Hybrid AI model for predicting the risk of baby falls, based on its experimental findings and data collection procedure. Its predictions are compared with those of the AI-1 model (PE + RF Classifier) to assess correctness. This model outperforms AI-1 and AI-2 in reducing false alarms. Therefore, its effectiveness is also compared to the most recent, cutting-edge research in this area. The first-generation Intel Core Ultra 7 155U (1.7 GHz) with 32 GB of RAM was used to run Python simulations on Windows 11 Enterprise.

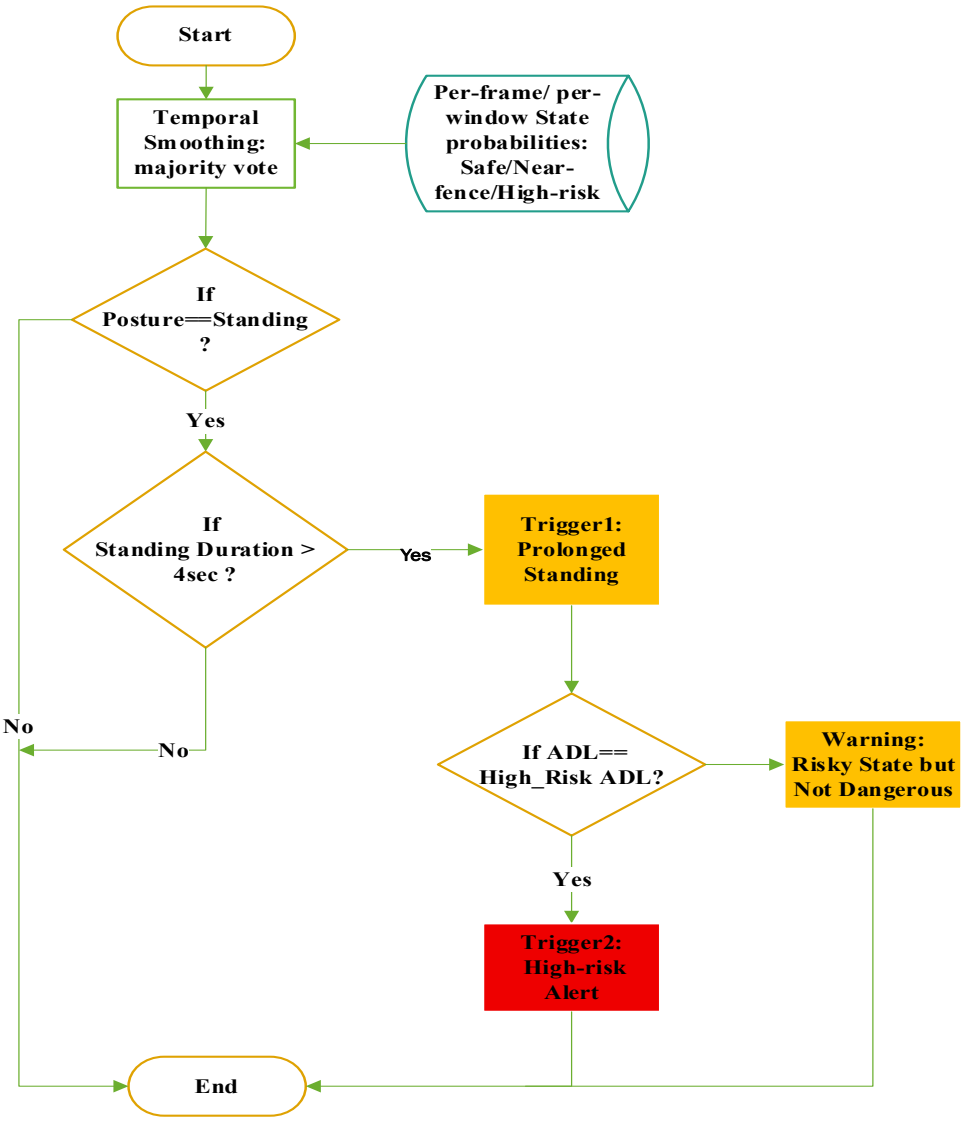


Figure 6.5: Flowchart showing the mechanism of Dual-Trigger Alert Logic

6.5.1 Data Preparation and Evaluation Metrics

The suggested framework is assessed using information taken from the various platforms and generated using various online sources. In this research, the dataset was created by crawling videos from online sources such as Vimeo, Dailymotion, YouTube, and TikTok. During data preparation, we faced many challenges, as the available online videos for the required ADLs were insufficient. For this reason, we have also created videos using an AI tool for testing purposes. These videos are chopped into 5 to 10 second segments for different ADLs. We have a total of 101 videos with 2 different scenarios of Safe and Unsafe, and 8 ADLs. The summary statistics for the customized benchmark dataset are shown in Table 6.1. A few conditions related to selecting videos and preparing the dataset have been taken into consideration. We have done a systematic analysis of domain bias, camera viewpoints, lighting conditions, and crib variability. The camera viewpoints were categorized into top-down, side, and oblique views on a stratified subset of 88 videos.

Table 6.1: Summary of the Statistics of the customized benchmark dataset

Dataset Statistics	Values
Total Videos	101
Total ADLs	8
Total Frames	16987
Total Samples	3111
Unsafe Videos	50
Safe Videos	51

Top-down views account for 12%, side views 75%, and oblique views 13%, indicating that the dataset is dominated by side views, which may limit generalization to other camera setups. The videos are roughly categorized as bright vs. dim, natural vs. artificial, and stable vs. flickering or variable lighting. Thus, we estimated lighting conditions by measuring frame-wise brightness and contrast, grouping videos into bright, medium, and low-light categories. Approximately 7% of videos are low-light, 25% medium, and 68% bright. The AI-generated videos for the unseen test dataset exhibit more uniform lighting than real videos. Crib variability is generally assessed by crib color, bar spacing, the presence of additional objects such as toys and blankets, and crib orientation within the frame. In our annotated subset, 50% of cribs are light-colored with narrow vertical bars, 38% exhibit mesh or solid walls, and 12% exhibit horizontal bars or wide spacing, suggesting limited coverage of certain crib designs commonly found in real homes.

The ability of the model to accurately discriminate between safe conditions and fall risks was assessed using established evaluation metrics, namely accuracy, precision, sensitivity, specificity, and the F_1-score (Equations 6.1-6.5). In all cases, higher metric values denote enhanced diagnostic performance of the classifier.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{6.1}$$

$$\text{Precision} = \frac{TP}{TP+FP} \tag{6.2}$$

$$\text{Sensitivity} = \frac{TP}{TP+FN} \tag{6.3}$$

$$\text{Specificity} = \frac{TN}{TN+FP} \tag{6.4}$$

$$\text{F1-score} = \frac{2TP}{2TP+FP+FN} \tag{6.5}$$

The baby fall risk prediction approach treats Fall Risk as the positive class and Safe as the negative class. In this evaluation, the metrics of the confusion matrix resulting from this classification structure is presented below:

- TP: A fall-risk event is correctly predicted as Fall Risk.
- TN: A safe event is correctly predicted as Safe.
- FP: A safe event is incorrectly predicted as Fall Risk (false alarm).
- FN: A fall-risk event is incorrectly predicted as Safe (missed risk)

6.6 Assessment of the Created Dual-Trigger Hybrid AI-3 Model

We proposed a novel hybrid model comprising MediaPipe Pose and a deep neural network with a GRU and an attention layer for time-series input. The model is trained with a sequence length of 30. The learning rate of an adaptive optimizer is 1 × 10^-3, and 40 epochs are used with the hyperparameters specified and optimized as indicated in Table 6.2. During compilation, the categorical cross-entropy loss is used for multiclass problems with one-hot labels. Accuracy is the metric used to track classification performance during training. During training, the model converged at epoch 12, as shown in Table 6.3, with transitions in accuracy and loss across the training and validation datasets. Using the Spatio-Temporal GRU model, we classified the test dataset into three groups. The three-class classification table, with performance metrics for each class, is shown in Table 6.4. As a result, Class 2 is the dominant class. Displaying the confusion matrix (Figure 6.6) further supports the claim. In multiclass problems, TP/FP/FN/TN are computed per class using a one-vs-rest view, where each class's diagonal entry is TP, its row gives FN, its column gives FP, and all remaining cells are TN.

Table 6.2: Optimized Hyperparameter Settings

Hyperparameter	**Optimized value**
Optimizer	Adam
Learning Rate	1×10^{-3}
Epochs	40
Time_Steps/SEQ_LEN	30
Loss function	Categorical cross entropy

Table 6.3: Model Convergence with epochs

Epochs	Training		Validation	
	Accuracy	Loss	Accuracy	Loss
1	0.9334	0.2208	0.9807	0.0511
12	0.9931	0.0156	0.9893	0.0206

Table 6.4: Classification Performance of the Spatial-Temporal GRU Model on the Testing Data

Class	Accuracy	Precision	Sensitivity (Recall)	Specificity	F1-score	Support
Safe (0)	0.5000	0.5000	1.0000	0.9979	0.6666	1
Near_fence (1)	1.0000	0.9952	0.9881	0.9574	0.9916	420
High risk (2)	0.9201	0.9166	0.9565	0.9905	0.9361	46
Macro avg	-	0.8040	0.9804	1.0000	0.8648	467
Weighted avg	-	0.9864	0.9910	NC*	0.9855	467

In addition to the multiclass confusion matrix, one-vs-rest confusion matrices were generated for each ADL class to compute class-specific sensitivity and specificity, as depicted in Figure 6.7 (a), (b), and (c). We have also shown comparisons of the proposed model's metrics with those of the previous two models to demonstrate the model’s modest reduction in false alarms. However, the support is different in each case. Furthermore, the suggested model allows 120 frames per window, whereas AI-1 and AI-2 used frames as the model's training input. Thus, Table V shows the least false alarm rate for binary

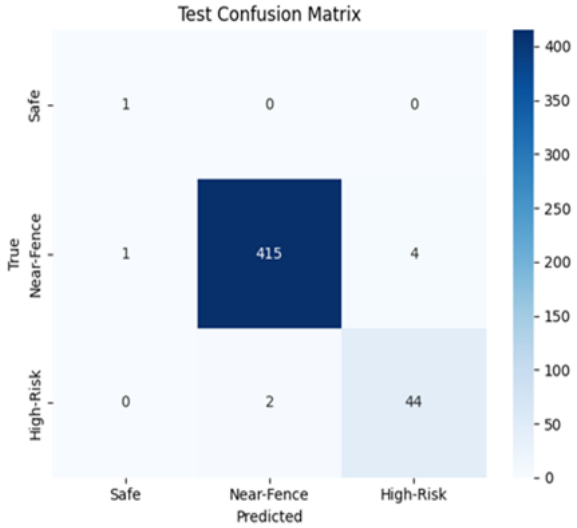


Figure 6.6: Confusion matrix

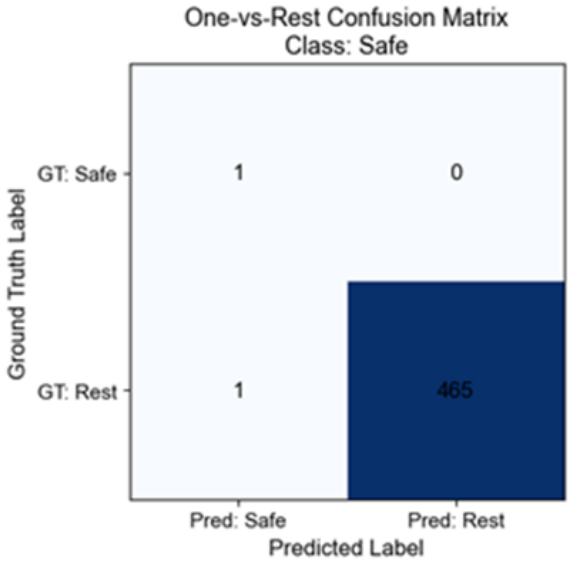


Figure 6.7 (a) Confusion matrix for the Safe class: One vs. Rest

classification, whereas the proposed model performs well in multiclass classification. To overcome this discrepancy, we added a dual-trigger event to the evaluation stage, along with a frame-level event.

The proposed method has achieved 99.76% sensitivity and 100% specificity, demonstrating a modest but promising improvement in reducing false alarms. Table 6.5 compares the proposed model with

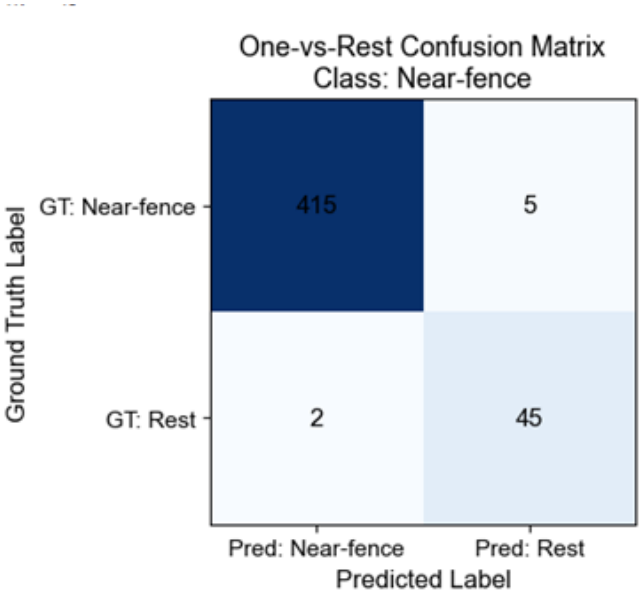


Figure 6.7 (b) Confusion matrix for the Near-Fence class: One vs. Rest

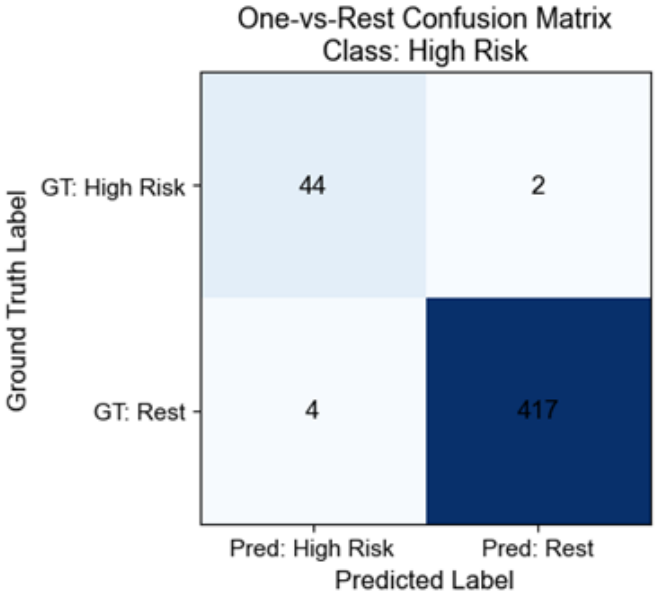


Figure 6.7 (c) Confusion matrix for the High-Risk class: One vs. Rest

2 previous state-of-the-art benchmark methods, as explained in Chapters 4 and 5, for baby fall prediction. The baby's behaviour is classified over a 10-second interval using the same AI-generated test video. Table 6.6 presents a temporal comparison between “UNSAFE” and Trigger-1/Trigger-2 for PE+RF (AI-1), SSCNN (AI-2) and the proposed model (PE+GRU+SSA). Few temporal periods were correctly identified as SAFE; however, it was unable to identify the correct UNSAFE scenarios. Nearly the whole interval is marked as dangerous by SSCN. In this case, early warning is considered the final alert. However, GRU+SSA is designed to prevent premature notifications. Trigger-1 suppresses transient postural noise by requiring prolonged standing beyond a 5-7-second temporal threshold, whereas Trigger-2 is triggered when a high-risk ADL occurs in a hazardous spatial context. Trigger 1

thus continues to deliver early warnings, whereas the proposed model raises Trigger 2 only at 40–80 and 100–110 seconds.

Table 6.5: Comparison of metrics of our current proposed method with AI-1 and AI-2

Models	**Methods**	**Features**	**Sensitivity (%)**	**Specificity (%)**
AI-1	PE + Random Forest	Keypoints	65.20	84.41
AI-2	CNN + SSA	Spatial	99.67	99.90
Current proposed method	PE + GRU + SSA	Keypoints + Temporal + Spatial	99.76	100.00

Table 6.6: Comparison of unsafe event detection using SSCNN and the proposed model

Time duration (secs)	**AI-1 PE + RF**	**AI-2 SSCNN**	**Proposed PE+GRU+SSA**	
	UNSAFE (HIGH RISK)	**UNSAFE (HIGH RISK)**	**TRIGGER-1[a] (WARNING)**	**TRIGGER-2[b] (HIGH RISK ALERT)**
0-10	SAFE	UNSAFE	✓	✗
10-20	SAFE	UNSAFE	✓	✗
20-30	SAFE	UNSAFE	✓	✗
30-40	UNSAFE	UNSAFE	✓	✗
40-50	SAFE	UNSAFE	✓	✓
50-60	SAFE	UNSAFE	✓	✓
60-70	SAFE	UNSAFE	✓	✓
70-80	SAFE	UNSAFE	✓	✓
80-90	SAFE	UNSAFE	✓	✗
90-100	SAFE	UNSAFE	✓	✗
100-110	SAFE	UNSAFE	✓	✓
110-120	UNSAFE	UNSAFE	✓	✗
120-130	SAFE	UNSAFE	✓	✗

Figure. 6.8 shows the graph of the escalation of Trigger-1 and Trigger-2. Because near fence + climbing/jumping/leaning triggers an alarm, while near fence + standing does not, the dual-trigger logic reduces false positives. This makes it easier to understand the reasons behind the alert's triggering, i.e., persistent dangerous status, and High-risk ADLs. Therefore, the model improves interpretability and usefulness by avoiding continuous dangerous flags and grouping risk into a small alert window.

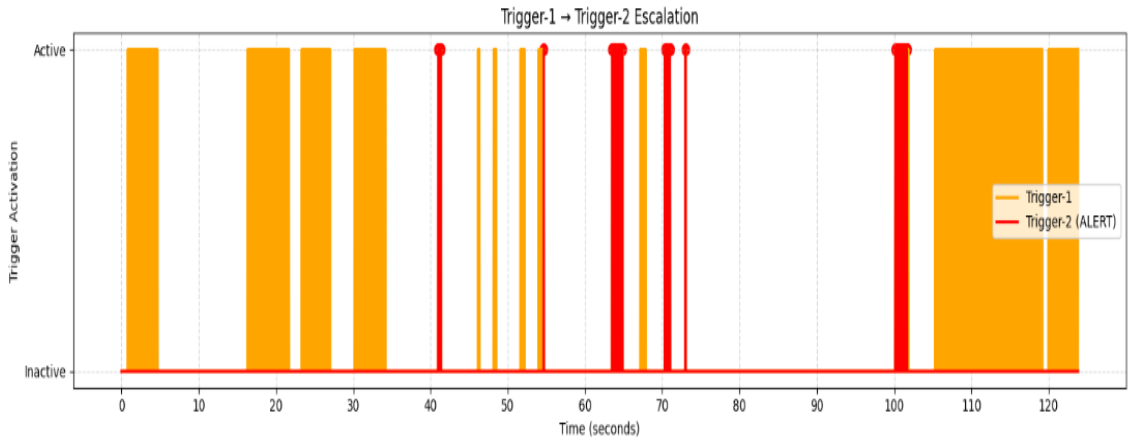


Figure 6.8: Graph showing Trigger-1 to Trigger-2 escalation

Table 6.7: Risk Density and Segment-level ADL Statistics

Segment No.	Time Interval	Frame Range	Sitting (0)	Standing (1)	High-risk (2)	Dominant ADL	Risk present	Risk Density (%)
0	0-10	0-300	0	157	143	Standing	Yes	47.67
1	10-20	300-600	0	179	121	Standing	Yes	40.33
2	20-30	600-900	0	248	52	Standing	Yes	17.33
3	30-40	900-1200	0	162	138	Standing	Yes	46
4	40-50	1200-1500	0	196	104	Standing	Yes	34.67
5	50-60	1500-1800	0	145	155	High-risk	Yes	51.67
6	60-70	1800-2100	0	136	164	High-risk	Yes	54.67
7	70-80	2100-2400	0	72	228	High-risk	Yes	76
8	80-90	2400-2700	0	36	264	High-risk	Yes	88
9	90-100	2700-3000	1	30	269	High-risk	Yes	89.67
10	100-110	3000-3300	0	225	75	Standing	Yes	25
11	110-120	3300-3600	0	299	1	Standing	Yes	0.33
12	120-130	3600-3716	0	116	0	Standing	No	0

Although standing is more common, there remains a risk, which necessitates the alerts. In most 10-second intervals, standing is the predominant activity, although there are sporadic short, high-risk behaviours. These brief but crucial occurrences are detected by the proposed dual-trigger system, which generates alarms even when high-risk activity does not predominate in the segment. According to the risk density shown in Table 6.7, warnings are associated with short-lived but semantically significant events that consume little time. Therefore, it is rule-based safety gating over learned representations, hierarchical decision fusion, and semantically conditioned risk escalation.

6.7 Comparison of Previous Studies and the Proposed Method

The performance measures of the various models employed in previous research are compared in this section. The research conducted in [18, 19] shows lower recall than the proposed method. Moreover, neither of the research studies mentioned specificity, whereas the proposed method achieved 100% specificity. A detailed comparison of prior AI-based studies with the baby fall prediction system is presented in Table 6.8. Almost every study focuses on the general state of baby safety, as well as the limitations of either action recognition or a one-stage warning mechanism. While the proposed method incorporates progressive warnings, it still needs to be tested in a real-world setting. Table 6.8 provides a detailed overview of performance metric comparisons across existing work.

Table 6.8: Summary of the comparison of AI-based studies for baby fall prediction in the crib

Research Work	**Purpose of the Study**	**Modality**	**Parameters used for predicting Fall Risk**	**Accuracy (%)**	**Sensitivity/Recall (%)**	**Precision (%)**	**Limitations**
[13]	To develop a safety monitoring system for a baby by detecting and monitoring various activities of the baby in the crib or at home.	Single modality (RGB video only)	Only jumping activity	90	87	86	No early prediction, only recognition, no explicit pose modelling and no spatial attention mechanism
[24]	To develop a learning-based model for early prediction	Multimodal (RGB video frames + motion)	only visual features	Not explicitly mentioned	80	53.68	No explicit pose modelling. Generalization issues are not fully addressed. Low precision, less interpretable as CNN features are black-box,

	of baby accidents						high computational complexity and not optimized for real-time deployment. Not even lighter and faster. No progressive warning system is used to predict the accident. No differentiation between imminent fall and risk state.
Our proposed work	To develop a two-stage trigger early prediction model for baby fall risk within the crib setting	RGB video + keypoints data	joint coordinates, visual features, and temporal features	98.5	99.76	80	Need to test the model in a real-world setting

6.8 Conclusions

The Semantic-conditioned Hybrid model introduced an innovative methodology for predicting baby falls from the crib, thereby reducing false alarms. The preceding SSCNN model predicted vision-based fall risk in 12- to 24-month-olds in a cot environment, resulting in false alarms. To address this issue, the proposed model's key contribution is to develop a pose-based spatio-temporal GRU with SSA and dual-trigger logic that combines risk persistence and high-risk ADL to manage alert escalation. This proposed system incorporates the PE technique using MPP to extract keypoints and calculate relevant parameters to estimate arbitrary poses and motions. GRU is employed to capture sequential dependencies, mapping and classifying the changing poses into three formulated classes within the crib environment. The attention network selectively focuses on crucial parameters, dynamically adjusting feature weights to achieve accurate predictions. High predictive accuracy is confirmed by final performance testing, with 98.5% accuracy, 100% specificity, and 99.76% sensitivity. These measures demonstrate the system's ability to correctly identify babies at risk and to maintain low false-positive and false-negative rates. 100% specificity ensures that those rated low risk pose little risk, and 99.76% sensitivity ensures that the system identifies babies in the crib who are at risk of falling. This research reveals strong sensitivity and near-perfect specificity, fewer ongoing dangerous signals, and more interpretable alerts at the event level.

Chapter 7. Conclusions and Future Directions

7.1 Introduction

This chapter summarizes the research results on the new ADL-conditioned TL-based hybrid fall risk prediction model, focusing on its key features and outcomes. To build a flexible, adaptable fall prediction model, significant focus is placed on cutting-edge technologies, including AI-based analysis, system integration, and software solutions. This advanced prediction system is designed to proactively assess the risk of baby falls, notifying parents and caregivers without the anxiety of false alarms. Thereby facilitating the implementation of preventive measures before a fall. This chapter also talks about where future research on the system's functionality should go.

The integration of real-time predictive models, which allow for immediate adaptation and modification in response to new data, is a key area for progress. This allows the model to improve its predictions over time using real-time ADLs. Future efforts will lead to a more advanced and complete fall-risk prediction system by enabling an ADL-conditioned hybrid deep learning model to integrate with the Internet of Things (IoT). Using pressure mats or sensors under the crib mattress will provide positional data and continually improve the crib's predictive accuracy. The objective is to develop a smart, self-learning system capable of refining and adapting its fall risk predictions. Thus, offering parents or caregivers more precise and tailored forecasts of future fall risks

7.2 Conclusions

This thesis served as the motivational foundation for developing an ADL-conditioned, TL-based hybrid fall-risk prediction system. Further, evaluating its viability in dynamically assessing and addressing potential fall threats among infants in a crib. Considering the problem space, the essential components required to enhance a viable, adaptive, and proactive fall prediction system are rigorously analyzed. It has focused on two domains: continuous activity monitoring and intelligent risk assessment. The research highlights the integration of postural data, spatial cues, motion dynamics, and temporal characteristics to develop a resilient predictive model. This should accurately identify at-risk babies and deliver timely warnings. Thus, the model will facilitate early intervention to avert falls and ensure pre-emptive care for the babies.

The first part concerns the proposed Baby Unsafe Scenarios Prediction Model, derived from the logic of PE and ML, and meticulously designed to identify dangerous and unsafe poses. The main motive for developing the model was to design an effective and efficient predictor of unsafe scenarios, i.e., potential fall risk. Thus, reducing the burden on parents who care for their babies all night. Additionally, addressing the imbalance in developmental outcomes related to fall injuries among babies. By employing a pose estimation technique that uses joint coordinates as keypoints, the model can identify dangerous or unsafe poses or potential fall risk and categorize the defined ADLs into safe or unsafe

outcomes. Key points enable the model to accurately identify postures by providing joint coordinate values, which are highly beneficial for addressing baby safety. The existing approaches rely on monitoring behaviour or activities of babies, addressing the activity recognition using camera or sensors, decoding non-verbal communication using pose features, or outside the crib. However, the developed model is distinct in its exclusive reliance on keypoints for pose estimation and in classifying the resulting feature vectors using an ML-based classifier. This will make it an effective and efficient tool for predicting potential falls. Compared with four ML classifiers, RF achieved the best average performance across six ADLs, with 51% accuracy and 47% recall. RF achieved the highest average accuracy for the ADL-based outcome, at 75%. These performance indicators indicate the model's efficiency and reliability in assessing potential falls and identifying optimal HPE classifiers for a baby's unsafe-scenarios prediction system.

Even though the developed AI-1 model is fairly accurate, it is still being talked about because the results aren't very promising yet due to the size of the datasets. To make the model even more accurate, the dataset size should be increased. Furthermore, AI-1 is a pose-based classification model that lacks background information, which can be directly learned from the pixels if the model is entirely end-to-end image-based. Acquiring only keypoints can estimate poses, compromising accuracy with respect to the baby's background, position, and location, motivating the development of a Deep Neural Network model leveraging spatial cues via a Soft spatial attention mechanism.

The second section concerns the proposed unsafe-scenario prediction model based on CNNs and SSA. It is specifically intended for babies in the crib and designed to detect the risk of falls or unsafe scenarios by continuously observing behavioural patterns, particularly for six specific ADLs. The model only predicts fall risks based on frame-based behavioural data. With an accuracy of 99.79%, sensitivity of 99.67%, and specificity of 99.90%, the model can accurately categorize babies at risk of falls from those who are not. A significant limitation of the current study is that the model only incorporates frames, with no continuous-time series input. Although it was trained on a large dataset of approximately 26000 frames, the model is limited to spatial features. The lack of continuous time-series data affects the model’s predictions for a few ADLs, such as standing and leaning. The frame-based evaluation model generates multiple alerts for all standing and leaning frames, considering them as unsafe. As a result, the model achieved high accuracy, indicating that including continuous time series can influence prediction results, either increasing or decreasing sensitivity and specificity. Hence, the outcome of the change is unknown. To generalize the model, it was evaluated on an unseen dataset, leveraging temporal coefficients and statistical parameters. The model with statistical parameters performed more efficiently than one without. This showed that a more effective and efficient model can be designed using a continuous-time series dataset and a temporal model.

Only keypoints can estimate poses, but if the pose features aren't engineered, accuracy is slightly lower. To amend the limitations of AI-1 and AI-2, we develop a third model, AI-3. This model incorporated

the eight ADL videos as continuous-time series inputs. After feature engineering, several parameters were derived, such as Torso angle, Center of Mass, Hand distance to the rail, etc., to analyze the baby's position within the crib and posture. The sequence of featured vectors helped the model in temporal smoothing. The decision system was built in two stages to minimize the false alarms. Thus, the model achieved increased sensitivity and specificity of 99.76% and 100%, respectively, thereby reducing the accuracy of 98.5%. Furthermore, the model shows effectiveness and efficiency by reducing false alarms. The feature representation (pose) was enhanced, and a temporal model (GRU) and an attention mechanism were added, rather than simply changing the loss function. This makes the system work better even with standard loss functions.

However, the solution proposed in this research is still in the simulation testing stage, and its feasibility in real-world scenarios needs further research and verification. The model needs to be trained and evaluated on a larger dataset. In addition, the model must be integrated with a mobile app to provide real-time alerts to parents or caregivers.

Despite these limitations, this proposed ADL-conditioned Hybrid model aims to transfer learning and outcomes from each block to the next and to learn from prior data to provide future fall-risk predictions with very high accuracy and reduced false alarms. Several advantages of this research work are:

- *Preserves Life* – mitigates the danger of subsequent falls and injuries
- *Parents/Caregivers Relaxation* – alleviates the constant load of round-the-clock vigilance
- *Non-Wearable* – purely AI-driven model mitigates the discomfort of wearing the devices.
- *Secure and Safe* – innocuous and alerts parents or caregivers in a two-tiered manner according to the severity of the incident.
- *Basic Infrastructure and affordability* – necessitate no intricate infrastructure, making them readily economical
- *Seamless Smart System Integration* – this research idea can be expanded to encompass various hazard detection systems or other health-related applications.

7.3 Future Directions

The contribution of the thesis necessitates further study to improve the overall concept. Future efforts to expand this research can be categorized into three stages.

Stage 1: Integrating the AI-3 model with the physical sensors beneath the mattress, enabling parents or caregivers to receive prompt notifications about precise location, posture, and fall risk. This AI system, integrated with IoT, will serve as a prototype to improve overall system functionality. This system can be further enhanced to provide real-time data collection using a camera aimed at the baby in the crib and pressure mats beneath the mattress. These sensors provide a comprehensive fall-risk assessment for babies by continuously recording key behavioural and positional parameters. By leveraging the

advantages of the suggested AI-based approach, including Pose Estimation, ADL-Conditioned, and Transfer Learning Network, the hybrid model can predict future fall risk. The integrated mobile application can provide either a warning or a high-risk alarm. Figure 7.1 illustrates the prospective integration of software with hardware, establishing an AIoT-based paradigm that provides timely alerts.

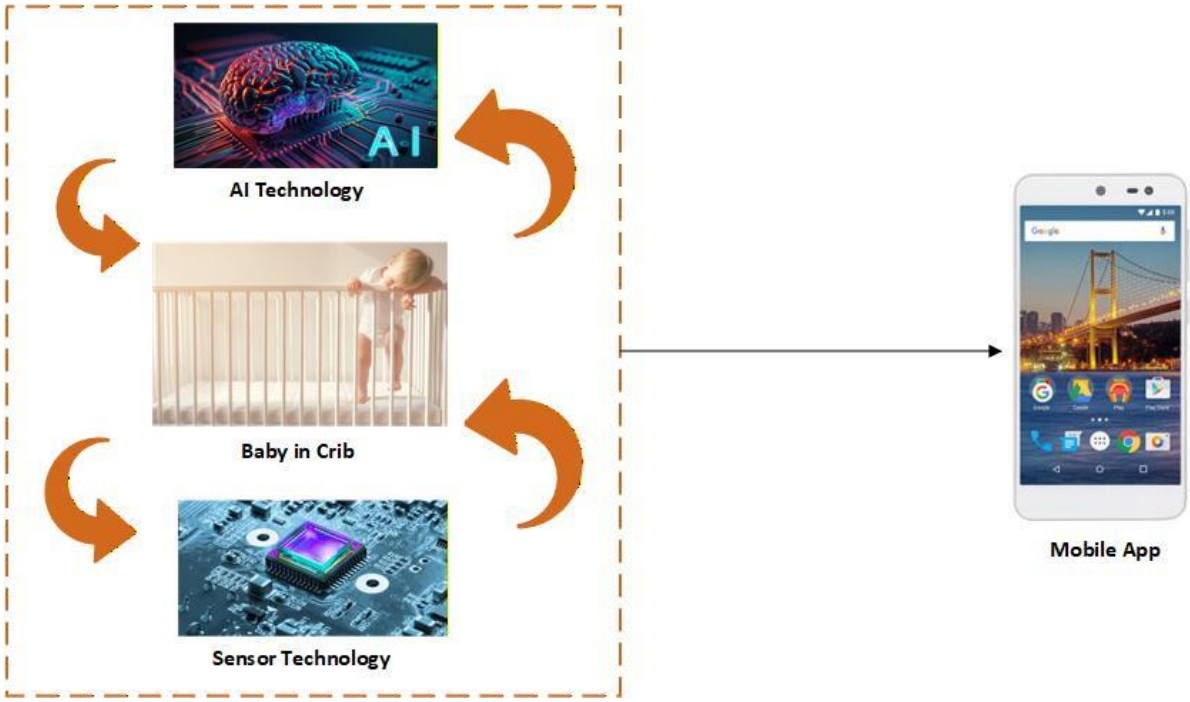


Figure 7.1: Integration of AI-3 Model with Sensor Technology and Smart App.

Stage 2: Like in stage 1, AI-3 can also work with sensors that measure the environment and the body to make the system stronger and more efficient. To make a smart mattress [132, 133], it would be a good idea to put physiological sensors like heart rate, breathing, environmental, movement, and sound sensors under the mattress within the crib. The suggested model with these extra features would make it easier to effectively estimate the likelihood of falls and provide parents or caregivers with more information about the baby's health. Also, voice models for excessive weeping can be used to set up an immediate voice alert, and immediate safety precautions can be put in place. The integrated arms of health monitoring, alerting, and predictive capabilities outlined above make babies in a crib even safer and more secure. They help prevent falls and promote health. This integrated solution makes a significant impact on fall prevention and safety by combining AI, non-wearable monitoring, and real-time health management. It also makes a smart crib. Figure 7.2 clearly shows the combined integration structure.

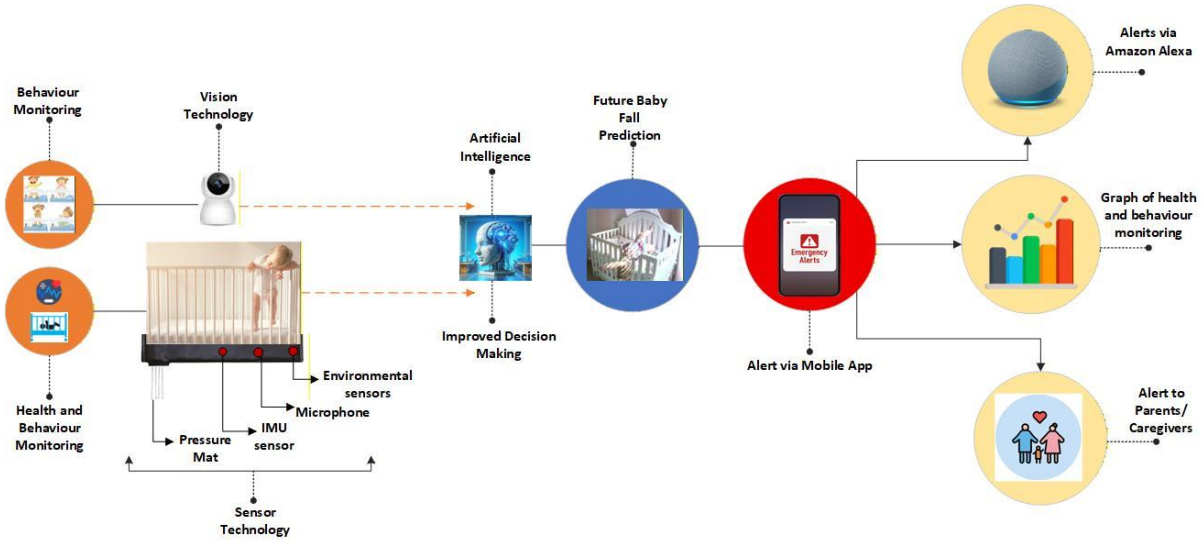


Figure 7.2: Combined Integration Structure of Future Work – Stage 2

Stage 3: To get ready for future changes, the model suggests adding a new development phase that combines AI-3, a system for real-time monitoring, with a transfer learning approach. This new feature will use key mobility and physical data from pressure mats to improve the model's accuracy and enable it to update its predictions in real time. Adding continuous learning algorithms will help the system learn more about the risk factors that lead to falls while maintaining high levels of accuracy, sensitivity, and specificity that can't be matched when predicting future falls. At this final stage, the primary things that are being worked on are as follows,

- Māori children with no conditions.
- Babies in New Zealand whose parents do not want childcare or a nanny.
- Babies in New Zealand with disabilities who need extra care.

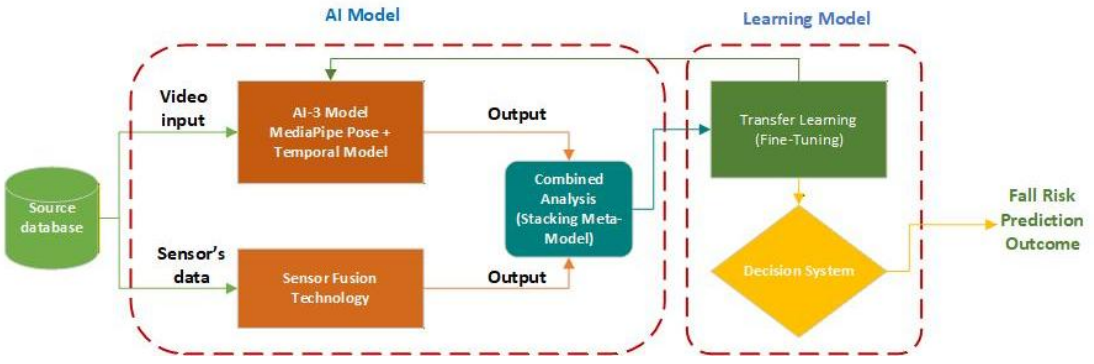


Figure 7.3: Future Directions - Baby Fall Risk Prediction Model using Transfer Learning

In future work, the system aims to develop an advanced, non-invasive, continuous fall-risk monitoring system for babies to provide caregivers with real-time alerts and mitigate the risk of fall-related injuries in and out of cribs. The model will utilize a multi-modal AIoT-based architecture comprising an

improved version of AI-3 (mentioned in stage 1) and another AI model using Transfer learning to integrate AI-3's output into a consolidated fall risk assessment. Figure 7.3 shows how the baby fall risk prediction system will work in the future, using AI and a learning model.

This architecture is based on the existing AI-3 model, which uses the Transfer Learning approach. Adding sensor fusion technology to AI-3 makes a stacking meta-model, an ensemble learning method. The output of the combined analysis will be fed to another AI model (Transfer learning Model), whose continuous learning capability will enable real-time detection of changes in position and behavioural patterns, supporting timely detection of high fall risk. It will produce a personalized, changing, and up-to-date fall risk profile for the baby. This is particularly beneficial, as it will provide doctors, parents, and caregivers with accurate, actionable insights to develop, test, and measure proactive, tailored fall-prevention interventions.

This approach is distinctive for its use of an existing standalone AI-3 model for posture estimation via keypoints, a spatial attention mechanism that enhances spatial features, and the incorporation of temporal dependencies from the temporal model. Moreover, sensor fusion technology enhances predictive precision and reduces false alarms by providing more precise baby placements, thereby refining the accuracy of the "center of mass" parameter computed from keypoints in the current AI-3 model. This method operates autonomously while facilitating interaction between the AI model and the learning model (the new AI model). The model will continually learn and refine its predictions using ground truth, historical data, and the most recent forecasts. The final prediction of fall risk will be analyzed in accordance with the decision system's rules.

The hybrid approach and the system's interactive nature enable continuous information sharing between the models. This approach underpins robust risk prediction through the synchronization of behavioural patterns and sensor data. In addition, the model would be able to learn continuously, enabling adaptation to change in input and facilitating the timely capture of changes in behaviour leading to fall risk. Following the phrase “Prevention is always better than a cure”, this research focuses on fall prevention, an area that has been underrepresented in previous studies. Henceforth, this research aims to create a safer, more reliable environment for babies. Moreover, owing to its AIoT-based transfer-learning methodology, the model learns from current and prior data, as well as its GT. It then yields more precise forecasts than existing techniques for fall risk assessment. The decision system relies on analysis of the AI model's and learning model's outputs to deliver an accurate fall-risk prediction. This study promises to provide a transformative tool for pediatric injury prevention, giving caregivers peace of mind and enhancing baby safety at home.

References

[1] S. Daley. "AI in Healthcare: Uses, Examples and Benefits." https://builtin.com/artificial-intelligence/artificial-intelligence-healthcare (accessed 13 January 2026, 2026).

[2] A. T. Özdemir, "An analysis on sensor locations of the human body for wearable fall detection devices: Principles and practice," *Sensors,* vol. 16, no. 8, p. 1161, 2016.

[3] Z. K. Lim, T. Connie, M. K. O. Goh, and N. I. B. Saedon, "Fall risk prediction using temporal gait features and machine learning approaches," *Frontiers in Artificial Intelligence,* vol. 7, p. 1425713, 2024.

[4] Z. Yang, B. Tsui, J. Ning, and Z. Wu, "Falling Detection of Toddlers Based on Improved YOLOv8 Models," *Sensors,* vol. 24, no. 19, p. 6451, 2024.

[5] "Nationwide roll out of artificial intelligence tool that predicts falls and viruses," 5 March 2025. [Online]. Available: https://www.england.nhs.uk/2025/03/nationwide-roll-out-of-artificial-intelligence-tool-that-predicts-falls-and-viruses/

[6] M. Javaid, A. Haleem, R. P. Singh, R. Suman, and S. Rab, "Significance of machine learning in healthcare: Features, pillars and applications," *International Journal of Intelligent Networks,* vol. 3, pp. 58–73, 2022.

[7] T. H. Harahap *et al.*, "An artificial intelligence approach to predict infants' health status at birth," *International Journal of Medical Informatics,* vol. 183, p. 105338, 2024.

[8] "Labor Force participation rate, female (% of female population ages 15+)." World Bank Data. https://data.worldbank.org/indicator (accessed 16 April, 2022).

[9] "Fast facts: Mothers in the Workforce (2019)." The American Association of University Women. https://www.aauw.org/resources/article/fast-facts-working-moms (accessed 13 April, 2022).

[10] B. S. Cheridan Christnacht. "The Choices Working Mother Makes." United Stated Census Bureau. https://www.census.gov/library/stories/2020/05/the-choices-working-mothers-make (accessed 13 April, 2022).

[11] "World Health Organisation (2020) Child Mortality (under 5 Years)." World Health Organisation. https://www.who.int/news-room/fact-sheets/detail/levels-and-trends-in-child-under-5-mortality (accessed 12 April, 2022).

[12] "COVID-19: Case Demographics." New Zealand Ministry of Health. https://www.health.govt.nz/covid-19- novel-coronavirus/covid-19-data-and-statistics/covid-19-case-demographics (accessed 13 April, 2022).

[13] "Child Unintentional Deaths and Injuries in New Zealand and Prevention Strategies." University of Otago. http://ipru3.otago.ac.nz/niqs (accessed 28 November, 2022).

[14] "Medium to long-term employment outlook: Looking ahead to 2028." https://www.mbie.govt.nz/business-andemployment/employment-and-skills/labour-market-reports-dataand-analysis/archives/labour-market-forecasting/medium-to-longterm-employment-outlook-looking-ahead-to-2028/ (accessed 27 November, 2024).

[15] J. Gerritsen. "ECE providers 'struggling' amid falling enrolments." RNZ Education. https://www.rnz.co.nz/news/national/503561/ece-providers-struggling-amid-falling-enrolments (accessed 11 April 2024.

[16] P. W. Āwhina. "Safety Around The Home." https://www.plunket.org.nz/caring-for-your-child/safety-around-the-home/safety-at-home/ (accessed 21/07/2025, 2025).

[17] N. Khan, D. Z. Al-Hamid, P. H. Chong, F. Mirza, and A. Kumar, "Systematic review of sleep monitoring systems for babies: research issues, current status, and future challenges," *Health & Social Care in the Community,* vol. 2024, no. 1, p. 5510164, 2024.

[18] V. Adewopo, N. Elsayed, and K. Anderson, "Baby physical safety monitoring in smart home using action recognition system," in *SoutheastCon 2023*, 2023: IEEE, pp. 142–149.

[19] P.-J. Wang, S.-F. Lien, and M.-S. Lee, "A learning-based prediction model for baby accidents," in *2019 IEEE International Conference on Image Processing (ICIP)*, 2019: IEEE, pp. 629–633.

[20] S. Foundation, "Falls," 2020. [Online]. Available: www.starship.org.nz/safekids/reference-cards.

[21] S. Sahni, S. Jain, and S. K. Saritha, "A novel ensemble model for fall detection: leveraging CNN and BiLSTM with channel and temporal attention," *Automatika,* vol. 66, no. 2, pp. 103–116, 2025.
[22] M. Nasr, M. M. Islam, S. Shehata, F. Karray, and Y. Quintana, "Smart healthcare in the age of AI: recent advances, challenges, and future prospects," *IEEE access,* vol. 9, pp. 145248–145270, 2021.
[23] D. Han and K. E. Adolph, "The impact of errors in infant development: Falling like a baby," *Developmental science,* vol. 24, no. 5, p. e13069, 2021.
[24] K. Kokulu, A. Algın, S. Özdemir, and H. Ş. Akça, "Characteristics of injuries among infants who fall from bed," *Injury,* vol. 52, no. 2, pp. 281–285, 2021.
[25] S. Chaudhary *et al.*, "Pediatric falls ages 0–4: understanding demographics, mechanisms, and injury severities," *Injury epidemiology,* vol. 5, no. Suppl 1, p. 7, 2018.
[26] "Preventing falls in babies (0 to 12 months)." [Online]. Available: https://www2.hse.ie/babies-children/child-safety/at-home/preventing-falls-babies-0-12-months/
[27] "Preventing falls in toddlers and children (1 to 5 years old)." [Online]. Available: https://www2.hse.ie/babies-children/child-safety/at-home/preventing-falls/
[28] "Preventing Injuries From Falling, Climbing, and Grabbing." [Online]. Available: https://kidshealth.org/en/parents/safety-falls.html
[29] (2013). *Keeping Kids Safe*. [Online] Available: https://www.consumerprotection.govt.nz/assets/PDFs/keeping-kids-safe.pdf
[30] (2011). *73-040,* [Online] Available: https://www.cpsc.gov/Newsroom/News-Releases/2012/CPSC-Sets-Crib-Safety-Standards
[31] J. R. Duncan and R. W. Byard, "Sudden infant death syndrome: an overview," 2018.
[32] E. A. Mitchell, "SIDS: past, present and future," *Acta Paediatrica,* vol. 98, no. 11, pp. 1712–1719, 2009.
[33] M. V. Ridenour, "How do children climb out of cribs?," *Perceptual and motor skills,* vol. 95, no. 2, pp. 363–366, 2002.
[34] A. N. Hassan, A. A. Mire, N. M. Mohamud, A. I. Aadan, M. A. Ali, and A. A. Siyad, "IoT-Enabled Smart Crib for Baby Health Monitoring, Sleep Tracking, and Parental Notification System," in *International Conference on Artificial Intelligence and Smart Energy*, 2025: Springer, pp. 139–149.
[35] D. Bigueta, I. Canonicato, A. David, M. Thomas, and D. Tran, "Tenshi Baby Crib," 2021.
[36] S. Joseph, A. Kumar, and M. H. Babu, "IOT based baby monitoring system smart cradle," in *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)*, 2021, vol. 1: IEEE, pp. 748–751.
[37] N. B. Prasetyo, D. Rahmawati, W. Priharti, and M. Dhalhaz, "Enhancing Infant Safety: Performance Analysis of Deep Learning Method on Development Board for Real-Time Monitoring," in *International Conference on Electronics, Biomedical Engineering, and Health Informatics*, 2023: Springer, pp. 285–302.
[38] S. Zhu *et al.*, "CribNet: Enhancing Infant Safety in Cribs Through Vision-Based Hazard Detection," in *2024 IEEE 18th International Conference on Automatic Face and Gesture Recognition (FG)*, 2024: IEEE, pp. 01–08.
[39] J. Yin, J. n. Zhao, and X. Zhang, "Design of a hazard prediction system with intelligent multimodal fusion based on artificial intelligence & internet of things technology: taking a crib as an example," *PeerJ Computer Science,* vol. 10, p. e2404, 2024.
[40] N. Abed, S. E. Aber, N. A. Malek, and S. Bentrad, "Genius Nest: A Multimodal AI and IoT Smart Crib for Infant Safety Monitoring," in *2025 International Conference on Artificial Intelligence and Innovative Applications (AIIA)*, 2025: IEEE, pp. 1–8.
[41] T. Khan, "An intelligent baby monitor with automatic sleeping posture detection and notification," *Ai,* vol. 2, no. 2, pp. 290–306, 2021.
[42] K. Deepti, G. Akshita, K. V. Swamy, R. Devi, G. Eshwar, and M. Vaishnavi, "An Automated Infant Baby Monitoring and Response System using Deep Learning," in *2025 5th IEEE International Conference on Applied Electromagnetics, Signal Processing, & Communication (AESPC)*, 2025: IEEE, pp. 1–6.

[43] H. Alam *et al.*, "IoT based smart baby monitoring system with emotion recognition using machine learning," *Wireless Communications and Mobile Computing,* vol. 2023, no. 1, p. 1175450, 2023.
[44] A. Moodbidri and H. Shahnasser, "Child safety wearable device," in *2017 International Conference on Information Networking (ICOIN)*, 2017: IEEE, pp. 438–444.
[45] J. Suud, F. H. Suliman, and K. H. Shahari, "Enhancing Safety: A Smart Baby Monitoring System for Deaf Parents," *Borneo Engineering & Advanced Multidisciplinary International Journal,* vol. 4, no. 1, pp. 31–38, 2025.
[46] M. M. E. Yurtsever and S. Eken, "BabyPose: Real-time decoding of baby's non-verbal communication using 2D video-based pose estimation," *IEEE Sensors Journal,* vol. 22, no. 14, pp. 13776–13784, 2022.
[47] M. M. Ali and S. I. Mohamed, "A pose estimation for motion tracking of infants cerebral palsy," *Multimedia Tools and Applications,* pp. 1–26, 2024.
[48] F. A. Khan and A. Dey, "Towards Enhancing Child Safety: A Deep Learning Approach to Detect Child Safe and Unsafe Objects," in *2024 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE)*, 2024: IEEE, pp. 123–128.
[49] J. H. Tan and C. P. Goh, "Enhancing Child Safety: Computer Vision-Based Accident Detection for Infants and Toddlers," in *2024 3rd International Conference on Digital Transformation and Applications (ICDXA)*, 2024: IEEE, pp. 1–5.
[50] Z. Yang, B. Tsui, and Z. Wu, "Assessment system for child head injury from falls based on neural network learning," *Sensors,* vol. 23, no. 18, p. 7896, 2023.
[51] K. Sivakrishna, J. S. Vaishnavi, A. Sreelekha, S. Sindhu, C. Varshitha, and G. Kalakoti, "Object Detection and Hazard Alert System for Child Safety on Robot using YOLO," in *2025 5th International Conference on Expert Clouds and Applications (ICOECA)*, 2025: IEEE, pp. 1094–1098.
[52] R. Ahmed *et al.*, "Smart Child Monitoring System with Real-Time Danger Detection Using Artificial Intelligence," in *2025 Twelfth International Conference on Intelligent Computing and Information Systems (ICICIS)*, 2025: IEEE, pp. 771–776.
[53] P. Kalpana, M. Chithritha, J. Divyakarthikaa, and N. Gayathri, "AI-Driven Child Safety System: Real-Time Fear Detection and Emergency Response using Enhanced Deep Learning Algorithms," in *2025 2nd International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE)*, 2025: IEEE, pp. 1–8.
[54] A. Maiti, C. Dutta, J. S. Banerjee, and P. Sarigiannidis, "AI for infant well-being: advanced techniques in cry interpretation and monitoring," *J. Mech. Contin. Math. Sci,* vol. 19, no. 2, 2024.
[55] K. Ravindran, M. Ibrahim, H. Lee, U. Qureshi, and K. A. Hafeez, "Design & development of an AI-powered baby monitoring system," in *2021 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, 2021: IEEE, pp. 1–6.
[56] Y. Lee, K. K. Kim, and J. H. Kim, "Prevention of safety accidents through artificial intelligence monitoring of infants in the home environment," in *2019 International Conference on Information and Communication Technology Convergence (ICTC)*, 2019: IEEE, pp. 474–477.
[57] S. Daley. "AI in Healthcare: Uses, Examples and Benefits." https://builtin.com/artificial-intelligence/artificial-intelligence-healthcare (accessed 13 January 2026, 2026).
[58] "Births and deaths: Year ended March 2025." Stats NZ. https://www.stats.govt.nz/information-releases/births-and-deaths-year-ended-march-2025/ (accessed 26 February, 2026).
[59] R. R. Kumar, P. Sanjeeva, D. Reddy, and A. Reddy, "Smart Monitoring System for Physically Handicapped Infants using IoT," in *E3S Web of Conferences*, 2023, vol. 391: EDP Sciences, p. 01161.
[60] M. Amutha, V. Sudha, B. Abinanthan, and S. Dinesh, "Empowering child safety through smart monitoring and environmental insights," in *Applications of AI in Smart Technologies and Manufacturing*: CRC Press, 2025, pp. 352–359.
[61] C. Bhelkar, A. Tripathi, S. Mishra, L. Malviya, and S. Awachat, "Deep Learning-Based Pose Estimation and Real-Time Toddler Fall Detection System," in *International Conference on*

Computational Electronics for Wireless Communications (ICCWC), 2023: Springer, pp. 141–150.

[62] L. H. Leong, A. B. Hussain, M. A. Zulkifley, and W. M. D. W. Zaki, "Camera-Based Toddler Fall Detection System by using Kalman Filter," *Journal of Theoretical and Applied Information Technology,* vol. 81, no. 2, p. 383, 2015.

[63] H. Kakara *et al.*, "Development of childhood fall motion database and browser based on behavior measurements," *Accident Analysis & Prevention,* vol. 59, pp. 432–442, 2013.

[64] L. Hu, H. Meng, Z. Xu, and Y. Wang, "A self-powered wearable sensor for infant fall detection based on triboelectric nanogenerator," *Applied Physics A,* vol. 131, no. 3, p. 164, 2025.

[65] S. D. Wipulasundara, "A novel concept for remotely monitoring babies," in *2017 Moratuwa Engineering Research Conference (MERCon)*, 2017: IEEE, pp. 299–303.

[66] N. L. Pratap, K. Anuroop, P. N. Devi, A. Sandeep, and S. Nalajala, "IoT based smart cradle for baby monitoring system," in *2021 6th International Conference on Inventive Computation Technologies (ICICT)*, 2021: IEEE, pp. 1298–1303.

[67] W. A. Jabbar, H. K. Shang, S. N. Hamid, A. A. Almohammedi, R. M. Ramli, and M. A. Ali, "IoT-BBMS: Internet of Things-based baby monitoring system for smart cradle," *IEEE Access,* vol. 7, pp. 93791–93805, 2019.

[68] T. Hussain, K. Muhammad, S. Khan, A. Ullah, M. Y. Lee, and S. W. Baik, "Intelligent baby behavior monitoring using embedded vision in IoT for smart healthcare centers," *Journal of Artificial Intelligence and Systems,* vol. 1, no. 1, pp. 110–124, 2019.

[69] V. P. Hotur, P. Abhishek, R. Chethan, and A. Bhatta, "Internet of things-based baby monitoring system for smart cradle," in *2021 International Conference on Design Innovations for 3Cs Compute Communicate Control (ICDI3C)*, 2021: IEEE, pp. 265–270.

[70] D. B. D. Amos, J. Jablonski, and F. Heisler, "Python Basics: A Practical Introduction to Python 3," 4th Edition ed.: Real Python, 2020, p. 98. [Online]. Available: https://static.realpython.com/python-basics-sample-chapters.pdf

[71] V. K. Sharma, "Python Programming: A Practical Approach," First ed. ResearchGate: Taylor & Francis, 2022. [Online]. Available: file:///C:/Users/hcb1500/Downloads/Python_Programming_A_Practical_Approach.pdf

[72] B. Heinold, *A Practical Introduction to Python Programming*, 2012. [Online]. Available: https://www.brianheinold.net/python/A_Practical_Introduction_to_Python_Programming_Heinold.pdf. Accessed on: 25/02/2026.

[73] C. Nielsen, "Practical Python Programming." [Online]. Available: https://www.kufunda.net/publicdocs/Practical%20Python%20Programming.pdf

[74] "Introduction to Jupyter Notebooks." National Initiatives for Open Science in Europe. https://training.ni4os.eu/pluginfile.php/4242/mod_resource/content/1/Jupyter%20notebook%20NI4OS.pdf (accessed 24/03/2026.

[75] C. Zheng *et al.*, "Deep learning-based human pose estimation: A survey," *ACM Computing Surveys,* vol. 56, no. 1, pp. 1–37, 2023.

[76] E. Samkari, M. Arif, M. Alghamdi, and M. A. Al Ghamdi, "Human pose estimation using deep learning: a systematic literature review," *Machine Learning and Knowledge Extraction,* vol. 5, no. 4, pp. 1612–1659, 2023.

[77] S. Dubey and M. Dixit, "A comprehensive survey on human pose estimation approaches," *Multimedia Systems,* vol. 29, no. 1, pp. 167–195, 2023.

[78] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "Openpose: Realtime multi-person 2d pose estimation using part affinity fields," *IEEE transactions on pattern analysis and machine intelligence,* vol. 43, no. 1, pp. 172–186, 2019.

[79] A. Onodera, R. Ishioka, Y. Nishiyama, and K. Sezaki, "Multilabel Classification Model for Infant Activity Recognition Using Single Inertial Sensor," *IEEE Pervasive Computing,* vol. 23, no. 4, pp. 20–29, 2024.

[80] K. Sindhu Meena and S. Suriya, "A survey on supervised and unsupervised learning techniques," in *International Conference on Artificial Intelligence, Smart Grid and Smart City Applications*, 2019: Springer, pp. 627–644.

[81] S. Usmani, A. Saboor, M. Haris, M. A. Khan, and H. Park, "Latest research trends in fall detection and prevention using machine learning: A systematic review," *Sensors,* vol. 21, no. 15, p. 5134, 2021.
[82] B. Cheng, Y. Liu, and Y. Jia, "Evaluation of students' performance during the academic period using the XG-Boost Classifier-Enhanced AEO hybrid model," *Expert Systems with Applications,* vol. 238, p. 122136, 2024.
[83] J.-W. Kim, J.-Y. Choi, E.-J. Ha, and J.-H. Choi, "Human pose estimation using mediapipe pose and optimization method based on a humanoid model," *Applied sciences,* vol. 13, no. 4, p. 2700, 2023.
[84] D. Rai, A. Kumar, and A. Baghel, "Pose detection using opencv and media pipe," in *2024 International Conference on Integrated Circuits, Communication, and Computing Systems (ICIC3S)*, 2024, vol. 1: IEEE, pp. 1–6.
[85] P. Nandi *et al.*, "Analysis of ML Algorithm for Geriatric Fall Detection Due to the Effects of Various User Characteristics," 2023.
[86] G. Premalatha and V. T. Bai, "Design and implementation of intelligent patient in-house monitoring system based on efficient XGBoost-CNN approach," *Cognitive Neurodynamics,* vol. 16, no. 5, pp. 1135–1149, 2022.
[87] A. Jahanjoo, M. N. Tahan, and M. J. Rashti, "Accurate fall detection using 3-axis accelerometer sensor and MLF algorithm," in *2017 3rd International Conference on Pattern Recognition and Image Analysis (IPRIA)*, 2017: IEEE, pp. 90–95.
[88] M. Hanzla, W. Wahid, and A. Jalal, "Robust Human Pose Estimation and Action Recognition over Multi-level Perceptron," in *2024 3rd International Conference on Emerging Trends in Electrical, Control, and Telecommunication Engineering (ETECTE)*, 2024: IEEE, pp. 1–6.
[89] L. Dann. "Unemployment rate hits 5.1%, biggest annual employment decline in 15 years." The New Zealand Herald. https://www.nzherald.co.nz/business/unemployment-rate-hits-51/BO7L3ECJKVHLVFWBJXUPS5XGUU/ (accessed 21/07/2025, 2025).
[90] "Monthly Economic Review." https://www.parliament.nz/mi/pb/library-research-papers/monthly-economic-review/monthly-economic-review-february-2025/#employment (accessed 21/07/2025, 2025).
[91] S. Marshal, S. A. Raj, T. J. Jebaseeli, and S. Niranjan, "An image-based fall detection system for the elderly using yolov5," in *2023 2nd International Conference on Automation, Computing and Renewable Systems (ICACRS)*, 2023: IEEE, pp. 493–498.
[92] Z. Zhu, T. Liu, G. Li, T. Li, and Y. Inoue, "Wearable Sensor Systems for Infants," *Sensors,* vol. 15, no. 2, pp. 3721–3749, 2015. [Online]. Available: https://www.mdpi.com/1424-8220/15/2/3721.
[93] Y. El Marhraoui, S. Bouilland, M. Boukallel, M. Anastassova, and M. Ammi, "CNN-based self-attention weight extraction for fall event prediction using balance test score," *Sensors,* vol. 23, no. 22, p. 9194, 2023.
[94] R. Madhu, U. Anbazhagu, R. S. Krishnan, K. Paulraj, E. E. Nithila, and R. Umesh, "IoT-Enabled Deep Learning Framework for Smart Infant Health Monitoring," in *2025 International Conference on Intelligent Computing and Control Systems (ICICCS)*, 2025: IEEE, pp. 420–427.
[95] A. Prathyanga, P. Shyaminda, P. Chamikara, S. Lakshan, S. Thelijjagoda, and D. Kasthurirathna, "Intelligent Daycare: Enhancing Child Safety with IoT and Machine Learning Innovations," in *2024 9th International Conference on Communication and Electronics Systems (ICCES)*, 2024: IEEE, pp. 530–538.
[96] C. S. Mulligan, S. Adams, D. Tzioumi, and J. Brown, "Injury from falls in infants under one year," *Journal of paediatrics and child health,* vol. 53, no. 8, pp. 754–760, 2017.
[97] S. Choi, S. Yun, and B. Ahn, "Implementation of automated baby monitoring: CCBeBe," *Sustainability,* vol. 12, no. 6, p. 2513, 2020.
[98] L. Alzubaidi *et al.*, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *Journal of big Data,* vol. 8, no. 1, p. 53, 2021.
[99] D. Bhatt *et al.*, "CNN variants for computer vision: History, architecture, application, challenges and future scope," *Electronics,* vol. 10, no. 20, p. 2470, 2021.

[100] J. Ayeni, "Convolutional neural network (CNN): the architecture and applications," *Applied Journal of Physical Science,* vol. 4, no. 4, pp. 42–50, 2022.
[101] A. A. Elngar *et al.*, "Image classification based on CNN: a survey," *Journal of Cybersecurity and Information Management,* vol. 6, no. 1, pp. 18–50, 2021.
[102] M. Xiao *et al.*, "Addressing overfitting problem in deep learning-based solutions for next generation data-driven networks," *Wireless Communications and Mobile Computing,* vol. 2021, no. 1, p. 8493795, 2021.
[103] A. P. Piotrowski and J. J. Napiorkowski, "A comparison of methods to avoid overfitting in neural networks training in the case of catchment runoff modelling," *Journal of Hydrology,* vol. 476, pp. 97–111, 2013.
[104] P. Thanapol, K. Lavangnananda, P. Bouvry, F. Pinel, and F. Leprévost, "Reducing overfitting and improving generalization in training convolutional neural network (CNN) under limited sample sizes in image recognition," in *2020-5th International Conference on Information Technology (InCIT)*, 2020: IEEE, pp. 300–305.
[105] R. O. Ogundokun, R. Maskeliunas, S. Misra, and R. Damaševičius, "Improved CNN based on batch normalization and adam optimizer," in *International Conference on Computational Science and Its Applications*, 2022: Springer, pp. 593–604.
[106] C. F. G. D. Santos and J. P. Papa, "Avoiding overfitting: A survey on regularization methods for convolutional neural networks," *ACM Computing Surveys (Csur),* vol. 54, no. 10s, pp. 1–25, 2022.
[107] D. Kibet, M. S. So, H. Kang, Y. Han, and J.-H. Shin, "Sudden Fall Detection of Human Body Using Transformer Model," *Sensors,* vol. 24, no. 24, p. 8051, 2024.
[108] L. Le, A. N. M. Kabir, C. Ji, S. Basodi, and Y. Pan, "Using transfer learning, SVM, and ensemble classification to classify baby cries based on their spectrogram images," in *2019 IEEE 16th International Conference on Mobile Ad Hoc and Sensor Systems Workshops (MASSW)*, 2019: IEEE, pp. 106–110.
[109] C. Aliferis and G. Simon, "Overfitting, underfitting and general model overconfidence and under-performance pitfalls and best practices in machine learning and AI," *Artificial intelligence and machine learning in health care and medical sciences: Best practices and pitfalls,* pp. 477–524, 2024.
[110] B. Roth, P. H. Luz de Araujo, Y. Xia, S. Kaltenbrunner, and C. Korab, "Specification overfitting in artificial intelligence," *Artificial Intelligence Review,* vol. 58, no. 2, p. 35, 2024.
[111] S. Pothuganti, "Review on over-fitting and under-fitting problems in Machine Learning and solutions," *Int. J. Adv. Res. Electr. Electron. Instrum. Eng,* vol. 7, no. 9, pp. 3692–3695, 2018.
[112] N. Khan, P. H. Chong, F. Mirza, and D. Z. Al-Hamid, "Baby Safety Monitoring: Improving the CNN Model Using Batch Normalization and SGD," in *2025 10th International Conference on Communication, Image and Signal Processing (CCISP)*, 2025: IEEE, pp. 40–45.
[113] "Standard Deviation." The University of Texas at San Antonio. https://mathresearch.utsa.edu/wiki/index.php?title=Standard_Deviation#:~:text=%7B%5Cdisplaystyle%20%5Csigma%20=%7B%5Csqrt%20%7B%7B%5Cfrac%20%7B1%7D%7BN%7D%7D%5Csum%20_%7Bi=1%7D%5E%7BN%7D(x_%7Bi%7D%2D%5Cmu%20)%5E%7B2%7D%7D%7D%2C%7B%5Ctext%7B%20where%20%7D%7D%5Cmu%20=%7B%5Cfrac%20%7B1%7D%7BN%7D%7D%5Csum%20_%7Bi=1%7D%5E%7BN%7Dx_%7Bi%7D.%7D (accessed 12/12/2025.
[114] V. Eskov, V. Eskov, Y. V. Vochmina, D. Gorbunov, and L. Ilyashenko, "Shannon entropy in the research on stationary regimes and the evolution of complexity," *Moscow university physics bulletin,* vol. 72, no. 3, pp. 309–317, 2017.
[115] J. E. Contreras-Reyes, "Jensen-autocorrelation function for weakly stationary processes and applications," *Physica D: Nonlinear Phenomena,* vol. 470, p. 134424, 2024.
[116] O. Keskes and R. Noumeir, "Vision-based fall detection using st-gcn," *IEEE Access,* vol. 9, pp. 28224–28236, 2021.
[117] C. Lai and L. Jiang, "An intelligent baby care system based on IoT and deep learning techniques," *Int J Electron Commun Eng,* vol. 12, pp. 81–5, 2018.

[118] J. Ranta, E. Ilén, K. Palmu, J. Salama, O. Roienko, and S. Vanhatalo, "An openly available wearable, a diaper cover, monitors infant's respiration and position during rest and sleep," *Acta Paediatrica,* vol. 110, no. 10, pp. 2766–2771, 2021.
[119] J. Xie, X. Long, R. A. Otte, and C. Shan, "Convolutional neural networks for audio-based continuous infant cry monitoring at home," *IEEE Sensors Journal,* vol. 21, no. 24, pp. 27710–27717, 2021.
[120] R. Jahangir, "CNN-SCNet: A CNN net-based deep learning framework for infant cry detection in household setting," *Engineering Reports,* vol. 6, no. 6, p. e12786, 2024.
[121] K. Cho, B. Van Merriënboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder-decoder approaches," *arXiv preprint arXiv:1409.1259,* 2014.
[122] V. G. Bazarevsky, I., "On-Device, Real-Time Body Pose Tracking with MediaPipe BlazePose, Google Search." [Online]. Available: https://research.google/blog/on-device-real-time-body-pose-tracking-with-mediapipe-blazepose/
[123] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *arXiv preprint arXiv:1412.3555,* 2014.
[124] "Gated Recurrent Unit Networks." GeeksforGeeks. https://www.geeksforgeeks.org/machine-learning/gated-recurrent-unit-networks/ (accessed 13/04/2026.
[125] M.-H. Guo *et al.*, "Attention mechanisms in computer vision: A survey," *Computational visual media,* vol. 8, no. 3, pp. 331–368, 2022.
[126] S. Xu, Y. Cheng, K. Gu, Y. Yang, S. Chang, and P. Zhou, "Jointly attentive spatial-temporal pooling networks for video-based person re-identification," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 4733–4742.
[127] J. Zhang, C. Wu, and Y. Wang, "Human fall detection based on body posture spatio-temporal evolution," *Sensors,* vol. 20, no. 3, p. 946, 2020.
[128] V. Mnih, N. Heess, A. Graves, and K. Kavukcuoglu, "Recurrent models of visual attention," *Advances in neural information processing systems,* vol. 27, 2014.
[129] M. Jaderberg, K. Simonyan, and A. Zisserman, "Spatial transformer networks," *Advances in neural information processing systems,* vol. 28, 2015.
[130] J. Hu, L. Shen, S. Albanie, G. Sun, and A. Vedaldi, "Gather-excite: Exploiting feature context in convolutional neural networks," *Advances in neural information processing systems,* vol. 31, 2018.
[131] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7794–7803.
[132] S. Mejía, I. C. Muñoz, F. A. Castaño, and A. M. Hernández, "Non-Invasive Contactless Tracking of Respiratory Rate and Heart Rate During Sleep," *Sensors,* vol. 26, no. 4, p. 1082, 2026. [Online]. Available: https://www.mdpi.com/1424-8220/26/4/1082.
[133] I. Sadek and B. Abdulrazak, "Contactless remote monitoring of sleep: evaluating the feasibility of an under-mattress sensor mat in a real-life deployment," *Health Systems,* vol. 12, no. 3, pp. 264–280, 2023.

List of Abbreviations

AI	Artificial Intelligence
ML	Machine Learning
NHS	National Health Service
IoT	Internet of Things
PE	Pose Estimation
HM	Health Monitoring
HR	High Risk
ADL	Activities of Daily Living
WHO	World Health Organization
DL	Deep Learning
RGB	Red Green Blue
LSTM	Long Short-Term Memory
GRU	Gated Recurrent Unit
HAR	Human Action Recognition
DNN	Deep Neural Networks
CNN	Convolutional Neural Networks
BMS	Baby Monitoring System
BHSMS	Baby Health & Safety Monitoring System
BFDS	Baby Fall Detection System
BFPS	Baby Fall Prediction System
TBI	Traumatic Brain Injury
SIDS	Sudden Infant Death Syndrome
ConvLSTM	Convolutional Long Short-Term Memory

MCU	Micro Controller Unit
DC	Direct Current
DHT11	Digital Humidity and Temperature
ESP32	Espressif32
MobileNetv2	Mobile Network version 2
VGA	Video Graphics Array
mAP	Mean Average Precision
CribHD	Crib-Hazardous Dataset
CAM	Camera
CRNN	Convolutional Recurrent Neural Network
VGG19	Visual Geometry Group – 19 layers
SpO2	Saturation of Peripheral Oxygen
GPS	Global Positioning System
GSM	Global System for Mobile Communication
UV	Ultra Voilet
OLED	Organic Light-Emitting Diode
MAX9814	Microphone Amplifier with Low-Noise
DeepSORT	Deep Simple Online and Realtime Tracking
KAPAO	Keypoints And Poses As Objects
GMA	General Movement Analysis
MINI-RGBD	Moving Infants in RGB-D
RVI-38	RGB Video sequences for Infant movement
ResNet	Residual Network
CCTV	Closed Circuit Television
3D	3-Dimensional
YOLO	You Only Look Once

HIC	Head Injury Criterion
KNN	K-Nearest Neighbour
MFCC	Mel-Frequency Cel Cepstrum
XGB	Extra Gradient Boost
MP3	Medial Player 3
WAV	Waveform
FT	Fourier Transform
ETRI	Electronics Telecommunications Research Institute
APDS9960	Digital RGB Light and Gesture Sensor
NEO-6M	GPS Module
UP-Fall	University of Pisa- fall Detection dataset
ROI	Region of Interest
CV	Computer Vision
TENG	Tribo-Electric Nano Generator
Logo TD	Logo Text Display
TP-LSTM	Temporal Pyramid - Long Short-Term Memory
EEL	Early Exponential Loss
UCF-101	University of Central Florida – 101Action Categories
2D	2-Dimensional
I3D	Inflated 3 Dimensional
FPR	False Positive Rate
FNR	False Negative Rate
RNN	Recurrent Neural Network
SVM	Support Vector Machine
MLP	Multi-Layer Perceptron
RF	Random Forest

IDE	Integrated Development Environment
GB	GigaBytes
RAM	Random Access Memory/Recurrent Attention Model
PE-ML	Pose Estimation-Machine Learning
USPA	Unsafe Scenario Prediction Algorithm
HPE	Human Pose Estimation
AR	Augmented Reality
VR	Virtual Reality
CP	Cerebral Palsy
DDH	Development Dysplasia of the Hip
GBDT	Gradient Boosted Decision Trees
CSV	Comma-Separated Values
ACC	Accuracy
PRC	Precision
REC	Recall
FFCV	Five-Fold Cross Validation
LOS	Line of Sight
BUSPS	Baby Unsafe Scenarios Prediction System
GELAN	Generalized Efficient Layer Aggregation Network
GHT	Generalized Hough Transform
ChildSUn	Child Safe Unsafe
SSA	Soft Spatial Attention
CIFAR-10	Canadian Institute For Advanced Research (10 classes)
SGD/S	Stochastic Gradient Descent
BCNN	Benchmark Convolutional Neural Network
BN/B	Batch Normalization

DO	Dropout
R	Regularization
GAP/G	Global Average Pooling
LR	Learning Rate
TCNN	Traditional Convolutional Neural Network
SSCNN	Soft Spatial Convolutional Neural Network
TP	True Postives
TN	True Negatives
FP	False Positives
FN	False Negatives
ent_{mean}	Mean of entropies
IMU	Inertial Measurement Unit
ST-GCN	Spatial-Temporal Graph Convolutional Network
Bi-LSTM	Bidirectional-Long Short-Term Memory
KFall	Large Scale Open Motion dataset
SisFall	Fall and Movement dataset
API	Application Interface
MPP	MediaPipe Pose
CPU	Central Processing Unit
KPs	Keypoints
STN	Spatial Transformer Networks
GE-Net	Gather Excite Network
SEQ_LEN	Sequence Length
TL	Transfer Learning
AIoT	Artificial Intelligence-based Internet of Things
GT	Ground Truth

Appendices:

A. Model Development Analysis

A.1. AI-2 Model – Unsafe Scenarios Prediction using Temporal Coefficients

A comprehensive process for predicting unsafe scenarios using statistical parameters, as outlined in Section 5.7.3, is presented here.

Unsafe Scenarios Prediction: This research uses ADL data to predict unsafe scenarios for a baby in the crib. The description of unsafe scenarios uses temporal coefficients derived from statistical parameters, such as autocorrelation, entropy, and standard deviation.

Algorithm S2: Baby Unsafe Scenarios Prediction Algorithm (BUSPA): Pseudocode of the prediction of unsafe scenarios using time-dependent coefficients

---

Input: probabilities: p; $\bar{p} = \frac{1}{N}\sum_{i=1}^{N} p_i$ N: total no. of observations in the time series,

i = individual data values

Output: Threshold: T, Result: R, as a result, predicting safe or unsafe for every 10-second video segment

Phase 1: Initialization

1. Load model
2. Initialize train_dir and test_video_path.

Phase 2: Compute and collect statistics from probabilities based on datasets

3. Compute the mean of the autocorrelation:

$$autocorr_mean = \frac{1}{N}\sum_{i=1}^{N-1} p_i . p_{i+1}$$

4. Compute Entropy for each probability p_i and the mean of all entropies:

$$entropy = -\sum_{i=1}^{N} p_i \log_2 p_i$$

$$ent_{mean} = \frac{1}{N}\sum_{i=1}^{N} entropy(pi)$$

5. Compute the mean of the standard deviation (σ_p):

$$std_mean = \sqrt{(\frac{1}{N}\sum_{i=1}^{N}(p_i - \bar{p})^2)}$$

6. Append “label”: 0 if label == safe, else 1
7. Append autocorr_mean, ent_mean, std_mean

Phase 3: Compute thresholds

8. Input data frame containing calculated statistics and binary class labels
9. Filter into two subsets:

Safe class (label =0), Unsafe class (label = 1)

10. Calculate the mean value for each statistic for both classes.
11. Calculate thresholds:

$T_{std}= \frac{Mean_{std,safe}+ Mean_{std,unsafe}}{2}$

$T_{entropy}=\frac{Mean_{entropy,safe}+ Mean_{entropy,unsafe}}{2}$

$T_{autocorr}=\frac{Mean_{autocorr,safe}+ Mean_{autocorr,unsafe}}{2}$

12. Return thresholds, T.

Phase 4: Final output Prediction

13. Call the predict_video_segment function ()
14. Return R as a result, predicting safe or unsafe for every 10-second video segment

A.2 Predict Video Segments using Model Probabilities

To predict the safe and unsafe labels, as shown in Table 5.7 outlined in Section 5.7.3 in Chapter 5, from the 10-second video segments using the temporal coefficients calculated through the algorithm presented in Section A.1 and the model probabilities.

Predict Video Segments Using Model Probabilities (P) Algorithm: Pseudocode (Table 5.7)

Input: Input Video: V, frames: f

Pretrained model, T = { T_{std}, $T_{entropy}$, $T_{autocorr}$} the thresholds,

Segment duration (in seconds): *S,* Standard deviation σ, Entropy H, Autocorrelation A

Output: Results: R as Safe or Unsafe label for video segments

Phase 1: Initialization

1. Open V; obtain FPS *f.*
2. Calculate frames per segment: Fs = $f\ x\ S$
3. Set segment index k = 1.

Phase 2: Data Preprocessing

For each frame *I in V:*

4. Preprocess frame: resize to input dimension, normalize

Phase 3: Extraction of Temporal Features

5. Predict unsafe probability: pi = M ($frame_i$)
6. Collect p_i into a segment, set P_k.
7. If i mod Fs = 0:
   - Compute statistics: st_k= ($\sigma_{P_k}, H_{P_k}, A_{P_k}$)
   - Unsafe if ($\sigma_{P_k} > T_{std}$) OR ($H_{P_k} > T_{entropy}$) OR ($A_{P_k} > T_{autocorr}$); else Safe
8. Increment k; reset P_{k+1}
9. Repeat steps 4 to 8 for any remaining P in the leftover frames.

Phase 4: Final Output Prediction

10. Return R as a table of Safe or Unsafe labels for video segments.

B. Results and Evaluation

B.1. AI-2 Model: Attention Map Visualization over Test Datasets

This study presents test results aimed at visualizing the attention map, including safe and unsafe predictions, as illustrated below in the figures of visualizing the attention map (a) Original image, (b) Attention map, (c) Overlaid attention and prediction outcome for safe, (d) Original image, (e) Attention map, (f) Overlaid attention and prediction outcome for unsafe.

