SPEECH ANALYSIS AS A DECISION SUPPORT SYSTEM IN HEALTHCARE FOR DETECTING MILD TRAUMATIC BRAIN INJURY

A THESIS SUBMITTED TO AUCKLAND UNIVERSITY OF TECHNOLOGY
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF COMPUTER AND INFORMATION SCIENCES

Supervisor

Dr. Samaneh Madanian

2024

By

Minh Hoang Quach

School of Engineering, Computer and Mathematical Science

Abstract

Mild Traumatic Brain Injury (mTBI), also known as concussion, is a prevalent neurological condition with significant public health consequences. Accurate and timely diagnosis of mTBI is crucial for effective management and prevention of long-term complications. However, current diagnostic methods, including clinical assessments and neuroimaging techniques, have limitations in terms of subjectivity, sensitivity, and accessibility. This thesis explores the potential of speech analysis as a non-invasive and accessible tool for mTBI diagnosis. The feature selection process highlighted the importance of specific speech tasks and acoustic features in concussion detection, particularly those related to timing variability, prosody, pitch variation, intensity deviation, and speech motor control. The findings of this research suggest that subtle changes in speech patterns, often undetectable to human ear, can be detected and analysed using Machine Learning algorithms to aid in concussion diagnosis. By analysing selected acoustic features, two Machine Learning models, Support Vector Machine (SVM) and Multilayer Perceptron (MLP), were developed to differentiate between concussed and non-concussed individuals. Although the MLP model slightly outperformed the SVM model in terms of accuracy (81.59% vs. 80.75%), the SVM model's marginally higher AUC-ROC value (89.77% vs. 88.45%) suggests a potentially better overall performance due to its ability to better distinguish between the two classes across various classification thresholds. This study represents a significant step towards developing a more accurate, efficient, and accessible diagnostic tool for mTBI, with the potential to improve patient outcomes and reduce the long-term consequences of this injury.

Contents

List of Tables

List of Figures

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 nor material which to a substantial extent has been accepted for the qualification of any other degree or diploma of a university or other institution of higher learning.

_____

Signature of student

13 June 2024

Acknowledgments

First and foremost, I want to express my sincere gratitude to Auckland University of Technology for providing me with a great graduate experience. I am deeply appreciative of the support and guidance from my professors and classmates, who have been helpful in my academic growth and development. Their unwavering encouragement and invaluable feedback have empowered me to evolve from a novice in academic writing to the stage of understanding how academic writing should be expected.

Secondly, I would like to express my appreciation to my supervisor, Dr. Samaneh Madanian, for her exceptional mentorship throughout my research journey. Her knowledge and support have been invaluable. Dr. Madanian's dedication to fostering both academic and practical skills in her students is truly commendable. She consistently provided clear direction, insightful feedback, and encouragement, enabling me to navigate the complexities of research and thesis writing with confidence. Her significant investment of time and effort in my education and research is deeply appreciated. Additionally, I would like to express my sincere gratitude to the US team, including Professor Christian Poellabauer (Florida International University), Dr John Michael Templeton (University of Southern Florida), and Professor Sandra Schneider (Saint Mary's College), for their provision of dataset that made this research possible.

Finally, I would like to express my gratitude to the specialists and academics who examined this work for their time and effort.

Chapter 1 - Introduction

This chapter provides the background and context for this research. Start by discussing the prevalence and impact of TBI, mTBI and the existing diagnostic methods. Highlight the limitations of current approaches and the need for more accurate and accessible diagnostic tools. This chapter also states research objectives and questions, emphasizing the benefits of this research, and provides the thesis structure.

1.1 Background and Motivation

Traumatic Brain Injury (TBI) is a consequence of head trauma, encompassing incidents such as impacts, bumps, or the penetration of foreign objects through the skull, reaching the brain (National Institute of Neurological & Stroke, n.d.). Based on National Institute of Neurological Disorders and Stroke (NINDS), the classification of TBIs distinguishes between closed injuries, where skull fractures are absent, and open injuries involving a breach in the skull, Notable, approximately 80% to 90% of all TBIs fall under the category of concussion, commonly referred to as mild Traumatic Brain Injury (mTBI) (Skandsen et al., 2019). These concussions typically result from forceful head impacts and manifest varying symptoms in terms of severity (Agarwal et al., n.d.)

In the short term, mTBI can lead to immediate cognitive deficits and memory impairments (Mouzon et al., 2012). These immediate effects can be exacerbated by factors such as hyperthermia, which has been shown to result in long-term cognitive deficits (Titus et al., 2015).

In the long term, mTBI has been associated with chronic neuroinflammation,

changes in synaptic plasticity, and cognitive deficits (Aungst et al., 2014). These deficits can manifest as persistent problems with memory, attention, and executive function, impacting daily activities, work or school performance, and overall quality of life. A study by Hawley et al. (2004) found that children who had experienced brain injuries, including concussions, often faced challenges returning to school life. These children reported difficulties with schoolwork and experienced attention or memory problems. Along with the cognitive impairment, mTBI symptoms often involve headaches, disrupted sleep, dizziness, cognitive slowing, poor concentration, memory issues, anxiety, and depression (Heegaard & Biros, 2007; Kelly & Rosenberg, 1998). These symptoms can manifest as slower reaction times, reduced energy, balance difficulties, challenges with multitasking, interpersonal issues, and noticeable personality changes. Research indicates that cognitive impairments following a single mTBI can result in detectable long-term deficits, including slower processing speed and memory problems. This supports the growing theory that repeated mTBIs may cause cumulative brain damage, leading to cognitive issues that could worsen over time and affect memory and learning (Nakadate et al., 2016). Moreover, mTBI has been linked to an increased risk of subsequent mishaps, indicating potential long-term implications for occupational health and safety (Whitehead et al., 2014).

As highlighted by Lubbers et al. (2024), mTBI is a prevalent neurological condition encountered frequently in emergency departments, with an estimated annual incidence of 100-300 cases per 100,000 individuals globally. Furthermore, based on Accident Compensation Corporation (n.d.), every year, around 35000 individuals in New Zealand experience TBI. Among these cases, mTBI accounts for 95%, equivalent to 33,250 cases. However, only 22000 claims are submitted, indicating that some individuals do not seek medical evaluation for their injuries. This discrepancy underscores the importance of raising awareness about the

significance of seeking medical attention following a TBI, especially in cases of mTBI.

1.1.1 Overview of current clinical methods

Diagnosing mTBI requires prompt assessment by a qualified professional (National Academies Press, 2019), encompassing a comprehensive neurological examination covering various aspects like motor skills (Young et al., 2023), sensory functions (Li et al., 2020), speech, balance (Alashram et al., 2020), mental state, mood (Howlett et al., 2022), and cognitive abilities (Sun et al., 2017). Based on NINDS, for mTBI, standardized tools like the Centers for Disease Control and Prevention's Acute Concussion Evaluation (CDC-ACE) form or Sport Concussion Assessment Tool (SCAT) are often used, evaluating symptoms including amnesia, seizures, and physical, cognitive, emotional, and sleep-related aspects. NINDS also stated that medical providers employ Computerized Tomography (CT) scans for moderate to severe TBIs and Magnetic Resonance Imaging (MRI) for subtle brain changes, neuropsychological tests assess cognitive functions, and the Glasgow Coma Scale measures consciousness levels. In sports, NINDS informed that establishing baseline brain function tests before the season and repeating them every one to two years, or after suspected concussions, helps evaluate readiness to return to normal activities (National Institute of Neurological & Stroke, n.d.).

Limitations of current clinical methods

Nevertheless, clinical diagnostic methods for mTBI come with several limitations that healthcare providers should consider. Clinical assessments, including neurological examinations and neuropsychological tests, often rely on subjective patient report, which can probably lead to an incomplete understanding of the full spectrum of symptoms (National Academies Press, 2019). Furthermore, imaging techniques like CT scan and MRI exhibit varying sensitivities (Bigler, 2023); CT

scans are excellent for identifying structural damage, but may miss subtle injuries (Bigler, 2023), while MRIs, which are more sensitive, might not be universally available or affordable in all healthcare settings.

Advanced imaging techniques like MRI, while valuable for diagnosing brain injuries, are often costly. For example, ACC review revealing that New Zealand pays significantly more for scans like MRIs compared to Australia and the United Kingdom and some other countries (Newshub, 2022). This high cost, coupled with limited availability in certain regions, can create barriers to access, especially in rural areas or for individuals with limited financial resources. For instance, despite acquiring an additional MRI machine in 2018, the Counties Manukau District Health Board still faced significant backlogs and wait times for scans, with only 23% of patients receiving them within six weeks (Bhullar et al., 2021). This delay affects the “golden hour”. The "golden hour" - that critical first 60 minutes after a severe injury (Maurya et al., 2022) - highlights a key principle for even milder concussions: acting fast makes a difference. Immediate evaluation, education, and preventative measures can drastically alter the course of recovery. Even a "mild" brain injury can worsen over time if left untreated, leading to lasting cognitive problems. Additionally, the absence of dependable biomarkers capable of rapidly and unobtrusively identifying concussion indicators (Toth et al., 2020).

1.1.2 The Role of Technology

Transitioning to non-invasive diagnostic methods, particularly those involving speech analysis, holds significant promise in addressing the limitations associated with TBI diagnosis. As Poellabauer et al. (2015) stated, in the past few decades, an abundance of research findings has supplied proof that speech production and the analysis of speech signals can reveal traces of neurological disorders. The core concept behind utilizing speech as a biomarker lies in the fact that brain injuries

frequently reveal their presence by impacting the synchronization and timing of the speech motor system, which subsequently becomes apparent in speech patterns such as altered vowels (Ferreres et al., 2003), increased nasal resonance (Thompson & Murdoch, 1995), less precise consonants (Marchman et al., 1991). Despite the commonality of physical and cognitive symptoms seen in both concussion and more severe head injuries, there has been limited investigation into how concussion affects speech production.

Changes in speech are generally not incorporated into widely used symptom assessments for milder injuries or sports-related sideline evaluations (Asken et al., 2020; Schatz et al., 2006). Nevertheless, recent initial evidence has unveiled significant alterations in speech rate (Salvatore et al., 2019), acoustic characteristics (Daudet et al., 2017), articulatory precision (Chong et al., 2021), and fluency (Toldi & Jones, 2021) in cases of concussion. These initial findings strongly advocate for further exploration to establish a specific pattern of speech changes linked to concussion. While the intricacies of speech analysis, especially in real-world scenarios with varying audio signal quality and accents, pose challenges, there is promising potential in leveraging Artificial Intelligence (AI) techniques, including Machine Learning (ML) and Deep Learning (DL), to overcome these complexities. AI can play a pivotal role in enhancing the robustness of speech analysis, offering a valuable solution for diagnosing concussions and monitoring their effects.

1.2 Research objective and Research questions

This research focuses on the role of AI in diagnosing concussions. By looking closely at specific ways that speech changes after a mild head injury, this work aims to explore speech features that can be used for concussion. This way could revolutionize how we assess and treat this common injury, making diagnosis more

accessible, efficient, and less costly. Furthermore, this research compares different speech analysis techniques, evaluating their accuracy for detecting mTBI. This comparative analysis will provide valuable insights for healthcare professionals, specifically in the context of concussion diagnosis.

Research questions:

- RQ1: What specific speech biomarkers should be considered when developing speech-based diagnostic tools to support clinical decision-making?
- RQ2: How do AI methods of Support Vector Machine(SVM), and Multilayer Perceptrons (MLP) detect the concussion patient?

1.3 Significance of research

This research will benefit health professionals and patients who may suffer from mTBI.

For patients: Earlier and more accurate diagnoses lead to faster interventions and treatments specifically tailored to each person's needs. This means potentially reducing the severity of symptoms, lowering the risk of long-term problems, and getting back to normal life sooner.

For health professionals: Speech analysis could offer a valuable addition to their toolkit. Finding specific speech markers that change after a concussion could mean earlier and more accurate diagnosis, which is crucial for getting the right treatment as quickly as possible. With objective, reliable data from speech analysis, doctors and specialists could make better decisions about each patient's care, tailoring treatment plans for the best possible outcomes. Furthermore, it is about accessibility. Speech analysis tools are far less expensive than traditional methods like MRI scans, and they do not require fancy equipment. Imagine being able to quickly assess someone for a concussion right on the sports field or in a remote

area - that is the kind of access this technology could provide. It also means less reliance on self-reported symptoms, which are not always reliable.

1.4 Thesis Structure

To achieve mentioned objective, the thesis structure comprises six chapters, each contributing significantly to the research objectives:

Chapter 1 - Introduction: This chapter provides the background and motivation for this research. It begins by introducing the prevalence and impact of mTBI and the current diagnostic methods. It highlights the limitations of these approaches and the need for more accurate and accessible diagnostic tools. This chapter also states the research objectives and questions, emphasizing the benefits of this research, and provides an overview of the thesis structure.

Chapter 2 – Background of the Study: The second chapter discusses the concepts and theories utilized in this research. It begins with an overview of current concussion diagnostic approaches. Then, it presents speech analysis as a promising diagnostic tool, supported by references that validate its efficacy in various healthcare contexts. This section examines the speech analysis pipeline, exploring the different stages involved and the potential techniques utilized in each stage: Data Preprocessing, Feature Selection, Model Training, and Model Evaluation.

Chapter 3 - Methodology: This chapter shows the research design and methodology used in this study. It outlines the processes involved in data collection, preprocessing, feature extraction, and feature selection for subsequent modelling. The chapter provides an in-depth explanation of the data analysis techniques, with a particular focus on justifying the chosen methods and their application to identify and extract remarkable features from speech data for training ML and DL models for concussion diagnosis.

Chapter 4 – Model Implementation: This chapter presents the algorithm implementation of SVM and MLP models. It describes the training process, and the hyperparameter tuning methods used to optimize the models for concussion detection.

Chapter 5 - Discussion: This chapter presents the findings of the research, obtained through the application of the techniques outlined in the Methodology and Model Implementation chapters. It details the specific speech biomarkers that demonstrate the strongest potential for inclusion in concussion diagnostic tools. Moreover, the chapter analyses how AI techniques contribute to enhancing the accuracy, robustness, and real-world applicability of speech-based concussion diagnosis tools. Acknowledging potential study limitations, this chapter discusses how they might influence the interpretation and generalizability of the results.

Chapter 6 - Conclusion: The final chapter marks the conclusion of the thesis. The primary findings are summarized here, and the research's unique contribution to the healthcare domain is underscored, particularly in the context of enhancing concussion diagnosis and aiding clinical decision-making. The chapter concludes with reflective insights derived from the study, reaffirming the pivotal role of speech analysis and AI in reshaping the landscape of healthcare diagnostics. Finally, the chapter outlines how these findings can be improved in future research, laying a foundation for the development of practical and effective speech-based diagnostic tools in healthcare.

Chapter 2 - Background of the Study

This chapter reviews current concussion diagnostic approaches, analysing their strengths and limitations. It also explores the emerging field of speech analysis in healthcare. By examining various techniques used in each stage of the speech analysis pipeline, this chapter aims to provide the foundation for understanding the potential application of speech analysis in concussion diagnosis.

2.1 Current diagnostic methods

Reviewing the current diagnostic methods employed in the assessment and diagnosis of TBI is an important first step. Understanding these methods is essential to contextualize the potential role of speech analysis as a clinical decision support tool. This exploration encompasses clinical assessment techniques as well as neuroimaging methods, highlighting both their strengths and limitations.

2.1.1 Clinical assessment techniques

Clinical assessments play a critical role in the initial evaluation of mTBI, serving as the cornerstone for diagnosis and treatment planning. These assessments involve a multi-faceted approach, combining various tools and procedures to gauge the extent of neurological impact and inform subsequent interventions. One of the most widely utilized tools in concussion assessment is the Glasgow Coma Scale (GCS), a standardized 15-point scale designed to evaluate a patient's level of consciousness by assessing their eye-opening response, verbal response, and motor response (Jain & Iverson, 2018). By assigning a numerical score to each of these responses, the GCS provides a quick and objective measure of the severity of brain injury, helping clinicians prioritize care and determine the need for further

evaluation.

In addition to the GCS, healthcare providers often employ standardized questionnaires like the Sport Concussion Assessment Tool 5th Edition (SCAT5) and the Post-Concussion Symptom Scale (PCSS) (Danielli et al., 2023). These questionnaires provide a structured framework for collecting information about a wide range of concussion-related symptoms, including headache, dizziness, fatigue, cognitive difficulties, and emotional changes. By systematically assessing and tracking these symptoms over time, clinicians can gain valuable insights into the trajectory of recovery and tailor treatment plans to the individual's specific needs.

Beyond the aforementioned tools, additional assessments like the Standardized Assessment of Concussion (SAC) and the Balance Error Scoring System (BESS) are frequently employed (Bell et al., 2011; McCrea et al., 1998). The SAC offers a more in-depth evaluation of cognitive function, specifically targeting memory, attention, and processing speed. The BESS, on the other hand, focuses on assessing balance and coordination, which are often affected by concussions.

Clinical assessments are the first line of defence in diagnosing mTBI, offering accessibility and ease of use. However, their reliance on subjective patient report and clinician interpretation can lead to inconsistencies and potential misdiagnosis (Table 1), as highlighted by research on the complex relationship between clinical impairments and self-reported limitations (Mactaggart et al., 2016). While valuable for assessing overt symptoms, these evaluations may miss subtle neurological changes and underlying damage (Shenton et al., 2012). Additionally, variability in healthcare provider expertise can lead to discrepancies in mTBI diagnosis and delayed treatment, as even experienced physicians may lack confidence in recognizing and managing this complex condition (Theadom et al.,

2021).

2.1.2 Neuroimaging Methods

Neuroimaging techniques give a detailed view inside the brain, helping doctors identify structural or functional changes caused by a concussion. CT scans use X-rays to create cross-sectional images of the brain, providing detailed information about bones, blood vessels, and tissues. This makes them particularly useful for quickly detecting bleeding or skull fractures after a head injury (Mayo Clinic, 2022). While, MRI scans use magnetic fields and radio waves to create even more detailed images of soft tissues and brain structures (Mayo Clinic, 2023). This makes them a powerful tool for diagnosing subtle brain abnormalities that might not be visible on a CT scan.

Besides static images, techniques like functional MRI (fMRI) allow us to see the brain in action, measuring changes in blood flow to reveal which parts of the brain are most active during specific tasks (Radiological Society of North & American College of, 2022). Diffusion Tensor Imaging (DTI), another specialized MRI technique, helps us understand how different brain regions are connected by tracking the movement of water molecules (Le Bihan et al., 2001). Even when the brain is at rest, resting-state fMRI can show how different brain networks communicate with each other, revealing insights into overall brain function.

Neuroimaging spectroscopy, like Magnetic Resonance Spectroscopy (MRS), goes even deeper, measuring the concentrations of chemicals in the brain that can signal cell damage or dysfunction (Rajvanshi et al., 2021). This can help us understand the metabolic changes that occur after a concussion and potentially track the healing process.

MRI and CT scans are, in general, non-invasive and provide objective, visual representations of the brain's structure and function. They offer quantitative data

that can be standardized and compared across individuals and time, making them valuable tools for diagnosing and understanding concussions. However, they still have limitations (Table 1). MRI scans, while generally safer than CT scans due to the absence of radiation, are costly and often unavailable outside major medical centers, posing access challenges in rural New Zealand (Jacobs & Henwood, 2013). They also require longer scan times and strict safety protocols, making them less ideal for urgent situations. CT scans, though more accessible, expose patients to radiation and are less sensitive to subtle injuries. Even the most advanced techniques have limitations: fMRI might miss rapid brain changes, and DTI is sensitive to movement during scanning (Battaglia, 2003; Li et al., 2013).

Table 1: Strengths and Limitations of Current Diagnostic Methods

Method	**Strengths**	**Limitations**
Clinical assessments	Accessible, easy to use, provide initial evaluation	Subjective, may miss subtle changes, variability in expertise (Mactaggart et al., 2016; Shenton et al., 2012; Theadom et al., 2021)
CT Scans	Quick, effective for detecting bleeding or fractures	Radiation exposure (Mayo Clinic, 2022), less sensitive to subtle injuries

Table 1: Strengths and Limitations of Current Diagnostic Methods (cont.)

Method	**Strengths**	**Limitations**
MRI Scans	Detailed images, detect subtle abnormalities	Expensive, less accessible, longer scan times (Jacobs & Henwood, 2013)
fMRI	Reveals active brain regions during tasks	May miss rapid changes, sensitive to movement (Battaglia, 2003)
DTI	Understands brain region connectivity	Sensitive to movement, complex analysis (Le Bihan et al., 2001)
MRS	Measures chemical concentrations, signals damage	High cost, requires specialized equipment, less accessible (Rajvanshi et al., 2021)

2.1.2 The Potential of AI

The limitations of both clinical assessments and neuroimaging techniques highlight the need for a more reliable, affordable and non-invasive approach. Combining patient-centred insights from clinical assessments with the objective data from neuroimaging can create a more complete picture of the injury's effects. This approach, strengthened by the power of AI, holds the potential to transform concussion care. AI can identify subtle patterns and biomarkers that might be not detected by traditional methods, potentially leading to more accurate diagnoses and personalized treatment plans.

2.2 AI-enabled Clinical Decision Support System

AI is revolutionizing healthcare, demonstrating exceptional ability in analysing complex patterns within large datasets. This technology surpasses traditional methods in various applications, from analysing medical images to predicting disease outbreaks (Fogel & Kvedar, 2018). AI has already proven its worth in detecting diseases like diabetic retinopathy and skin cancer with remarkable accuracy (Gulshan et al., 2016). It is also making progress towards personalized medicine, helping doctors create treatment plans tailored to each individual patient, which can lead to better outcomes and lower costs (Ahmed & Adil, 2023).

The real power of AI lies in its ability to process massive amounts of data that would take humans years to process (Hossain et al., 2022). This encourages exciting new research in areas such as understanding disease patterns and predicting the spread of epidemics. During the COVID-19 pandemic, researchers utilized AI models to forecast the virus's spread, predict peak infection times, and assess the effectiveness of lockdown measures in different regions of China (Feng et al., 2021). These examples illustrate AI's potential in enhancing healthcare delivery and public health management.

Building on the foundation of these AI advancements, clinical decision support systems (CDSS) are emerging to empower healthcare professionals with more accurate and personalized patient care. These tools act as assistants, providing doctors and other clinicians with tailored recommendations and assessments based on the unique characteristics of each patient (Kawamoto et al., 2005). By making the most of a wealth of knowledge - from the patient's medical history to the latest research – AI-enabled CDSS can help clinicians make informed decisions and improve how healthcare is delivered (Sutton et al., 2020). These systems are becoming more and more common in hospitals and clinics, promising a future where AI works together with healthcare providers to improve both efficiency and patient outcomes. Gaube et al. (2023) found that AI-generated

advice with visual annotations improved doctors' diagnostic accuracy on X-rays, especially for junior doctors. Notably, the study also found that physicians rated AI advice higher than human advice, but confidence was not significantly impacted. This highlights the potential of AI to enhance clinical decision-making when designed to be explainable and user-friendly.

Building on the advancements of AI in healthcare and CDSS, another promising area is speech analysis. This technique leverages the natural characteristics of speech to provide insights into an individual's health status, particularly in the context of brain injuries such as concussions.

2.3 Speech Analysis in Healthcare

Speech is a fundamental and versatile mean that holds significant value in various domains, including healthcare, communication, and technology (Jiang et al., 2006). Speech is non-invasive, easily collectible, transmittable, and storable, making it a convenient and assessable source of information (McCullough, 2001). Because speech production is linked to our nervous system, mental state, and individual traits, it can reveal biological, emotional, and even pathological states (Ali et al., 2015). This makes changes in speech patterns valuable markers for identifying and tracking brain injuries, making it a promising way for concussion assessment.

Speech analysis is a multifaceted field that involves examining various characteristics of speech signals to gain insights into a speaker's physical and mental state. Acoustic analysis focuses on properties such as pitch, rhythm, and loudness. Linguistic analysis examines word choice, grammar, and syntax, while content analysis looks at the meaning and themes conveyed through language.

By examining changes in articulation, fluency, and prosody, speech can provide a non-invasive window into subtle neurological deficits, potentially

revealing symptoms missed by traditional methods (Chong et al., 2021; O'Brien, 2020; Taylor et al., 2020). The development of portable, AI-powered tools based on temporal and frequency analysis holds promise for improving concussion diagnosis in various settings (Daudet et al., 2017). AI's ability to analyse complex speech patterns could transform detection by identifying subtle biomarkers imperceptible to the human ear (Wall et al., 2022). Additionally, speech analysis can aid in understanding and addressing persistent post-concussion symptoms, potentially leading to more targeted therapies (Chong et al., 2021; Hoover et al., 2017).

2.4 Speech Analysis for Concussion

Speech analysis offers a non-invasive window into potential neurological deficits caused by mTBI, as even subtle changes in brain function can impact speech production and language processing. These impacts can manifest in various ways, ranging from observable alterations in articulation, fluency, and word choice to more nuanced shifts in prosody, such as rhythm and intonation (Chong et al., 2021; O'Brien, 2020). Research has revealed specific speech characteristics that may be indicative of mTBI. For instance, studies have found that individuals with mTBI often exhibit slower speech rates, altered pause patterns, and changes in pitch variability compared to healthy controls (Daudet et al., 2017; Salvatore et al., 2019). The work of Daudet et al. (2017) has led to the development of portable mTBI assessment tools, including smartphone-based applications that leverage temporal and frequency analysis to detect these changes. Such tools have the potential to change concussion diagnosis by making it more accessible in settings like sports fields or remote areas, where immediate medical evaluation may be challenging. Falcone et al. (2013)'s study demonstrated the feasibility of using mobile devices to record speech samples from athletes after boxing matches, isolating vowel sounds, and extracting acoustic features. By training ML

algorithms on these features, they achieved a remarkable 98% accuracy in detecting concussions. Wall et al. (2022) employed a DL model (Bi-LSTM-A) to identify specific speech biomarkers, such as pitch period entropy, articulation rate, and silent/filled pause rate, that could differentiate between concussed and non-concussed individuals.

2.5 Speech Biomarkers

Speech biomarkers can be categorized into linguistic and acoustic features, each reflecting distinct aspects of speech production and carrying specific information (Table 2).

Table 2: Types of Speech Biomarkers

Feature type	**Examples**	**Description**
Linguistic	Lexical richness, sentence complexity	Reveal cognitive shifts, early signs of neurological changes
Acoustic	Pitch, rhythm, intensity, fundamental frequency, jitter, shimmer, Harmonics-to-Noise Ratio (HNR), formant frequencies	Indicates health of vocal cords, neurological changes

2.5.1 Linguistic Features

Linguistic features focus on the meaning and structure of speech, offering a rich source of information for detecting subtle neurological changes. For example, lexical richness, which refers to the variability and complexity of vocabulary used, can indicate cognitive function. A richer vocabulary might suggest better cognitive health. Similarly, sentence complexity, which involves the use of complex

grammatical constructions and varied sentence structures, reflects cognitive processing abilities. More complex sentences suggest healthier cognitive function. Studies have shown that individuals with early cognitive impairments often exhibit reduced lexical diversity and simpler sentence structures (Vincze et al., 2016).

Although these features are significant in the broader context of speech analysis, they fall outside the primary scope of this thesis. In this research, our interest is purely on signal features ignoring the speech content and spoken language. This thesis analyzes the speech signal, with an emphasis on acoustic features, as they are directly measurable and relevant to understanding the physiological and neurological aspects of mTBI.

2.5.2 Acoustic Features

Acoustic features, derived from the physical properties of speech signals, provide insights into various aspects of speech production. They capture information related to the timing, frequency, and overall quality of the voice (Ekberg et al., 2023). These features mirror aspects of human hearing and offer insights into how the vocal tract produces speech (Daudet et al., 2017). The acoustic parameters can be classified into four types (Ekberg et al., 2023) (Table 3):

1. Frequency features include parameters such as the fundamental frequency (F0) and the frequency of the formants (Ekberg et al., 2023). The fundamental frequency represents the pitch of the voice and can be influenced by emotional state and neurological conditions. Formant frequencies, which are the resonance frequencies of the vocal tract, are important for vowel production and can indicate changes in speech articulation.
2. Temporal features encompass the timing, duration, and rhythm of speech, including prosodic patterns such as intonation, stress, and phrasing (Ekberg et al., 2023). Prosody plays a crucial role in conveying meaning and emotion, and

alterations in these features may reflect underlying neurological or cognitive changes. Temporal features also include parameters related to changes over time, such as the length of voiced and unvoiced segments and the timing of amplitude and spectral balance variations (Ekberg et al., 2023).

3. Amplitude features encompass measures of loudness, shimmer, jitter, and Harmonics-to-Noise Ratio (HNR) (Ekberg et al., 2023). Loudness varies with emotional state and vocal health (Ekberg et al., 2023), while shimmer measures the variability in amplitude between consecutive vocal cycles, providing insights into the stability of voice production (Teixeira et al., 2013). Jitter measures the cycle-to-cycle variations in the F0 of the voice, offering insights into the frequency stability and regularity of vocal fold vibrations, which can indicate neurological or vocal cord health issues (Teixeira et al., 2013). The HNR is a measure of the ratio of harmonic sound to noise in the voice, which can indicate vocal quality and health (Teixeira et al., 2013).
4. Spectral balance features describe the distribution of energy across various frequency ranges (Ekberg et al., 2023). An example is the Hammarberg Index, which compares the energy in the lower versus higher frequency bands of the voice spectrum. This index can provide information about voice quality and the presence of vocal strain or pathology.

By leveraging advanced AI techniques, these features can be systematically analysed to develop robust, non-invasive clinical decision support tools. The integration of AI in speech analysis allows for the identification of complex patterns and subtle biomarkers that might be missed by human analysis, enhancing the accuracy and reliability of concussion diagnosis and monitoring.

Table 3: Acoustic Feature Types

Acoustic Feature Type	**Examples**
Frequency	Fundamental frequency (F0), the frequency of the formants
Temporal	Timing, duration, and rhythm of speech Prosodic patterns such as intonation, stress, and phrasing Length of voiced and unvoiced segments The timing of amplitude and spectral balance variations
Amplitude	Loudness, shimmer, jitter, and HNR
Spectral balance	Hammarberg Index

2.6 The Process of Speech Technology Based System

This section outlines the systematic process involved in developing a speech technology-based system. The steps include voice recording, noise reduction, feature extraction, feature scaling, feature selection, dimensionality reduction, training of AI algorithms, algorithm validation, and device integration (Figure 1).

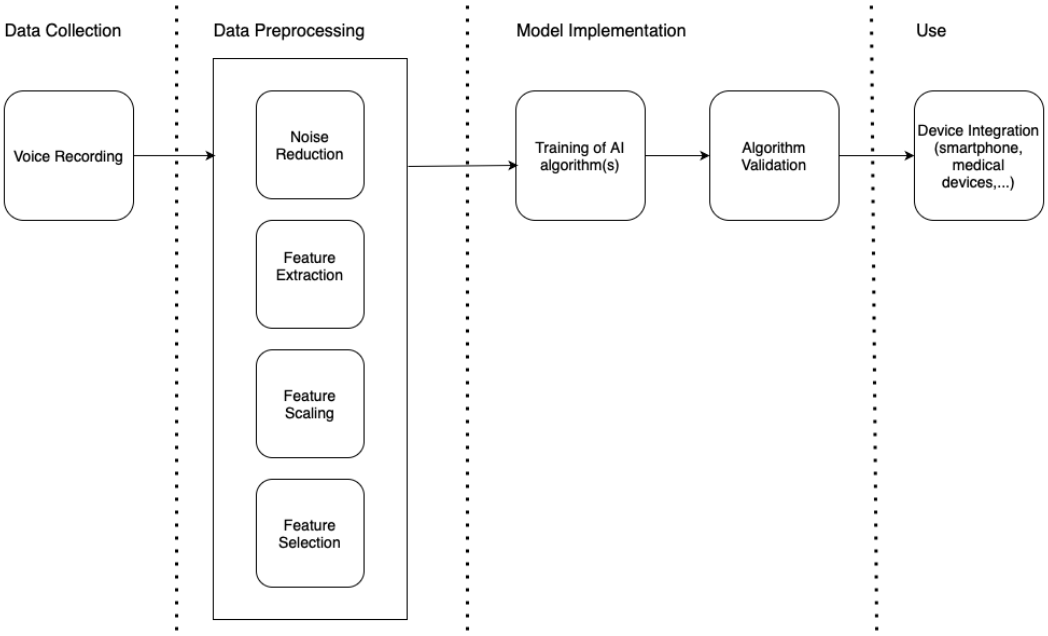

Figure 1: The process of speech technology-based system

2.6.1 Data Collection

Speech data collection is the crucial first step in any speech technology-based system. Researchers have various options regarding the types of tasks they ask participants to perform and the methods they use to record the speech. The best approach depends on the specific research questions or diagnostic goals. It is essential to align the task, recording method, and research objectives to ensure the collected speech data is useful and relevant.

Different speech tasks can provide distinct insights into a person's voice and health. Controlled tasks, such as reading a prepared text or describing a picture, are highly structured. For example, a task involving reading a standard passage can help assess articulation and fluency consistently across participants (Brenk et al., 2022).

On the other hand, semi-spontaneous tasks, such as describing a familiar topic, allow for more natural speech (Boschi et al., 2017). This type of task helps researchers observe typical speech patterns, vocabulary usage, and speaking style.

It provides a more realistic representation of how individuals communicate in everyday settings.

Other tasks target specific aspects of voice and speech. Sustained vowel phonations, where participants hold a single sound like "ah" for example, are useful for assessing overall voice quality (Dedry et al., 2022). These tasks can reveal characteristics such as vocal strength and stability. Diadochokinetic (DDK) tasks, which involve rapidly repeating sounds, test the coordination of mouth muscles, highlighting the ability to produce clear and rapid speech sequences (Lancheros et al., 2023).

Once the speech task is chosen, selecting the right recording method becomes important. Each method presents a unique set of advantages and drawbacks that researchers must carefully weigh against their specific needs and available resources. Studio-based recordings, conducted in specialized sound booths, offer the highest quality audio with minimal background noise (Fagherazzi et al., 2021). This perfect sound capture is ideal for detailed acoustic analysis, but the cost and limited accessibility of these facilities often make them impractical for large-scale studies. Telephone-based recordings, while convenient for reaching a large number of participants, sacrifice some audio quality due to factors like handset variability, background noise, and limited bandwidth (Fagherazzi et al., 2021). However, they can be a valuable option for studies where remote data collection is essential. Web-based recordings offer similar benefits and drawbacks to telephone recordings, with the added challenge of potential internet connectivity issues (Fagherazzi et al., 2021). While accessible to anyone with an internet connection, the quality and reliability of the recordings can be less predictable. Smartphone-based recordings strike a balance between quality and accessibility. Due to the ubiquity of smartphones and their increasingly high-quality microphones, this method offers a cost-effective way to gather speech data (Fagherazzi et al., 2021). However,

ensuring consistent recording conditions and managing data transmission can be challenging, and privacy considerations must be carefully addressed.

2.6.2 Data Preprocessing

The Data Preprocessing stage is essential for preparing raw voice recordings for subsequent analysis and model development. This stage focuses on enhancing the quality of the audio data and transforming it into a format suitable for feature extraction and analysis. In the context of concussion diagnosis using speech analysis, data preprocessing is particularly important due to the subtle nature of the speech changes associated with mTBI.

Effective data preprocessing can significantly improve the accuracy and reliability of the subsequent analysis and modelling steps (Miller, 2019). By removing noise and irrelevant information, data preprocessing enhances the signal-to-noise ratio, making it easier to extract meaningful features from the speech signal. In the context of this research, data preprocessing involved several key steps, including noise reduction, feature extraction, and feature scaling. These steps were essential for preparing the raw speech recordings for analysis and ensuring the quality and consistency of the data used for model development. The specific techniques employed in each step were carefully chosen to address the unique challenges posed by the dataset and the research objectives.

Noise Reduction

Noise reduction is designed to improve the clarity and quality of the signal by attenuating or removing unwanted background noise (Boll, 1979). This process ensures that the acoustic and linguistic features within the speech signal are not obscured or distorted. The selection of an appropriate noise reduction technique depends on the specific characteristics of the noise present in the recording environment.

Various noise reduction techniques exist, each with its strengths and limitations depending on the characteristics of the interfering noise (Table 4). Traditional linear filtering methods, such as high-pass filters (Niederjohn & Grotelueschen, 1976), are effective when the noise source is stationary and exhibits relatively constant spectral characteristics. High-pass filters are simple and computationally efficient, removing low-frequency noise like hum. However, they may distort low-frequency components of speech if the cutoff frequency is too high.

Spectral subtraction (Boll, 1979) is another widely used technique, assessing the noise spectrum during quiet periods or low speech activity and removing it from the noisy signal spectrum. This method is effective for both stationary and non-stationary noise, but its performance can be sensitive to variations in the signal-to-noise ratio (SNR).

Wiener filtering (Benesty et al., 2005) adopts a statistical approach, minimizing the mean-squared error between the clean and noisy signal. This method can adapt to varying noise conditions if accurate estimates of the noise characteristics are available but may struggle in rapidly changing environments.

Table 4: Noise Reduction Techniques

Method	**Strength**	**Limitation**
High-Pass Filter (Niederjohn & Grotelueschen, 1976)	Simple, computationally efficient. Removes low-frequency noise (e.g., hum).	May distort low-frequency components of speech if cutoff frequency is too high.

Table 4: Noise Reduction Techniques (cont.)

Method	**Strength**	**Limitation**
Spectral Subtraction (Boll, 1979)	Effective for both stationary and non-stationary noise	Sensitive to variations in SNR. May introduce musical noise artifacts
Wiener Filter (Benesty et al., 2005)	Adaptable to varying noise conditions. Minimizes mean square error between clean and noisy signals.	Requires accurate estimation of the power spectral density of the signal and noise. Performance may degrade in rapidly changing noise environments

Feature Extraction

Following noise reduction, feature extraction is the process of converting the raw audio data into a set of quantifiable values, or features, that represent key characteristics of the speech signal. These numerical features encapsulate the essential qualities of the sound, enabling structured analysis and interpretation of the data. In the context of concussion diagnosis, these features serve as potential biomarkers, offering insights into how the injury may have affected speech production.

For the purpose of this thesis, the focus was on extracting interpretable acoustic features that have shown promise in Daudet et al. (2017). Below table shows the acoustic metrics used in this research, categorizing them into temporal and frequency features, and providing descriptions.

Table 5: Temporal and Frequency Acoustic Metrics (Daudet et al., 2017)

Acoustic Metric	**Type**	**Description**
Average Duration	Temporal	Average duration taken to say a word in the test.
Standard Deviation in Duration	Temporal	The standard deviation in durations of words being spoken.
Stressed Word Duration	Temporal	The time taken to say a word while stressing it.
Stress Pause	Temporal	Pause time before saying the stressed word.
Average Syllable Duration	Temporal	The average syllable duration in a continuous passage of speech.
Average Pause Duration	Temporal	The average pause duration in a continuous passage of speech. Notable pauses indicates a possible concussion.
Average DDK Rate	Temporal	The number of consonant-vowel (C-V) vocalizations per second.
Standard Deviation in DDK Period	Temporal	This is the standard deviation of the DDK period (in ms).

Table 5: Temporal and Frequency Acoustic Metrics (Daudet et al., 2017) (cont.)

Acoustic Metric	**Type**	**Description**
Coefficient of Variation in DDK Period	Temporal	This parameter measures the degree of rate variation in the period (%). If the C-V vocalization is repeated with little variation in rate, then this number is very small. However, as a speaker varies the rate of DDK durng the seven-second-analysis window, this number increases. This parameter is assessing the participants ability to maintain a constant rate of C-V combinations.
Average Pitch	Frequency	Average pitch from a speech sample.
Pitch Standard Deviation	Frequency	The standard deviation of the pitch in a speech sample.

Table 5: Temporal and Frequency Acoustic Metrics (Daudet et al., 2017) (cont.)

Acoustic Metric	**Type**	**Description**
Pitch Variation	Frequency	How many times the pitch goes above or below the pitch average in a speech sample, weighted by how much it deviates from that average.
Average Pitch Variation	Frequency	Average of the pitch variation.
Pitch Variation Standard Deviation	Frequency	Standard deviation of the pitch variation.
Frequency of Pitch Variation	Frequency	This metric is computed by adding together all the weights for the time component of the pitch variance
Standard Deviation of the Frequency of Pitch Variation	Frequency	Standard deviation of the time between fluctuations in pitch.
Average Power	Frequency	Average power from a speech sample.
Power Variation	Frequency	How many times the power deviates from the power average in a speech sample, weighted by how much it deviates from that average.

Table 5: Temporal and Frequency Acoustic Metrics (Daudet et al., 2017) (cont.)

Acoustic Metric	**Type**	**Description**
Power Variation Standard Deviation	Frequency	Standard deviation of the power variation.
Frequency of Power Variation	Frequency	This metric is computed by adding together all the weights for the time component of the power variance similarly
Standard Deviation of the Frequency of Power Variation	Frequency	Standard deviation of the time between fluctuations in power.

Feature Scaling

After feature extraction, feature scaling scales the data to a consistent range (*Feature Engineering: Scaling, Normalization, and Standardization*, 2023), reducing irrelevant variability caused by factors such as recording volume or speaker loudness. This ensures that the extracted speech features are comparable across different recordings, which is essential for building reliable AI algorithms capable of accurately detecting subtle patterns related to concussion, rather than being misled by differences in recording conditions or individual speaking styles.

Several feature scaling techniques (Table 6) exist to normalize speech data. Min-Max Scaling (*Feature Engineering: Scaling, Normalization, and Standardization*, 2023), the simplest method, adjusts the data to fit within a predetermined range. While easy to use, it can be influenced by unusually high or low values. Z-Core Normalization (*Feature Engineering: Scaling, Normalization, and Standardization*, 2023) rescales the data around zero, making it easier to compare across different recordings. This method is less affected by outliers and

works well for data that follows a bell curve distribution. Robust Scaling (*Feature Engineering: Scaling, Normalization, and Standardization*, 2023), a variation of min-max scaling, is less sensitive to extreme values and can be a good choice for messy data.

Both standardization (Z-Score Normalization) and normalization aim to rescale features in a dataset, but they achieve this in different ways. Standardization rescales data so that the mean becomes 0 and the standard deviation becomes 1, which is beneficial for algorithms that either presume a Gaussian distribution or are influenced by the scale of the features. Normalization, on the other hand, typically scales data to a specific range (often 0 to 1), preserving the relative relationships between data points but making it more susceptible to the influence of outliers. The choice between these techniques often depends on the specific algorithm's requirements and the characteristics of the dataset, with experimentation sometimes necessary to determine the optimal approach.

Table 6: Normalization Techniques

Method	**Strength**	**Limitation**
Min-Max Scaling (*Feature Engineering: Scaling, Normalization, and Standardization*, 2023)	Simple and easy to implement. Preserves the relative relationships between data points	Sensitive to outliers May not handle non-Gaussian distribution well.

Table 6: Normalization Techniques (cont.)

Method	**Strength**	**Limitation**
Z-Score Normalization (*Feature Engineering: Scaling, Normalization, and Standardization*, 2023)	Less sensitive to outliers compared to Min-Max Scaling. Suitable for algorithms that assume Gaussian-distributed data	Does not ensure a specific range for the data. The resulting distribution may still not be Gaussian for non-Gaussian input data
Robust Scaling (*Feature Engineering: Scaling, Normalization, and Standardization*, 2023)	Less sensitive to outliers than Min-Max scaling. Suitable for datasets with extreme values.	Like Min-Max scaling, may not handle non-Gaussian distributions well.

Feature Selection

Following normalization, feature selection helps researchers identify the most informative subset of features extracted from the signal (Sanjyal, 2022). By selecting the most informative features, the analysis is more focused, potentially improving the accuracy of models while reducing the time and resources needed to train them. This becomes especially important with large, complex datasets, where including irrelevant or redundant features can compromise the results.

Feature selection techniques (Table 7) can be employed to identify the most informative features in speech analysis. Correlation Coefficient (Gupta, 2023), a visualization tool, quickly identifies linear relationships between features, highlighting potential redundancy or multicollinearity. However, it overlooks non-

linear associations and only considers pairwise interactions, potentially missing more complex relationships between features. The Variance Threshold (Gupta, 2023) is another straightforward filter method that eliminates features with very low variance, assuming they contribute little information. While incredibly fast, this approach might discard potentially relevant features if their variance happens to be low by chance. Recursive Feature Elimination (RFE) (Gupta, 2023), offer a more sophisticated approach that evaluates feature subsets based on their performance in a predictive model. RFE begins by considering all features and progressively eliminates the least significant ones until an optimal subset is identified. This iterative process considers feature interactions and directly measures the impact of feature subsets on model performance. However, it is computationally expensive, particularly with complex models and large datasets, as it requires retraining the model multiple times. LASSO (L1 regularization) (Gupta, 2023), perform feature selection as an integral part of the model training process. LASSO adds a penalty term to the model's objective function, encouraging the model to select only a subset of the most relevant features. This results in a sparse model where many feature coefficients are set to zero, effectively eliminating them from the model. While effective in high-dimensional datasets, LASSO requires careful tuning of the regularization parameter to avoid underfitting or overfitting.

Table 7: Feature Selection techniques

Method	**Strength**	**Limitation**
Correlation Coefficient (Gupta, 2023)	Simple, fast. Easily visualizes linear relationships between features.	Ignores non-linear relationships. Only considers pairwise correlations, not interactions.

Table 7: Feature Selection techniques (cont.)

Method	**Strength**	**Limitation**
Variance Threshold (Gupta, 2023)	Very fast. Removes features with minimal variance (potential noise).	Can discard informative features if their variance is low by chance. Does not consider relationship to the target variable.
RFE (Gupta, 2023)	Considers feature interactions by evaluating them in the context of a model.	Computationally expensive, especially with complex models. Requires retraining models repeatedly.
L1 Regularization (Gupta, 2023)	Performs feature selection as part of model training. Encourages sparsity (many features set to zero).	Requires careful tuning of regularization parameter. Can be computationally expensive with very large datasets.

2.7 Model Training

Following data preprocessing, the refined dataset serves as the foundation for training AI models to detect meaningful patterns indicative of concussion. Given the research's focus on classifying speech patterns as either concussion-related or not, the primary emphasis is on supervised learning methods. Supervised ML techniques utilize knowledge gained from previous and present data, along with labels, to make predictions (Saravanan & Sujatha, 2018). This approach commences with a training phase, during which the ML model constructs a function to forecast output values. Upon sufficient training, the system can generate predictions for new input data. The ML algorithm then compares these

predictions with the actual outcomes to identify errors and refine the model accordingly.

Supervised learning methods - Support Vector Machines (SVM) and Multilayer Perceptron (MLP) - are widely recognized for their effectiveness in data classification tasks, particularly when dealing with clearly labelled datasets like the binary classification of speech patterns as concussed or non-concussed (Saravanan & Sujatha, 2018). SVMs excel at finding optimal decision boundaries in high-dimensional data, making them suitable for analysing the numerous acoustic features extracted from speech. This strength has been demonstrated in research like Ali et al. (2016) , where SVM was successfully used for voice pathology detection, showcasing its robustness in handling varying numbers of voice signals and achieving promising accuracy. Similarly, MLPs have proven effective in capturing complex relationships in speech data, as demonstrated by Arya et al. (2021), who achieved high accuracy in speech-based emotion recognition using MLP. Given that concussions can impact emotional and cognitive processes, which may manifest in speech patterns, the ability of MLPs to model such complexities makes them a promising candidate for concussion detection. The selection of SVM and MLP for this research was thus based on their proven success in speech analysis, their capacity to handle high-dimensional data, and their potential for accurate concussion detection.

Given the importance of these methods for this research, the following sections shows these specific supervised learning techniques in detail:

2.7.1 Support Vector Machines (SVM)

SVM, introduced by Vapnik and coworkers in 1992 (Boser et al., 1992), is widely used for classification problems, especially in scenarios with high-dimensional data. The fundamental concept behind SVMs, as illustrated in Figure 2, is to find

an optimal hyperplane (the solid black line) that best separates two classes of data points (blue triangles and green circles). This hyperplane serves as a decision boundary, effectively classifying new data points based on which side of the line they fall.

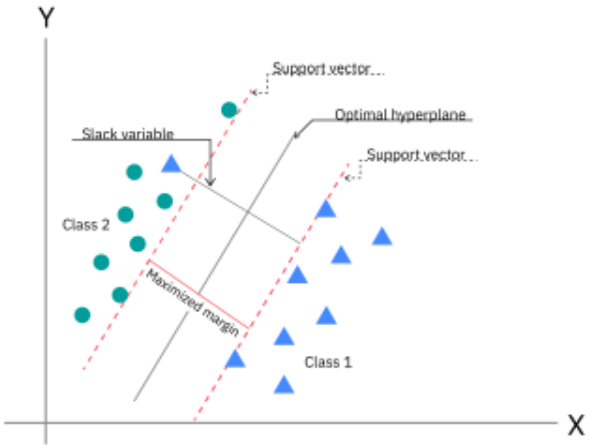

Figure 2: Illustration of SVM (IBM, 2023)

The strength of SVM lies in their ability to not only find a separating hyperplane but also to maximize the margin around it. This margin, the region defined by the red dashed lines on either side of the hyperplane, represents the distance to the closest data points from each class, referred to as support vectors. By maximizing the margin, the SVM ensures that the decision boundary is as far away as possible from both classes, increasing the model's confidence in its classifications and reducing the risk of overfitting to noise in the training data.

In real-world scenarios, data points often cannot be perfectly divided into distinct classes using a simple straight line. To overcome this, the "kernel trick" is employed. Kernel functions project the data into a higher-dimensional space, in which it becomes easier to separate them linearly. The specific kernel function chosen, such as linear, polynomial, or radial basis function (RBF), depends on how complex the relationship between the features is and how intricate the model needs to be. This flexibility allows SVMs to be applied to a wide variety of classification problems, ranging from straightforward linear separations to more

complex, non-linear patterns.

SVM has demonstrated its value in various healthcare domains, particularly those involving the analysis of speech and audio data. For example, in a study on pathological voice detection, SVM showed high accuracy in differentiating between normal and disordered voices using a novel voice intensity-based approach (Ali et al., 2016). Beyond diagnosis, SVMs have also been utilized in speech recognition tasks, such as identifying spoken words or phonemes, and in speaker recognition systems for biometric authentication purposes (Campbell et al., 2006). This broad range of applications highlights the adaptability of SVM to various challenges in speech and audio analysis, making them a promising tool for developing new diagnostic and therapeutic interventions.

2.7.2 Multilayer Perceptron (MLP)

Artificial Neural Networks (ANNs), inspired by the intricate workings of the human brain, are a class of ML models capable of tackling both linear and nonlinear problems by learning complex relationships within data. One particularly popular type of ANN, the MLP, has emerged as a tool for pattern recognition (Taud & Mas, 2018).

To understand the MLP, we have to consider its building block: the single neuron perceptron (Figure 3). This simplest neural network has multiple inputs, each weighted according to its importance, that are summed and then passed through an activation function to produce a single output.

MLP takes this concept further by introducing one or more hidden layers between the input and output layers, as illustrated in Figure 4. Each node in the hidden layer receives weighted inputs from all nodes in the preceding layer, computes a weighted sum, and then applies an activation function. This layered structure, with non-linear activation functions at each node, allows MLP to learn

complex, non-linear patterns in the data.

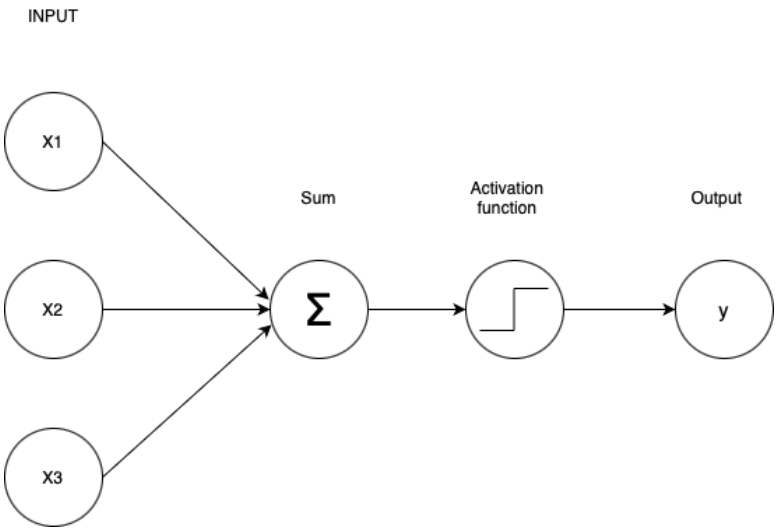

Figure 3: Single Neuron Perceptron

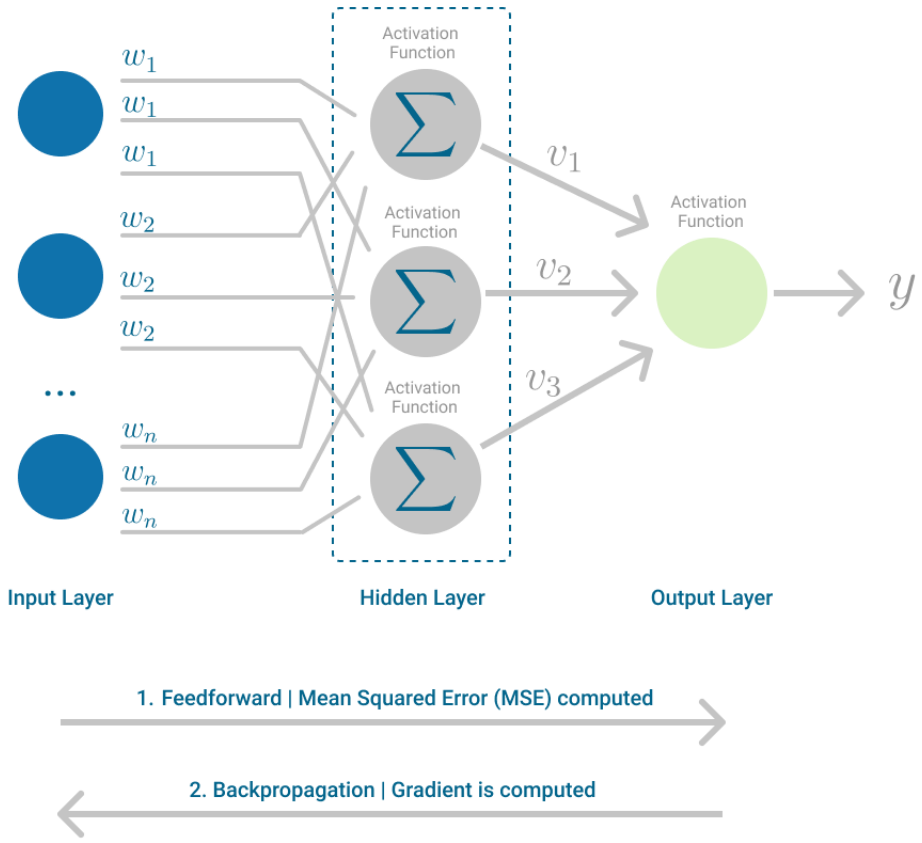

Figure 4: Multilayer Perceptron (Carolina, 2021)

The true power of MLP lies in their ability to learn and adapt through a process called backpropagation. This iterative process involves calculating the difference, or error, between the model's predicted output and the actual output. Beginning at the output layer and progressing towards the input layer, this error is then propagated backward through the network. The error signal serves as a guide, instructing the model to adjust the weights associated with each connection in a way that minimizes the overall error. This gradual fine-tuning of the weights (1), represented by the change in weight Δw(t) at each iteration t, allows the MLP to learn the complex relationships within the data and improve its prediction accuracy over time. The magnitude and direction of this weight adjustment are determined by the learning rate ε, which manages the step size of the updates, and the gradient of the error $\frac{\mathrm{dE}}{\mathrm{dw(t)}}$ with respect to the weights at that specific iteration. A momentum term α is often incorporated into the update rule, combining a fraction of the previous weight change Δw(t-1) with the current gradient to accelerate convergence and smooth out oscillations in the learning process.

$$\Delta w(t) = -\varepsilon \frac{dE}{dw(t)} + \alpha \Delta w(t-1) \qquad (1)$$

A crucial requirement for backpropagation to function effectively is that the functions used within the neurons—the weighted sum of inputs and the activation function—must be differentiable. This ensures that the gradient of the error can be calculated, providing the necessary information to guide the weight adjustments. Widely used activation functions used in MLPs include sigmoid, ReLU (Rectified Linear Unit), and tanh (hyperbolic tangent) (*Activation functions in Neural Networks*, 2024). Each of these functions adds a non-linear element to the model, allowing it to learn intricate relationships that linear models cannot represent. The choice of activation function often depends on the specific task and dataset, with each function offering different trade-offs in terms of computational efficiency,

learning dynamics, and the types of patterns they can effectively capture.

MLP has established themselves as a useful tool in the realm of speech and audio analysis, particularly in healthcare application. For instance, their ability to model complex patterns and relationships within data has proven valuable for tasks such as recognizing spoken digits (Ahad et al., 2002) and diagnosing diseases like Parkinson's disease based on speech characteristics (Bakar et al., 2010).

2.8 Evaluation Metrics

Different evaluation metrics are available to assess the performance of prediction.

Table 8: Evaluation Metrics

Metrics	**Description**	**Formula**
Confusion Matrix	TP: True Positive TN: True Negative FP: False Positive FN: False Negative	
Accuracy	Measures the overall correctness of predictions, calculated as the ratio of correctly predicted instances to the total instances.	Accuracy $= \frac{TP + TN}{TP + TN + FP + FN}$

Table 8: Evaluation Metrics (cont.)

Metrics	**Description**	**Formula**
Precision	Indicates the proportion of true positive predictions among all positive predictions.	$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$
Recall (Sensitivity)	Measures the proportion of actual positives correctly predicted by the model.	$\text{Recall (Sensitivity)} = \frac{\text{TP}}{\text{TP} + \text{FN}}$
Specificity	Measures the proportion of actual negatives correctly predicted by the model.	$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$
F1 Score	Harmonic mean of precision and recall, offering a balanced metric.	$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$
AUC-ROC	Represent the Are Under the ROC (Receiver Operating Characteristics) curve, capturing the trade-off between True Positive Rate and False Positive Rate	

A confusion matrix (Table 9) is a simple table that helps visualize the performance of a classification model. It presents the number of true positives (correctly identified concussions), true negatives (correctly identified non-concussions), false positives (incorrectly diagnosed concussions), and false negatives (missed concussions).

Table 9: Example of Confusion Matrix

		Predicted Value	
		Positive	Negative
Actual Value	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

Accuracy is the most basic metric for evaluating machine learning models. It simply evaluates the proportion of correct predictions out of all the predictions. This is easy to understand but can be misleading when one class (like concussed individuals) is much less common than the other (non-concussed). Accuracy can be high even if the model misses most of the less common cases.

Precision and recall (also known as sensitivity) offer a more nuanced view. Precision tells us how often a positive prediction is correct, which is important when false alarms are undesirable. Recall measures how well the model captures all the true positives, critical in concussion diagnosis where missing a concussion can have serious consequences. Specificity is the opposite of recall; it measures how many true negatives (people who do not have concussions) are correctly identified, ensuring a low false positive rate.

To get a balanced picture, we often use the F1 score, which combines precision and recall into a single value. This is especially useful for imbalanced datasets. Another helpful tool is the AUC-ROC curve, which shows how well the

model distinguishes between classes at different thresholds. A higher AUC-ROC indicates better overall discriminatory power.

The choice of which metrics to focus on depends on the specific goals of the analysis. In concussion diagnosis, both minimizing missed cases (high recall) and avoiding unnecessary alarm (high precision) are important. Therefore, a combination of these metrics, along with the AUC-ROC, will provide a thorough evaluation of the model's performance.

2.9 Research Gap

While the current diagnostic methods for mTBI, including clinical assessments and neuroimaging, have provided valuable insights, significant gaps remain to detect accurately and efficiently diagnose concussion. Clinical assessments, while accessible and relatively inexpensive, rely heavily on subjective patient reporting, introducing variability and potential for misdiagnosis (Mactaggart et al., 2016; Shenton et al., 2012; Theadom et al., 2021). These methods often rely on manual interpretation by experts, which can be time-consuming and prone to subjectivity (Theadom et al., 2021). Additionally, these assessments may not capture subtle neurological changes or underlying brain damage that may not manifest as overt symptoms (Shenton et al., 2012). Neuroimaging techniques, such as CT and MRI scans, offer objective visualization of brain structure and function but are costly, often inaccessible, and may not always be sensitive to subtle mTBI-related changes (Jacobs & Henwood, 2013).

The limitations of current methods highlight the need for a more objective, accessible, and affordable diagnostic tool for mTBI. Speech analysis, with its non-invasive nature and potential for capturing subtle neurological changes, emerges as a promising alternative (Poellabauer et al., 2015). The integration of AI into speech analysis addresses these limitations by automating the analysis process and enabling the detection of subtle patterns and biomarkers that may be missed by

human observation (Daudet et al., 2017). However, research in this field is still in its early stages, and there is a need for further investigation to establish the reliability and validity of AI-powered speech analysis for mTBI diagnosis.

This research aims to fill this gap by identifying specific speech biomarkers that are indicative of mTBI and developing AI models that can accurately classify speech patterns as concussed or non-concussed.

Chapter 3 - Methodology

This chapter presents the research design and methodology used to investigate the potential of speech analysis for concussion detection. It encompasses the entire pipeline, from data preparation to model development, highlighting the steps taken to ensure the rigor and reproducibility of the study. The chapter is organized as follows: the first section provides an overview of the research design, justifying the choice of a quantitative approach for this study. The second section describes the dataset and the data collection process, including details about the participants, recording equipment, and speech tasks. The subsequent sections cover data preprocessing techniques such as noise reduction, feature extraction, normalization, and feature selection. These steps ensure the data is suitable for analysis and model training. Specific details of the model implementation, training, and validation will be addressed in dedicated following chapter.

3.1 Research Design

A research design is a plan for finding an answer to a research question through empirical data collection and analysis. It serves as a comprehensive roadmap that outlines the steps a researcher will take to achieve their research objectives. One of the fundamental choices in crafting this design is selecting between a qualitative or quantitative approach (McCombes, 2021). While qualitative research delves into subjective experiences, beliefs, and concepts to gain a deeper understanding of a particular context, quantitative research emphasizes objective measurement, statistical analysis, and hypothesis testing (McCombes, 2021).

For this research, a quantitative cross-sectional study is deemed most suitable (*Quantitative study designs*, 2024). This choice is driven by the aim of the research

questions, which try to identify specific speech biomarkers for concussion diagnosis and evaluate the performance of ML models using selected speech biomarkers. A quantitative approach aligns with these objectives as it allows for precise measurement of various acoustic features in speech, enabling the use of statistical techniques to identify significant differences between individuals with and without concussions. This approach facilitates rigorous statistical testing to determine if the identified speech biomarkers are significantly different between the two groups. Additionally, the performance of ML models can be quantitatively assessed using metrics such as accuracy, sensitivity, and specificity, providing objective evidence of their diagnostic potential.

3.2 Data Description

This research utilizes a dataset containing speech recordings from 2708 student-athletes attending colleges and high schools across the eastern and midwestern United States. The data, originally collected in Fall 2014 by Yadav (2015) and provided by my supervisor Dr. Samaneh Madanian, were obtained through an iPad application designed to assess speech characteristics.

All baseline tests were conducted under the supervision of trained staff, including athletic trainers. To familiarize participants with the testing procedure, each athlete completed the test twice, with only the second recording retained for analysis. This helps to minimize the impact of unfamiliarity with the task on speech performance. Post-baseline recordings were classified as either "normal" (from healthy individuals) or "concussion suspected" (from individuals within 48 hours of a suspected concussive event). The underlying hypothesis is that speech patterns in individuals with a concussion will differ significantly from those in the healthy control group.

The speech data were collected using a standardized test administered on iPad

mini devices equipped with a Shure SM10A[1] low-impedance microphone, ideal for close-talk applications. The audio was sampled at 44.1kHz, 16-bit, mono, ensuring high fidelity. To maintain data quality, a noise management technique, which was proposed in the work of Yadav et al. (2014), was implemented on the device to reject recordings with low SNR ratio, ensuring that only clear and usable speech samples were included in the dataset.

To gather a diverse range of speech samples, participants completed a set of speech tasks designed to assess various aspects of speech production. These tasks captured both temporal and prosodic features, such as speech rate, rhythm, intonation, and articulation. The specific tasks are detailed in Table 10, which outlines the text displayed on the screen, the display duration, and the description of each task.

Table 10: Speech Tasks (Yadav, 2015)

Test	**Text Displayed on Screen**	**Display Duration**	**Description**
1	Application, Participate, Education, Difficulty, Congratulations, Possibility, Mathematical, Opportunity	1.5 seconds per word	Participant reads out multisyllabic words. Test designed to study articulation rate and word duration

[1] https://www.shure.com/en-US/products/microphones/sm10

Table 10: Speech Tasks (Yadav, 2015) (cont.)

Test	**Text Displayed on Screen**	**Display Duration**	**Description**
2	PUT the book here Put the BOOK here Put the book HERE	10 seconds	Participants reads and stresses different parts of the sentence, i.e., put, book, here, test designed to capture intonation stimulability
3	We saw several wild animals	5 seconds	Participant reads simple sentence to test standard syllabic rate
4	pa	5 seconds	Participants repeat the pa sound as quickly as possible. Alternating motion rate is captured
5	ka	5 seconds	Participants repeat the ka sound as quickly as possible. Alternating motion rate is captured

Table 10: Speech Tasks (Yadav, 2015) (cont.)

Test	**Text Displayed on Screen**	**Display Duration**	**Description**
6	Pa-ta-ka	5 seconds	Participants repeat the pa-ta-ka sound as quickly as possible. Sequential motion rate is captured
7	Aaaahhhhh	5 seconds	Jitter and duration captured as participants are asked to sustain the *Aaaahhhhh sound*

Table 10 shows various speech tasks for audio collection. The word production (test 1) assessed the participant's ability to articulate words of varying complexity, starting with simple monosyllabic words and progressing to more challenging multisyllabic words. Sentence repetition (test 2 and 3), specifically reading aloud sentences with different emphasized words, examined the prosodic aspects of speech, exploring how concussions might disrupt the rhythm, stress, and intonation patterns that convey meaning and emotion. To assess speech fluency and rhythm in a more natural context, participants were asked to read aloud a standard sentence, providing an opportunity to measure syllable duration and syllabic rate. The battery of tasks also included DDK exercises (tests 4, 5, and 6), requiring participants to rapidly repeat specific syllable sequences. These tasks are known to be sensitive indicators of motor coordination and speech motor control, potentially revealing subtle deficits in articulation and fluency. Lastly, the

sustained vowel phonation task (test 7), where participants held an "Aaaahhhhh" sound, was used to evaluate vocal stability, pitch control, and muscle tone.

To ensure the quality of the recordings, each speech recording went through a SNR calculation and threshold comparison phase, intended to guarantee the recording meets the quality standards. This method, proposed by Yadav et al. (2014), involved an automatic rejection of low-quality recordings based on their SNR. Yadav's noise management technique estimated the SNR for each recording by quantifying the levels of the speech signal and background noise. The SNR is calculated as:

$$SNR = 10log\frac{\text{PeakSpeechPower}}{\text{MeanNoisePower}}$$

A threshold SNR value was set to differentiate between acceptable and unacceptable recordings. Based on the recommendations in Yadav et al. (2014), a voiced SNR of above 38 dB and an unvoiced SNR over 22 dB were advised for accurate voice recognition. Therefore, recordings with an average SNR below these thresholds were automatically rejected to ensure clarity.

If the recording was noisy or too quiet, the athlete was asked to retake the test, potentially adjusting the microphone, speaking louder, or relocating to a quieter area. Once a satisfactory recording was achieved, it was transmitted to a remote server for further analysis. Baseline tests were conducted while participants were at rest and not influenced by concussion or intense exercise. Tests conducted after such conditions were categorized accordingly and underwent the same quality checks as baseline tests.

Note that out of the initial 2,708 subjects, many were excluded because they only provided baseline recordings without a corresponding post-baseline recording. Additionally, several more subjects were eliminated due to noise and other quality issues. Consequently, the dataset used for feature extraction included

only 702 unique athletes. This careful selection process ensured that only high-quality and relevant data were used for further analysis, enhancing the reliability of the study's findings.

By implementing this robust noise management technique, the study of Yadav (2015) ensured that the dataset comprised high-quality recordings. Therefore, noise reduction here was not necessary. This simplifies the preprocessing pipeline and focuses efforts on other critical aspects such as feature extraction and normalization.

This carefully collected dataset provides a rich resource for investigating the potential of speech analysis in concussion diagnosis. The subsequent sections of this chapter will detail the steps taken to prepare, analyse, and model this data to uncover speech biomarkers that could aid in the identification of mTBI.

3.3 Experimental Process

The experimental process for this study involves several key stages, starting with noise reduction and proceeding through feature extraction, data preprocessing, feature selection, model development, and evaluation. The overall goal was to prepare the data, select features, develop models, and evaluate their performance, as illustrated in Figure 5.

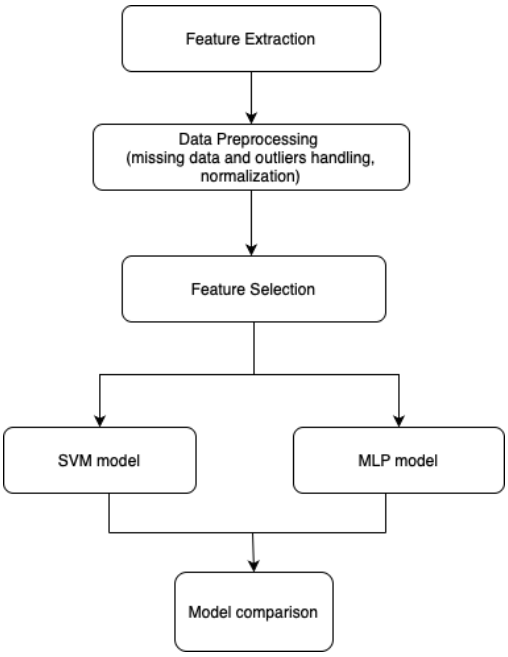

Figure 5: Research experiment process

First, relevant speech features were extracted from the recordings. Using the 'ConcussionDataset_ID_to_Files_Mapping' Excel file, which labelled each recording as concussed or non-concussed, 41 acoustic features were identified. These features were considered potential biomarkers for concussion diagnosis and included various temporal and frequency-based metrics. Following feature extraction, the data underwent preprocessing to handle missing values, manage outliers, and normalize the features. This step ensured that the data was suitable for analysis and enhanced the performance of the ML models. Subsequently, feature selection was performed to identify the most informative subset of features. This process helped to improve model accuracy and efficiency by focusing on the most relevant data.

For model development, two types of ML models were built: a SVM and a MLP. These models were trained using the processed and selected features, aiming to classify the speech recordings accurately. Finally, the models were evaluated and compared based on various performance metrics. This comparison helped to determine which model was more effective in detecting concussion-related speech

patterns. The results of this comparison were discussed in depth to draw meaningful conclusions about the best approach for concussion detection using speech analysis.

3.4 Feature Extraction

After receiving the audio files, which had undergone noise reduction, the next step was to extract relevant speech features. The dataset included an Excel file named "ConcussionDataset_ID_to_Files_Mapping", which contained all the recordings corresponding to each athlete, labelled 1 for concussed and 0 for non-concussed. There were 702 unique athletes in total. Based on the corresponding recordings for each athlete in the mapping file, Python code was executed in Jupyter Lab to extract acoustic features according to the idea used in the study of Daudet et al. (2017). These recordings were then processed to extract 41 acoustic features considered as potential biomarkers for concussion diagnosis. The extracted features are detailed in Table 11, which outline the feature name, acoustic metric, and description.

The feature extraction process involved parsing each recording, segmenting it as required by the specific feature extraction methods, and computing the relevant metrics. The extracted features span various aspects of speech, including temporal and frequency-based metrics. The key acoustic features extracted in this study include timing and rate measurements, pitch and frequency measurements, and amplitude and intensity measurements. Timing and rate measurements can be directly calculated in the time domain after identifying the boundaries of words, syllables, or phonemes. These features are crucial as they can reflect changes in motor control and coordination, which are often impaired following a concussion (Salvatore et al., 2019). Examples include average duration, standard deviation of durations, and pause durations. Pitch and frequency measurements require analysis in the frequency domain and include fundamental frequency (F0), pitch movement,

and pitch rate. These features are important because they can reveal alterations in vocal fold function and control (Wang & Song, 2022), which may be affected by a head injury. Amplitude and intensity measurements involve analysing the energy and intensity variations in the speech signal, such as intensity deviation and standard deviation of peak intensity. These features are relevant as they can indicate changes in respiratory and phonatory effort, potentially reflecting underlying neurological deficits (Wang & Song, 2022).

To ensure consistency and accuracy in feature extraction, several key functions were implemented using the Librosa library (Khare, 2024). The `detect_words` function identifies non-silent intervals in the audio signal, helping to isolate spoken words from pauses and background noise. The `find_optimal_top_db` function dynamically determines the appropriate top_db value for detecting non-silent intervals, ensuring the best possible signal quality for feature extraction. The `extract_features_1` function calculates the average duration and standard deviation of word durations, while the `extract_stress_pause_and_word_durations` function measures pause durations before stressed words like "PUT," "BOOK," and "HERE." The `extract_pitch_movement` and `extract_pitch_rate` functions estimate the pitch movement and rate for specified words, providing insights into pitch variations. The `extract_amp_intensity_deviation` function calculates the deviation in energy intensity for specified words. The `extract_avg_syllable_duration` and `extract_avg_pause_duration` functions compute the average syllable and pause durations, respectively. The `extract_avg_ddk_rate_period`, `extract_avg_ddk_rate`, and related functions measure DDK rates, providing metrics for speech motor control. Finally, the `extract_f0_features` function calculates various F0-related metrics, including average, standard deviation, range, and coefficient of variation.

Table 11: Extracted Acoustic Features and Descriptions

Test	**Feature**	**Acoustic Metric**	**Description**
Test 1	Time	Average Duration	Average duration of words spoken in test
Test 1	Time	Standard Deviation Duration	Standard Deviation in the words spoken in the test
Test 2-PUT/ Test 2-BOOK/ Test2-HERE	Time	Stressed Word Duration	Time taken to say the stressed word "PUT"/"BOOK"/"HERE"
Test 2-PUT/ Test 2-BOOK/ Test2-HERE	Time	Stress Pause	Pause time before saying the stressed word
Test 2-PUT/ Test 2-BOOK/ Test2-HERE	Pitch	F0 Movement	Pitch Movement

Table 11: Extracted Acoustic Features and Descriptions (cont.)

Test	**Feature**	**Acoustic Metric**	**Description**
Test 2-PUT/ Test 2-BOOK/ Test2-HERE	Amplitude	Intensity Deviation	Deviation in energy intensity
Test 3	Time	Average Syllable Duration	This is the syllable duration for the passage (ms). Many dysarthric speakers have slower rates of speech and the duration increases.
Test 3	Time	Average Pause Duration	This is the pause duration for the passage (ms). This passage should have no pauses. Therefore, any significant pause time is a variation from normal speech patterns.

Table 11: Extracted Acoustic Features and Descriptions (cont.)

Test	**Feature**	**Acoustic Metric**	**Description**
Test 3	Time	Average Diadochokinetic Rate Period	Average DDK period of the subject during this vocalization (ms). The average period is the average time between the consonant-vowel vocalizations. The period is inversely related to the rate.
Test 4/5/6	Time	Average DDK Rate	The average DDK rate is the number of the consonant-vowel (i.e., “pa”) vocalizations per second. The rate is inversely related to the average period. Many motor disordered speakers show reduced DDK rates due to decreased articulatory motility.

Table 11: Extracted Acoustic Features and Descriptions (cont.)

Test	**Feature**	**Acoustic Metric**	**Description**
Test 4/5/6	Time	Standard Deviation in DDK Period	This is the standard deviation of the DDK period (ms). A normal speaker can maintain periodic repetitions while many disordered voices show more variability in their repetition rate, therefore increased DDK.
Test 4/5/6	Time	Coefficient of Variation in DDK Period	Degree of rate variation in the period (%). If the consonant-vowel vocalization in repeated with little variation in rate, then this number is very small. However, as a speaker varies the rate of DDK during the seven-second-analysis window, this number increases.
Test 4/5/6	Amplitude	Standard Deviation in DDK Peak Intensity	This is the standard deviation of the DDK peak intensity (dB).

Table 11: Extracted Acoustic Features and Descriptions (cont.)

Test	**Feature**	**Acoustic Metric**	**Description**
Test 4/5/6	Amplitude	Coefficient of Variation of DDK Peak Intensity	Degree of intensity variation in the peak of each C-V vocalization
Test 7	Pitch	Average F0	Arithmetic mean of the fundamental frequency
Test 7	Pitch	Standard Deviation F0	This is a measure of variability in the data.
Test 7	Pitch	F0 Range	This is a measure of the difference between the maximum and minimum pitch values (in Hz) in the active window (time frame saying the “ah” sound).
Test 7	Pitch	Coefficient of Variation of F0 (vF0)	The vF0 is defined as the standard deviation F0 divided by the arithmetic mean
Test 7	Amplitude	Standard Deviation F0	Energy measure
Test 7	Amplitude	Coefficient of Variation of F0 (vF0)	Coefficient of variation related to energy measure.

After extracting these features into a DataFrame using Pandas (Python Pandas DataFrame, 2024), they were saved along with the corresponding labelled value

for each athlete into a separate CSV (Comma Separated Values) file for later use, named “concussion_dataset”. The saved feature names are listed in Table 12. The name of speech features followed the format “<test>_<feature>_<acoustic metric>” in Table 11, along with the target feature ‘concussionsuspected’.

Table 12: Feature Name in CSV dataset

Column Name	**Data Type**
athlete_id	object
test1_time_avg_duration	float64
test1_time_std_duration	float64
test2_time_PUT_stressed_word_duration	float64
test2_time_BOOK _stressed_word_duration	float64
test2_time_HERE_stressed_word_duration	float64
test2_time_PUT_stress_pause	float64
test2_time_BOOK_stress_pause	float64
test2_time_HERE_stress_pause	float64
test2_pitch_PUT_f0_movement	float64
test2_pitch_BOOK_f0_movement	float64
test2_pitch_HERE_f0_movement	float64
test2_pitch_HERE_f0_rate	float64
test2_pitch_HERE_f0_rate	float64
test2_pitch_HERE_f0_rate	float64
test2_amp_PUT_intensity_deviation	float64

Table 12: Feature Name in CSV dataset (cont.)

Column Name	**Data Type**
test2_amp_BOOK_intensity_deviation	float64
test2_amp_HERE_intensity_deviation	float64
test3_time_avg_syllable_duration	float64
test3_time_avg_pause_duration	float64
test3_time_avg_DDK_rate_period	float64
test4_time_avg_DDK_rate	float64
test4_time_std_DDK_period	float64
test4_time_coef_of_var_DDK_period	float64
test4_amp_std_DDK_peak_intensity	float64
test4_amp_coef_of_var_DDK_peak_intensity	float64
test5_time_avg_DDK_rate	float64
test5_time_std_DDK_period	float64
test5_time_coef_of_var_DDK_period	float64
test5_amp_std_DDK_peak_intensity	float64
test5_amp_coef_of_var_DDK_peak_intensity	float64
test6_time_avg_DDK_rate	float64
test6_time_std_DDK_period	float64
test6_time_coef_of_var_DDK_period	float64
test6_amp_std_DDK_peak_intensity	float64
test6_amp_coef_of_var_DDK_peak_intensity	float64

Table 12: Feature Name in CSV dataset (cont.)

Column Name	**Data Type**
test7_pitch_avg_F0	float64
test7_pitch_std_F0	float64
test7_pitch_F0_range	float64
test7_pitch_vF0	float64
test7_amp_std_ F0	float64
test7_amp_vF0	float64
concussionsuspected	int64

3.6 Data Preprocessing

Once the relevant speech features were extracted, the next step involved data preprocessing. This stage is important for preparing the data to be suitable for ML model development. The data preprocessing steps included handling missing values, managing outliers, and normalizing the data to ensure consistency and readiness for analysis.

3.6.1 Handling Missing Values

In the dataset, for some athletes, there were some missing audio files, which resulted from various reasons, but mainly from not being qualified after applying noise reduction. To determine the appropriate method for imputing missing data, it was necessary to consider the distribution of the features in the dataset. Two common imputation methods are mean and median imputation (Brownlee, 2020b). Mean imputation is best suited for normally distributed data, where the distribution is symmetric. However, it is sensitive to outliers, which can skew the mean and result in inaccurate imputations. On the other hand, median imputation is more

appropriate for skewed data or data with outliers, as the median is robust and not affected by extreme values.

Upon examining the histograms of the features, it was observed that most features exhibited skewed distributions, particularly right-skewed distributions. For instance, features such as `test1_time_avg_duration`, `test1_time_std_duration`, and `test2_time_PUT_stressed_word_duration` (Figure). These skewed distributions suggested that median imputation would be more appropriate, as it would not be influenced by the presence of outliers and would provide a more accurate representation of the central tendency of the data.

A few features, such as `test2_pitch_PUT_f0_movement`, and `test2_pitch_BOOK_f0_movement`, showed symmetric distributions (Figure 7). For these features, mean imputation could be considered suitable. However, given the overall skewed nature of the dataset, median imputation was generally recommended for consistency and robustness against outliers.

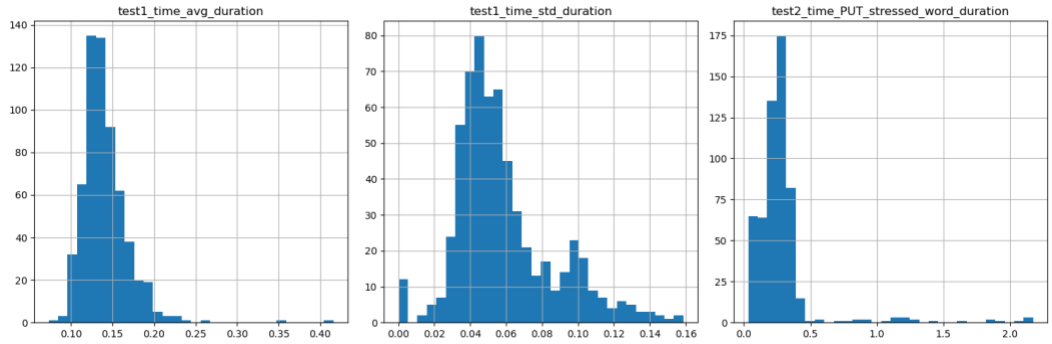

Figure 6: Histograms of test1_time_avg_duration, test1_time_std_duration, and test2_time_PUT_stressed_word_duration

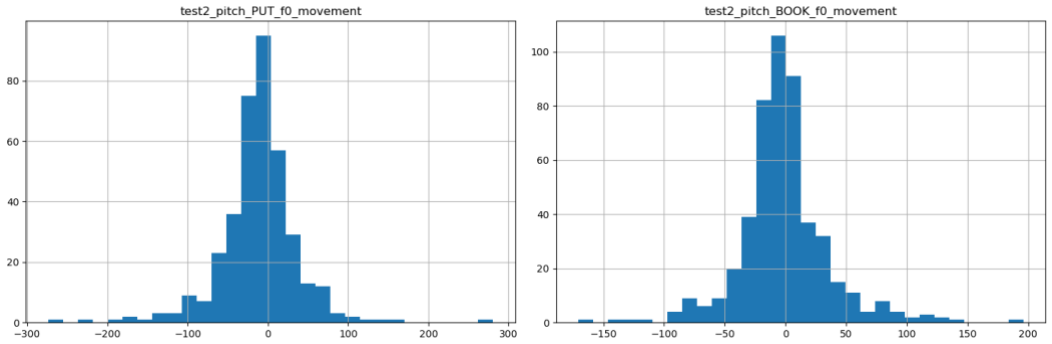

Figure 7: Histograms of test2_pitch_PUT_f0_movement, and test2_pitch_BOOK_f0_movement

The `SimpleImputer` class from the Scikit-Learn library was imported to handle missing values (Brownlee, 2020b). It initializes a `SimpleImputer` object with the imputation strategy set to 'median', meaning that any missing values in the dataset will be replaced with the median value of the respective feature. Then, it applies the imputer to the numeric dataframe `df_numeric`, transforming it by replacing missing values with the median. The transformed data is then converted back into a Pandas DataFrame, `df_imputed`, with the same column names as the original DataFrame. This ensures that the dataset is complete and ready for further analysis.

3.6.2 Handling Outliers

After analysing the distribution of the dataset using histograms and boxplots, it became evident that several features contain outliers (**Error! Reference source not found.** shows example of 4 features), which could impact the effectiveness of our predictive models. Among the various methods available to handle outliers, Winsorization was chosen for this analysis due to its balanced approach in addressing outliers without discarding any data points (Horsch, 2021).

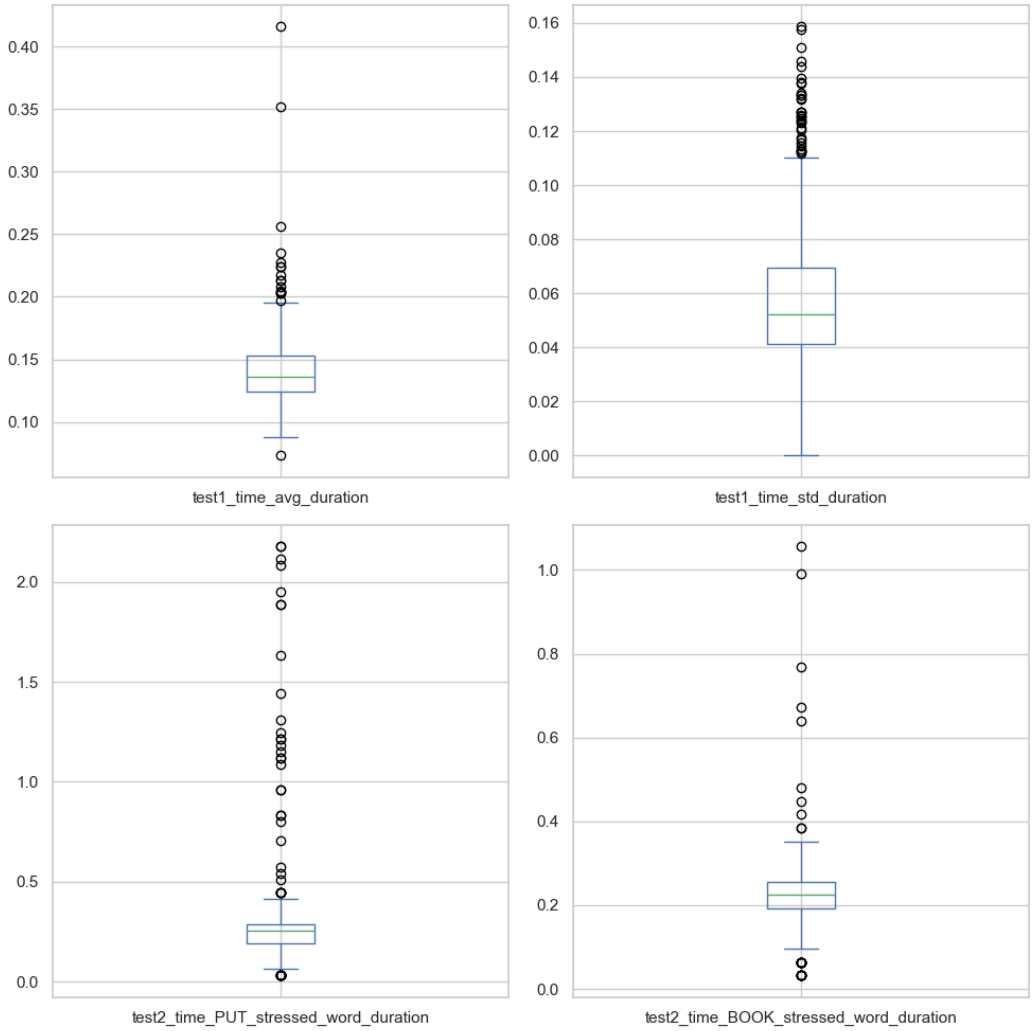

Figure 8: Example Boxplots to show outliers

Winsorization is a technique that transforms extreme values in the data to reduce the effect of outliers (Horsch, 2021). Unlike other methods such as complete removal of outliers or applying robust statistical methods, Winsorization limits the impact of outliers by capping extreme values to a specific percentile value. This ensures that all data points are retained, thereby preserving the dataset's size and structure. By capping the outliers, Winsorization prevents them from having an unjustified influence on the model while maintaining the overall distribution of the data. This method also ensures that important but extreme values are not outrightly discarded, preserving valuable information. Moreover, Winsorization is straightforward to implement and interpret. It involves setting a

threshold (typically using percentiles) beyond which data values are capped. This method also does not make assumptions about the underlying distribution of the data, making it a versatile choice for different types of datasets. In the case of this research, I chose to set the limits at 1% for both tails of the distribution, ensuring that the most extreme 1% of values are adjusted to fall within the 1st and 99th percentiles, respectively. The 1% threshold for Winsorization was chosen based on its application in various studies across different fields. For instance, Hostetler et al. (2020) employed this threshold to develop injury risk curves in the context of motor vehicle crashes, while Waalwijk et al. (2022) utilized it to examine the impact of prehospital time intervals on mortality rates in injured patients. Additionally, Lauharatanahirun et al. (2023) applied this threshold to investigate risk-related brain activation and its association with adolescent health risk behaviors. This approach ensures that the majority of the data remains intact and the effect of extreme values is minimized, leading to more robust and reliable predictive models for concussion detection.

From the `scipy.stats.mstats` module (*Winsorization*, 2021), `winsorize` function was imported to define a function `winsorize_dataframe` that applies Winsorization to each column in a DataFrame. The `limits` parameter specifies the percentiles for Winsorization (1% in this case). The function creates a copy of the DataFrame, applies Winsorization to each column, and returns the modified DataFrame. The `winsorize_dataframe` function is then used to Winsorize the imputed DataFrame `df_imputed`, resulting in the Winsorized DataFrame `df_winsorized`.

3.6.3 Feature Scaling

After handling missing values and outliers, the next step in data preprocessing is feature scaling. Feature scaling adjusts the data to a consistent range (*Feature*

Engineering: Scaling, Normalization, and Standardization, 2023). Due to the the right skewed distribution of many features in concussion_dataset (mentioned in section 3.6.1 Handling Missing Values), normalization (Min-Max Scaling) was chosen. Normalization is particularly useful when the distribution of the data is not known or when it is known to be non-Gaussian. It is beneficial for datasets with varying scales (*Feature Engineering: Scaling, Normalization, and Standardization*, 2023). By scaling the features to a fixed range, usually [0, 1] or [-1, 1], normalization ensures that each feature carries the same weight in the model's predictions, thus improving the model's overall performance and stability. The default range for Min-Max Scaling in Scikit-Learn is [0, 1] (*Feature Engineering: Scaling, Normalization, and Standardization*, 2023).

The `MinMaxScaler` was imported from Scikit-Learn to do feature scaling (*Feature Engineering: Scaling, Normalization, and Standardization*, 2023). It starts by separating the features and the target variable, initializing the `MinMaxScaler` object, fitting and transforming the features, and then converting the scaled features back into a DataFrame. Finally, the target variable is added back to the scaled DataFrame, ensuring that the data is ready for subsequent analysis.

3.6.3 Feature Selection

Feature selection is the process of identifying the most informative subset of features in the dataset to improve model performance and reduce computational complexity (Sanjyal, 2022). This step is crucial for ensuring that the ML models focus on the most relevant data, which enhances their accuracy and efficiency. Additionally, in this research, this helps filter the most significant acoustic features for concussion. Feature selection involved various techniques to evaluate and select the most significant features.

Multicollinearity

Multicollinearity happens when two or more features in the dataset are highly correlated, which can lead to redundancy and instability in the machine learning models (Tate, 2023). To mitigate this issue, Pearson's correlation matrix was employed to identify and remove highly correlated features (Tate, 2023).

Pearson's correlation matrix provides a measure of the linear relationship between pairs of features, with correlation coefficients spanning from -1 (perfect negative correlation) to 1 (perfect positive correlation). A coefficient of 0 signifies no linear correlation. In this analysis, a threshold of 0.7 was established, meaning that any pair of features with an absolute correlation coefficient surpassing 0.7 would be deemed highly correlated.

A threshold of 0.7 is commonly used in feature selection as it strikes a balance between these extremes (Tate, 2023). It allows to identify and remove features that are strongly correlated while retaining those that provide unique information. This threshold ensures that the features remaining in the dataset are sufficiently distinct. The highly correlated features, which has threshold above 0.7, are listed in Table 13. The second feature in each pair is removed, resulting in a reduced DataFrame `df_reduced`.

Table 13: Highly Correlated Features

Feature 1	**Feature 2**	**Correlation Value**
test4_time_std_DDK_period	test4_time_coef_of_var_DDK_period	0.9161
test5_time_std_DDK_period	test5_time_coef_of_var_DDK_period	0.9393
test6_time_std_DDK_period	test6_time_coef_of_var_DDK_period	0.9634
test7_pitch_std_F0	test7_pitch_F0_range	0.9621
test7_pitch_std_F0	test7_pitch_vF0	0.9236
test7_pitch_F0_range	test7_pitch_vF0	0.9514

Recursive Feature Elimination (RFE)

After addressing multicollinearity using Pearson's correlation matrix, RFE was chosen for feature selection. RFE is particularly advantageous because it considers feature interactions by evaluating features in the context of a ML model (Brownlee, 2020a). This is essential for understanding the combined effect of multiple speech biomarkers on concussion diagnosis. Additionally, RFE works by recursively removing the least significant features based on model performance (Brownlee, 2020a), which aligns well with the goal of identifying the most predictive speech biomarkers for concussion diagnosis. This model-based selection process ensures that the final set of features is highly relevant and informative.

While RFE is computationally expensive and requires the model to be retrained multiple times, the benefits of obtaining a robust set of features outweigh these drawbacks, especially in a research setting where accuracy and reliability are

crucial. The computational burden is further mitigated by the prior reduction of the dataset through the handling of multicollinearity. Moreover, RFE's flexibility allows it to be used with various models, such as SVM and MLP, which are planned for use in this research.

In this research, feature selection used RFE with a SVM model from Scikit-Learn library. The SVM model, initialized with a linear kernel, is used as the estimator for RFE. The linear kernel SVM calculates a coefficient for each feature, representing its importance in the classification task. Features with larger absolute coefficient values are considered more important, as they contribute more to the decision boundary that separates the two classes (concussed and non-concussed). The RFE process begins after removing highly correlated features from the dataset. It ranks the remaining features based on their coefficients in the SVM model and eliminates the feature with the smallest absolute coefficient. The SVM model is subsequently retrained using the smaller set of features, and this process is repeated until the target number of features (10 in this instance) is achieved. After the selected features are identified, a new DataFrame is created with these selected features. Finally, the target variable is added back to the DataFrame with the selected features, ensuring that the data is ready for subsequent modelling. Figure 9 shows the 10 selected features after applying RFE.

Feature	Coefficient
test5_amp_coef_of_var_DDK_peak_intensity	0.000097
test2_time_BOOK_stress_pause	-0.000006
test3_time_avg_syllable_duration	-0.000053
test7_pitch_avg_F0	-0.000137
test2_pitch_PUT_f0_rate	-0.000283
test4_amp_std_DDK_peak_intensity	-0.000288
test1_time_std_duration	-0.000290
test4_amp_coef_of_var_DDK_peak_intensity	-0.000474
test2_time_PUT_stressed_word_duration	-0.000517
test2_amp_BOOK_intensity_deviation	-0.000696

Figure 9: Selected Features after applying RFE with SVM model

3.7 Data Split

After data preprocessing and feature selection, the dataset was partitioned into two subsets: 80% for training and 20% for testing. The training set is utilized to optimize the model by adjusting its parameters to minimize errors within the training data itself. In this study, both the SVM and MLP models underwent training using this designated training set. Conversely, the test set functions as an independent dataset to evaluate the final model's performance, simulating real-world scenarios with new, unseen data. By evaluating the model on the test set, its ability to generalize and make accurate predictions can be assessed.

In this study, a validation set was not used because the hyperparameter tuning process was performed using. GridSearchCV is a function within the Scikit-learn library that automates the process of hyperparameter tuning for ML models (Team, 2024). Hyperparameters are parameters that are set prior to the start of the model training process., and finding the optimal combination of these values is crucial for achieving the best model performance. GridSearchCV systematically explores a predefined set of hyperparameter values, training and evaluating the model for each combination. This allows for the identification of the best-performing set of hyperparameters, ultimately leading to a more accurate and effective model.

When optimizing the hyperparameters of a model, like the C value in an SVM, there is a risk of overfitting to the test set if parameters are repeatedly adjusted for optimal performance. This can lead to the model learning the specifics of the test set, making evaluation metrics unreliable for measuring how well the model generalizes to new data. To address this, a validation set is normally used. However, splitting the data into three sets (training, validation, and testing) reduces the amount of data available for training, , and the results can vary depending on the random (*3.1. Cross-validation: evaluating estimator performance, n.d.*). Cross-validation offers a solution (*3.1. Cross-validation: evaluating estimator*

performance, n.d.), and GridSearchCV incorporates K-Fold cross-validation (Shah, 2024). While a test set is still set aside for the final evaluation, the validation set is not required. In K-fold cross-validation (Figure 10), the training set is split into K smaller sets. For each fold, a model is trained on k-1 folds and validated on the remaining fold. The final performance metric is the average of the values from each fold (e.g., AUC-ROC). Although computationally intensive, this method maximizes the use of available data, which is particularly advantageous when dealing with small sample sizes. In this research, a 5-fold cross validation was used for both models SVM and MLP, as that is the default value of GridSearchCV class from Scikit Learn

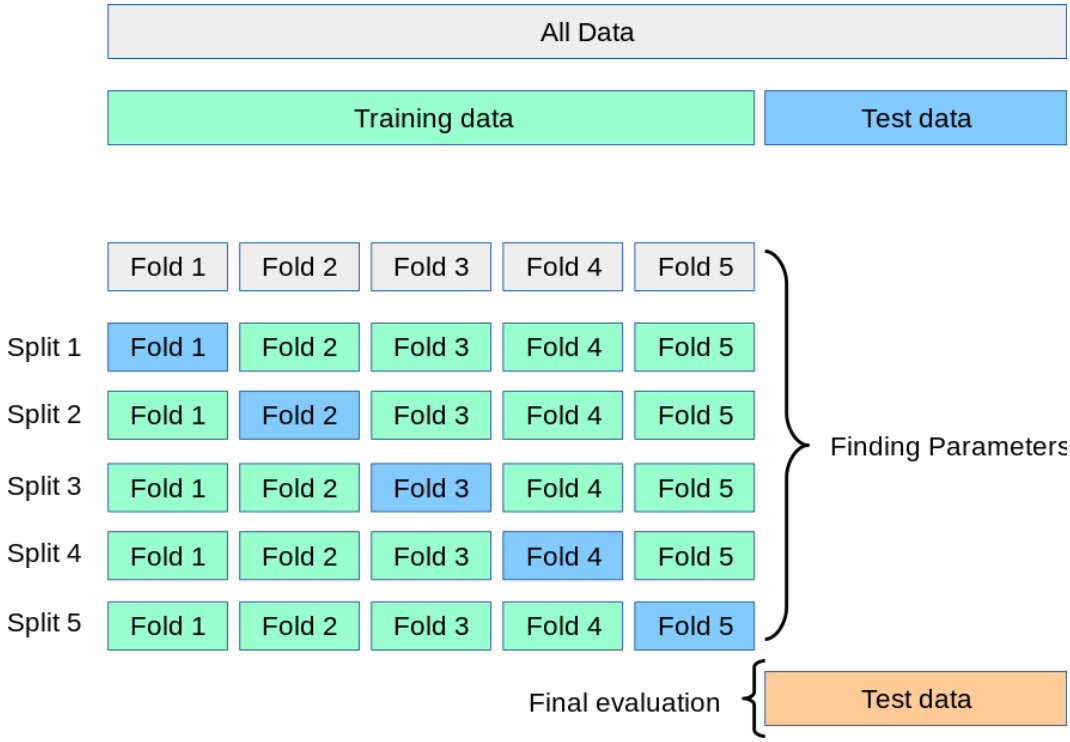

Figure 10: K-Fold Cross-validation (3.1. Cross-validation: evaluating estimator performance, n.d.)

3.7.1 Handling Class Imbalance

After preprocessing the data, it was observed that the dataset is highly imbalanced, with 596 athletes having concussions and only 106 without concussions (Figure 11). Such class imbalance can lead to biased models that are more likely to predict the majority class. To solve this issue, SMOTE (Synthetic Minority Over-sampling Technique) was employed (Basha et al., 2022).

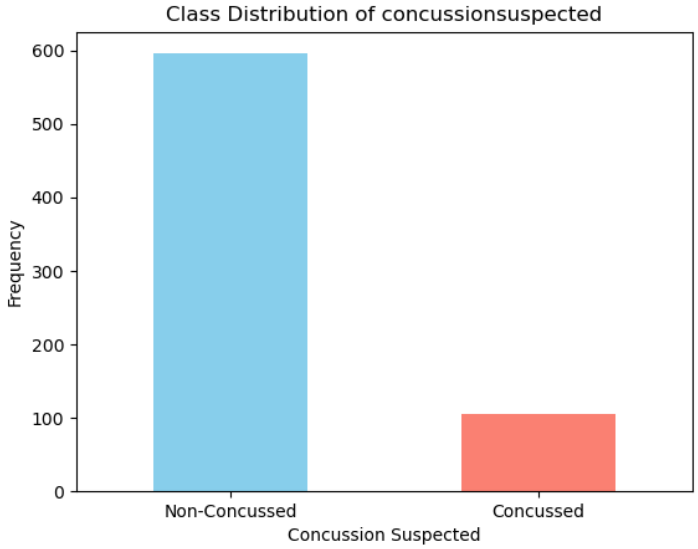

Figure 11: Dataset Balance Checking

SMOTE is used to generate synthetic samples for the minority class, thereby balancing the dataset. This technique improves the model's ability to learn and recognize patterns in both classes, reducing the bias towards the majority class. In the context of concussion detection, addressing class imbalance is critical as it ensures the model does not overlook concussed individuals, which is essential for accurate diagnosis and treatment. SMOTE technique works by creating synthetic examples rather than by over-sampling with replacement. It creates a new instance in the feature space by selecting two or more similar instances from the minority class and interpolating between them. This approach helps in generating more informative synthetic instances and improves the model's robustness.

In this work, handling class imbalance used the `SMOTE` method from the

imbalanced-learn library (*Imbalanced-Learn module in Python*, 2020). It starts by separating the features and the target variable from the DataFrame. The `SMOTE` object was then initialized and applied to this dataset, generating synthetic samples for the non-concussed class until the number of instances in both classes was equal. As depicted in Figure 12, the application of SMOTE successfully balanced the class distribution, resulting in 596 instances for each class.

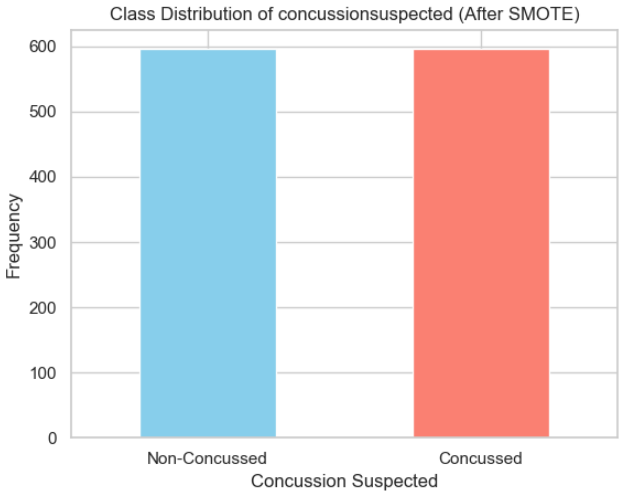

Figure 12: Dataset after applying SMOTE

This balanced dataset was then split into training and testing sets, maintaining an 80:20 ratio while ensuring that the balanced class distribution was preserved across both sets. This approach guarantees that the data is appropriately split for model training, and evaluation, maintaining a balanced representation of both classes across all sets.

3.8 Model Evaluation Criteria

As presented in Section 2.8 Evaluation Metrics, various metrics can be employed to evaluate the efficacy of a classification algorithm, including Accuracy, Confusion Matrix, Precision, Recall, F1 Score, and AUC-ROC. While all these metrics will be utilized in the evaluation process, the primary criterion for assessing the model's performance will be the AUC-ROC, with the remaining

metrics serving as supplementary references.

In the context of concussion diagnosis, the cost of misclassification can have significant consequences. Specifically, we are more concerned with identifying people who have suffered a concussion (true positives) rather than incorrectly classifying healthy athletes as concussed (false positives). This focus cannot be adequately measured by Accuracy alone.

The ROC curve illustrates the trade-off between the false positive rate (FPR) and the true positive rate (TPR). It essentially represents the trade-off between benefits and costs. The goal is to maximize TPR (sensitivity) while minimizing FPR (1-specificity), reflecting a balance where the classifier correctly identifies concussed athletes without misclassifying healthy ones. The point on the ROC curve where the true positive rate (TPR) is 1 and the false positive rate (FPR) is 0 represents a perfect classification, signifying the best possible model performance. Therefore, the AUC-ROC is the primary evaluation metric for our models. A higher AUC-ROC indicates a better-performing classifier.

Chapter 4 - Model Implementation

This chapter primarily presents the implementation of SVM and MLP models, along with the experimental procedures. The goal is to use the preprocessed dataset with 10 selected features to predict whether a person has concussed or not. First, the implementation of the SVM model using `sklearn.svm` is described. Then, the MLP implementation using `sklearn.neural_network` is presented.

4.1 Implementation of SVM

In this research, the SVM model was developed using the Scikit-Learn library in Python, a widely used ML framework.

4.1.1 Model Building and Hyperparameter Tuning

The SVM model was constructed using the `SVC` class from the `sklearn.svm` module in Scikit-Learn. This class provides a versatile interface for implementing SVM classifiers with different kernel functions and hyperparameters. The `SVC` class is specifically designed for classification tasks and offers various kernel functions to handle different types of data relationships.

```
from sklearn.svm import SVC
```

To enhance the SVM model's performance, a grid search with cross-validation was utilized to systematically explore various hyperparameter combinations and identify the optimal configuration for achieving the best model performance. The hyperparameters evaluated in this grid search included the regularization parameter (`C`), the type of kernel function (`kernel`), and the kernel

coefficient for the RBF kernel (`gamma`).

```
param_grid = {
    'C': [0.1, 1, 10, 100],
    'kernel': ['linear', 'rbf'],
    'gamma': ['scale', 'auto']
}
```

The regularization parameter controls the trade-off between maximizing the margin (the separation between the decision boundary and the closest data points of each class) and minimizing classification errors. A larger `C` value allows for a more complex model that may fit the training data more closely but could lead to overfitting. Conversely, a smaller `C` value encourages a simpler model with a wider margin, potentially improving generalization but risking underfitting. In this study, the grid search explored `C` values of 0.1, 1, 10, and 100.

The `kernel` function is employed to transform the input data into a higher-dimensional space, where it may be simpler to distinguish between the classes. In this study, both linear and radial basis function (RBF) kernels were considered. The linear kernel is appropriate for data that is linearly separable, while the RBF kernel can deal with non-linear relationships between features (Hue, 2019).

The `gamma` parameter determines how much impact a single training instance has on the shape of the decision boundary. A low gamma value results in a wider influence, leading to a smoother decision boundary, while a high gamma value results in a narrower influence, leading to a more complex decision boundary. In this study, the grid search explored 'scale' and 'auto' options for gamma. From Scikit-Learn library for `SVC`, '`scale`' sets gamma $\frac{1}{\text{n\_features} * \text{X.var()}}$, where n_features refers to the number of features in dataset and X.var() calculates the variance of the features in dataset

, and 'auto' sets gamma as $\frac{1}{n\_features}$.

```
svm_model = SVC(probability=True)

grid_search = GridSearchCV(estimator=svm_model,
param_grid=param_grid, scoring='roc_auc', cv=5,
n_jobs=-1)
```

In the above code, a SVM model is initialized with the `probability=True`, enabling the model to estimate class probabilities, which are essential for generating ROC curves and calculating AUC-ROC scores. The `param_grid` dictionary defines the range of values for each hyperparameter to be explored. The `GridSearchCV` class performs an exhaustive search over this grid, evaluating each combination of hyperparameters using 5-fold cross-validation (`cv=5`). The scoring parameter is set to `'roc_auc'`, indicating that the area under the receiver operating characteristic curve will be used as the metric to evaluate model performance. The `n_jobs=-1` argument allows the grid search to utilize all available processors for parallel computation, speeding up the process.

4.1.2 Model Training and Validation

The grid search is then trained on the training set to find the optimal hyperparameters:

```
grid_search.fit(X_train, y_train)
```

After training, the best estimator, representing the SVM model with the optimal hyperparameters, is retrieved. Figure 13 shows the results of best parameters for SVM model.

```
best_svm_model = grid_search.best_estimator_
```

The model was then tested on the test set to evaluate its generalization capability:

```
y_test_pred_svm = best_svm_model.predict(X_test)

y_test_prob_svm =
best_svm_model.predict_proba(X_test)[:, 1]
```

4.1.3 Result

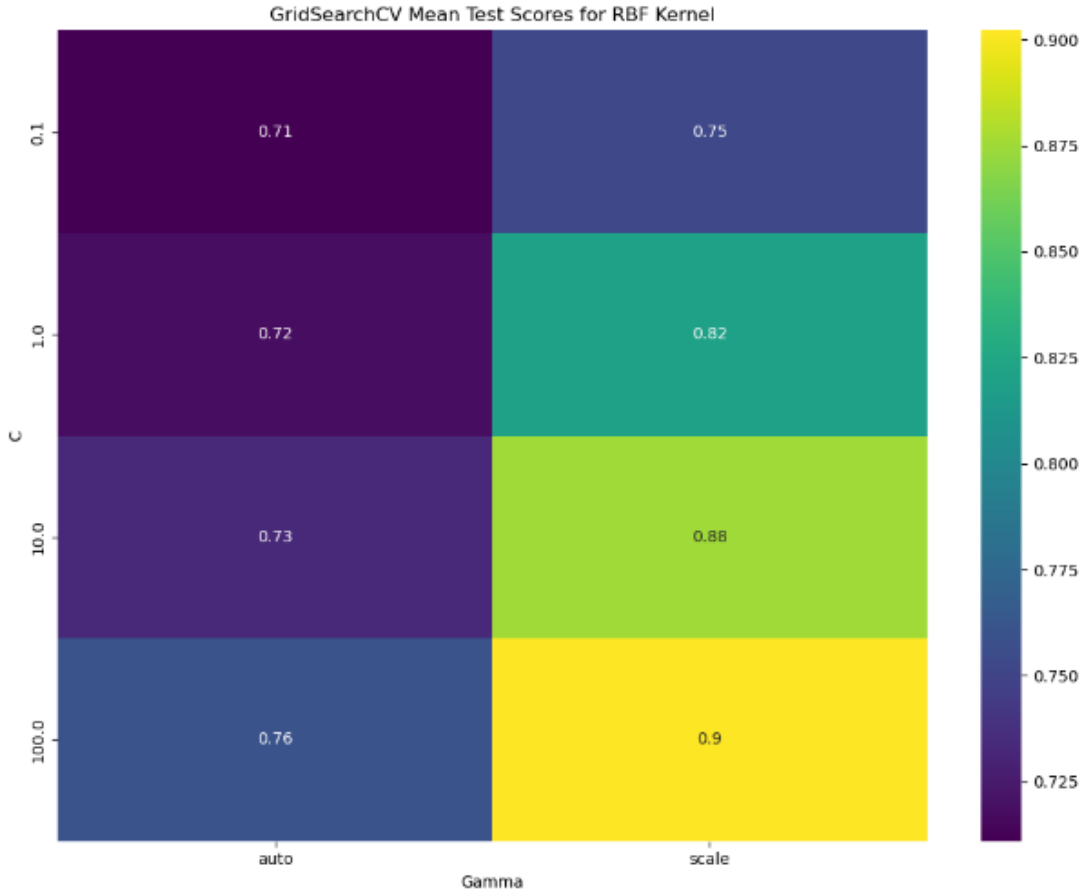

Figure 13: Using GridSearchCV to select the best parameters for SVM

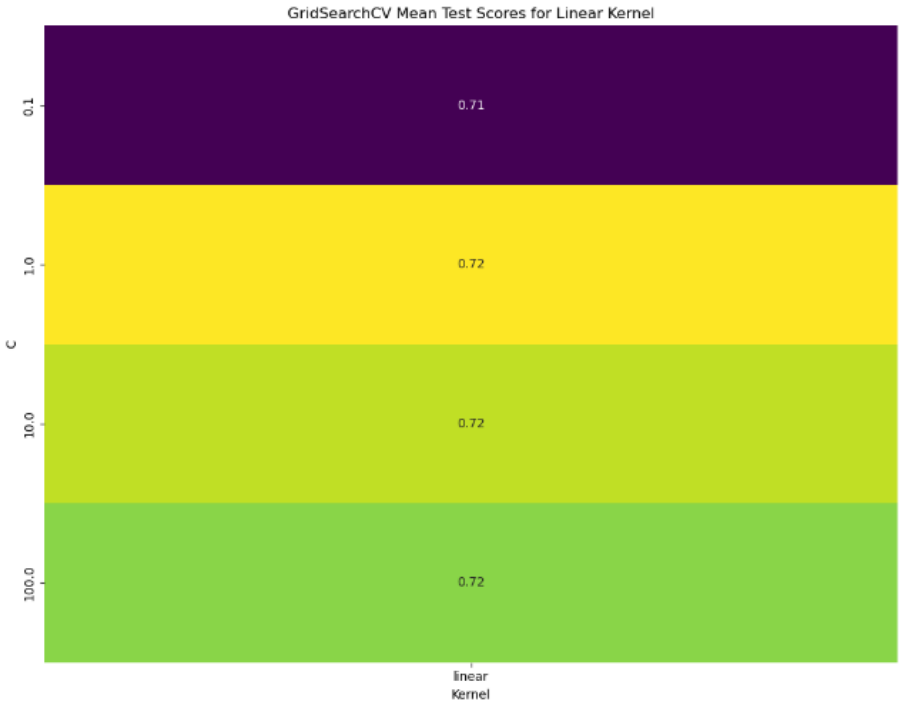

Figure 13 (cont.): Using GridSearchCV to select the best parameters for SVM

It can be seen from Figure 13 SEQ Figure * ARABIC that the best parameters for SVM are: `C` as 100, `kernel` as RBF and `gamma` as scale. And with the best parameters, the best SVM model was used to predict on test set and the result are listed in Table 14. The accuracy is 0.8075, the precision is 0.7448, the recall is 0.9231, the F1-score is 0.8244, and AUC-ROC is 0.8977. ROC curve for SVM is presented in Figure 14.

Table 14: Evaluation metrics result of SVM

Evaluation Metric	Result
Accuracy	0.8075
Precision	0.7448
Recall	0.9231
F1 Score	0.8244
AUC-ROC	0.8977

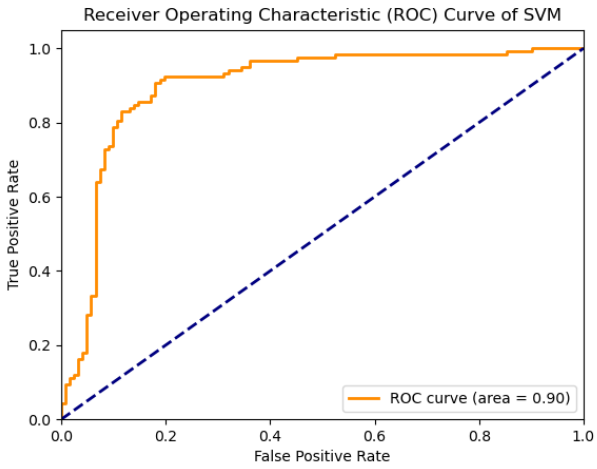

Figure 14: ROC curve for SVM

4.2 Implementation of MLP

In this research, a MLP model was implemented using the Scikit-Learn library.

4.2.1 Model Building and Hyperparameter Tuning

The MLP model was made using the `MLPClassifier` class from the `sklearn.neural_network` module in Scikit-Learn. This class provides a convenient way to build and train MLP models with various architectures and hyperparameters.

```
from sklearn.neural_network import MLPClassifier
```

Similar to the SVM model, a grid search with cross-validation was employed to optimize the MLP model's performance. This approach systematically explored different combinations of hyperparameters to identify the configuration that yielded the highest AUC-ROC score. The hyperparameters considered in this grid search included the number of neurons in each hidden layer (`hidden_layer_sizes`), the activation function used in the hidden layers (`activation`), the optimization algorithm for weight updates during training

(`solver`), the L2 regularization parameter (`alpha`), the learning rate for weight updates (`learning_rate`), and the maximum number of iterations for the solver to converge (`max_iter`).

```
param_grid = {
    'hidden_layer_sizes': [(50,), (100,), (50, 50),
(100, 100)],
    'activation': ['tanh', 'relu'],
    'solver': ['sgd', 'adam'],
    'alpha': [0.0001, 0.05],
    'learning_rate': ['constant', 'adaptive'],
    'max_iter': [5000]
}
```

The `hidden_layer_sizes` parameter defines the architecture of the hidden layers in the MLP. Different combinations of hidden layer sizes were explored to find the optimal network architecture that balances complexity and generalization. In this study, the grid search considered configurations with one or two hidden layers, with 50 or 100 neurons in each layer.

The `activation` function enables the network to model non-linear relationships within the data, facilitating the learning of intricate patterns. In this study, both the hyperbolic tangent ('`tanh`') and rectified linear unit ('`relu`') activation functions were considered. The '`tanh`' function is a smooth, S-shaped function that maps inputs to the range [-1, 1], while the '`relu`' function is a piecewise linear function that returns 0 if the input is negative and returns the input itself for positive inputs (Brownlee, 2021).

The `solver` parameter specifies the method used to adjust the model's weights during training. In this study, both '`sgd`' (stochastic gradient descent) and '`adam`' optimizers were considered. SGD is a simple and widely used optimization

algorithm (Stojiljković, n.d.). It updates the model's parameters iteratively based on the gradients of the loss function calculated on small batches of training data. While SGD is simple and computationally efficient, it can sometimes struggle to converge to the optimal solution, especially when the loss function has a complex landscape (Stojiljković, n.d.). While Adam is a more advanced optimization algorithm. Adam dynamically adjusts the learning rate for each parameter by utilizing estimations of the gradients' first and second moments. This adaptive learning rate mechanism enables Adam to achieve faster and potentially more reliable convergence compared to SGD, particularly in scenarios with sparse gradients or gradients of varying magnitudes (Park, 2021).

The `alpha` controls the strength of the L2 regularization, which mitigates overfitting by introducing a penalty term to the loss function that limits excessively large weights. L2 regularization, also known as ridge regression, reduces model complexity by adding a penalty term to the loss function that is proportional to the sum of the squares of the model's coefficients (Tewari, 2021). This the model to learn less complex relationships, preventing overfitting to the training data and enhancing its capacity to generalize to new, unseen data. In this study, the grid search explored alpha values of 0.0001 and 0.05.

This `learning_rate` determines how the learning rate (the step size at which the optimizer updates the weights) changes during training. In this study, both 'constant' and 'adaptive' learning rate schedules were considered. A constant learning rate remains the same throughout training, while an adaptive learning rate decreases over time, potentially leading to better convergence.

The `max_iter` parameter sets the maximum number of iterations (epochs) that the solver will run to try to find the optimal weights. In this study, the maximum number of iterations was set to 5000.

```
mlp_model = MLPClassifier()

grid_search_mlp = GridSearchCV(estimator=mlp_model,
param_grid=param_grid, scoring='roc_auc', cv=5,
n_jobs=-1)
```

In the above code, an MLP model is initialized using the `MLPClassifier` class. The `param_grid` dictionary defines the range of values for each hyperparameter to be explored. The GridSearchCV class performs an exhaustive search over this grid, evaluating each combination of hyperparameters using 5-fold cross-validation (`cv=5`). The scoring parameter is set to `'roc_auc'`, indicating that the area under the receiver operating characteristic curve (AUC-ROC) will be used as the metric to evaluate model performance. The `n_jobs=-1` argument allows the grid search to utilize all available processors for parallel computation, speeding up the process.

4.2.2 Model Training and Validation

Similar to SVM model, the grid search is then trained on the training set to find the optimal hyperparameters:

```
grid_search.fit(X_train, y_train)
```

After training, the best estimator, representing the MLP model with the optimal hyperparameters, is retrieved. **Error! Reference source not found.** shows the results of best parameters for MLP model.

```
best_mlp_model = grid_search.best_estimator_
```

The best MLP model was then tested on the test set to evaluate its generalization capability:

```
y_test_pred_svm = best_mlp_model.predict(X_test)

y_test_prob_svm =
best_mlp_model.predict_proba(X_test)[:, 1]
```

4.3.3 Result

The optimal hyperparameters identified by the grid search were 'relu' activation, an alpha value of 0.0001, two hidden layers with 100 neurons each, a constant learning rate, and the 'adam' solver. With these parameters, the MLP model achieved an accuracy of 0.8159, precision of 0.7920, recall of 0.8462, F1 score of 0.8182, and an AUC-ROC of 0.8845 on the test set (Table 15). The ROC curve for the MLP model is depicted in Figure 16.

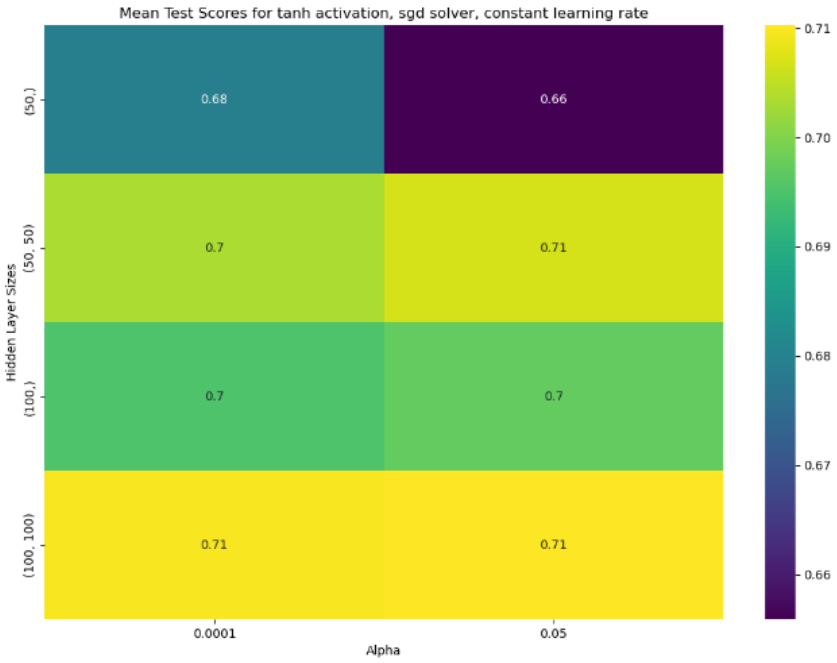

Figure 15: Using GridsearchCV to select the best parameters for MLP model

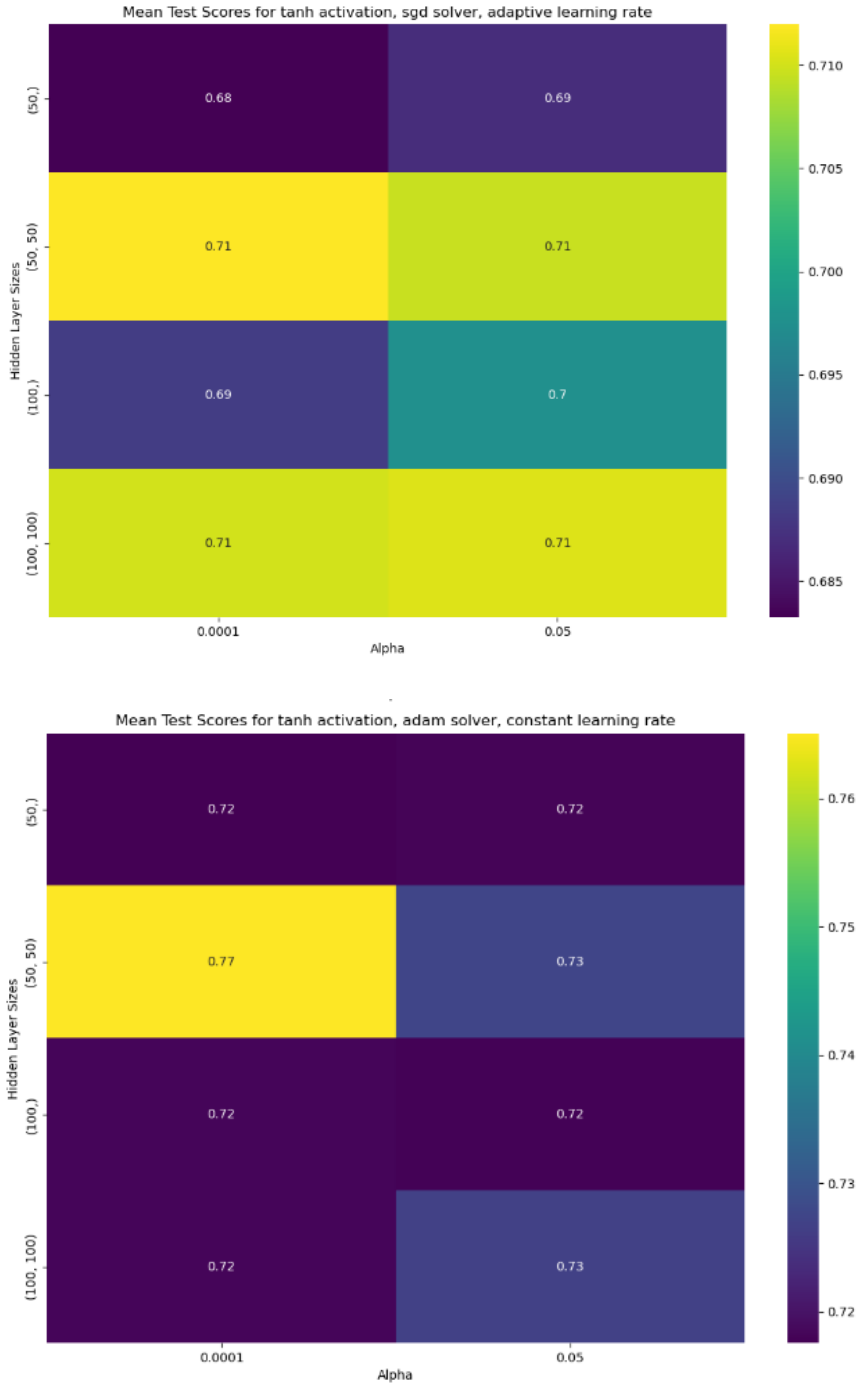

Figure 15 (cont.): Using GridsearchCV to select the best parameters for MLP model

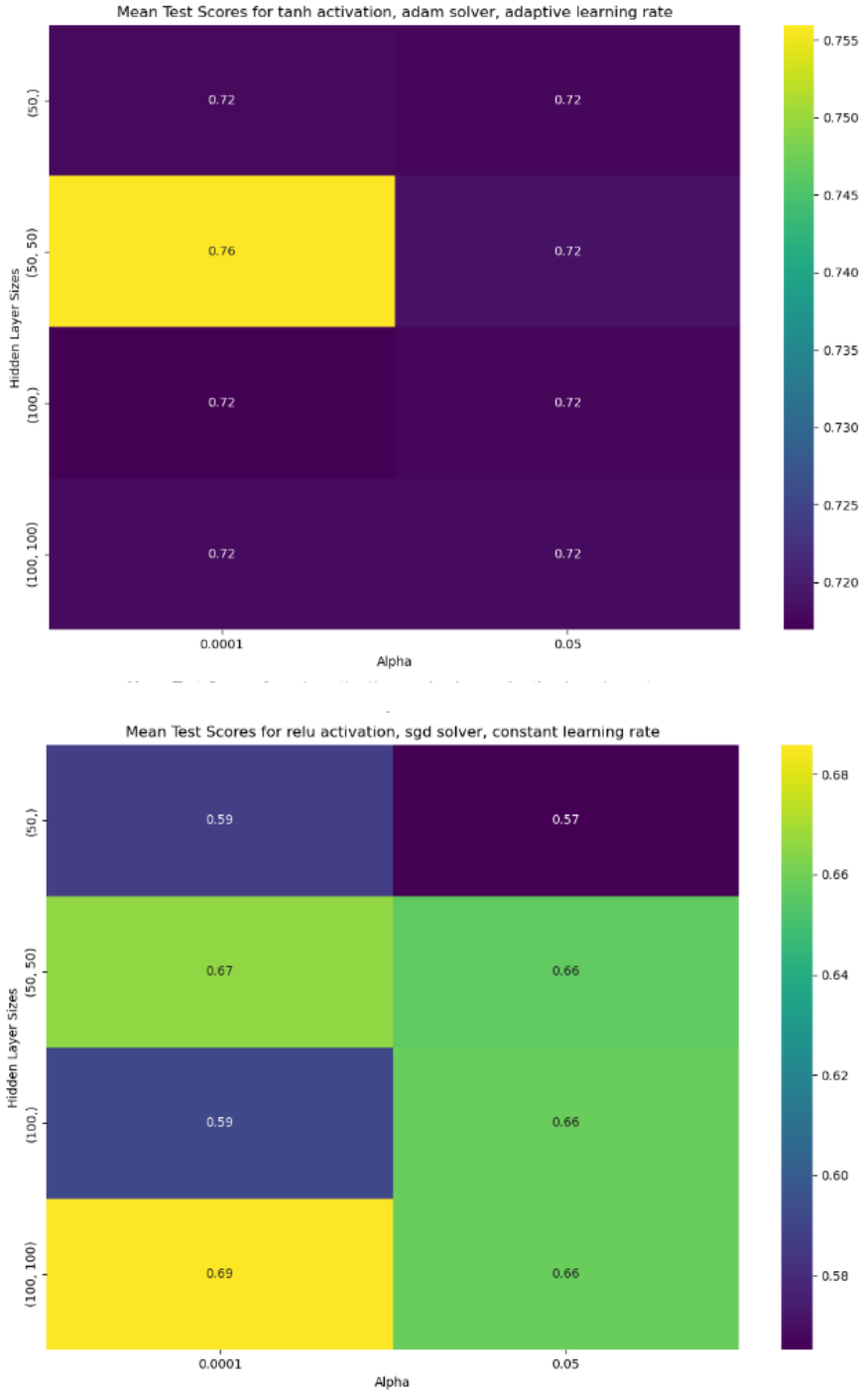

Figure 15 (cont.): Using GridsearchCV to select the best parameters for MLP model

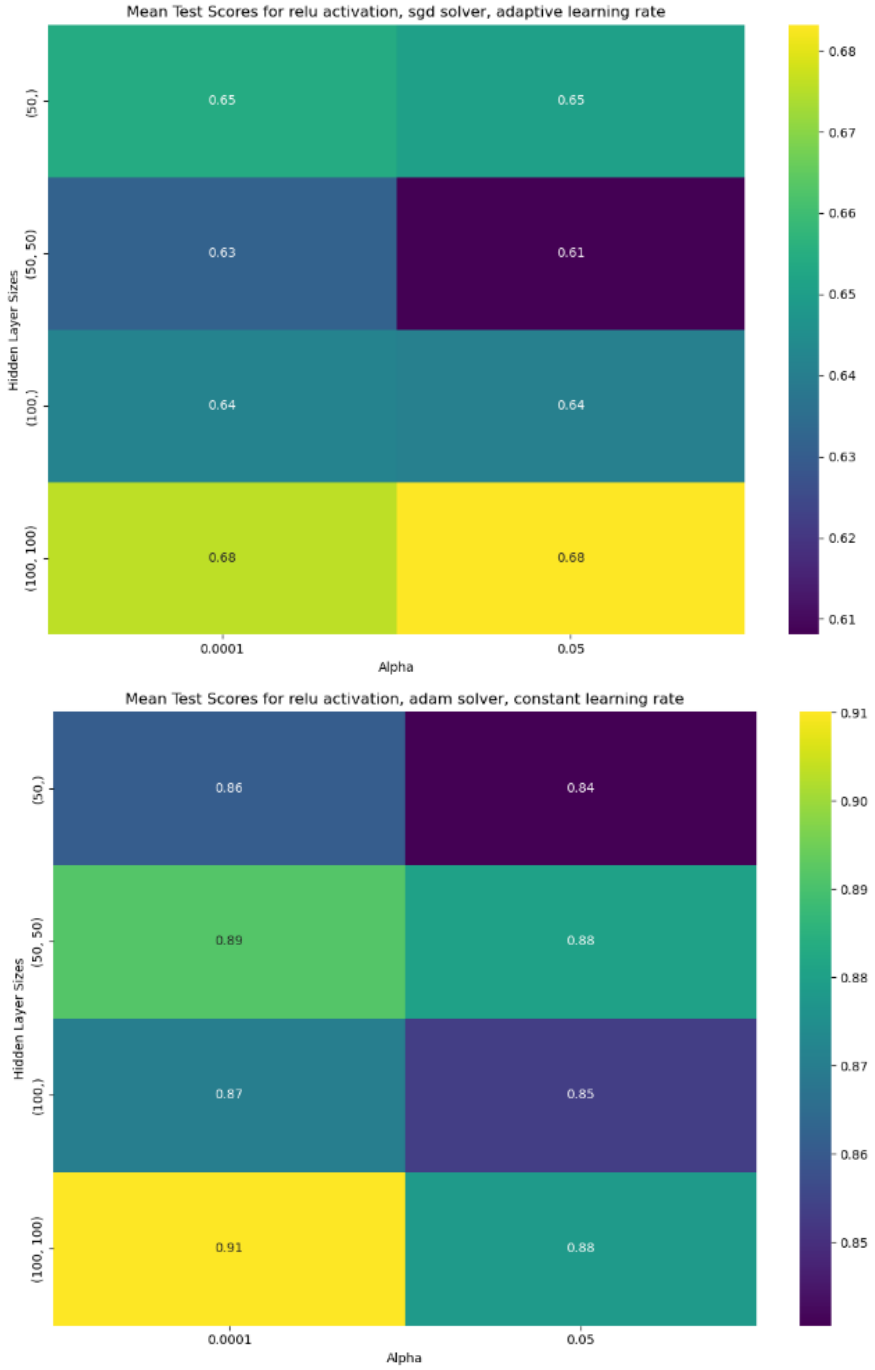

Figure 15 (cont.): Using GridsearchCV to select the best parameters for MLP model

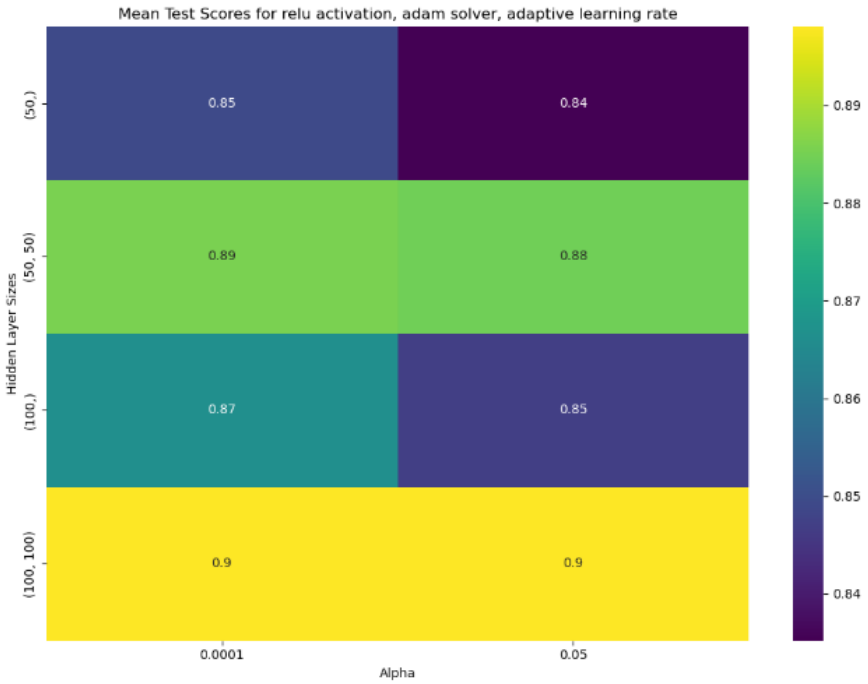

Figure 15 (cont.): Using GridsearchCV to select the best parameters for MLP model

Table 15: Evaluation metrics result of MLP

Evaluation Metric	Result
Accuracy	0.8159
Precision	0.7920
Recall	0.8462
F1 Score	0.8182
AUC-ROC	0.8845

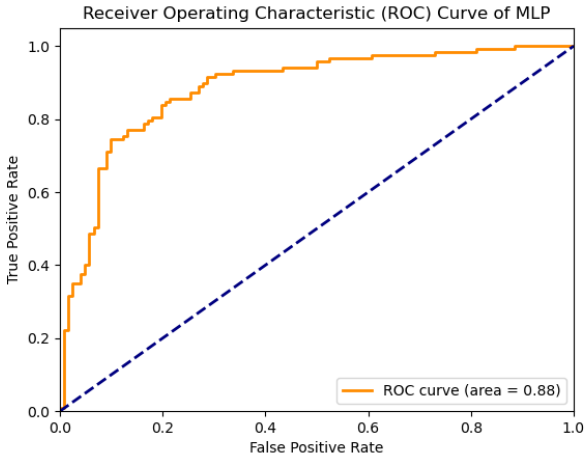

Figure 16: ROC curve for MLP

Chapter 5 - Discussion

This chapter first compares two models of SVM, and MLP to see the performance of prediction of concussion on the test data set. Next, it interprets the results in depth along with 10 selected acoustic features out of 41 extracted features. Limitations of this research are also presented here in this chapter.

5.1 Model Performance Comparison

The performance of the SVM and MLP models in predicting concussion using speech data was compared. Both models were trained and evaluated on the same preprocessed dataset with 10 selected acoustic features. Table 16 shows all the performance index values of SVM and MLP models. However, in Chapter 3, the AUC-ROC was selected as the primary metric for evaluating model performance. Therefore, the ROC curves for both models are illustrated in Figure 17.

Table 16: Comparison of results for two models

	SVM	**MLP**
Accuracy	0.8075	0.8159
Precision	0.7448	0.7920
Recall	0.9231	0.8462
F1-score	0.8244	0.8182
AUC-ROC	0.8977	0.8845

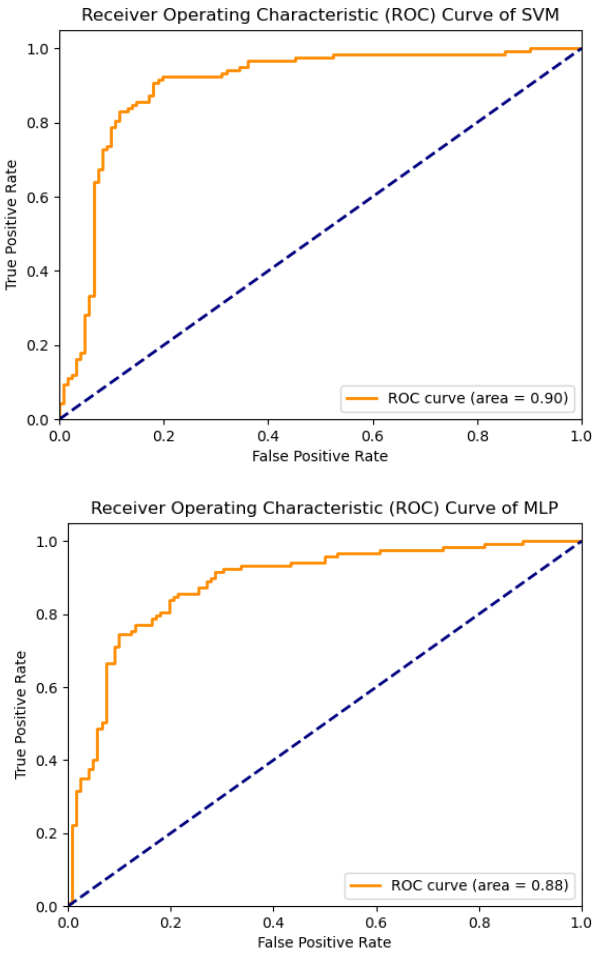

Figure 17: ROC curve comparison

Both the SVM and MLP models demonstrated high accuracy in classifying individuals with and without concussions, exceeding 80%. The MLP model slightly outperformed the SVM model in overall accuracy (81.59% vs. 80.75%), and was more precise in identifying true positive cases, with fewer false positives (79.20% vs. 74.48%). However, the SVM model excelled in recall, capturing a higher proportion of actual positive cases with fewer false negatives (92.31% vs. 84.62%). The F1-scores, indicating the balance between precision and recall, were comparable for both models (82.44% for SVM and 81.82% for MLP). The SVM model achieved a marginally higher AUC-ROC score (0.8977 vs. 0.8845), suggesting a slightly better overall ability to discriminate between the two classes.

5.2 Interpretation of Results

To enhance the models' performance and interpretability, a rigorous feature selection process was undertaken. Initially, the dataset contained 41 acoustic features extracted from the speech recordings. However, many of these features exhibited high correlations, a phenomenon known as multicollinearity, which can negatively impact model stability and interpretability. To address this, Pearson's correlation matrix was employed to identify and eliminate highly correlated features. A correlation threshold of 0.7 was set, and one feature from each pair exceeding this threshold was removed, resulting in a reduced set of features. Following the removal of highly correlated features, RFE was applied to further refine the feature set. RFE is a wrapper-style feature selection method that recursively eliminates the least important features based on their impact on model performance. In this study, RFE was used in conjunction with a linear SVM model to select the 10 most informative features. These selected features are listed in Figure 9.

The selected features encompass a mix of temporal and frequency acoustic metrics, reflecting various aspects of speech production that are potentially affected by concussion. Notably, these features were derived from specific speech tasks (Table 10). For instance, `test1_time_std_duration`, a measure of timing variability in multisyllabic word production (Test 1), was identified as a key feature. This suggests that alterations in the consistency of word pronunciation may be indicative of concussion. Similarly, features such as `test2_time_PUT_stressed_word_duration` and `test2_time_BOOK_stress_pause`, which assess stress and pause durations in sentence repetition tasks (Test 2), highlight the importance of prosodic elements in concussion detection. Changes in the timing and rhythm of speech, as captured by these features, could reflect underlying neurological disruptions.

Frequency-based features also emerged as significant contributors to the models' predictive power. `test2_pitch_PUT_f0_rate`, a measure of pitch variation during sentence repetition (Test 2), and `test7_pitch_avg_F0`, representing the average fundamental frequency during sustained vowel phonation (Test 7), underscore the relevance of pitch-related characteristics in identifying concussions. Alterations in pitch control and vocal stability may be indicative of neurological impairment. Additionally, `test2_amp_BOOK_intensity_deviation`, which quantifies variations in intensity during sentence repetition (Test 2), suggests that changes in vocal effort or loudness could also be associated with concussion.

The inclusion of features related to DDK tasks (Tests 4 and 5), such as `test4_amp_std_DDK_peak_intensity` and `test5_amp_coef_of_var_DDK_peak_intensity`, further emphasizes the importance of speech motor control and coordination in concussion assessment. These features capture the variability and consistency of amplitude during rapid syllable repetition, potentially revealing subtle deficits in motor function that may not be apparent in other speech tasks. Notably, Test 6, also a DDK task, did not give any selected features, suggesting that the specific syllable sequence used in this test might be less sensitive to concussion-related speech changes compared to the sequences used in Tests 4 and 5.

In this research, two ML models, SVM and MLP, were designed to predict the possibility of concussion based on speech data. The dataset, derived from a study by Yadav (2015), consisting 702 unique athletes, is considered substantial within the context of medical research, especially for a condition like concussion that often lacks large-scale datasets.

As shown in Table 16, both models demonstrated promising results in concussion prediction. The SVM model exhibited a marginally superior overall

discriminative ability, as evidenced by its slightly higher AUC-ROC score of 89.77% compared to the MLP's 88.45%. As discussed in Chapter 3, the AUC-ROC is a crucial metric in this context, as it represents the model's ability to distinguish between individuals with and without concussions across all possible classification thresholds. A higher AUC-ROC score suggests the model's enhanced capability to differentiate between these two groups, irrespective of the chosen diagnostic threshold.

In this study, the emphasis on AUC-ROC as the primary evaluation metric stems from the importance of minimizing both false positives (incorrectly classifying healthy individuals as concussed) and false negatives (incorrectly classifying concussed individuals as healthy). The AUC-ROC encapsulates this trade-off, providing a comprehensive measure of the model's diagnostic capability. While the MLP model shows a slightly higher accuracy (81.59% vs. 80.75%), its lower AUC-ROC score suggests that the SVM model might be more reliable across a wider range of decision thresholds. This is particularly relevant in a clinical setting, where the choice of threshold can significantly impact the number of individuals correctly or incorrectly diagnosed with a concussion. Furthermore, the SVM model's higherr recall (92.31% vs. 84.62%) underscores its potential for concussion diagnosis. A high recall indicates that the model is effective at identifying a large proportion of true concussion cases, which is crucial for ensuring that individuals receive appropriate medical attention. Although the MLP model demonstrates higher precision (79.20% vs. 74.48%), its lower recall suggests that it might miss some concussion cases, which could have negative consequences for the affected individuals.

To summarize, the results highlight the potential of both SVM and MLP models for concussion classification based on speech analysis. However, the SVM model's marginally higher AUC-ROC score and superior recall suggest that it

might be a more reliable and effective tool for concussion diagnosis, particularly in scenarios where identifying all potential cases is of paramount importance.

This research has some advantages and limitations, which are discussed as follow.

5.2.1 Strengths

The rigorous feature selection process, involving the removal of highly correlated features and the use of RFE, ensured that the most informative and relevant acoustic features were used for model training. This not only improved the models' performance but also enhanced their interpretability by focusing on the most remarkable speech characteristics associated with concussion. The use of a variety of evaluation metrics, including accuracy, precision, recall, F1-score, and especially AUC-ROC, provided a comprehensive assessment of the models' performance. This allowed for a nuanced understanding of the strengths and weaknesses of each model, facilitating informed decision-making in clinical applications. The study utilized a dataset collected from a real-world population of student-athletes, enhancing the potential applicability of the findings to clinical settings. The use of standardized speech tasks and recording procedures further strengthens the study's relevance to practical concussion diagnosis.

5.2.2 Limitations

In this research, a proof of concept of AI application for speech analysis is developed to support mTBI diagnosis and rehabilitation. However, some limitations should be acknowledged. The dataset used in this study was collected from a specific population of student-athletes, which restricts the generalizability of the findings to other populations. The inclusion of a more diverse population, beyond student-athletes, could potentially enhance the generalizability and applicability of the results across different demographic groups.

This study primarily focused on acoustic features, neglecting other potentially relevant aspects of speech such as linguistic features. Integrating linguistic features and analysis of the spoken language and the content of speech could potentially take into consideration future research to compare and contrast the accuracy of the outcome.

Furthermore, The study did not include a longitudinal component, preventing the assessment of the long-term effects of concussion on speech patterns and the potential for tracking recovery trajectories. The data collection was conducted in a controlled setting, which may not fully reflect the real-world conditions in which concussions occur and are diagnosed. Finally, it is important to note that the ethical considerations related to the use of AI in healthcare for concussion, particularly concerning data privacy and informed consent, were beyond the scope of this thesis.

Chapter 6 - Conclusion

This chapter summarizes the research findings, offering insights into potential future directions and proposing avenues for further exploration in this domain.

6.1 Conclusion

TBI, particularly mTBI or concussion, is a prevalent neurological condition with significant public health implications. Accurate and timely diagnosis is crucial for effective management and prevention of long-term complications. However, current diagnostic methods, including clinical assessments and neuroimaging techniques, have limitations in terms of subjectivity, sensitivity, and accessibility.

This thesis explored the potential of speech analysis as a non-invasive and accessible tool for concussion diagnosis. The feature selection process highlighted the importance of specific speech tasks and acoustic features in concussion detection. Temporal features related to word and syllable durations, pauses, and DDK rates, along with frequency features related to pitch and intensity variations, emerged as key indicators of concussion. These findings suggest that subtle changes in speech patterns, often imperceptible to the human ear, can be detected and analysed using ML algorithms to aid in concussion diagnosis.

By analyzing acoustic features extracted from speech recordings, two machine learning models, SVM and MLP, were developed to differentiate between concussed and non-concussed individuals. Both models demonstrated promising results, with high AUC-ROC values indicating their strong discriminatory power. While the MLP model exhibited slightly higher accuracy (81.59% vs. 80.75%) and precision (79.20% vs. 74.48%), the SVM model achieved a marginally higher

AUC-ROC (89.77% vs. 88.45%) and recall (92.31% vs. 84.62%). As discussed in Chapter 3, the AUC-ROC is the primary evaluation metric in this study, as it represents the model's ability to distinguish between individuals with and without concussions across all possible classification thresholds. Although the differences between the two models are small, the SVM model's higher AUC-ROC suggests a slightly better overall ability to distinguish between the two classes, making it potentially more reliable across a wider range of diagnostic thresholds.

6.2 Future work

The findings of this research open up several opportunities for future investigation. For future work, both the dataset and the model need to be improved:

To enhance the generalizability of the models, future studies should include more diverse populations, encompassing individuals from various age groups, genders, socioeconomic backgrounds, and cultural contexts. This would ensure that the models are robust and applicable across a wider range of individuals, improving their clinical utility.

Incorporating linguistic features, such as lexical diversity and syntactic complexity, could further enhance the models' predictive power and provide a more comprehensive understanding of the cognitive and linguistic impairments associated with concussion. This could lead to the development of more sophisticated diagnostic tools that consider both acoustic and linguistic aspects of speech.

Longitudinal studies are essential to track changes in speech patterns over time and assess the long-term effects of concussion. By collecting speech data at multiple time points following a concussive event, researchers can gain insights into the trajectory of recovery and evaluate the effectiveness of interventions. This information could be invaluable for developing personalized treatment plans and

monitoring the long-term neurological health of individuals with concussions.

Future research should also explore the potential for integrating speech analysis with traditional diagnostic methods, such as neuroimaging and clinical assessments. Combining these established techniques with AI-driven speech analysis could create a more comprehensive diagnostic framework for mTBI, enhancing the accuracy of diagnoses and providing a holistic view of the patient's neurological condition.

While this thesis primarily focuses on ML models, future work should extend these efforts by developing and incorporating more advanced technologies such as Deep Learning, Generative AI, or Video Transformer. These cutting-edge techniques could be particularly effective in identifying complex patterns in speech data, leading to more sophisticated and precise diagnostic tools.

Moreover, exploring the feasibility of collecting speech data in more naturalistic settings, such as on the sidelines of sporting events or in emergency departments, would be crucial for validating the models' performance in real-world scenarios. The development of portable, user-friendly speech analysis tools could revolutionize concussion diagnosis, making it more accessible and efficient in various settings. These tools, integrated into CDSS, could provide healthcare professionals with objective, data-driven insights to complement their clinical judgment, leading to more accurate and timely diagnoses of concussions. By leveraging the power of AI and speech analysis, we can strive towards a future where concussion diagnosis is not only more accurate and efficient but also readily available to individuals in diverse settings, ultimately improving patient care and reducing the burden of this widespread injury.

Finally, future work should also address the ethical implications of applying AI in healthcare for mTBI. The focus should be given to issues such as data privacy, informed consent, and the potential biases in AI models. Ensuring that

these ethical considerations are thoroughly explored and addressed will be crucial for the responsible development and deployment of AI-driven diagnostic tools.

References

3.1. Cross-validation: evaluating estimator performance. (n.d.). Scikit Learn. https://scikit-learn.org/stable/modules/cross_validation.html

Accident Compensation Corporation. (n.d.). *Reducing traumatic brain injuries (TBI)*. Accident Compensation Corporation. .

Activation functions in Neural Networks. (2024). GeeksforGeeks. https://www.geeksforgeeks.org/activation-functions-neural-networks/

Agarwal, N., Thakkar, R., & Than, K. (n.d.). *Concussion*. American Association of Neurological Surgeons. https://www.aans.org/en/Patients/Neurosurgical-Conditions-and-Treatments/Concussion

Ahad, A., Fayyaz, A., & Mehmood, T. (2002, 16-17 Aug. 2002). Speech recognition using multilayer perceptron. IEEE Students Conference, ISCON '02. Proceedings.,

Ahmed, A., & Adil, K. (2023). Transforming Healthcare Systems with Artificial Intelligence: Revoluti onizing Efficiency, Quality, and Patient Care.

Alashram, A. R., Annino, G., Raju, M., & Padua, E. (2020). Effects of physical therapy interventions on balance ability in people with traumatic brain injury: A systematic review. *NeuroRehabilitation*, *46*(4), 455-466.

Ali, Z., Alsulaiman, M., Elamvazuthi, I., Muhammad, G., Mesallam, T. A., Farahat, M., & Malki, K. H. (2016). Voice pathology detection based on the modified voice contour and SVM. *Biologically Inspired Cognitive Architectures*, *15*, 10-18. https://doi.org/https://doi.org/10.1016/j.bica.2015.10.004

Ali, Z., Elamvazuthi, I., Alsulaiman, M., & Muhammad, G. (2015). Detection of Voice Pathology using Fractal Dimension in a Multiresolution Analysis of Normal and Disordered Speech Signals. *Journal of medical systems*, *40*(1), 20. https://doi.org/10.1007/s10916-015-0392-2

Arya, R., Pandey, D., Kalia, A., Zachariah, B. J., Sandhu, I., & Abrol, D. (2021, 24-25 Oct. 2021). Speech based Emotion Recognition using Machine Learning. 2021 IEEE Mysore Sub Section International Conference (MysuruCon),

Asken, B. M., Houck, Z. M., Bauer, R. M., & Clugston, J. R. (2020). SCAT5 vs. SCAT3 symptom reporting differences and convergent validity in collegiate athletes. *Archives of Clinical Neuropsychology*, *35*(3), 291-301. https://doi.org/10.1093/arclin/acz007

Aungst, S. L., Kabadi, S. V., Thompson, S. M., Stoica, B. A., & Faden, A. I. (2014). Repeated mild traumatic brain injury causes chronic neuroinflammation, changes in hippocampal synaptic plasticity, and

associated cognitive deficits. *Journal of Cerebral Blood Flow & Metabolism*, *34*(7), 1223-1232.

Bakar, Z. A., Tahir, N. M., & Yassin, I. M. (2010, 21-23 May 2010). Classification of Parkinson's disease based on Multilayer Perceptrons Neural Network. 2010 6th International Colloquium on Signal Processing & its Applications,

Basha, S. J., Madala, S. R., Vivek, K., Kumar, E. S., & Ammannamma, T. (2022, 4-5 March 2022). A Review on Imbalanced Data Classification Techniques. 2022 International Conference on Advanced Computing Technologies and Applications (ICACTA),

Battaglia, A. (2003). Neuroimaging studies in the evaluation of developmental delay/mental r etardation.

Bell, D., Guskiewicz, K., Clark, M. A., & Padua, D. (2011). Systematic Review of the Balance Error Scoring System. *Sports Health*, *3*, 287 - 295. https://doi.org/10.1177/1941738111403122

Benesty, J., Chen, J., Huang, Y., & Doclo, S. (2005). Study of the Wiener Filter for Noise Reduction. In J. Benesty, S. Makino, & J. Chen (Eds.), *Speech Enhancement* (pp. 9-41). Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-27489-8_2

Bhullar, H., County, B., Barnard, S., Anderson, A., & Seddon, M. E. (2021). Reducing the MRI outpatient waiting list through a capacity and demand time series improvement programme. *The New Zealand Medical Journal*, *134*(1537), 27-35.

Bigler, E. D. (2023). Volumetric MRI Findings in Mild Traumatic Brain Injury (mTBI) and Neuropsychological Outcome. *Neuropsychology Review*, *33*(1), 5-41. https://doi.org/10.1007/s11065-020-09474-0

Boll, S. (1979). Suppression of acoustic noise in speech using spectral subtraction. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *27*(2), 113-120. https://doi.org/10.1109/TASSP.1979.1163209

Boschi, V., Catricala, E., Consonni, M., Chesi, C., Moro, A., & Cappa, S. F. (2017). Connected Speech in Neurodegenerative Language Disorders: A Review. *Front Psychol*, *8*, 269. https://doi.org/10.3389/fpsyg.2017.00269

Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. Annual Conference Computational Learning Theory,

Brenk, F. v., Stipancic, K., Kain, A., & Tjaden, K. (2022). Intelligibility Across a Reading Passage: The Effect of Dysarthria and Cued Speaking Styles. *American Journal of Speech-Language Pathology*, *31*(1), 390-408. https://doi.org/doi:10.1044/2021_AJSLP-21-00151

Brownlee, J. (2020a). *Recursive Feature Elimination (RFE) for Feature Selection in Python*. Machine Learnring Mastery. https://machinelearningmastery.com/rfe-feature-selection-in-python/

Brownlee, J. (2020b). *Statistical Imputation for Missing Values in Machine Learning*. Machine Learning Mastery.

https://machinelearningmastery.com/statistical-imputation-for-missing-values-in-machine-learning/

Brownlee, J. (2021). *How to Choose an Activation Function for Deep Learning*. Machine Learrning Mastery. https://machinelearningmastery.com/choose-an-activation-function-for-deep-learning/

Campbell, W. M., Sturim, D. E., & Reynolds, D. A. (2006). Support vector machines using GMM supervectors for speaker verification. *IEEE Signal Processing Letters*, *13*(5), 308-311. https://doi.org/10.1109/LSP.2006.870086

Carolina, B. (2021). *Multilayer Perceptron Explained with a Real-Life Example and Python Code: Sentiment Analysis*. Towards Data Science. https://towardsdatascience.com/multilayer-perceptron-explained-with-a-real-life-example-and-python-code-sentiment-analysis-cb408ee93141

Chong, C. D., Zhang, J., Li, J., Wu, T., Dumkrieger, G., Nikolova, S., Ross, K., Stegmann, G., Liss, J., Schwedt, T. J., & others. (2021). Altered speech patterns in subjects with post-traumatic headache due t o mild traumatic brain injury. *The Journal of Headache and Pain*, *22*, 1-12. https://doi.org/10.1186/s10194-021-01296-6

Danielli, E., Simard, N., DeMatteo, C. A., Kumbhare, D., Ulmer, S., & Noseworthy, M. D. (2023). A review of brain regions and associated post-concussion symptoms. *Front Neurol*, *14*, 1136367. https://doi.org/10.3389/fneur.2023.1136367

Daudet, L., Yadav, N., Perez, M., Poellabauer, C., Schneider, S., & Huebner, A. (2017). Portable mTBI Assessment Using Temporal and Frequency Analysis of Speech. *IEEE J Biomed Health Inform*, *21*(2), 496-506. https://doi.org/10.1109/JBHI.2016.2633509

Dedry, M., Maryn, Y., Szmalec, A., Lith-Bijl, J. v., Dricot, L., & Desuter, G. (2022). Neural Correlates of Healthy Sustained Vowel Phonation Tasks: A Systematic Review and Meta-Analysis of Neuroimaging Studies. *Journal of Voice*. https://doi.org/https://doi.org/10.1016/j.jvoice.2022.02.008

Ekberg, M., Stavrinos, G., Andin, J., Stenfelt, S., & Dahlström, Ö. (2023). Acoustic Features Distinguishing Emotions in Swedish Speech. *Journal of Voice*. https://doi.org/https://doi.org/10.1016/j.jvoice.2023.03.010

Fagherazzi, G., Fischer, A., Ismael, M., & Despotovic, V. (2021). Voice for Health: The Use of Vocal Biomarkers from Research to Clinical Practice. *Digital Biomarkers*, *5*(1), 78-88. https://doi.org/10.1159/000515346

Falcone, M., Yadav, N., Poellabauer, C., & Flynn, P. (2013, 26-31 May 2013). Using isolated vowel sounds for classification of Mild Traumatic Brain Injury. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing,

Feature Engineering: Scaling, Normalization, and Standardization. (2023). GeeksforGeeks. https://www.geeksforgeeks.org/ml-feature-scaling-part-2/

Feng, S., Feng, Z., Ling, C., Chang, C., & Feng, Z. (2021). Prediction of the

COVID-19 epidemic trends based on SEIR and AI models. *PLOS ONE*, *16*(1), e0245101. https://doi.org/10.1371/journal.pone.0245101

Ferreres, A., López, C.-V., & China, N. N. (2003). Phonological alexia with vowel–consonant dissociation in non-word read ing. *Brain and Language*, *84*, 399-413. https://doi.org/10.1016/S0093-934X(02)00559-X

Fogel, A. L., & Kvedar, J. C. (2018). Artificial intelligence powers digital medicine. Npj Digital Medicine, 1 (1).

Gaube, S., Suresh, H., Raue, M., Lermer, E., Koch, T. K., Hudecek, M. F. C., Ackery, A. D., Grover, S. C., Coughlin, J. F., Frey, D., & others. (2023). Non-task expert physicians benefit from correct explainable AI advice when reviewing X-rays. *Scientific reports*, *13*(1), 1383.

Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., Kim, R., Raman, R., Nelson, P. C., Mega, J. L., & Webster, D. R. (2016). Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *Jama*, *316*(22), 2402-2410. https://doi.org/10.1001/jama.2016.17216

Gupta, A. (2023). *Feature Selection Techniques in Machine Learning (Updated 2024)*. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2020/10/feature-selection-techniques-in-machine-learning/

Hawley, C. A., Ward, A. B., Magnay, A. R., & Mychalkiw, W. (2004). Return to school after brain injury. *Arch Dis Child*, *89*(2), 136-142. https://doi.org/10.1136/adc.2002.025577

Heegaard, W., & Biros, M. (2007). Traumatic Brain Injury. *Emergency Medicine Clinics of North America*, *25*(3), 655-678. https://doi.org/https://doi.org/10.1016/j.emc.2007.07.001

Hoover, E. C., Souza, P. E., & Gallun, F. J. (2017). Auditory and Cognitive Factors Associated with Speech-in-Noise Complaints following Mild Traumatic Brain Injury. *J Am Acad Audiol*, *28*(4), 325-339. https://doi.org/10.3766/jaaa.16051

Horsch, A. (2021). *Detecting and Treating Outliers In Python — Part 3*. Towards Data Science. https://towardsdatascience.com/detecting-and-treating-outliers-in-python-part-3-dcb54abaf7b0

Hossain, M. S., Bilbao, J., Tobón, D. P., Muhammad, G., & Saddik, A. E. (2022). Special issue deep learning for multimedia healthcare. *Multimedia Systems*, *28*(4), 1147-1150.

Hostetler, Z. S., Hsu, F.-C., Barnard, R., Jones, D. A., Davis, M. L., Weaver, A. A., & Gayzik, F. S. (2020). Injury risk curves in far-side lateral motor vehicle crashes by AIS level, body region and injury code. *Traffic Injury Prevention*, *21*(sup1), S112-S117. https://doi.org/10.1080/15389588.2021.1880006

Howlett, J. R., Nelson, L. D., & Stein, M. B. (2022). Mental health consequences of traumatic brain injury. *Biological psychiatry*, *91*(5), 413-420.

Hue, A. (2019). *Kernel Support Vector Machines from scratch*. Towards Data

Science. https://towardsdatascience.com/support-vector-machines-learning-data-science-step-by-step-f2a569d90f76

IBM. (2023). *What are support vector machines (SVMs)?* IBM. https://www.ibm.com/topics/support-vector-machine

Imbalanced-Learn module in Python. (2020). GeeksforGeeks. https://www.geeksforgeeks.org/imbalanced-learn-module-in-python/

Jacobs, P. K., & Henwood, S. (2013). Investigating the experiences of New Zealand MRI technologists: Explor ing intra-orbital metallic foreign body safety practices. *Journal of Medical Radiation Sciences*, *60*(4), 123-130.

Jain, S., & Iverson, L. M. (2018). Glasgow coma scale.

Jiang, J. J., Zhang, Y., & McGilligan, C. (2006). Chaos in voice, from modeling to measurement. *Journal of Voice*, *20*(1), 2-17.

Kawamoto, K., Houlihan, C. A., Balas, E. A., & Lobach, D. F. (2005). Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *Bmj*, *330*(7494), 765.

Kelly, J. P., & Rosenberg, J. H. (1998). The development of guidelines for the management of concussion in sports. *J Head Trauma Rehabil*, *13*(2), 53-65. https://doi.org/10.1097/00001199-199804000-00008

Khare, Y. (2024). *Hands-On Guide To Librosa For Handling Audio Files*. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2024/01/hands-on-guide-to-librosa-for-handling-audio-files/

Lancheros, M., Friedrichs, D., & Laganaro, M. (2023). What Do Differences between Alternating and Sequential Diadochokinetic Tasks Tell Us about the Development of Oromotor Skills? An Insight from Childhood to Adulthood. *Brain Sciences*, *13*(4), 655. https://www.mdpi.com/2076-3425/13/4/655

Lauharatanahirun, N., Maciejewski, D. F., Kim-Spoon, J., & King-Casas, B. (2023). Risk-related brain activation is linked to longitudinal changes in adolescent health risk behaviors. *Developmental Cognitive Neuroscience*, *63*, 101291. https://doi.org/https://doi.org/10.1016/j.dcn.2023.101291

Le Bihan, D., Mangin, J.-F., Poupon, C., Clark, C. A., Pappata, S., Molko, N., & Chabriat, H. (2001). Diffusion tensor imaging: concepts and applications. *Journal of Magnetic Resonance Imaging: An Official Journal of the Inte rnational Society for Magnetic Resonance in Medicine*, *13*(4), 534-546.

Li, J., Cheng, Q., Liu, F.-K., Huang, Z., & Feng, S.-S. (2020). Sensory stimulation to improve arousal in comatose patients after trau matic brain injury: a systematic review of the literature. *Neurological Sciences*, *41*, 2367-2376.

Lubbers, V. F., van den Hoven, D. J., van der Naalt, J., Jellema, K., van den Brand, C., & Backus, B. (2024). Emergency Department Risk Factors for Post-Concussion Syndrome After Mild Traumatic Brain Injury: A Systematic Review. *Journal of neurotrauma*.

https://doi.org/10.1089/neu.2023.0302
Mactaggart, I., Kuper, H., Murthy, G. V. S., Oye, J., & Polack, S. (2016). Measuring disability in population based surveys: the interrelationshi p between clinical impairments and reported functional limitations in Cameroon and India. *PLOS ONE*, *11*(10), e0164470.
Marchman, V., Miller, R., & Bates, E. (1991). Babble and first words in children with focal brain injury. *Applied Psycholinguistics*, *12*, 1 - 22. https://doi.org/10.1017/S0142716400009358
Maurya, V. P., Mishra, R., Moscote-Salazar, L. R., Janjua, T., Cincu, R., & Agrawal, A. (2022). Neurotrauma Care,"Golden Hour" or "Golden Sixty Minutes". *Journal of Neurointensive Care*, *5*(2), 44-47. https://doi.org/10.32587/jnic.2022.00542
Mayo Clinic, S. (2022). CT scan - Mayo Clinic. https://www.mayoclinic.org/tests-procedures/ct-scan/about/pac-20393675
Mayo Clinic, S. (2023). MRI - Mayo Clinic. https://www.mayoclinic.org/tests-procedures/mri/about/pac-20384768
McCombes, S. (2021). *What Is a Research Design | Types, Guide & Examples*. Scribbr. https://www.scribbr.com/methodology/research-design/
McCrea, M., Kelly, J., Randolph, C., Kluge, J., Bartolic, E., Finn, G., & Baxter, B. (1998). Standardized Assessment of Concussion (SAC): On-Site Mental Status Eva luation of the Athlete. *Journal of Head Trauma Rehabilitation*, *13*, 27-35. https://doi.org/10.1097/00001199-199804000-00005
McCullough, A. (2001). Viability and effectiveness of teletherapy for pre-school children wit h special needs. *International journal of language & communication disorders*, *36*(S1), 321-326.
Miller, R. (2019). *Data Preprocessing: what is it and why is important*. CEOWORLD Magazine. https://ceoworld.biz/2019/12/13/data-preprocessing-what-is-it-and-why-is-important/
Mouzon, B., Chaytow, H., Crynen, G., Bachmeier, C., Stewart, J., Mullan, M., Stewart, W., & Crawford, F. (2012). Repetitive mild traumatic brain injury in a mouse model produces learn ing and memory deficits accompanied by histological changes. *Journal of neurotrauma*, *29*(18), 2761-2773. https://doi.org/https://doi.org/10.1089/neu.2012.2498
Nakadate, H., Kurtoglu, E., Shirasaki, S., & Aomura, S. (2016). Repetitive stretching enhances the formation of neurite swellings in c ultured neuronal cells. *Integrative Molecular Medicine*, *3*(4), 723-728.
National Academies Press. (2019). *Diagnosis and assessment of traumatic brain injury*. National Academies Press.
National Institute of Neurological, D., & Stroke. (n.d.). *Traumatic Brain Injury (TBI)*. National Institutes of Health. https://www.ninds.nih.gov/health-information/disorders/traumatic-brain-injury-tbi
Newshub. (2022). *Significantly higher costs for radiology scans in NZ than Australia, UK*. Newshub. https://www.newshub.co.nz/home/money/2022/09/significantly-higher-

costs-for-radiology-scans-in-nz-than-australia-uk.html

Niederjohn, R., & Grotelueschen, J. (1976). The enhancement of speech intelligibility in high noise levels by high-pass filtering followed by rapid amplitude compression. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *24*(4), 277-282. https://doi.org/10.1109/TASSP.1976.1162824

O'Brien, K. H. (2020). Overcoming Knowledge Barriers for Inclusion of School-Based Speech-Language Pathologists in the Management of Students with Mild Traumatic Brain Injury. *Semin Speech Lang*, *41*(2), 195-208. https://doi.org/10.1055/s-0040-1701687

Park, S. (2021). *A 2021 Guide to improving CNNs-Optimizers: Adam vs SGD*. Medium. https://medium.com/geekculture/a-2021-guide-to-improving-cnns-optimizers-adam-vs-sgd-495848ac6008

Poellabauer, C., Yadav, N., Daudet, L., Schneider, S. L., Busso, C., & Flynn, P. J. (2015). Challenges in Concussion Detection Using Vocal Acoustic Biomarkers. *IEEE Access*, *3*, 1143-1160. https://doi.org/10.1109/ACCESS.2015.2457392

Quantitative study designs. (2024). Deakin University. https://deakin.libguides.com/quantitative-study-designs/cross-sectional

Radiological Society of North, A., & American College of, R. a. (2022). Functional MRI (FMRI). https://www.radiologyinfo.org/en/info/fmribrain

Rajvanshi, N., Bhakat, R., Saxena, S., Rohilla, J., Basu, S., Nandolia, K. K., Agrawal, S., Bhat, N. K., & Chacham, S. (2021). Magnetic Resonance Spectroscopy in Children With Developmental Delay: Time to Look Beyond Conventional Magnetic Resonance Imaging (MRI). *Journal of Child Neurology*, *36*(6), 440-446.

Salvatore, A. P., Cannito, M. P., Hewitt, J., Dolan, L. D., King, G., Brassil, H. E., Salvatore, A. P., Cannito, M. P., Hewitt, J., Dolan, L. D., & others. (2019). Motor speech and motor limb status in athletes following a concussion. http://dx.doi.org/10.21849/cacd.2019.00150

Sanjyal, A. (2022). *Dimensionality Reduction VS Feature Selection*. medium. https://medium.com/@asanjyal81/dimensionality-reduction-vs-feature-selection-e68f91aa8724

Saravanan, R., & Sujatha, P. (2018, 14-15 June 2018). A State of Art Techniques on Machine Learning Algorithms: A Perspective of Supervised Learning Approaches in Data Classification. 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS),

Schatz, P., Pardini, J. E., Lovell, M. R., Collins, M. W., & Podell, K. (2006). Sensitivity and specificity of the ImPACT Test Battery for concussion in athletes. *Archives of Clinical Neuropsychology*, *21*(1), 91-99. https://doi.org/10.1016/j.acn.2005.08.001

Shah, R. (2024). *Tune Hyperparameters with GridSearchCV*. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2021/06/tune-hyperparameters-with-gridsearchcv/

Shenton, M. E., Hamoda, H. M., Schneiderman, J. S., Bouix, S., Pasternak, O.,

Rathi, Y., Vu, M. A., Purohit, M. P., Helmer, K., Koerte, I., & others. (2012). A review of magnetic resonance imaging and diffusion tensor imaging fi ndings in mild traumatic brain injury. *Brain imaging and behavior*, *6*, 137-192.

Skandsen, T., Nilsen, T. L., Einarsen, C., Normann, I., McDonagh, D., Haberg, A. K., & Vik, A. (2019). Incidence of mild traumatic brain injury: a prospective hospital, emer gency room and general practitioner-based study. *Frontiers in Neurology*, *10*, 638. https://doi.org/10.3389/fneur.2019.00638

Stojiljković, M. (n.d.). *Stochastic Gradient Descent Algorithm With Python and NumPy*. Real Python. https://realpython.com/gradient-descent-algorithm-python/

Sun, H., Luo, C., Chen, X., & Tao, L. (2017). Assessment of cognitive dysfunction in traumatic brain injury patients : a review. *Forensic sciences research*, *2*(4), 174-179.

Sutton, R. T., Pincock, D., Baumgart, D. C., Sadowski, D. C., Fedorak, R. N., & Kroeker, K. I. (2020). An overview of clinical decision support systems: benefits, risks, and strategies for success. *NPJ digital medicine*, *3*(1), 17.

Tate, A. (2023). *Detecting and Remedying Multicollinearity in Your Data Analysis*. Hex Technologies. .

Taud, H., & Mas, J. F. (2018). Multilayer Perceptron (MLP). In M. T. Camacho Olmedo, M. Paegelow, J.-F. Mas, & F. Escobar (Eds.), *Geomatic Approaches for Modeling Land Change Scenarios* (pp. 451-455). Springer International Publishing. https://doi.org/10.1007/978-3-319-60801-3_27

Taylor, C., Bell, J., Breiding, M., Xu, L., Langlois, J., Rutland-Brown, W., Wald, M., Lumba-Brown, A., Yeates, K., Sarmiento, K., & others. (2020). Overcoming knowledge barriers for inclusion of school-based speech-language pathologists in the management of students with mild traumatic b ain injury.

Team, G. L. (2024). *Hyperparameter Tuning with GridSearchCV*. Great Learning. https://www.mygreatlearning.com/blog/gridsearchcv/

Teixeira, J. P., Oliveira, C., & Lopes, C. (2013). Vocal Acoustic Analysis – Jitter, Shimmer and HNR Parameters. *Procedia Technology*, *9*, 1112-1122. https://doi.org/https://doi.org/10.1016/j.protcy.2013.12.124

Tewari, U. (2021). *Regularization — Understanding L1 and L2 regularization for Deep Learning*. Medium. https://medium.com/analytics-vidhya/regularization-understanding-l1-and-l2-regularization-for-deep-learning-a7b9e4a409bf

Theadom, A., Hardaker, N., Bray, C., Siegert, R., Henshall, K., Forch, K., Fernando, K., King, D., Fulcher, M., Jewell, S., & others. (2021). The brain injury screening tool (BIST): tool development, factor struc ture and validity. *PLOS ONE*, *16*(2), e0246512.

Thompson, E., & Murdoch, B. (1995). Disorders of nasality in subjects with upper motor neuron type dysarth ria following cerebrovascular accident. *Journal of communication disorders*, *28 3*, 261-276.

https://doi.org/10.1016/0021-9924(94)00013-P

Titus, D. J., Furones, C., Atkins, C. M., & Dietrich, W. D. (2015). Emergence of cognitive deficits after mild traumatic brain injury due to hyperthermia. *Experimental neurology*, *263*, 254-262. https://doi.org/https://doi.org/10.1016/j.expneurol.2014.10.020

Toldi, J., & Jones, J. (2021). A Case of Acute Stuttering Resulting after a Sports-related Concussion. *Current Sports Medicine Reports*, *20*(1), 10-12. https://doi.org/10.1249/jsr.0000000000000795

Toth, D., Tamas, A., & Reglodi, D. (2020). The neuroprotective and biomarker potential of PACAP in human traumati c brain injury. *International journal of molecular sciences*, *21*(3), 827. https://doi.org/10.3390/ijms21030827

Vincze, V., Gosztolya, G., Tóth, L., Hoffmann, I., Szatlóczki, G., Bánréti, Z., Pákáski, M., & Kálmán, J. (2016). Detecting mild cognitive impairment by exploiting linguistic informati on from transcripts.

Waalwijk, J. F., van der Sluijs, R., Lokerman, R. D., Fiddelers, A. A. A., Hietbrink, F., Leenen, L. P. H., Poeze, M., van Heijl, M., & Pre-hospital Trauma Triage Research, C. (2022). The impact of prehospital time intervals on mortality in moderately and severely injured patients. *J Trauma Acute Care Surg*, *92*(3), 520-527. https://doi.org/10.1097/TA.0000000000003380

Wall, C., Powell, D., Young, F., Zynda, A. J., Stuart, S., Covassin, T., & Godfrey, A. (2022). A deep learning-based approach to diagnose mild traumatic brain injury using audio classification. *PLOS ONE*, *17*(9), e0274395. https://doi.org/10.1371/journal.pone.0274395

Wang, T. V., & Song, P. C. (2022). Neurological voice disorders: a review. *International Journal of Head and Neck Surgery*, *13*(1), 32-40.

Whitehead, C. R., Webb, T. S., Wells, T. S., & Hunter, K. L. (2014). Airmen with mild traumatic brain injury (mTBI) at increased risk for s ubsequent mishaps. *Journal of safety research*, *48*, 43-47.

Winsorization. (2021). GeeksforGeeks. https://www.geeksforgeeks.org/winsorization/

Yadav, N. (2015). Portable Concussion Assessment Using Speech Biomarkers.

Yadav, N., Daudet, L., Poellabauer, C., & Flynn, P. (2014, 8-10 Oct. 2014). Noise management in mobile speech based health tools. 2014 IEEE Healthcare Innovation Conference (HIC),

Young, D., Cawood, S., Mares, K., Duschinsky, R., & Hardeman, W. (2023). Strategies supporting parent-delivered rehabilitation exercises to imp rove motor function after paediatric traumatic brain injury: A systema tic review. *Developmental Medicine & Child Neurology*.

Appendix

Glossary

BESS The Balance Error Scoring System is a standardized clinical test used to assess static postural stability, or balance, in individuals with suspected concussions. It involves a series of timed stances on both firm and foam surfaces, with varying foot positions and eyes closed. Errors, such as hands lifting off the hips, opening eyes, stumbling, or falling, are recorded and summed to provide a quantitative measure of balance impairment.

CT Scan Computerized Tomography scans utilize X-rays to create cross-sectional images of the brain, providing detailed information about bones, blood vessels, and tissues.

DTI Diffusion Tensor Imaging is another specialized MRI technique that examines the white matter tracts in the brain, which are responsible for communication between different brain regions.

fMRI Functional MRI is a specialized MRI technique that measures brain activity by detecting changes in blood flow. It provides insights into which brain regions are activated during specific tasks.

GCS The Glasgow Coma Scale is a standardized neurological scale used to assess

the level of consciousness in individuals with brain injuries. It evaluates three key responses: eye-opening, verbal communication, and motor function. Each response is assigned a numerical score, and the total score provides a quick and objective measure of the severity of brain injury, aiding clinicians in prioritizing care and determining the need for further evaluation.

MRI Scan Magnetic Resonance Imaging scans utilize magnetic fields and radio waves to create detailed images of the brain's soft tissues and structures, making them valuable for diagnosing subtle brain abnormalities that might not be visible on a CT scan.

MRS Magnetic Resonance Spectroscopy is a technique that measures the concentration of various chemicals in brain.

PCSS The Post-Concussion Symptom Scale is a standardized questionnaire used to assess and track the severity of 22 common concussion-related symptoms, including headache, dizziness, fatigue, cognitive difficulties, and emotional changes. It is often used to monitor the recovery process and evaluate the effectiveness of treatment interventions

SAC The Standardized Assessment of Concussion is a brief neuropsychological test used to evaluate cognitive function in individuals who have experienced a suspected concussion. It specifically targets orientation, immediate memory, concentration, and delayed recall, providing a quick assessment of cognitive impairment following a head injury. The SAC is often used in conjunction with other clinical assessments and neuroimaging techniques to aid in the diagnosis and management of concussions.

SCAT5 The Sport Concussion Assessment Tool 5th Edition is a standardized tool used to evaluate a wide range of concussion-related symptoms, including headache, dizziness, fatigue, cognitive difficulties, and emotional changes. It also

assesses physical signs, balance, and coordination, providing a comprehensive evaluation of an individual's condition following a suspected concussion.