A Speech-Based Classification Model for Mild Cognitive Impairment Screening

By

Ahindukha Noela Shisiali

121432

Submitted in Partial fulfilment of the Requirements for the Degree of Master of Science in Information Technology at Strathmore University

School of Computing & Engineering Sciences

Strathmore University

Nairobi, Kenya

June, 2025

Declaration and Approval

Declaration

I declare that this work has not been previously submitted and approved for the award of a degree by this or any other University. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.

Ahindukha Noela Shisiali

May, 2025

Approval

The thesis of Ahindukha Noela Shisiali was reviewed and approved by the following:

Prof. Bernard Shibwabo

Associate Professor, School of Computing & Engineering Sciences,

Strathmore University

Dr. Julius Butime,

Dean, School of Computing & Engineering Sciences

Strathmore University

Prof. Bernard Shibwabo,

Director of Graduate Studies,

Strathmore University

Abstract

Mild Cognitive Impairment (MCI) is a condition that often presents symptoms similar to those of normal aging, which makes distinguishing between the two challenging. While some cognitive decline is expected as people age, MCI involves more noticeable memory and thinking difficulties that do not yet interfere significantly with daily life. Depending on its underlying cause, MCI may progress to dementia, a neurodegenerative condition that leads to a significant decline in cognitive function and quality of life. Dementia has become a global public health issue, with increasing numbers of cases due to aging populations worldwide. Early intervention at the MCI stage is critical, as it provides an opportunity to slow down or possibly prevent the progression of dementia, improving patient outcomes and reducing the burden on healthcare systems. Various cognitive screening tools are currently used in clinical settings to assess cognitive function and detect early signs of impairment, but they have exhibited several challenges. This study leveraged Rapid Application Development due to its flexibility and iterative structure to develop an effective model for Mild Cognitive Impairment Screening. The study used an open-source database containing audio samples from MCI patients and healthy controls to build the model. Consequently, the study aimed to improve classification performance by extracting universal features such as Mel-Frequency Cepstral Coefficients (MFCC), jitter, shimmer, and fundamental frequency. Various models were trained, including Random Forest, convolutional neural network (CNN), Convolutional Neural Network, Long Short-Term Memory (CNN-LSTM) with Deep Neural Network (DNN), achieving the highest accuracy of 84%. This research demonstrated that universal features can effectively support the early detection of Mild Cognitive Impairment using deep learning, offering a non-invasive and scalable screening alternative for clinical settings. It also provided a foundation for future research into speech biomarkers for cognitive disorders and encourages the integration of machine learning in health technology applications. Based on the results, further improvements are recommended, such as exploring additional audio features and applying transfer learning to enhance the model's robustness.

Keywords: Mild Cognitive Impairment, Speech Analysis, Speech-Based Detection, Early Diagnosis, Classification Model

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ACER -R	Addenbrooke's Cognitive Examination-Revised
CNNs	Convolutional Neural Networks
CT	Computed Tomography
DNN	Deep Neural Networks
DSM-5-TR	Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Text Revision
LSTM	Long short-term memory
MCI	Mild Cognitive Impairment
MFCC	Mel Frequency Cepstral Coefficients
MMSE	Mini-Mental State Examination
MoCA	Montreal Cognitive Assessment
MRI	Magnetic Resonance Imaging
PET	Positron Emission Tomography
PCP	Primary Care Physicians
RAD	Rapid Application Development
RAVLT	Rey Auditory Verbal Learning Test
RNN	Recurrent Neural Networks
SVF	Semantic Verbal Fluency

Definitions of Terms

Acoustic features These are measurable attributes of sound waves that characterize speech or audio signals (Ramanarayanan et al., 2022).

Alzheimer disease It is a progressive brain disorder that gradually destroys memory, thinking skills, and eventually the ability to perform even the simplest of tasks (Breton et al., 2019).

Amyloid plaques Refers to abnormal clusters of protein fragments that accumulate in the brains of individuals with Alzheimer disease (Castellano et al., 2024).

Cortical atrophy Refers to the loss of neurons and the connections between them in the cerebral cortex, which can lead to a decrease in brain volume (Castellano et al., 2024).

Dementia This is a general term that is used to describe a range of neurological conditions that lead to a decline in cognitive functioning (Anderson, 2019).

Acknowledgements

I would like to express my heartfelt gratitude to God for His guidance and strength throughout this thesis. I am deeply grateful to my supervisor, Prof. Benard Shibwabo for his unwavering support and insightful feedback which have been invaluable in shaping this thesis.

Dedication

I dedicate this thesis to my parents; their unending love and support have been a strong foundation of my journey in this research. To my siblings, thank you for your constant presence and encouragement. I also dedicate this thesis to families and individuals who are facing challenges of cognitive impairment, in the hope that this research may contribute to understand and ease their struggles.

Chapter 1: Introduction

1.1 Background of the study

According to the United Nations Report (2023), the global population aged 65 and older is expected to rise significantly, from 761 million in 2021 to an additional of approximately 839 million in 2050. As the aging population grows, age-related conditions such as Mild Cognitive Impairment (MCI) are becoming increasingly prevalent, underscoring the importance of understanding and diagnosing this syndrome (Sabbagh et al., 2020). MCI is defined as a neuropsychological syndrome marked by noticeable cognitive decline that is a transitional phase between normal aging and dementia (Anderson, 2019).

It is characterized by a decline in brain function, where cognitive abilities begin to deteriorate, yet the impairment is not severe enough to interfere significantly with daily activities (Sabbagh et al., 2020). Research shows that 10% to 20% of individuals aged 65 years and older are diagnosed with MCI (Anderson, 2019). The diagnosis of MCI requires a comprehensive, multi-dimensional approach, integrating both clinical evaluations and advanced imaging modalities. Cognitive function is primarily assessed through cognitive assessments such as Montreal Cognitive Assessment (MoCA) and Mini-Mental State Examination (MMSE), which are essential in identifying subtle cognitive deficits (Islam et al., 2023).

Complementing these assessments, brain imaging methods, including magnetic resonance imaging (MRI) and positron emission tomography (PET), provide critical insights into structural and functional changes, such as cortical atrophy and the presence of amyloid plaques, both of which are often linked to MCI progression to dementia (Castellano et al., 2024). Screening of MCI involves a thorough evaluation of all the six cognitive functions; learning and memory, language, executive function, complex attention, perceptual-motor control and social cognition. Comprehensive assessment of these cognitive domains allows for a deeper understanding of cognitive decline, facilitating accurate diagnosis.

The progression of MCI varies depending on the underlying cause, but notable, 10% to 15% of those diagnosed with MCI have developed dementia, particularly Alzheimer disease, emphasizing the seriousness of the condition (Breton et al., 2019). With approximately 55 million people diagnosed with dementia worldwide, and 10 million new cases reported each year, the public health challenge posed by cognitive impairment is extensive and growing

(WHO, 2023). The term dementia is referred to as major cognitive impairment by the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Text Revision (DSM-5-TR).

While screening techniques for MCI have advanced, challenges remain. Cognitive tests like the MOCA and the MMSE are highly subjective to the clinic examiner and can be influenced by external factors like a patient's emotional state and educational level, potentially leading to inconsistent results (Adana Díaz et al., 2021). Additionally, the overlap between MCI symptoms and normal age-related cognitive changes further complicates the differentiation between healthy aging and early cognitive decline (Alzheimer's Association, 2023). The lack of definitive biomarkers for MCI presents another challenge, as current imaging techniques such as MRI and PET may show variability in finding making a conclusive diagnosis difficult.

Furthermore, the high cost and limited accessibility of advanced imaging options, such as PET scans, restrict their routine use in clinical practice (Al-Sharify et al., 2020). These limitations in the screening tools emphasize the pressing need for continued research focused on developing more precise, cost effective, and accessible tools for the early detection of MCI. Current screening methods, while useful in identifying cognitive decline at later stages, are inadequate in capturing early subtle changes in speech patterns (Isaacson & Saif, 2020)

Speech alterations are among the earliest indicators of cognitive decline; however they are often overlooked in conventional cognitive assessments, which primarily focus on memory and executive function. MCI patients and early-dementia patients typically exhibit reduced slower speech, increased hesitations and longer pauses, which reflect difficulties in word retrieval (Wang et al., 2022). The absence of sensitive tools to these early markers contributes to delayed diagnosis, preventing timely intervention strategies that may slow the progression of MCI to more severe cognitive impairments, such as dementia.

1.2 Problem Statement

There is a pressing need to develop an effective screening tool capable of distinguishing between normal aging and MCI syndrome (Isaacson & Saif, 2020). There are some associated symptoms such as memory lapses, attention lapses, slower information processing that overlap between MCI and normal aging making it challenging to differentiate between individuals with MCI and those experiencing normal aging. The key difference lies in the severity and frequency of these symptoms which can be detected using specialized assessments tools such as neuropsychological assessments or cognitive assessments such as MMSE, MoCA, brain imaging techniques, neurological assessments among others.

These tools can be used as screening tools that assist healthcare professionals to accurately diagnose an individual having MCI. These screening tools, techniques and methods, though instrumental, often fall short in detecting subtle cognitive changes at the early stages of MCI to ensure timely intervention (Isaacson & Saif, 2020). Many of these screening tools can be influenced by factors such as education level where individuals with lower educational levels perform worse than individuals with higher educational background. This can lead to misinterpretation of results and misdiagnosis of impairment (Cardoso et al., 2024). In addition to this, the poor sensitivity experienced in existing screening techniques where subtle early symptoms are not detected in a timely manner.

Many individuals seek medical treatment when they experience the early symptoms of MCI. These early symptoms may last for months to several years before progressing to serious conditions (Judge et al., 2019). The delay often results in late-stage diagnosis and increases the risk of progressing to dementia. While self-screening tools such as ACE-R have been developed to help individuals assess their cognitive health without the presence of healthcare professionals, they face similar limitations as other screening techniques.

1.3 Research Objectives

i. To investigate the characteristics of Mild Cognitive Impairment (MCI)
ii. To review the existing methods and approaches in screening for Mild Cognitive Impairment (MCI)
iii. To develop a speech-based classification model for screening MCI
iv. To validate the proposed model

1.4 Research Questions

i. What are the characteristics of Mild Cognitive Impairment
ii. What are the existing methods and approaches used in screening of Mild Cognitive Impairment
iii. How can a speech-based classification model be developed
iv. How well can the model detect MCI from speech

1.5 Justification

MCI symptoms may be mistaken for normal aging, therefore leading to delayed diagnosis or misdiagnosis. Speech classification offers a promising approach for differentiating between normal aging and the MCI syndrome, as speech deterioration is among the early symptoms of MCI (Wang et al., 2022). Research also indicated that individuals with dementia usually exhibit challenges in the language cognitive domain long before a formal diagnosis is made (Huang et al., 2024). By analyzing the features and patterns in speech, it may assist in pinpointing those at risk for MCI, therefore allowing an early intervention aimed at preventing the progression to dementia. Dementia is a syndrome that affects not only the patients but also their families. It is the 7th leading cause of death and one of the leading causes of dependence and dementia globally (Gustavsson et al., 2023).

The patients tend to forget their loved ones, and they experience a shift in their personality where they might become more withdrawn or irritable. The patient's family goes through emotional distress and financial challenges, especially when the patient is a breadwinner. Unfortunately, many patients receive a diagnosis at an advanced stage, as dementia is a progressive and irreversible condition. However, it is critical to understand that speech classification should be viewed as one component of a thorough diagnostic evaluation, supplemented by other recognized clinical assessments to achieve an accurate and reliable diagnosis.

1.6 Scope

This research focuses majorly on MCI and partly on early dementia due to the similarities in speech that these syndromes exhibit at the early stages. It concentrates on speech features, that is, exploring how speech can be utilized as a potential screening tool to identify individuals who are in the early stages of MCI. Its focus is deliberately on screening rather than diagnosis due to the multi-factor aspect that comes in to provide an accurate diagnosis of MCI. Screening is essential to provide an early indication of cognitive issues, enabling timely interventions, while diagnosis requires a more detailed assessment to confirm a specific condition.

1.7 Limitations

One limitation of the study is time constraints. Due to the strict deadlines of this research, it is not possible to implement the research in diverse environments or contexts. This results in the study not being able to fully explore how factors such as different cultural, linguistic, and clinical settings of a place might influence the outcomes. The study is also limited to individuals who do not have speech impairments, language barriers, or other physical conditions, as these factors might affect how they produce sound, leading to inaccurate results. In addition to this, factors such as the individual's emotional state, fatigue might be misinterpreted as indicators of cognitive impairment, leading to inaccurate results.

Chapter 2: Literature Review

2.1 Introduction

The growing old population has led to more focused research done on neurocognitive diseases such as MCI, Dementia and Parkinson's disease, since they are more common in older persons (Scharre et al., 2021). Research has been done on this sector to ensure that MCI is diagnosed at an early stage to prevent its progression to more serious conditions such as dementia. Screening tools and techniques have been developed to assist in detecting the changes over time of these deficits.

2.2 Theoretical Literature Review

This type of literature focuses on evaluating existing theories and their applications within the field of neuropsychological disorders . It aims to analyze these theories and understand how they relate to the research problem, thereby assisting in comprehending the broader concept of the study.

2.2.1 Understanding the characteristics of MCI

Given MCI's transitional nature and its potential progression to Alzheimer disease, early detection and intervention are crucial in slowing cognitive decline. Traditional assessments primarily focus on memory and executive function: however, recent studies highlight that linguistic and speech changes serve as early indicators of cognitive impairment (Wang et al., 2022). Understanding the signs and symptoms of MCI is therefore essential for designing effective screening tools that support early detection.

2.2.1.1 Signs and symptoms of MCI

Cognitive symptoms are typical symptoms that primarily affect memory, executive function, attention and language, with variations depending on whether the condition is amnestic MCI or non-amnestic MCI. Amnestic MCI primarily affects memory while non-amnestic MCI affects other cognitive domains (Abbatantuono et al., 2023).

Memory dysfunction is one of the prominent symptoms of MCI where the episodic memory is affected. Research done by Pillny et al., (2022), described episodic memory as the ability to remember and mentally relieve specific events or experience from one's past. Unlike normal aging, where occasional forgetfulness is expected, individuals with MCI experience significant difficulty retrieving recent events or experiences. This often manifests as forgetting recent conversations, struggling to retain newly learned information, frequently misplacing objects, missing appointments, and having trouble retracing their steps. Memory dysfunction in MCI is strongly linked to the shrinkage of the hippocampal atrophy, a critical brain region involved in memory storage and retrieval (Zhu et al., 2023). According to studies done on structural MCI, individuals with MCI have a considerable loss in the hippocampus volume, placing them at a greater risk of developing Alzhemier disease (Zhao et al., 2023).

Executive dysfunction is characterized by the inability to perform a set of cognitive abilities that usually help individuals to focus, control their actions, remember things and solve problems. These skills primarily include paying attention, staying organized, thinking flexibly, and making decisions. Research has identified a linkage between executive dysfunction in MCI to reduced activity in the prefrontal cortex, a critical part of the brain involved in higher-order cognitive processes (Jobson et al., 2021). As the impairment worsens , individuals with MCI experience a gradual decline in executive function, making it harder for them to plan daily activities, manage money, or follow detailed instructions (Chehrehnegar et al., 2022).

Another symptom of MCI is attention deficits where an individual with MCI struggles in maintaining focus and shifting attention between activities. These deficits become evident in tasks that require prolonged attention, such as working in noisy environments, multitasking, reading or engaging in lengthy discussions (Derbie et al., 2022). This issue stems from the changes in the prefrontal cortex which are responsible for maintaining concentration and multitasking. Changes in this region cause the individual to react more slowly or suffer mental weariness more rapidly than those experiencing normal aging. These subtle changes in attention control can significantly affect daily life, particularly in work and social settings.

Individuals with MCI frequently struggle with word retrieval, which causes more pauses and hesitations during speaking (Mueller et al., 2018). Verbal fluency deteriorates, making it difficult for them to converse smoothly. Additionally, understanding and constructing complex sentences becomes difficult due to disruptions in semantic memory and language processing networks. Language and speech impairments are characterized by long pauses during speech, substitution of simpler words, slow speech production, altered articulation and increased hesitation (Sanborn et al., 2022). These alterations suggest an underlying loss in executive function, attention, and motor coordination, which might further contribute to challenges in effective communication (Wang et al., 2022).

2.2.2 Neuroimaging as a biomarker for MCI

According to Zhao et al., (2023), neuroimaging techniques together with neuropsychological test can be used alongside each other to diagnose MCI. The study argues that this integration enhances the accuracy in diagnosis and provides a more comprehensive examination of the cognitive function. The neuroimaging technique suggested is computed tomography (CT) which provides cross-sectional images of the brain using X-rays and computer technology. CT has been recorded to have a high sensitivity and high variability that makes it more effective in identifying individuals who are likely to progress to dementia. Longitudinal studies also show that the patients who tested positive using the CT, developed dementia within three years.

Zhao et al., (2023) suggest that CT can detect Alzheimer disease at an early stage by quantifying amyloid-beta deposition at various stages of Alzheimer disease. Amyloid-beta deposits accumulate before cognitive decline, become apparent. The study further states that by detecting the amyloid-beta deposits at the MCI stage, may lead to the early diagnosis of Alzheimer disease and timely intervention. One of the advantages of this study is that it combines CT scans and neuropsychological test providing a holistic view of the patient's cognitive health, thereby, improving diagnostic accuracy.

CT scans have been compared to other neuroimaging techniques such as MRIs and PET and has been found to be less expensive. MRI are more effective in visualization and more sensitive to subtle changes associated with MCI compared to CT scans. PET scans on the other hand can detect biochemical changes that are usually associated with neurodegeneration. In addition to that PET scans can inform a treatment strategy based on the visualized pathological changes in the brain (Shen et al., 2022). Despite these two neuroimaging techniques having vital advantages, they are very expensive, making them inaccessible to individuals who require them.

2.2.3 Speech as a biomarker of MCI

A study by Troger et al., (2022) evaluated the use of a digital speech biomarker known as ki:e SB-C that assesses the cognitive function of elderly personnel. The ki:e SB-C was derived from two neuropsychological tests known as the Semantic Verbal Fluency (SVF) task and the Rey Auditory Verbal Learning Test (RAVLT). SVF is a cognitive assessment tool that evaluates the way an individual produces words from a specific semantic within a limited time frame while RAVLT evaluates learning and memory where the participants listen to a list of words then ask to recall as many words as possible immediately. They are both used to evaluate aspects of speech and cognitive function. Speech from these tests is transcribed and relevant features extracted for analysis purposes. The ki:e SB-C leveraged automatic speech processing during transcription to provide an approach to detect early detection in a timely manner. Due to this, the study supports that speech can be used as a biomarker for MCI and early detection of dementia.

Wang et al., (2022) studied the silent pauses present in cognitive impairment. The study investigates the percent of silence duration (PSD) as a potential biomarker that can help in detecting MCI at an early stage. The focus on PSD provides an interesting perspective since it removes the focus on traditional linguistic parameters such as word count and fluency. It highlights the relationship between increased silence and cognitive decline where individuals with cognitive impairments exhibit inconsistency of PSD across different contexts. The study claims that PSD can be used as a biomarker for MCI across different languages, therefore can be adapted in diverse populations. Despite this ability, PSD cannot be able to capture all dimensions of cognitive decline as there are other factors such as an individual's emotional state which can play an important role.

According to a study by Ramanarayanan et al., (2022), it explores the potential of using speech as a biomarker for neurodegenerative diseases. It emphasizes how easily speech features can be interpreted and the challenges involved in implementing speech-based biomarkers in clinical settings. It acknowledges that speech is non-invasive and can be easily collected. Through remote speech remotely allows broader outreach, especially among older adults who may have preferred self-assessments to clinical assessments.

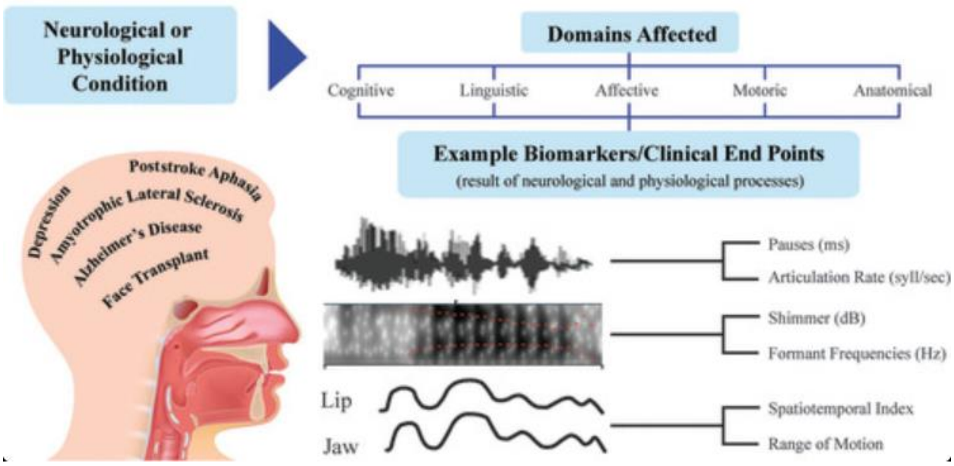


Figure 2. 1 : Neurological conditions affecting speech , adapted from (Ramanarayanan et al., 2022)

Figure 2.1 above depicts how various conditions such as Alzheimer disease affect the different cognitive domains including language and speech. The paper emphasizes the importance of identifying the best speech features that are best to indicate cognitive impairment. This involves studying acoustic parameters such as pitch, tone, pauses and linguistic characteristics like vocabulary diversity and sentence structures. The challenges in speech-based data identified were the difference in individual's speech patterns, background noise during recordings and different languages that could influence the results and reliability of speech data. The paper further emphasizes the need for verification against established clinical assessments and the need to track the changes of speech over time.

2.2.4 Machine Learning and AI application in cognitive assessments

The traditional cognitive assessments tools which are discussed in the next section rely on standardized neuropsychological tests . However, these tools have significant shortcomings, including subjectivity, and potential biases in test administration. To overcome these issues, AI-powered methods provide objective, automated, and scalable solutions for identifying cognitive deficits such as MCI. Machine learning has become more popular for improving cognitive evaluations by analyzing vast datasets and extracting significant cognitive indicators using pattern recognition, statistical modeling, and deep learning architectures(Penfold et al., 2022). These models are trained on a variety of data sources, including neuroimaging scans, behavioral patterns and voice recordings allowing them to identify minor cognitive impairments that standard tests may miss.

Research done by De Neys, (2021), described the Dual-Process Theory of Cognition, where thinking occurs in two ways: controlled and deliberate thinking (System1) and automatic and intuitive thinking (System 2). The former type of thinking is effortless and involves automatic speech production, such as everyday conversation, spontaneous responses while the latter type of thinking is effortful and involves complex reasoning, word selection and structuring sentences. Individuals with MCI experience impairments in both System 1 and System 2. Speech-based AI evaluations examine both linguistic and non-linguistic features of speech leveraging the Dual-Process Theory of Cognition. They analyze subtle shifts in unconscious speech and difficulties in deliberate cognitive processing to detect early signs of cognitive impairment.

Another theory that AI models leverage on is Information Processing Theory. This theory proposes that cognitive decline may be detected through changes in how the brain encodes, stores and retrieves information (Ball et al., 2023). The theory views cognition as a series of information handling steps and impairments at any stage, can indicate a neurological dysfunction. In MCI, disturbances in these processes emerge as difficulty recalling recent experiences, paying attention, and organizing ideas, making it a crucial framework for evaluating early cognitive decline (Huys et al., 2021)

2.3 Empirical Literature Review

Research studies and reports that are based on observed phenomena, presenting knowledge from actual experiments conducted. It focuses on real-world experiences and aims to establish the facts to answer specific research questions.

2.3.1 Studies on Screening tools for MCI

MCI screening tools and techniques have been used to support clinical decisions, but due to the challenges they face, the effectiveness in clinical practice has been impacted. These assessments are essential in identifying cognitive decline and ensuring timely and appropriate treatment. The following section provides an in-depth analysis of the literature done on the current assessment tools.

2.3.1.1 *Cognitive Assessments Tools*

These refer to tools designed to assess the cognitive health of individuals either in a clinical environment or self-administered by the individuals themselves. Healthcare professionals interprets the results produced from these tools and techniques to make an informed choice on what the diagnosis might be. This section delves into literature done on both clinical screening tools that is done with the presence of a healthcare professional and self-screening tools that can be administered by an individual themselves.

2.3.1.1.1 *Clinical Screening Tools*

These refer to standardized tools that are designed to be used in a clinical setting and administered by healthcare professionals who interpret the results and provide further intervention, if necessary, to improve the patient’s outcome. These tools provide a foundational use for more in-depth evaluations. Some of the most used screening tools include MMSE, MoCA and Addenbrooke's Cognitive Examination-Revised (ACE-R). The disadvantages and advantages of each tool are explained in the following sections

2.3.1.1.1.1 *Mini-Mental State Examination*

MMSE is a widely applied test for cognitive assessment in clinical medicine, comprising of eleven questions that evaluate five areas of cognitive functions, that is, orientation, registration, attention, recall and language. It also includes a basic physical examination and a review of the patient's medical history. According to Myberg (2020), this screening tool has both advantages and disadvantages. One of its main strengths is that it is quick and straight forward, featuring a standardized scoring system that ensures consistent use across different regions worldwide.

Despite these advantages, the study shows that MMSE is not effective in detecting MCI and fails to capture changes in severe dementia. There is a risk of the test producing false negatives, that is, when the test fails to identify a present condition due to its limited sensitivity to milder forms of cognitive impairment leading to inaccurate results (Pellicer-Espinosa & Díaz-Orueta, 2022). Educational attainment has also been shown to affect the validity of MMSE tool. Individuals with lower educational levels perform worse than individuals with higher educational background. This can lead to misinterpretation of results and misdiagnosis of impairment (Cardoso et al., 2024).

2.3.1.1.1.2 *Montreal Cognitive Assessment*

According to a study conducted by Dautzenberg et al., (2020), MoCA is a widely used brief screening tool that is useful in detecting MCI. The study aimed to validate that MoCA is a suitable screening tool and not a diagnostic tool for cognitive impairment. MoCA demonstrated strong validity achieving an impressive area under the curve of 0.93 when differentiating between individuals with cognitive impairment from those with normal cognition. According to Jia et al., (2021), MoCA's structured approach allows thorough evaluation and has higher sensitivity to subtle cognitive impairment compared to MMSE since it can differentiate the level of cognitive changes among those who were found to have normal cognitive in an MMSE assessment. A score of 26 points or higher is typically considered normal whereas a lower score may indicate cognitive impairment

However, while MoCA provides a quick and effective assessment, to administer it requires specific training and a mandatory certificate to ensure accurate administration of interpretation of results. This requirement adds a layer of complexity, as difference in the training and experience of those administering the test can lead to inconsistent outcomes. Similar to MMSE, the MoCA is also influenced by the educational background of the individual, potentially leading to misdiagnosis (Malek-Ahmadi & Nikkhahmanesh, 2024).

2.3.1.1.1.3 *Addenbrooke's Cognitive Examination-Revised*

The ACE-R is a 20-25 minutes cognitive assessment tool that is particularly sensitive to differentiate between different types of dementia. This tool is a revised version of a previous tool, Addenbrooke's Cognitive Examination (ACE). It was developed to improve accuracy by removing fewer sensitive items, making it easier to administer earlier in the diagnostic process. According to a study conducted by Tavares-Júnior et al., (2021), the improved version included additional tests for visuospatial abilities, which are especially useful for diagnosing conditions like Dementia with Lewy Bodies. These improvements made ACE-R more sensitive and increased its overall diagnostic accuracy.

The study's main goal was to determine the accuracy of ACE-R as a screening tool for older adults with lower levels of schooling, who have MCI and dementia, based in Brazil. The study involved 87 participants, where 9 were diagnosed with mild dementia, 49 were diagnosed with MCI while 29 were healthy control. The researchers found that although ACE-r was more sensitive compared to MMSE, education levels still influence the accuracy of the results. The study emphasized the introduction of lower cutoff scores that might work well across different educational backgrounds. This study highlighted the need for personalized approaches in cognitive screening.

2.3.1.1.2 *Self-Screening Tools*

These are tools that are used as preliminary tools by individuals to assess their cognitive function without the need for a trained professional. They typically consist of questionnaires and checklists that allow the evaluation of various conditions. They are designed to flag potential outcomes and assist the individual to seek further evaluation from trained professionals. These tools are essential, especially with the number of geriatric psychiatrists and neurologists expected to reduce by 2025 (Gould, 2023). They serve as an important tool for facilitating early intervention (Xavier et al., 2024). Some of the self-screening tools used for detecting MCI are listed in the following sections.

2.3.1.1.2.1 *Self-Administered Gerocognitive Examination (SAGE)*

Scharre et al., (2021) describe Self-Administered Gerocognitive Examination (SAGE) as a reliable tool that can be self-administered without any special equipment and is capable to identifying those with MCI or early dementia. The study compared longitudinal SAGE test scores with MMSE scores over an eight-year period in different cognitive sub-groups, that is, subjective cognitive decline, MCI, dementia converters of all types. They hypothesized that the SAGE test contains more challenging questions and evaluates executive functions better than MMSE in detecting mild cognitive decline. Although it takes longer to administer SAGE, it has four different versions to prevent individuals from memorizing and improving their scores.

Several activities of daily living of the participants were surveyed and any decline that would cause loss of function due to cognitive abilities was considered significant. Patients that were diagnosed with Alzheimer diseases or those with MCI eventually developed Alzheimer disease while some progressed to other forms of dementia based on the DSM-5-TR criteria. SAGE displayed a faster decline compared to MMSE when tested annually. For patients who developed Alzheimer's disease dropped by 1.91 points while MMSE scores dropped by 1.68 points, while for MCI patients who developed other forms of dementia SAGGE scores dropped by 2.33 points and MMSE scores dropped by 1.83 points.

Despite these advantages SAGE has some disadvantages. Being self-administered, there might be some differences in how the patients understand the results of the test potentially since explanations are not allowed. Individuals with low vision found it difficult to fully complete the test. SAGE tool results might also be affected by the individual's education backgrounds. Another notable challenge is personal bias and subjectiveness which can influence the accuracy of the results since some patients may overestimate their abilities, underreporting symptoms, while others may exaggerate their difficulties out of concern. All these disadvantages influence the results

2.3.1.1.2.2 *MyMemCheck*

Mansbach et al., (2020) did some study on integrating a self-assessment tool named MyMemCheck into the primary workflow. It was found to be a reliable and valid self-assessment for identifying older adults to assess their risk of mild cognitive impairment or dementia. The study focuses on finding the suitable cutoff score for detecting cognitive impairment. Two studies were carried out to test the accuracy level of MyMemCheck involving 59 participants in the first study and 357 participants in the second study. The results showed that the tool provided adequate reliability with a score of 67% and could effectively differentiate between individuals with normal cognition, MCI and dementia (Mansbach et al., 2020).

The tool was able to identify 80% of those who were diagnosed with MCI or dementia and 67% of those with normal cognition. However, there was a 43% possibility that an individual with MCI could receive a false negative result where they are thought to have normal cognition (Mansbach et al., 2020). This can lead to lack of necessary attention or treatment and may increase the likelihood of increasing dementia. Additionally, like SAGE, MyMemCheck may be subject to bias in self-reporting when self-assessment is done without a health professional. Factors such as personal perceptions or misunderstanding of the test items can influence their responses leading to incorrect results.

2.3.1.1.2.3 *Test Your Memory*

Brown et al., (2019) states that Test Your Memory (TYM) was designed for non-specialized clinics to assess cognitive functions in a timely manner, to differentiate those with Alzheimer disease from those with normal cognition. Test Your Memory for Mild Cognitive Impairment (TYM-MCI) based on the TYM was developed to focus specifically on improving the detection of MCI. The study involved 202 patients with cognitive issues and had passed the MMSE test. The TYM-MCI test was able to correctly distinguish between MCI and Alzheimer's Disease with a sensitivity of 79% and a specificity of 91%. The results show that the tool is valuable when distinguishing these conditions and works as well as ACE-R (Brown et al., 2019).

TYM was originally developed as a self-assessment tool but since it could only detect non-Alzheimer dementias, TYM-MCI was developed to be used in a clinical setting. The difference clinical setting used various methods, changing how cognitive results are understood (Brown et al., 2019). The results can also be influenced by lack of self-awareness where individuals can underreport their symptoms leading to inaccurate results. Unfortunately, this limitation is applied to screening tools that require paper.

2.3.2 Misdiagnosis of MCI cases

Judge et al., (2019) wrote a research article where they explored the challenges faced by physicians in diagnosing MCI and Alzheimer's Disease. 1365 physicians across different countries and medical specialties completed a survey which highlighted several barriers that contribute to delayed diagnosis or misdiagnosis, which can severely impact the patient's outcome. One of the barriers identified is patient-related barriers. 53% of the physicians agreed that families mistake cognitive symptoms and 50% of them noted that patients fail to disclose their symptoms. As earlier discussed, symptoms of MCI such as forgetfulness or confusion are also signs of normal aging but to a different degree. Families and patients may overlook this and not seek treatment leading to delayed proper diagnosis (Mansbach et al., 2020). Figure 2.2 shows the responses of the physician when asked about patient-related barriers.

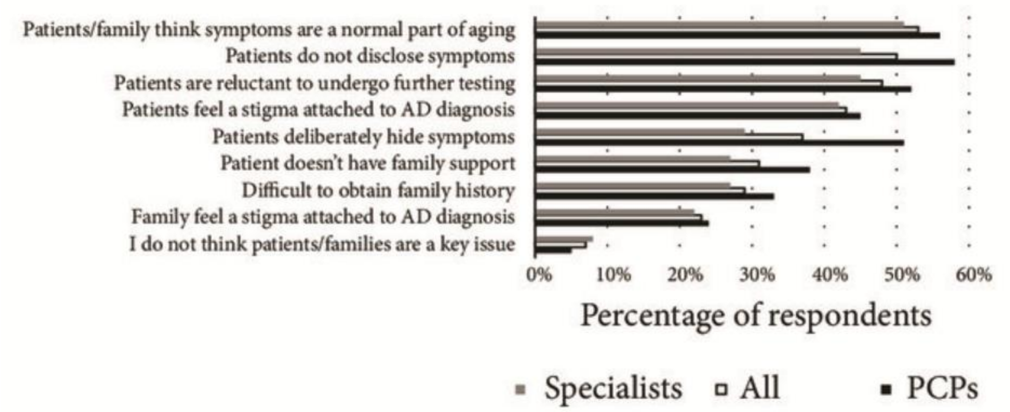


Figure 2. 2: Patient-related barriers survey adapted from Mansbach et al., (2020).

The study conducted by Judge et al., (2019), also identified clinical barriers and 43% of the respondents agreed that the lack of a definitive biomarker highlighting the need for more research on this area. 41% of the physicians noted that early symptoms often resemble normal aging thus aligning with an identified barrier under patient-related barriers. This shows that it is difficult to distinguish between normal aging and MCI. Figure 2.3 shows the percentage of respondents in the clinical barriers category

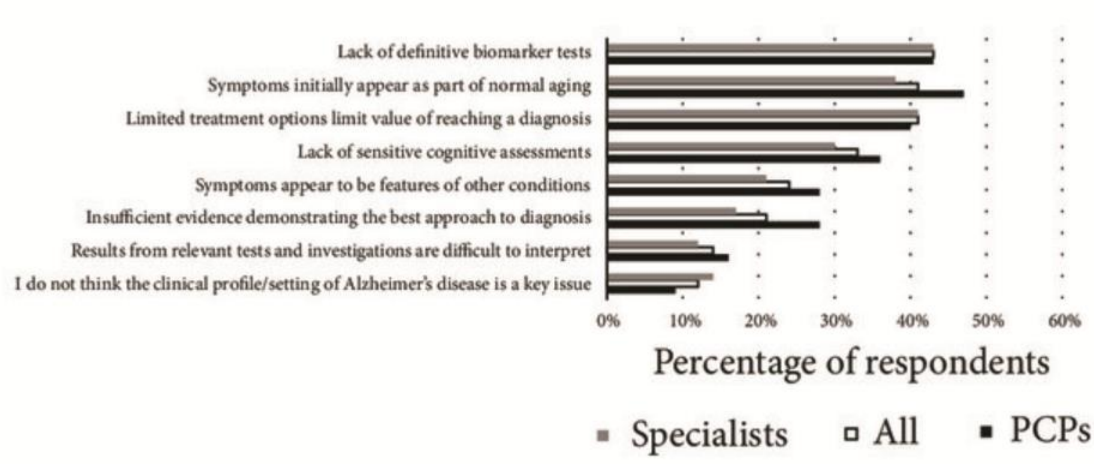


Figure 2. 3: Clinical related barriers survey adapted from (Mansbach et al., 2020).

Another category of diagnostic barriers is physician-related barriers where 42% of the specialist and 29% of the Primary Care Physicians (PCPs) indicated that they were not the issue. 37% of the physicians noted that they are concerned about how a diagnosis might affect the patient. This might lead to hesitation in giving the diagnosis leading to misdiagnosis since they might attribute the cognitive symptoms to other conditions that are less emotionally or socially impactful. Figure 2.4 shows the percentage of respondents in the physician-related barriers category.

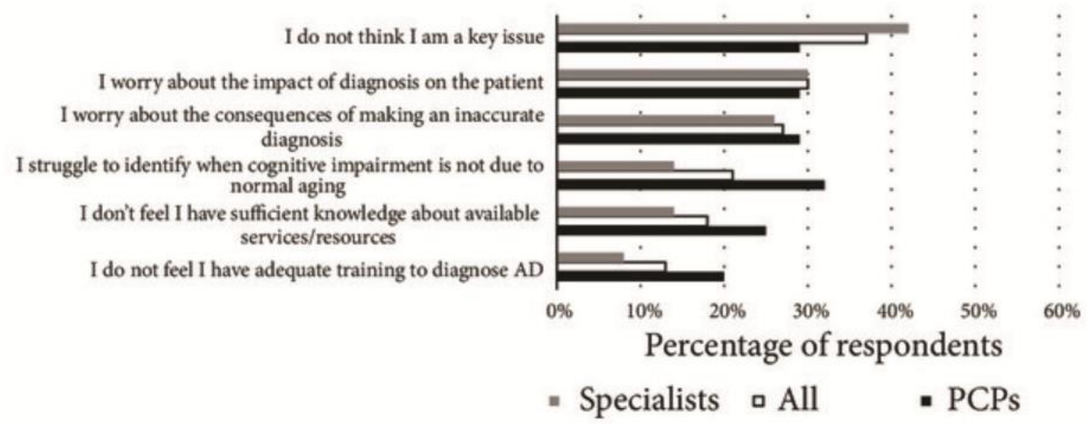


Figure 2. 4: Physician-related barriers survey adapted from (Mansbach et al., 2020).

The final diagnostic category explored was setting-related barriers. The most identified barrier in this category was long waiting lists for patients' appointments, with 43 % of the respondents identifying this issue while 33% noted that the time allocated for assessments was not enough for a proper diagnosis. Time is an important aspect when it comes to early intervention. When healthcare practitioners do not have enough time to perform thorough assessments it may lead to delayed diagnosis.

A study conducted by Sabbagh et al., (2020) supports the study by Judge et al., (2019), states that MCI is significantly misdiagnosed in the primary care setting. The study argues that PCPs are the initial point of contact therefore they should be equipped and made comfortable to monitor cognitive performance and make a referral when need be. One of the barriers identified by Judge et al., (2019) in his study is that there is a delay in referral from primary care, where 36% of the respondents agreed on that. The aspect of insufficient time and individuals overlooking their symptoms was also identified by the Sabbagh et al., (2020) study.

Both studies bring out the need for a tool that can effectively differentiate between normal aging and MCI to ensure timely intervention. A reliable self-assessment tool that encourages an individual to seek further medical attention, in this way addressing the issues and concerns on long waiting lines and tendency to overlook the symptoms as being normal aging. Furthermore, such a tool would aid clinical decision making by giving healthcare practitioners with early insights, allowing for more accurate diagnosis.

2.4 Architecture Designs

Research has been conducted on different studies and practices that are related to architectural design. It provides a deeper understanding of how solutions were designed, the impacts and advancements. The focus is on the solution's design process and the current trend to help inform future studies.

2.4.1 Gamification of MCI therapeutic process

Games have been used in the detection and screening of neuropsychological disorders such as MCI and dementia. In addition to that games have been used in the rehabilitation of patients with MCI (Lau & Agius, 2021). A lot of research has been done to ensure that games can be used as screening tools. According to a paper done by Lau and Agius (2021), an immersive serious game called A-go! that utilizes the use of gestures as controls to engage the patients in therapeutic tasks to improve their cognitive ability. These therapeutic tasks are targeted to improve cognitive skills such as attention, executive function and memory. The games involve motivational elements that enhance the player's willingness to participate in the therapeutic activities in the game to get a positive outcome. Figure 2.5 is an architecture framework for the A-go! game.

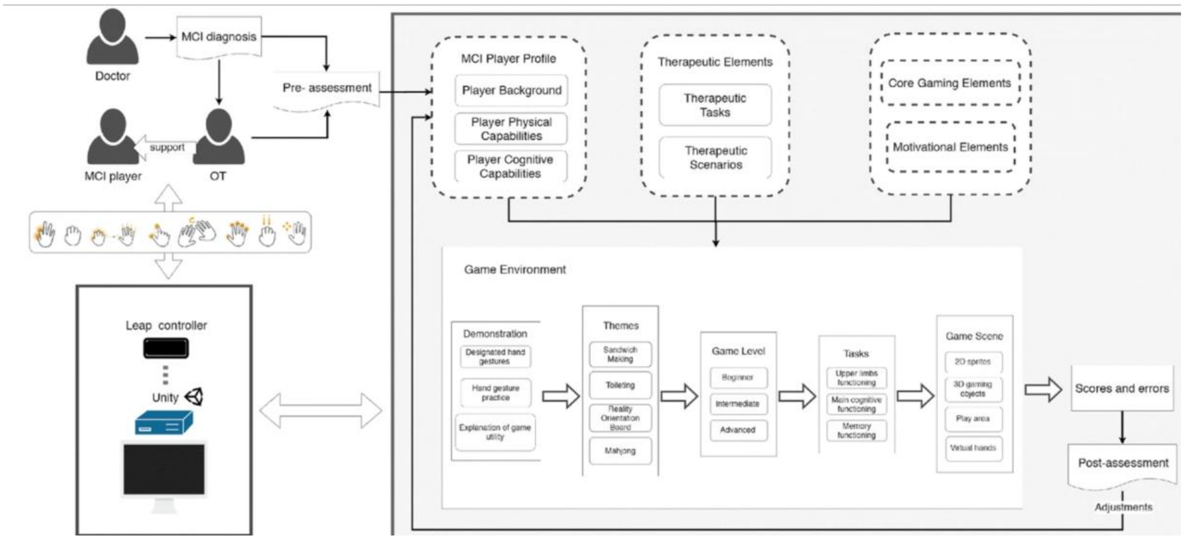


Figure 2. 5: Architecture of A-go! game , adapted from (Lau & Agius, 2021)

The architectural framework shows that the presence of a doctor who can be an occupational therapist is required to provide support to the player who has been diagnosed with MCI. At the pre-assessment stage, the baseline of the patients' cognitive health is understood and used to create the MCI player profile. The game involves therapeutic tasks as earlier discussed and core gaming elements and motivational elements are integrated to improve engagement (Lau & Agius, 2021). Before the game starts, the player goes through a demonstration phase that helps the user to learn about basic controls and what the game entails. Later, the player chooses a therapeutic task to work on, it can be a sandwich making task or a mahjong task.

The game has different levels that allows the progression of the player's abilities. The game can also focus on specific cognitive area and the game scene is set. After the tasks are completed, the scores and errors are recorded, and a post-assessment is done to evaluate the performance of the player. One of the advantages of this architecture is that it involves the longitudinal aspect where the records of each session are stored and can be monitored over time. This is important to understand whether the patient's condition is improving or deteriorating. In addition to that, it involved tasks that are similar to clinical tasks. The sandwich making task shares similar characteristics as the picture description tasks such as the Cookie Theft picture that requires an individual to explain a scene in an image.

2.4.2 Leveraging on Acoustic Features

A study conducted by Ding et al., (2024) aimed to explore whether acoustic features could accurately identify individuals with MCI among the participants from the study. They developed a structured design of a machine learning pipeline to analyze recordings, extracting relevant acoustic features so as to be able to effectively identify MCI patients. Figure 2.6 shows that machine learning pipeline developed. The voice recordings are separated based on the speaker labels, that is the participant and tester. This assists in focusing in only the participant's speech segments, which are fed into the feature extraction model. At this step, acoustic features such as pitch, tone, speech rate and filled pauses are extracted from the speech segment and the most relevant features are selected and fed to the classification model for appropriate categorization.

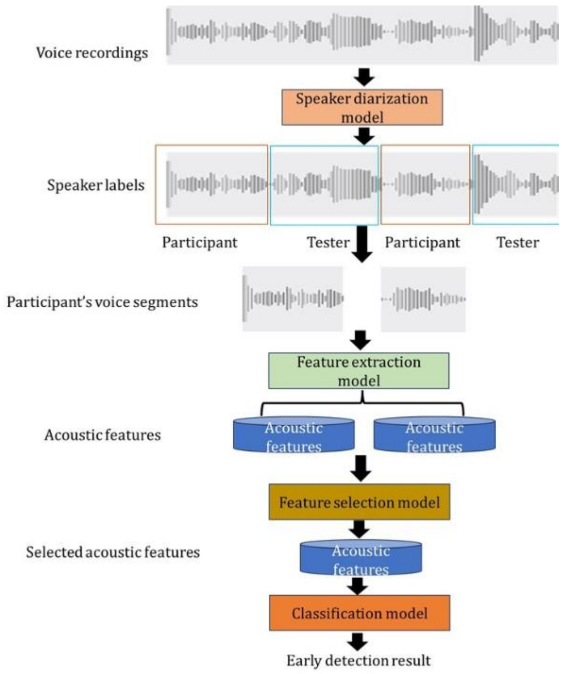


Figure 2. 6: Architecture design of a speech classification model , adapted from Ding et al., (2024)

The study acknowledges that the utilization of voice recording is a non-invasive method for screening cognitive impairment. This aspect makes it more accessible for patients and improves the ability to detect MCI at an early stage can facilitate timely intervention potentially delaying the progression to more severe cognitive decline. The architecture discussed in this paper offers robustness against noise using advanced signal processing making it suitable for real-world applications where recording conditions may vary. In addition to that, the architecture is easily scalable to accommodate larger populations allowing the widespread in clinical settings without requiring intensive resources. It also allows easy integration with other assessments to provide a more comprehensive view of the cognitive health of an individual.

Another study conducted by Huang et al., (2024), compared the effectiveness of using acoustic features, linguistic features and a combination of both, for developing an effective screening tool for Alzheimer's disease and MCI. The speech data is obtained from language tasks such as a picture description task, to extract various speech features. Comparison was made between acoustic and linguistic features to determine the advantages and disadvantages of both as indicators. Linguistic features need an in-depth analysis of the content therefore, it is highly dependent on the quality of the automatic speech recognition (ASR).

In addition to this, the robustness of the system is dependent on the uniqueness of the language. Depending on a specific language, limits transferability and prevents the system to be applied to other languages (Huang et al., 2024). Despite these limitations, using linguistic features achieves higher performance since they better reflect the changes that the patient's experience in language abilities. On the other hand, acoustic features have strong transferability and is language -independent therefore the solution can be generalized and can be applied to any language. However, such systems heavily rely on the quality of the audio since it can affect the detection of cognitive impairment.

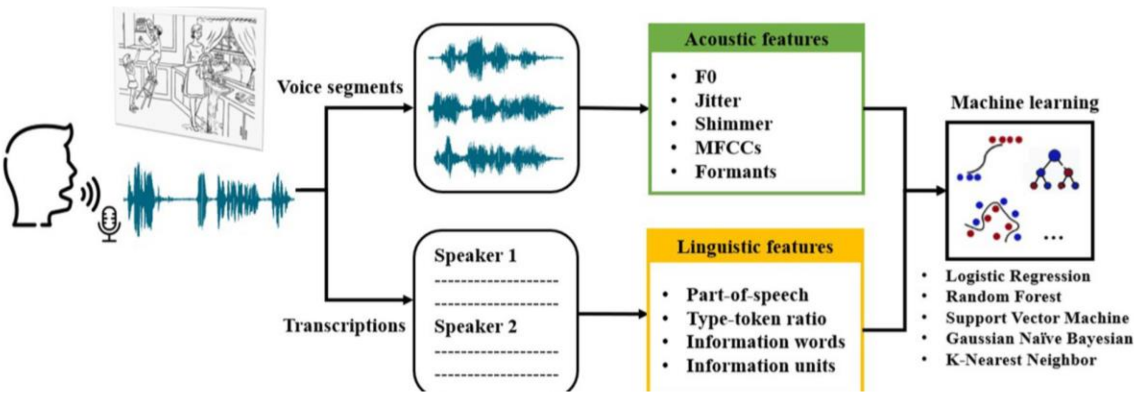


Figure 2. 7: An architecture design of a screening tool , adapted from Huang et al., (2024)

As depicted by Figure 2.7 above, the voice responses from picture description tasks were recorded and voice segments and transcriptions were recorded and voice segments and transcriptions are obtained. Specific acoustic features such as jitter, shimmer, Mel-frequency cepstral coefficients (MFCCs) and fundamental frequency (F0) are extracted to help capture the voice's pitch and quality. At the same time, linguistic features such as part-of-speech, information words and information units are extracted from the speech transcriptions. Both the linguistic and acoustic features are fed into various machine learning models to identify cognitive impairments.

2.5 Algorithms Literature Review

Machine learning algorithms have been used to develop solutions that can distinguish individuals with cognitive disorders, either classifying the severity of the impairment by identifying those affected. Through training of given datasets, the models can identify patterns that enable easy differentiation of cognitive impaired individuals from healthy controls. Some of the common classifiers are discussed in the following sections.

2.5.1 Traditional Machine Learning Classifiers

Decision trees are hierarchical structures that have nodes representing the input features from and branches indicating an occurrence of the testing done (Charbuty & Abdulazeez, 2021). They are simple and easy to interpret due to their visualization aspect. The root node represents the dataset which then splits to internal nodes based on the value of the feature until reaching the end node known as the leaf node which provides a final decision. The paper further explained that the decision tree is a classifier that uses measures like classification error and Gini Index to determine the best attribute to determine the next node after the split. This process is recursive until certain stopping criteria are met. Unlike other algorithms, the decision tree does not require extensive pre-processing of the dataset.

Despite these advantages, decision trees can face the issue of overfitting, where the model learns too much from the training data including unnecessary information such as noise therefore making it not able to adapt by predicting new changes. In a decision tree the concept of overfitting comes up when the final tree contains many branches and has a high accuracy when classifying the training data and low accuracy when classifying the testing data. The paper encourages the use of techniques like pruning or ensemble methods such as Random Forest to enhance the model's performance (Charbuty & Abdulazeez, 2021).

Vimal et al., (2021) supports the use of Random Forest algorithm for audio classification. Random Forest reduces the sensitivity to noise and outliers in data, therefore stabilizing the model's performance. This approach ensures that even if individual trees make inaccurate predictions, the overall accuracy is still high due to the majority voting mechanisms. This paper shows that Random Forest significantly reduces overfitting risk compared to a single decision tree. In addition to that, Random Forest ensures generalizability of unseen data by aggregating decisions or predictions from multiple decision trees. This advantage is vital since variability in audio data can lead to overfitting and making it more suitable for complex audio datasets with numerous features.

They are capable to provide more insights into the features by identifying which features contribute most to the classification tasks, highlighting the significance of various features. The paper further highlights that by combining Mel Frequency Cepstral Coefficients (MFCC) with Random Forests, the model's performance can be increased. MFCC can be utilized for feature extraction where the important audio characteristics are captured then fed into a Random Forest Classifier, which enhances the classification accuracy due to its ensemble nature (Vimal et al., 2021). The paper further emphasizes that Random Forest can evaluate the important MFCC features and which one among them contributes most to the final classification outcome or decision. Overall, MFCC and Random Forest integration has demonstrated effectiveness across various tasks such as sound classification and sound recognition.

2.5.2 Deep Learning

Neural networks have been found to be more advantageous compared to traditional classifiers such as the tree-based classifiers due to their ability to model complex data. They can achieve high levels of accuracy when trained on large datasets. For audio data to be analyzed using deep learning, it must be presented in a certain format either through spectrograms, or MFCC. Figure 2.8 displays the classification of different deep learning algorithms and architectures.

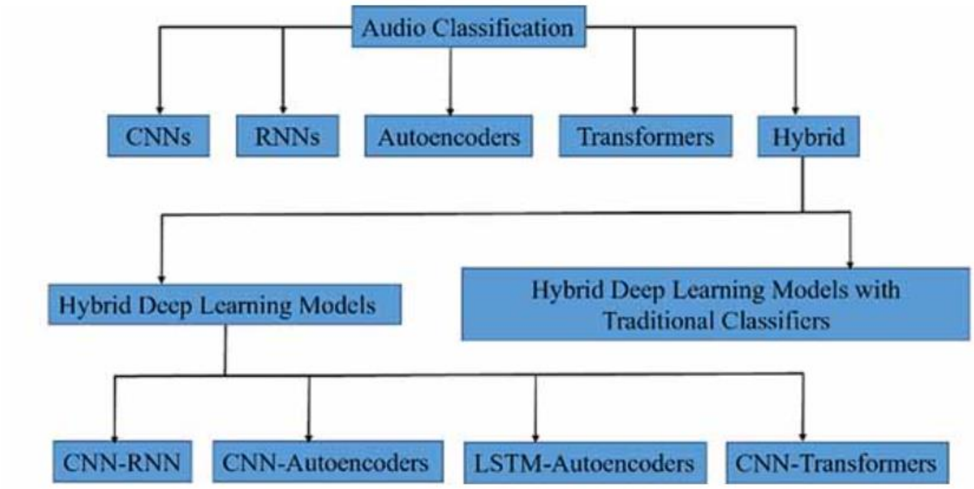


Figure 2. 8: Different deep learning architectures and algorithms adapted from Zaman et al., (2023)

Zaman et al., (2023) conducted a survey for the various deep learning methods used for audio classification such as Recurrent Neural Networks (RNN). RNN s are best for sequential data, therefore they are more effective on tasks that are time-dependent audio signals. They are designed to remember previous inputs, therefore making them better suited for ordered data like speech recognition or music genre classification. Based on the maintain a memory of past inputs they can process new ones. However, when provided with long sequences, they struggle to maintain previous information leading to poor performance on long tasks. This is known as the vanishing gradient issue (Zaman et al., 2023). To solve this issue LSTM was developed. The paper describes LSTM as a special type of RNN that controls information using gates, leading to better retention of crucial details on longer sequence.

Another type of deep learning method discussed in the is paper is Convolutional Neural Networks (CNNs). CNNs are power architecture used in classification of image data commonly. When it comes to audio data, the audio signals are represented as spectrograms. A spectrogram is a visual representation of the frequencies of an audio signal as it changes over time. CNNs can extract spatial features through convolutional layers, making them more suitable for identifying patterns in audio data. Its structure allows them to learn increasingly complex features at different levels of abstraction that is beneficial for audio that contain background noise (Zaman et al., 2023). The paper further highlights that CNNs have shown greater possibilities in audio applications enabling them to generalize well across various datasets.

Zaman et al., (2023) also talks about the combination of different classifiers. These hybrid methods combine the strength of two different classifiers. CNN-LSTM is a hybrid model that contains the strength from both CNN and RNN. In this hybrid, the architecture leverages on CNNs ability to extract features from spectrograms while also utilizing LSTMs to capture temporal dependencies on the features extracted. This combination enables a comprehensive overview of audio signals by addressing both temporal aspects and spatial aspects simultaneously. The paper concludes that hybrid methods perform much better in complex audio classification tasks, therefore it is suitable for tasks such as music analysis and speech processing. The paper highlights a study on music classification, the accuracy score gained by using CNN-LSTM was 92.1%.

2.6 Research gap

From the Literature review discussed in the previous section, a lot of research has been done on speech analysis to assist in the screening of cognitive impairments. However, these studies involve the use of smaller sample sizes that may introduce biases and affect generalizability of the results. Additionally, research papers have pointed out the need to develop a cost-effective solution that can easily be integrated into the clinical setting or used as a home-based screening tool for early detection of cognitive decline.

2.7 Conceptual Framework

The solution involves capturing audio recordings from individuals and stored in either a .WAV format or an MP3 format, then taken to the preprocessing stage where noise is reduced to enhance the quality of the audio. The improved quality audio goes through the feature extraction process using Mel-Frequency Cepstral Coefficients (MFCC). where universal acoustic features. This is because the solution is language-independent therefore the uniqueness of the language is not considered. By focusing on universal features, such as jitter, shimmer, fundamental frequency and MFCCs, the solution can be generalized to other languages (Huang et al., 2024). Jitter shows small changes in pitch, which is measured in percentages. Shimmer shows the small changes in loudness between voice cycles and is measured in decibels. Fundamental Frequency refers to the average pitch of the voice, measured in Hertz, while MFCCs are values that describe the shape of the sound and have no units of measurements. These features are then fed into a classification model which categorizes the speech as either MCI potential or Healthy control using DNN algorithm due to good performance that deep learning methods have achieved in previous research. Figure 2.9 is the proposed conceptual framework

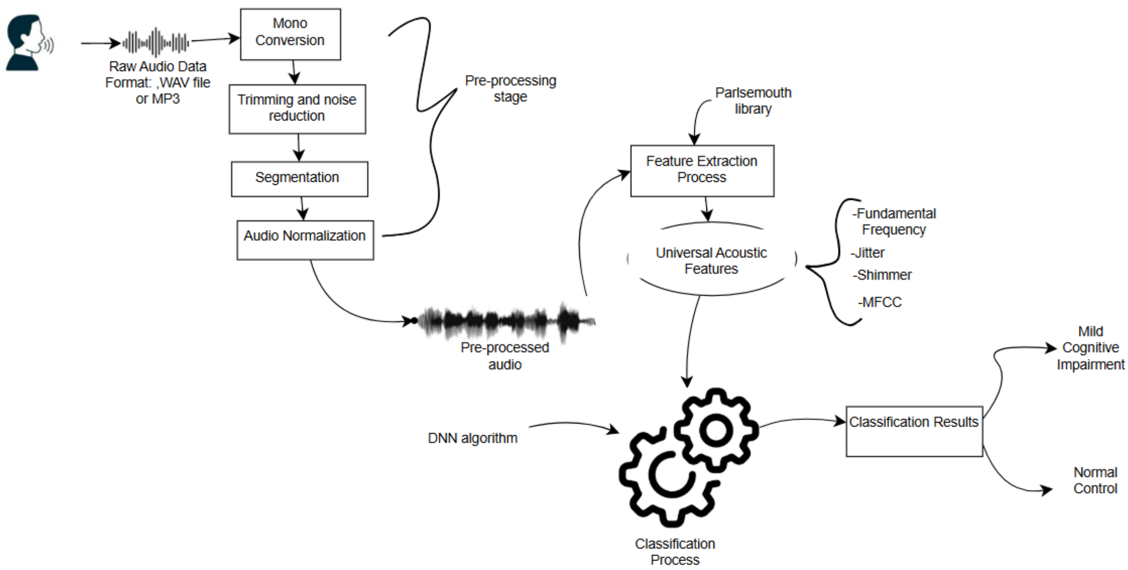


Figure 2. 9 Proposed conceptual framework

Chapter 3: Research Methodology

3.1 Introduction

Research methodology provides a clear and systematic plan used to classify Mild Cognitive Impairment from audio data. This chapter offers a detailed explanation of how the research was conducted to ensure the results' validity and reliability. The methodology is designed to systematically address the research objectives and support the development of an effective classification model that can distinguish between MCI and normal control subjects based on audio data.

3.2 Research Design

A research design is a structured plan that contains methods and approaches for collecting data and analyzing it to provide answers. It outlines how a research study should be conducted. This study leverages a combination of both applied and descriptive research to study MCI through speech analysis. Descriptive research refers to an approach that provides details depictions of the characteristics of a specific population (Siedlecki, 2020). It is useful when identifying trends and relationships within the dataset providing a clear understanding of the research area. It involves gathering data on speech patterns from individuals with normal control and those affected with MCI. By focusing on this, the study can identify the key features that might be useful for screening and how the displayed patterns relate to cognitive decline.

In addition to descriptive research, the study uses applied research, which focuses on developing a practical solution. By developing an effective screening tool that assists early detection and timely interventions through speech classification, this study can create a meaningful impact in the cognitive health field. This research design aims to address the real-world challenge of improving early diagnosis of MCI, therefore managing its progression to dementia. Insights from descriptive research and applied research are used to enable the creation of an effective screening tool that can assist healthcare professionals to identify MCI.

3.3 Target Population

The study population consists of elderly individuals aged sixty years and above, as this demographic is highly vulnerable to cognitive disorders. Age is a risk factor for cognitive decline, and based on the literature review the risk of having MCI or any other cognitive disorder increases with age. According to the United Nations Report (2023), older adults typically start at the age of sixty, and it is considered a milestone age when it comes to health research. MCI is higher among individuals aged sixty to seventy compared to those in their 40s or 50s. The age of sixty is also a transition age to retirement and this can impact cognitive health due to change in routines and lifestyle factors (Garrouste & Perdrix, 2022). The study majorly targets those with higher risks, and that is individuals aged sixty years and above. According to the World Health Organization (2023), the global population of these individuals is approximately 1.1 billion.

3. 4 Sampling

Out of this large population, a sample is required to make inferences. The study uses purposive sampling to select relevant dataset(s) based on specific criteria relevant to the research study. This allows the study to focus on a specific subset of the target population, therefore providing rich, detailed information making it easy to gather meaningful data. This type of sampling enables the selection of dataset(s) with different degrees of cognitive impairment as well as healthy participants. Being a non-random sampling method, the selection is based on a criterion, that is age, medical diagnosis, that is having MCI or normal cognitive health, and number of recordings.

To determine the appropriate sample size of audio files required, Cochran's formulae was applied as it is largely used for population that is too large or infinite. The formula is given below:

$$n_0 = \frac{Z^2 p(1-p)}{e^2}$$

Where :

- n0 is the required sample size,
- Z is the Z-score corresponding to the chosen confidence level (1.96 for 95%),
- p is the estimated population proportion (0.5 for maximum variability),
- e is the margin of error (0.05 for ±5%).

A 95% confidence level is used because it is commonly used to strike a balance between practicability and accuracy without overburdening the study with excessive data requirements. The margin of error is set to 0.05 as it provides a balance between precision and practicality ensuring reliable results. The calculations are as shown below:

$$n_0 = (1.96)^2 \times 0.5 \times (1 - 0.5) / (0.05)^2$$

$$n_0 = (3.8416 \times 0.25) / 0.0025$$

$$n_0 = 0.9604 / 0.0025$$

$$n_0 = 384$$

The computed sample size is 384, which is representative of the larger population and provides statistically reliable results.

When training the model 80% of the dataset is used for training and 20% used for testing. This allows the model to learn from a substantial amount of data, which is important in identifying patterns and trends within the dataset. The testing data is used for evaluation and provides a reliable estimate of the performance of the model, and how well the model can be generalized. This is meant to reduce the risk of overfitting since the training set is used to check whether the model is learning based on training data or learning to make predictions based on the underlying patterns.

3.5 Data Collection

The primary method for the data collection method is systematic review, a systematic examination of existing research related to the application of speech analysis in screening for cognitive impairment. It assists in gathering a comprehensive understanding of the current state of knowledge on the subject. The study identifies key research that pertains to the effectiveness of various screening tools and the significance of specific speech features which are vital. The findings are ordered thematically to ensure structure and better synthesis of information. By analyzing previous studies, the study aims to identify trends, challenges and effective practices that inform the development of a screening tool.

The study makes use of an open-source database known as DementiaBank. DementiaBank is a specialized resource that focuses on collecting and providing access to a wide range of data related to dementia. The dataset contains various types of data that capture the linguistic and cognitive characteristics of individuals with dementia. The study specifically selects a subset that contains at least 384 audio recordings. This selection aligns with the computed sample size, ensuring reliability while maintaining a balanced representation of individuals with MCI and cognitively normal individuals. The dataset chosen is based on availability and relevance to speech-based classification for MCI screening

3.6 Development Design

The study adopts an Agile methodology, specifically focusing on Rapid Application Development (RAD), due to its flexibility and iterative development therefore allowing quick adjustments based on the emerging insights. Agile methodology focuses on delivering small, functional and increments of a project through short cycles called sprints that adapts to changing requirements therefore promoting a dynamic environment. As new insights emerge from the analysis of speech data, the research process can be refined.

The fundamental of RAD is to deliver working software quickly through iterative cycles. Development is divided down into smaller phases, which allows for speedier feature delivery and early customer input. RAD makes extensive use of prototyping to get a basic, functioning version of the software into the hands of consumers as early as possible. In addition to this, RAD facilitates a more efficient use of resources, since the focus is on creating a prototype and collecting feedback.

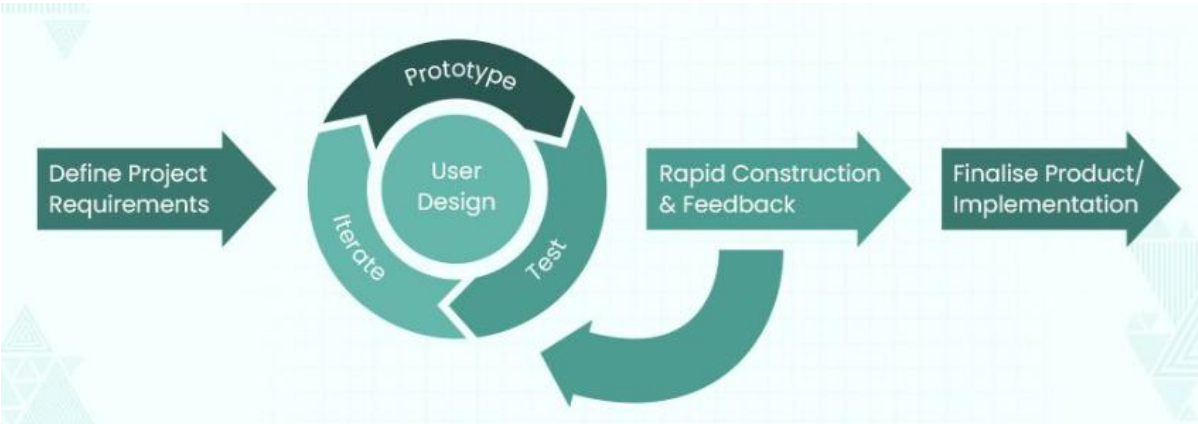


Figure 3. 1: Rapid Application Development adapted from Gadhavi (2022)

Figure 3.1 shows the different phases involved in RAD. The first stage involves the definition of the project's objectives, that is the goals and the schedule for the project. It then moves to the prototyping stage which is an iterative stage since each prototype created is tested and improved based on the requirements. The third phase involves gathering user feedback for prototype performance. Based on the feedback, the next step might be making final adjustments to the prototype or finalizing and implementing the product.

Python language is used because of its many machine learning packages, including Librosa, Keras and noisereduce, for model development. Python's ease of use makes it ideal for managing data processing, model training and evaluation. Google Colab, a cloud-based platform, facilitates model training, and effective machine learning experiments. In addition to that , Google Colab is an appropriate option due to its smooth collaboration features and pre-installed libraries. To develop the user interface, Visual Studio is used as a platform for creating web applications since it provides a suitable tool for performance optimization and user interface design. Flask language is used to develop the backend of the interface and assists in connecting the model to the interface.

3.7 Dissemination of Research Outcomes

Timely detection of cognitive impairment such as MCI is essential in delaying its progression to dementia, a syndrome which has affected over 10 million individuals worldwide yearly. The outcome of this research is a screening theses tool that is used to distinguish between normal cognition and MCI. This tool can be used as a self-screening tool for home-based assessments or be used as a clinical screening tool. The results are required to be interpreted by a healthcare practitioner due to the complexities involved in diagnosis of cognitive impairment.

The research outcome is available in an open-access repository of the university that scholars, students and researchers can freely access and utilize the findings. In addition to that, the study findings are also disseminated through seminars and workshops ensuring that the research findings are put into practice by offering chances for dialogue and possible cooperation. Furthermore, the findings are disseminated through conference proceedings, ensuring visibility within the broader research community, increasing the research's influence and legitimacy.

3.8 Ethical Considerations

When conducting research on speech analysis of individuals with MCI, confidentiality is key, and it is essential for health data. Measures such as anonymizing personally identifiable information should be taken. The dataset does not contain any information that could identify the participant. Additionally consent form the participants was required before collection of data. Access to the dataset is also limited to only authorized personnels due to the sensitivity of the data in the DementiaBank.

Chapter 4: System Analysis, Design and Architecture

4.1 Introduction

This chapter provides an overview of the model's requirements, both functional requirements and non-functional requirements. Additionally, it introduces the key diagrams, that illustrates how different components interact with each other and how information flows to satisfy the identified requirements.

4.2 Requirement Analysis

The model's requirements were gathered and analyzed to ensure their relevance to the developed system. These requirements are categorized into functional and non-functional requirements, which define the model's capabilities and constraints. The following sub-sections provide a detailed discussion of these requirements.

4.2.1 Functional Requirements

Functional requirements refer to functions that the system should do. Table 4.1 contains the functional requirements of the system.

Table 4. 1: Functional Requirements

No	Functional Requirement
i.	The user should be able to record audio using the interface
ii.	The interface should collect the audio recorded by the user
iii.	The interface should process the audio collected from the user
iv.	The interface should present the classification results to the users
v.	The interface should display past classification results

4.2.2 Non-functional Requirements

Non-functional requirements focus on the behavior of the system rather than the functions of the system. They ensure that the functional requirements work effectively under real-world conditions. Table 4.2 contains the identified, non-functional requirements of the system.

Table 4. 2: Non-functional requirements

No	Non-functional Requirements
i.	Authenticated users should be able to login
ii.	The system should have a user-friendly interface
iii.	The system should be reliable in identifying users with potential MCI
iv.	The system should encrypt personal identifiable information
v.	The system should have a database large enough to store all the required data

4.3 System Design Diagrams

4.3.1 Use Case Diagram

The case diagram provides a clear representation of how the user interacts with the system. Figure 4.1 is a use case diagram of the system that has a single actor who is the user. The user logs in into the system after registration and records audio. The recorded audio is then passed to the model for speech classification. After classification, the results are displayed to the user along with a brief recommendation that is based on the results received. In addition to this, the user should be able to review past results from the database and download their audio recording. The admin is responsible for managing the user accounts, where they can remove user accounts. In addition to that , the admin can update the machine learning model, monitor performance and retrain the model when needed.

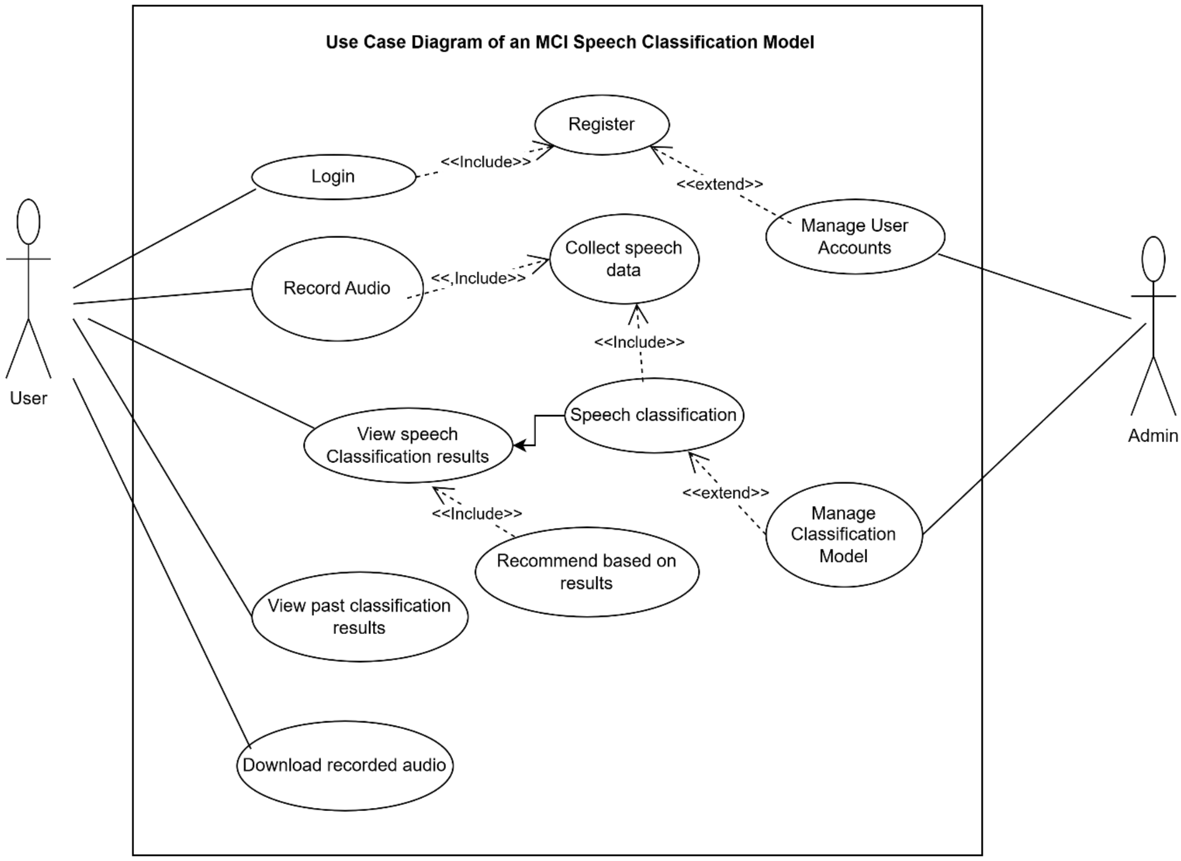


Figure 4. 1: Use Case Diagram

4.3.2 Sequence Diagram

Sequence diagram shows how elements in a system interact with each other over time. Figure 4.2 is a sequence diagram that captures the flow among four elements, that is, the user, user interface, speech classification model and the database. The process begins when the user logs in and records audio through the interface. The audio is passed to the speech classification model and classifies it to determine whether it indicates potential MCI or normal control. The classification results are then displayed on the user interface. To maintain real-time functionality, both the audio name and the results are sent to the database once the classification results are displayed. The user can retrieve past classification results that are stored in the database.

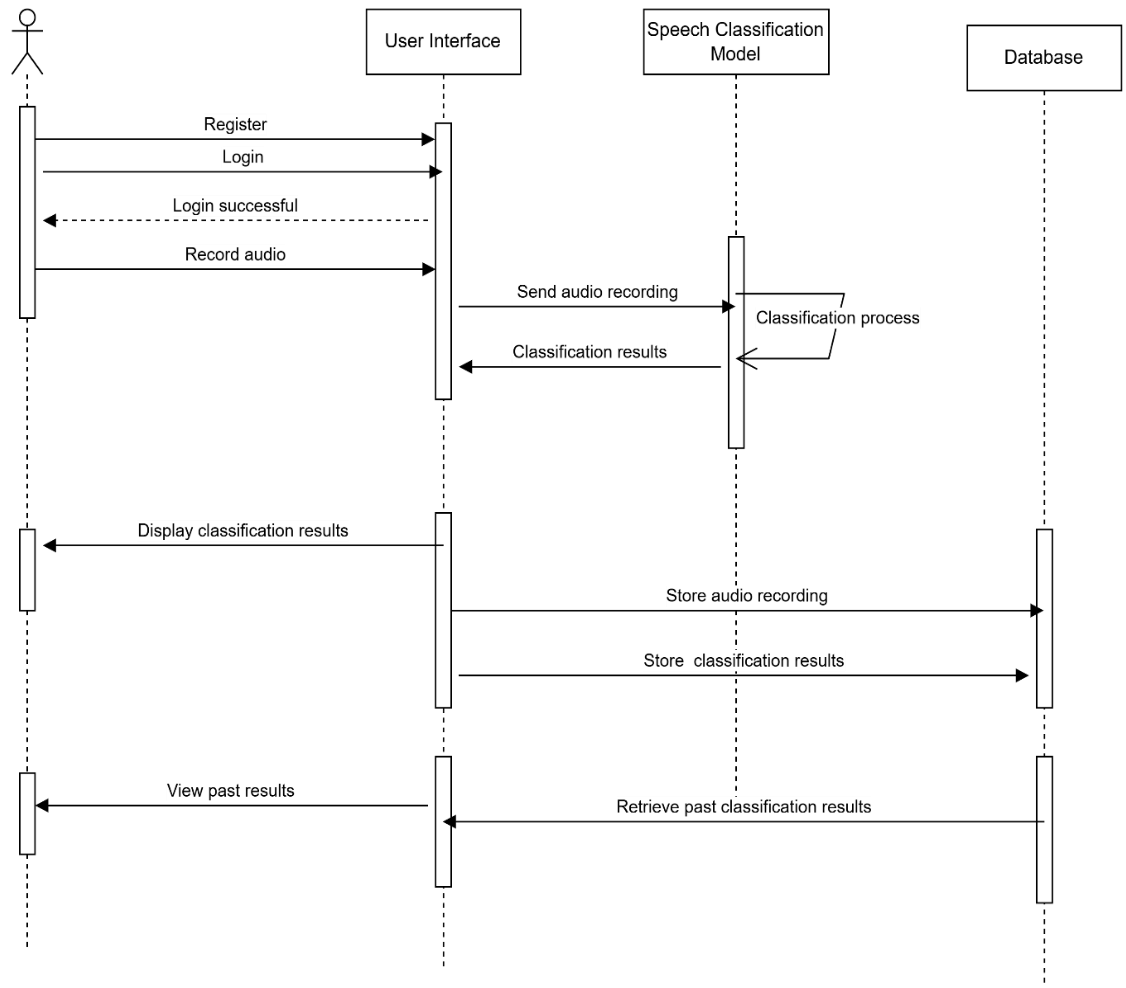


Figure 4. 2: Sequence diagram

4.3.3 Class Diagram

Figure 4.3 is a class diagram of the system having three classes. Each of the classes identified have attributes and the relationships shown in the figure below. To ensure privacy, the user class stores minimal personal identifiable information for login purposes. The recording class captures the audio from each data, while the result class contains the classification result from each recording.

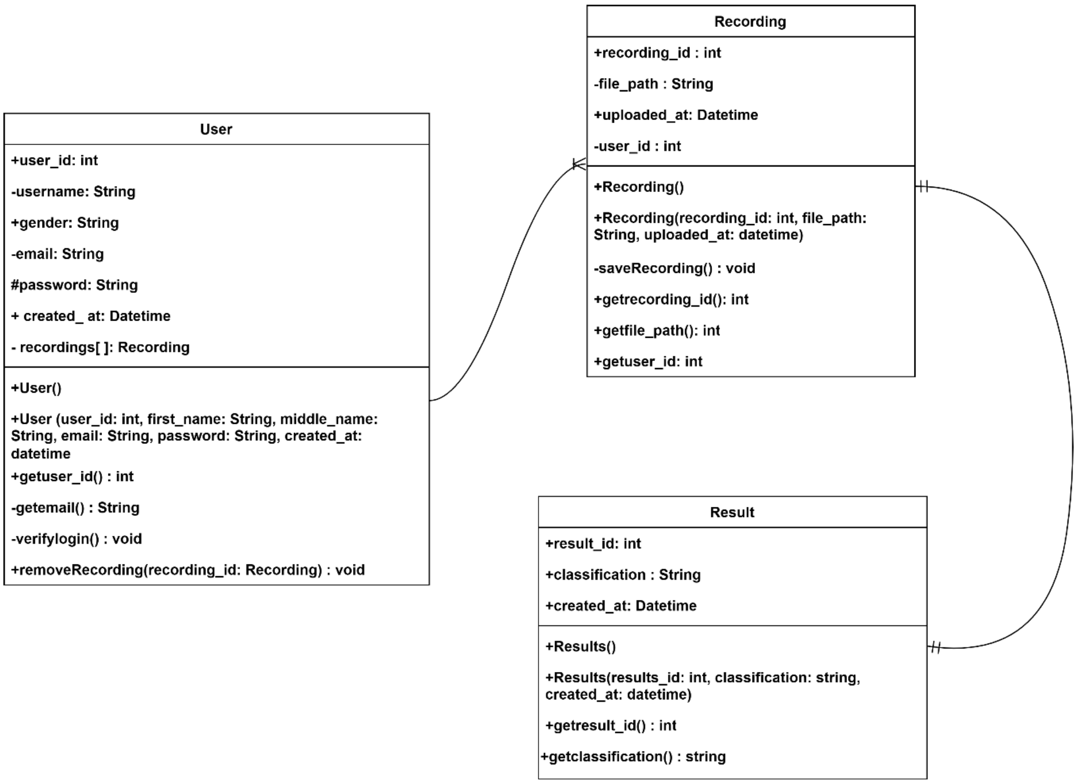


Figure 4. 3: Class diagram

4.3.4 Database Schema

A database schema depicts the structure of the system's database. In this design, the database holds three tables that work together to store all necessary information. The users table which contains basic details of each registered user, recordings table that holds the information of the uploaded audio files while the results table holds the information of the classification results from the model. Figure 4.4 displays the tables and the relationships between them, ensure that information can easily be retrieved and consistent.

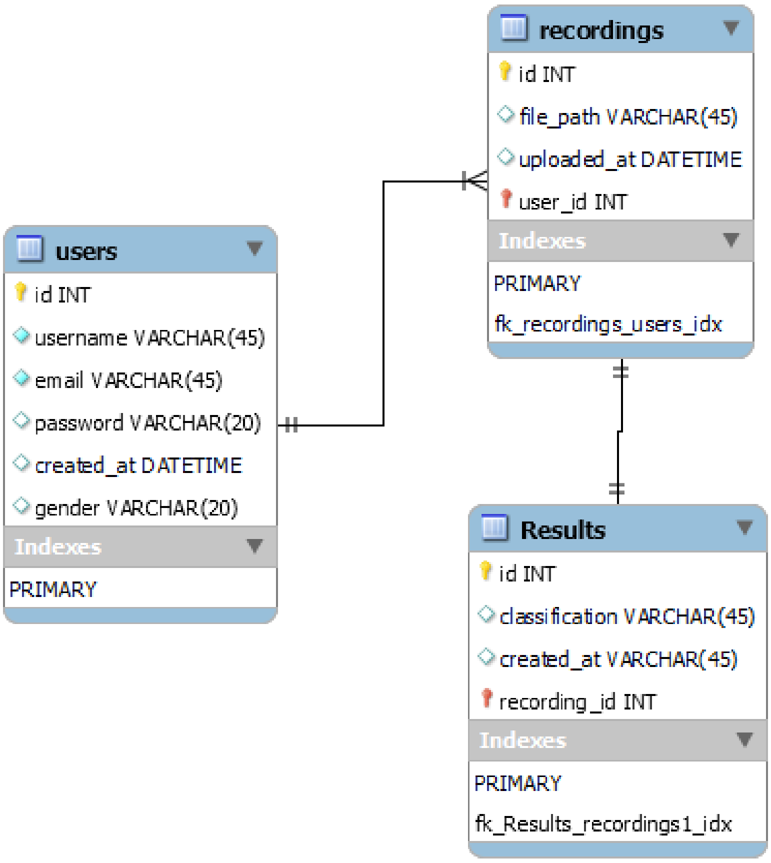


Figure 4. 4: Database Schema

Chapter 5: System Implementation and Testing

5.1 Introduction

This chapter briefly describes the actions taken to design and test models that can classify Mild Cognitive Impairment and normal controls by following a structured process. The chapter focuses on preparing the dataset where audio features were selected and preprocessed to ensure quality input for model training. Several models were tested, and the most accurate model was integrated into an interface that users can interact with.

5.2 Model Development

The process of developing the model starts with exploring the data to comprehend the dataset and examine the distribution of its existing features. Audio processing takes place to ensure that the audio recordings are cleaned and trimmed to extract valuable features from them. Once the dataset is cleaned, features are extracted, and the model is trained using these features. The subsequent sections elaborate on these steps in detail

5.2.1 Data exploration

The dataset contains 387 audio recordings from English and Chinese speakers, along with a csv file that detailed the gender and diagnosis of each patient. The identities of the speakers were anonymized to safeguard their personally identifiable information. The dataset compromised five columns: tkdname (the name of the audio), age, gender, MMSE(Mini-Mental State Examination), and dx (diagnosis). The MMSE column contained the scores of participants which ranged from 13 to 30. The histogram plot below examined the distribution of the mini-mental scores. The plot indicated that most participants had high scores, specifically between 27 and 30, while a smaller number of individuals had lower scores ranging from 12 to 24. Overall, the plot is positively skewed, indicating that there are fewer individuals with low MMSE scores

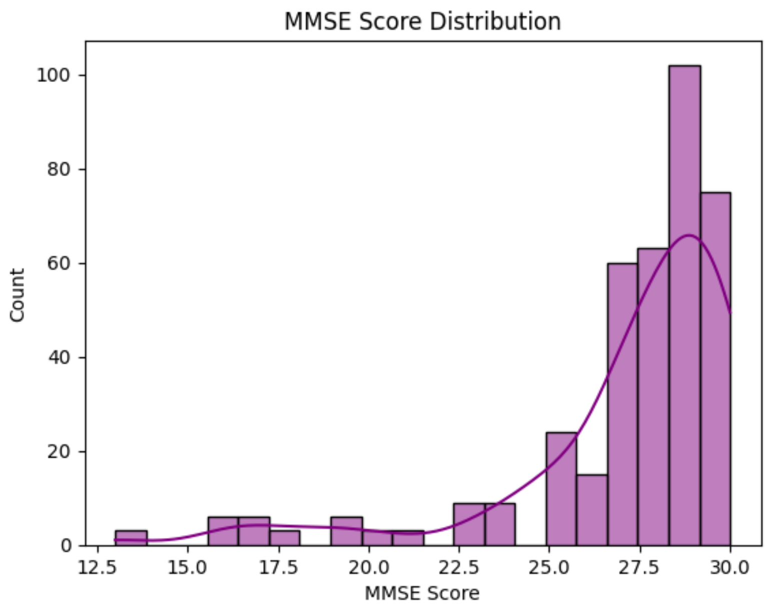


Figure 5. 1: Histogram examining MMSE Score Distribution

A different plot was created to illustrate the relationship between the MMSE scores and the diagnosis (dx column). The graph displayed a significant concentration of participants in the Normal Control group with scores around 28-30, while the MCI group predominantly falls on the lower end of the score distribution. Additionally, the graph reveals an overlap between the scores of 25 and 30, suggesting that while the MMSE is a valuable cognitive assessment tool, it may not clearly differentiate between Normal Control and MCI patients.

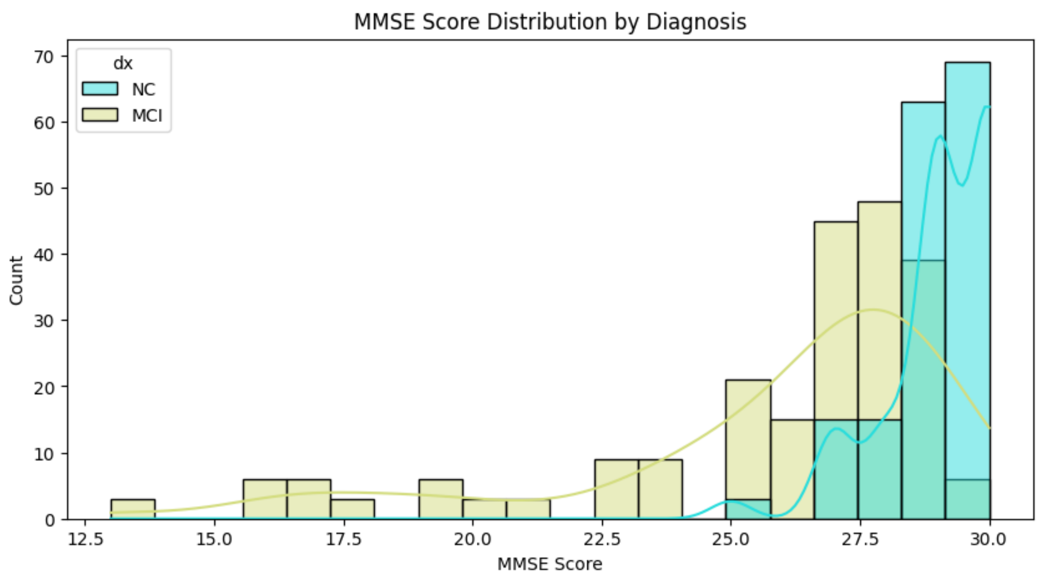


Figure 5. 2: Histogram examining MMSE and Diagnosis relationship

5.2.2 Audio Preprocessing

When working with audio data, it is important to consider the choice of audio channels. There are two audio channels, that is, mono or stereo. Mono audio uses a single channel, and it is mostly used in classification tasks and situations where lower data usage is a key priority. It contains one stream of data making it easier to process and analyze simplifying the algorithms designed for audio processing. To ensure that the audio files are using the mono channel, Librosa library was used . By default, Librosa converts the audio into a mono and sets the sample rate to 22050 Hz as it strikes a balance between audio quality and computational efficiency. A comparison was done using another library , wavfile, retains the original format of the audio, including its stereo channels and sample rate. With Librosa library the waveform of an audio can be displayed as shown in Figure 5.3.

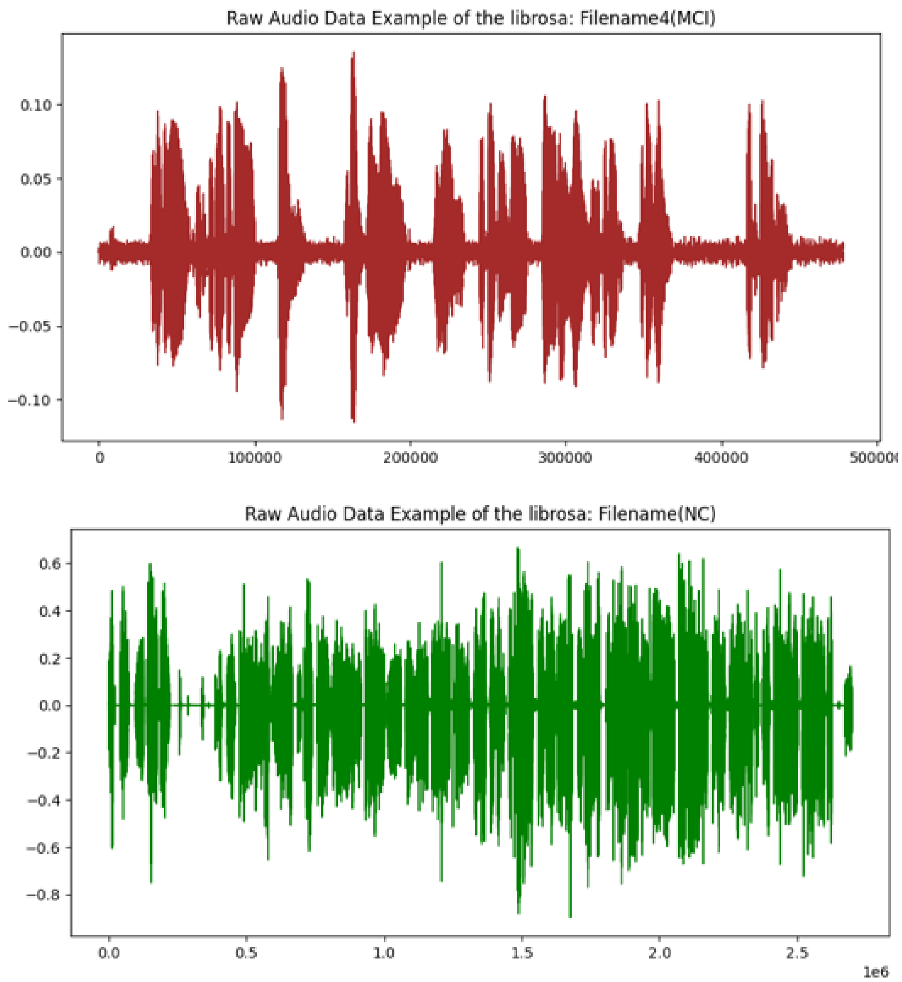


Figure 5. 3: MCI and NC audio waveform comparison

The difference observed in the two waveforms above can be seen in Figure 5.4, such as variations in amplitude, frequency. While these changes could indicate a change in speech patterns, they could be influenced by other factors such as the microphone quality, background noise and the speaker speaking manner. A thorough analysis is required.

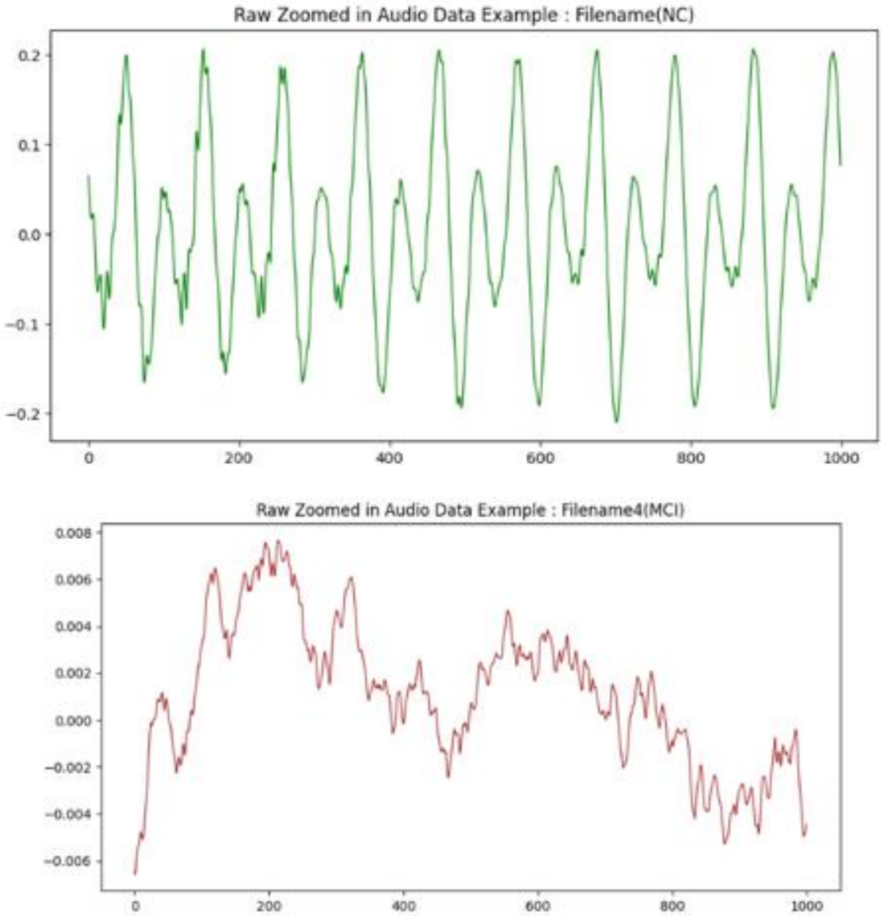


Figure 5. 4: Zoomed-in Audio MCI and NC comparison

In addition to waveforms, spectrograms were used to visualize the audio sample

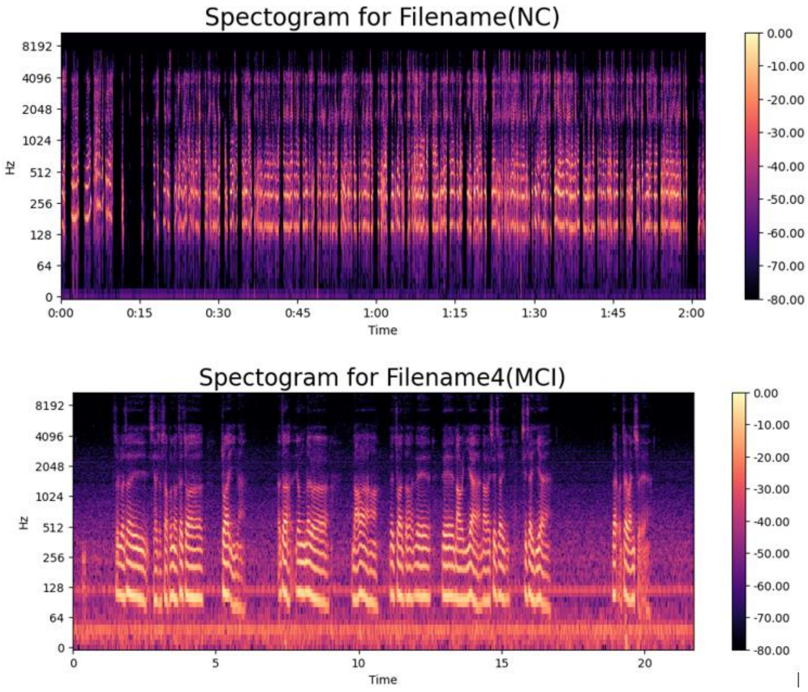


Figure 5. 5: MCI and NC spectrogram comparison

Just like waveforms, these spectrograms in Figure 5.5 reveal changes in speech patterns, but these variations could be influenced by different factors. In the last spectrogram, there seems to be some noise, noticeable through the scattered low-intensity frequencies across the spectrum. To ensure accurate analysis, it is necessary to clean the data.

5.2.2.1 Trim audio

Trimming audio is an important step to ensure that only the patient's voice is captured for analysis. In recordings longer than one minute, often included a doctor introducing the test, which became part of the audio and could interfere with the results. To address this a custom function called trim_audio was created to trim the first 20 seconds and last 5 seconds on longer recordings. For shorter audios only one second was trimmed, at the start and at the end. The final seconds were trimmed, since they mostly contained silence or electronic noise. Soundfile library was used to trim the audios, improving the quality of the audio but also making the analysis more efficient and reliable for the classification task. An example of an output of the trimming process is as shown below

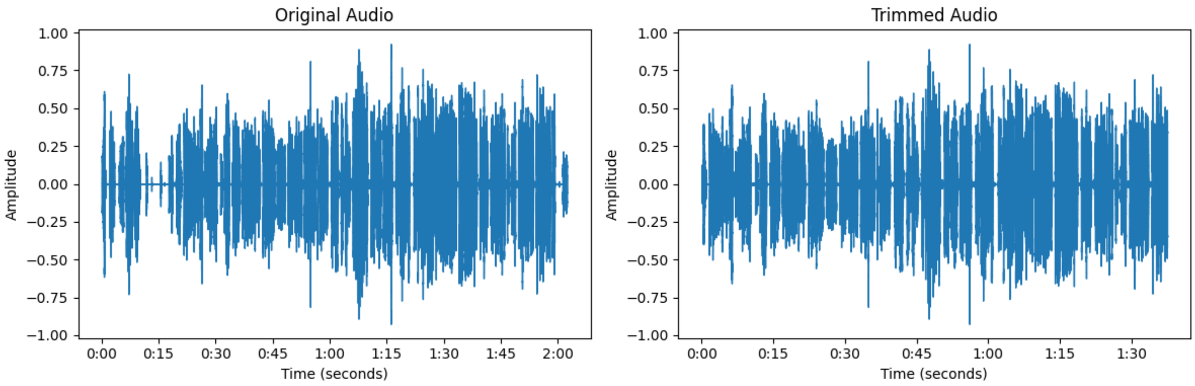


Figure 5. 6: Trimmed Recording

5.2.2.2 Noise reduction

After a successful trim, the next crucial step is noise reduction to enhance the clarity of the recordings. Background noise such as microphone static can interfere with accurate analysis. To address this, a custom function called noise_removal was created and it used the noisereduce library to filter out unwanted noise while preserving the key speech characteristics necessary for classification. Noise is removed by estimating a noise profile from the 0.5 seconds of audio and applying noisereduce.reduce_noise() to filter it out. An example of an output of the noise reduction process is as shown below

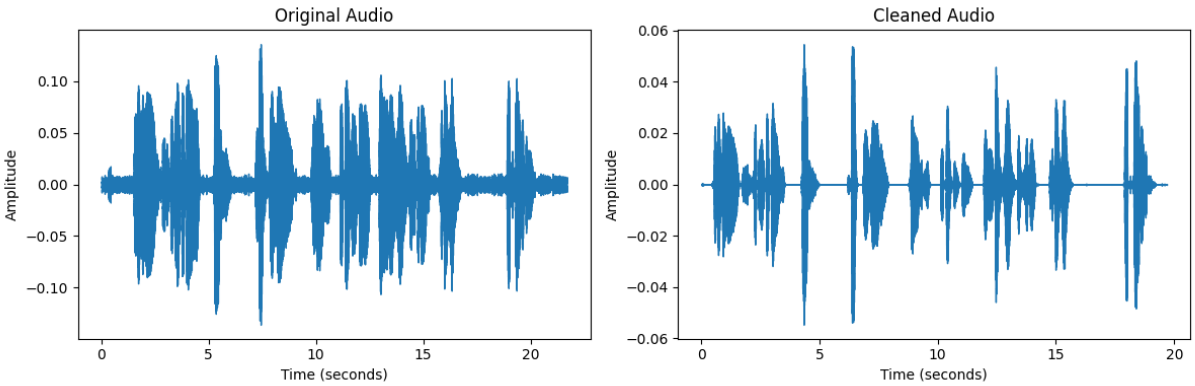


Figure 5. 7: Waveform depicting noise removal

5.2.2.3 Audio segmentation

Previous research suggests that 30 seconds is the optimal audio length for classification tasks, as it provides sufficient information for analysis while maintaining efficiency. Since the dataset contains audio recordings of varying lengths, standardizing them to 30 seconds ensures consistency. This balance helps preserve key speech features while avoiding excessively long or redundant segment . To achieve this, a custom function called split_audio was developed using the AudioSegment library, allowing each recording to be segmented into manageable 30-seconds clips. An output of the audio segmentation process is as shown below

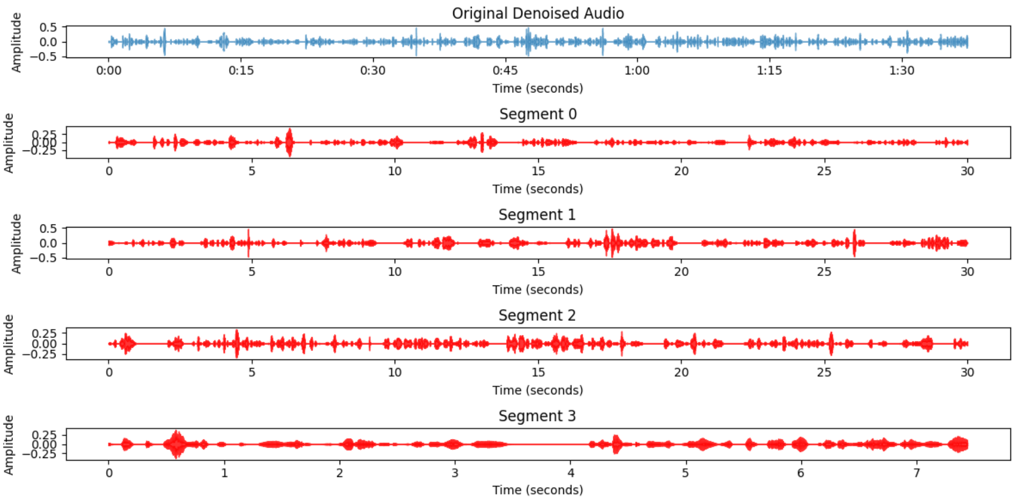


Figure 5. 8: Audio Segmentation

After segmenting the recordings, the total number of audio files increased to 825 with 438 clips being exactly 30 seconds long. The standardized 30-seconds segments were stored in a separate folder, and a new csv file was created to organize and track the updated dataset. Due to the class imbalance identification, data augmentation was done so as to improve the performance of the model. Shorter random splicing and smoothing…new samples from the shorter audio samples and smoothing it to make it more natural

5.2.2.4 Resampling and audio normalization

Resampling audio recordings to 16KHz , ensures consistency across all recordings and optimizes processing efficiency. Many speech-processing models are trained on 16 kHz audio, making it a standard choice for speech analysis and classification tasks. Higher sample rates capture more details, increasing computational complexity without improving performance. In addition to resampling, audio normalization must be done to ensure that all recordings have consistent loudness level, preventing variations that could affect the accuracy of classifications. This inconsistency can come up when different microphones are used in recording. By normalizing, we enhance the reliability of the dataset, making speech analysis more effective.

To normalize audio the target decibels are set to -3.0 as it provides a balance between loudness and avoiding distortion. Audio recordings often vary in loudness due to different recording environments or devices. Normalization ensures that all audio inputs are brought to a similar loudness level. In the normalize_audio custom function, the root mean square is calculated by measuring the average energy of the signal and is then converted to decibels to determine the current loudness. An adjustment factor is calculated based on the difference between the target decibels and the current loudness, which is then used to scale the audio waveform ensuring that all recordings maintain a standardized amplitude without distortion. Figure 5.9 below shows the function used to normalize the recordings. This step is important as it ensures that the model learns meaningful patterns that are consistent and clean. Without normalization the model might pick irrelevant volume differences as signals, leading to reduced accuracy.

```
[ ] target_dBFS = -3.0

    def normalize_audio(file_path, output_path):
        try:
            # Load audio
            audio, sr = librosa.load(file_path, sr=16000)
            rms = np.sqrt(np.mean(audio**2))
            current_dBFS = 20 * np.log10(rms)
            adjustment_factor = 10 ** ((target_dBFS - current_dBFS) / 20)
            normalized_audio = audio * adjustment_factor

            sf.write(output_path, normalized_audio, sr)


        except Exception as e:
            print(f"Error processing {file_path}: {e}")

    # Normalize all audio files in the folder
    for filename in os.listdir(segmented_audio_dir):
        if filename.endswith(".wav"):
            file_path = os.path.join(segmented_audio_dir, filename)
            output_path = os.path.join(segmented_audio_dir, filename)
            normalize_audio(file_path, output_path)

    print(" complete!")

    complete!
```

Figure 5. 9:Audio Normalization

5.2.3 Feature Extraction

Using the pre-processed recordings, the extracted features are extracted to populate a csv file that will be used in training the model. The universal features extracted are as follows: fundamental frequency, jitter, shimmer and Mel-frequency cepstral coefficient (MFCC). As mentioned in chapter 2, these features are affected with the MCI condition, making them suitable for a classification task. To ensure that the features used are efficient, the praat-parselmouth library is used.

The first feature extracted is the fundamental frequency (F0) which refers to the pitch of speech that is produced by the vibration of the vocal cords and varies based on a person's characteristics, such as, gender and vocal health. A custom function was created to extract the

F0 values from a recording using the parselmouth library. One audio can have more than 1000 fundamental frequency , as shown in the figure below. Processing all of them would be impractical. To make the data more manageable, the mean and standard deviation of the fundamental frequency was calculated. Mean F0 provides an overall measure of an individual speech , with higher value represents a higher overall pitch. Variability in F0 values , represented by the standard deviation, may indicate irregular speech patterns which can be important in detecting neurological conditions

```
The number of extracted f0 are 1749
First 10 Extracted Fundamental Frequency (F0) values:
275.79093574597005
275.99539596421084
84.84345033560895
87.78212599920056
86.13332282553924
84.71426753659796
164.18682115191305
164.12035201002092
167.78320062158554
171.40883728527044
Mean F0: 172.63427635197493
```

Figure 5. 10: Fundamental Frequency values

Jitter and shimmer were also extracted from the recordings alongside F0. Jitter refers to the cycle-to -cycle variations in F0 while shimmer refers to the fluctuations in amplitude. When extracting the jitter feature, the mean, standard deviation and 90^{th} percentile was calculated. These metrics assist in capturing useful information of the frequency. The meaning represents how pitch varies, standard deviation represents the spread of these variations while 90^{th} percentile represents extreme fluctuations, that are clinically detected as speech irregularities. For the shimmer feature, the focus is majorly on the mean and standard deviation. The mean shimmer provides the overall measure of the amplitude variation while the standard deviation captures the inconsistency. Lastly MFCC re extracted as a feature to capture the spectral characteristics of speech. 13 MFCCs were extracted to represent both low and high frequency components to capture changes in articulation, prosody and vocal stability, which are common indicators of MCI.

5.2.4 Model training

To dataset was split into an 80:20 ratio where 80% of the data was used for training and 20% reserved for testing to evaluate the model's performance. Splitting the data ensures that the model is not just memorizing the input but is learning to generalize. By splitting the data, the model can generalize to new users and provides an objective way to measure model accuracy, precision , and other performance metrics. Figure 5.11 displays the code that was used to split the dataset.

```
y = df["dx"]
X = df.drop(columns=["dx"])

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

Figure 5. 11: Training of the model

Normalization of data was done based on the distribution of the data. It is usually between 0 and 1 or -1 and -1 making it easier for machine learning to process. The Z-score normalization was used where data is standardized by adjusting the mean to 0 and the standard deviation to 1. This helps neural networks process inputs more efficiently, especially when features follow a normal distribution. It ensures consistent scaling, preventing certain features from dominating the learning process. The study focused on training two Neural Networks and performed a comparison with other models to get the baseline performance.

5.2.4.1 Random Forest

This model was trained to get a baseline to the model's performance. To build the Random Forest Classifier, a series of steps were carried out to train and evaluate the model. First, the classifier was initialized using the RandomForestClassifier function with the random_state parameter set to 42 ensuring consistent and reproducible results. After training the model prediction was made using the predict function. Figure 5.12 shows the training process of the Random Forest Model

```
rf_model = RandomForestClassifier(random_state=42)
rf_model.fit(X_train, y_train)
rf_preds = rf_model.predict(X_test)
rf_acc = accuracy_score(y_test, rf_preds)
```

Figure 5. 12: Random Forest Model

The classification report is as shown below. Precision indicates how many of the instances predicted to belong to a certain class is accurate. For class 0, which represents the label for MCI diagnosis, were accurately classified. On the other hand, recall measures the proportion of actual positive cases that were successfully identified by the model. For class 0, this metric reflects how effectively the model detected true cases of MCI.

```
Random Forest Accuracy: 0.8351
              precision    recall  f1-score   support

           0       0.80      0.90      0.85        49
           1       0.88      0.77      0.82        48

    accuracy                           0.84        97
   macro avg       0.84      0.83      0.83        97
weighted avg       0.84      0.84      0.83        97
```

Figure 5. 13: Random Forest Classification Results

5.2.4.2 Logistic Regression

Logistic Regression is a statistical method commonly employed in binary classification task, as it is designed to estimate the probability of an instance belonging to one of two possible classes. The Logistic Regression model is initialized with two key parameters, random_state and max_iter. Random_state ensures reproducibility while the max_iter parameter is used to set the maximum number of iterations for the algorithm to converge. The random_state was set to 42 and max_iter set to 1000

```
logreg = LogisticRegression(random_state=42, max_iter=1000)
logreg.fit(X_train, y_train)
y_pred = logreg.predict(X_test)
accuracy = accuracy_score(y_test, y_pred)
print(f"Logistic Regression Accuracy: {accuracy:.4f}")
print(classification_report(y_test, y_pred))
```

Figure 5. 14: Logistic Regression

The classification results for the Logistic Regression model, as shown in the report an overview of the model's performance on the test data. The overall accuracy of the model is 65.9%, meaning that approximately 66% of the instances in the test set were correctly classified

```
Logistic Regression Accuracy: 0.6598
              precision    recall  f1-score   support

           0       0.65      0.69      0.67        49
           1       0.67      0.62      0.65        48

    accuracy                           0.66        97
   macro avg       0.66      0.66      0.66        97
weighted avg       0.66      0.66      0.66        97
```

Figure 5. 15: Logistic Regression classification results

5.2.4.2 Support Vector Machine

Support Vector Machines offer insights into how effectively SVMs performed on the test data. In this study, the SVM model was initialized as shown in Figure 5.16. The SVM model looks for the optimal hyperplane by maximizing the margin between the support vectors, which are ethe data points that as closest to the boundary. The SVM model performed better than the Logistic Regression model by achieving an accuracy of 79.38%

```
svm_model = SVC()
svm_model.fit(X_train, y_train)
svm_preds = svm_model.predict(X_test)
svm_acc = accuracy_score(y_test, svm_preds)
```

Figure 5. 16: Support Vector Machine

```
SVM Accuracy: 0.7938
              precision    recall  f1-score   support

           0       0.65      0.69      0.67        49
           1       0.67      0.62      0.65        48

    accuracy                           0.66        97
   macro avg       0.66      0.66      0.66        97
weighted avg       0.66      0.66      0.66        97
```

Figure 5. 17: SVM model Analysis

5.2.4.3 CNN-LSTM Model

Th hybrid CNN-LSTM model combines the strength of Convolutional Neural Network and Long Short-Term Memory Networks, making it effective for sequential data tasks . This study used one dimensional (1D) convolutional layer, which applies 64 filters of size 2 to extract local patterns from the input data. This layer is followed by a Rectified Linear Unit (ReLU) activation function that introduced non-linearity helping the model to capture more complex patterns. The input shape was set to (19, 1), indicating 19 steps with a single feature per step. A Batch Normalization layer follows which normalizes the outputs, stabilizes training, and speeds up convergence. The MaxPoolingID layer selects the maximum value from each 2 element simplifying the data and helping mitigating overfitting. The model then transitions to the LSTM layer that has 128 units and is set to output only the final hidden state. To handle overfitting at this step, a Dropout layer is used to deactivate 50% of the neurons, The hyperparameter of the model were carefully adjusted to optimize performance and achieve higher accuracy and precision. However, due to the stochastic nature of model training, the model's accuracy varied between runs with the observed range accuracy documented at 60 percent to 68 percent.

```
#cnn-lstm
model = Sequential([
    Conv1D(filters=64, kernel_size=2, activation='relu', input_shape=(19, 1)),
    BatchNormalization(),
    MaxPooling1D(pool_size=2),
    LSTM(128, return_sequences=False),
    Dropout(0.5),
    Dense(32, activation='relu'),
    Dense(1, activation='sigmoid')
])
```

Figure 5. 18 :CNN-LSTM Model

```
Test Accuracy: 0.6392
4/4 ━━━━━━━━━━━━━━━━━━━━ 0s 75ms/step
              precision    recall  f1-score   support

           0       0.69      0.51      0.59        49
           1       0.61      0.77      0.68        48

    accuracy                           0.64        97
   macro avg       0.65      0.64      0.63        97
weighted avg       0.65      0.64      0.63        97
```

Figure 5. 19: CNN-LSTM model accuracy

5.2.4.4 Deep Neural Networks

Deep Neural Network is a type of artificial neural network that consists of multiple layers of interconnected neurons that enable it to learn complex patterns. In this study , the model consists of multiple layers, each contributing to learning and improving performance. It starts with an input layer where an input shape ensures that the input matches the number of features in the training data. The hidden layers have different numbers of neurons , each using the ReLU activation function that allows the model to learn complex data. BatchNormalization layer stabilizes the training process while the Dropout layer prevents the model from becoming reliant on the training data . This combination ensures that the model is balanced by reducing overfitting. The model achieved an accuracy of 84% with an F1 score of 0.84 that indicated the model was not biased towards a class. The accuracy range fluctuates between 78% to 86%.

```
improved_model = Sequential([
    Dense(128, activation='relu', input_shape=(X_train.shape[1],)),
    BatchNormalization(),
    Dropout(0.2),

    Dense(64, activation='relu'),
    BatchNormalization(),
    Dropout(0.2),

    Dense(32, activation='relu'),
    BatchNormalization(),
    Dropout(0.2),

    Dense(16, activation='relu'),
    BatchNormalization(),
    Dropout(0.2),

    Dense(1, activation='sigmoid')
])
```

Figure 5. 20: Deep Neural Network

```
Test Accuracy: 0.8454
4/4 ━━━━━━━━━━━━━━━━━━━━ 0s 60ms/step
              precision    recall  f1-score   support

           0       0.81      0.90      0.85        49
           1       0.88      0.79      0.84        48

    accuracy                           0.85        97
   macro avg       0.85      0.84      0.84        97
weighted avg       0.85      0.85      0.84        97
```

Figure 5. 21:DNN classification results

A confusion matrix , a performance evaluation tool used in classification, was created to visualize the accuracy of the DNN model's prediction. Figure 5.22 demonstrates a crucial tradeoff in early detection. The model was able to identify 42 cases out of 48 cases of MCI, minimizing the false negatives to 6, which is essential for early intervention. However, this affects the number of false positives, where healthy individuals were flagged for further testing. This tradeoff is necessary to ensure that potential MCI cases are not overlooked.

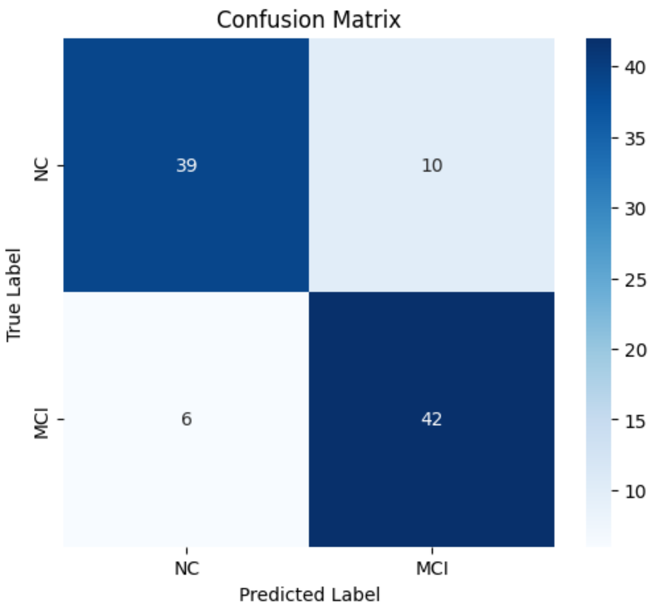
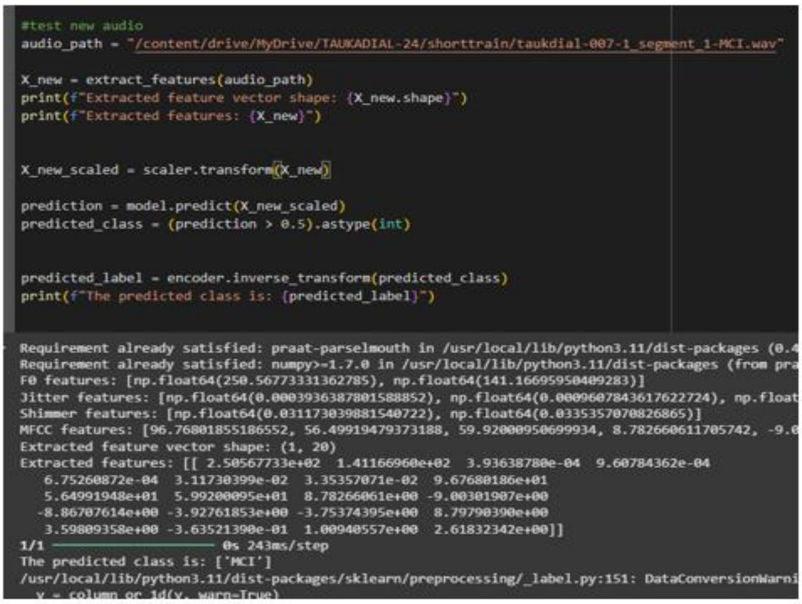


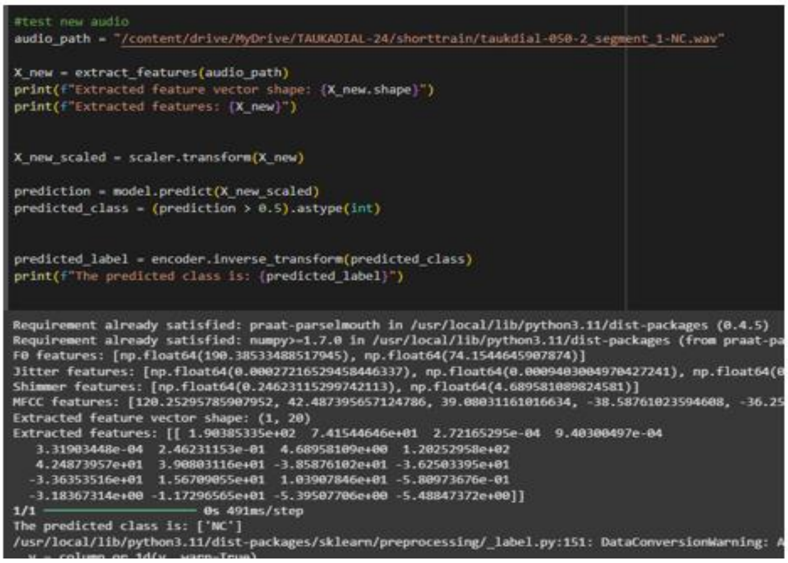
Figure 5. 22: DNN Confusion Matrix

The model's test predictions are presented on the figure below, with the diagnosis being appended on the audio file names for easier identification. The model demonstrated a strong ability to classify the provided audio samples accurately, effectively predicting the diagnosis based on the extracted features.



```
#test new audio
audio_path = "/content/drive/MyDrive/TAUKADIAL-24/shorttrain/taukdial-007-1_segment_1-MCI.wav"

X_new = extract_features(audio_path)
print(f"Extracted feature vector shape: {X_new.shape}")
print(f"Extracted features: {X_new}")


X_new_scaled = scaler.transform(X_new)

prediction = model.predict(X_new_scaled)
predicted_class = (prediction > 0.5).astype(int)


predicted_label = encoder.inverse_transform(predicted_class)
print(f"The predicted class is: {predicted_label}")
```

```
Requirement already satisfied: praat-parselmouth in /usr/local/lib/python3.11/dist-packages (0.4
Requirement already satisfied: numpy>=1.7.0 in /usr/local/lib/python3.11/dist-packages (from pra
F0 features: [np.float64(250.56773331362785), np.float64(141.16695950409283)]
Jitter features: [np.float64(0.0003936387801588852), np.float64(0.0009607843617622724), np.float
Shimmer features: [np.float64(0.031173039881540722), np.float64(0.0335357070826865)]
MFCC features: [96.76801855186552, 56.49919479373188, 59.92000950699934, 8.782660611705742, -9.0
Extracted feature vector shape: (1, 20)
Extracted features: [[ 2.50567733e+02  1.41166960e+02  3.93638780e-04  9.60784362e-04
   6.75260872e-04  3.11730399e-02  3.35357071e-02  9.67680186e+01
   5.64991948e+01  5.99200095e+01  8.78266061e+00 -9.00301907e+00
  -8.86707614e+00 -3.92761853e+00 -3.75374395e+00  8.79790390e+00
   3.59809358e+00 -3.63521390e-01  1.00940557e+00  2.61832342e+00]]
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 243ms/step
The predicted class is: ['MCI']
/usr/local/lib/python3.11/dist-packages/sklearn/preprocessing/_label.py:151: DataConversionWarni
  y = column_or_1d(y, warn=True)
```



```
#test new audio
audio_path = "/content/drive/MyDrive/TAUKADIAL-24/shorttrain/taukdial-050-2_segment_1-NC.wav"

X_new = extract_features(audio_path)
print(f"Extracted feature vector shape: {X_new.shape}")
print(f"Extracted features: {X_new}")


X_new_scaled = scaler.transform(X_new)

prediction = model.predict(X_new_scaled)
predicted_class = (prediction > 0.5).astype(int)


predicted_label = encoder.inverse_transform(predicted_class)
print(f"The predicted class is: {predicted_label}")
```

```
Requirement already satisfied: praat-parselmouth in /usr/local/lib/python3.11/dist-packages (0.4.5)
Requirement already satisfied: numpy>=1.7.0 in /usr/local/lib/python3.11/dist-packages (from praat-pa
F0 features: [np.float64(190.38533488517945), np.float64(74.1544645907874)]
Jitter features: [np.float64(0.0002721652945844637), np.float64(0.0009403004970427241), np.float64(0
Shimmer features: [np.float64(0.24623115299742113), np.float64(4.689581089824581)]
MFCC features: [120.25295785907952, 42.4873956571247 86, 39.08031161016634, -38.58761023594608, -36.25
Extracted feature vector shape: (1, 20)
Extracted features: [[ 1.90385335e+02  7.41544646e+01  2.72165295e-04  9.40300497e-04
   3.31903448e-04  2.46231153e-01  4.68958109e+00  1.20252958e+02
   4.24873957e+01  3.90803116e+01 -3.85876102e+01 -3.62503395e+01
  -3.36353516e+01  1.56709055e+01  1.03907846e+01 -5.80973676e-01
  -3.18367314e+00 -1.17296565e+01 -5.39507706e+00 -5.48847372e+00]]
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 491ms/step
The predicted class is: ['NC']
/usr/local/lib/python3.11/dist-packages/sklearn/preprocessing/_label.py:151: DataConversionWarning: A
  y = column_or_1d(y, warn=True)
```

Figure 5. 23: DNN Model's test classification

5.3 User Interface

An interactive interface was developed to enable users to record their voices, which are then uploaded to the model for analysis as shown in Figure 5.24. Upon logging in, users are presented with a screen that includes a Cookie Theft picture – a widely used tool in critical assessments. This image serves as a stimulus for eliciting speech, allowing the user to describe what they see. Their spoken response is recorded and analyzed by the model. Once the user stops the recording, the system processes the input and displays both the prediction results and appropriate recommendations.

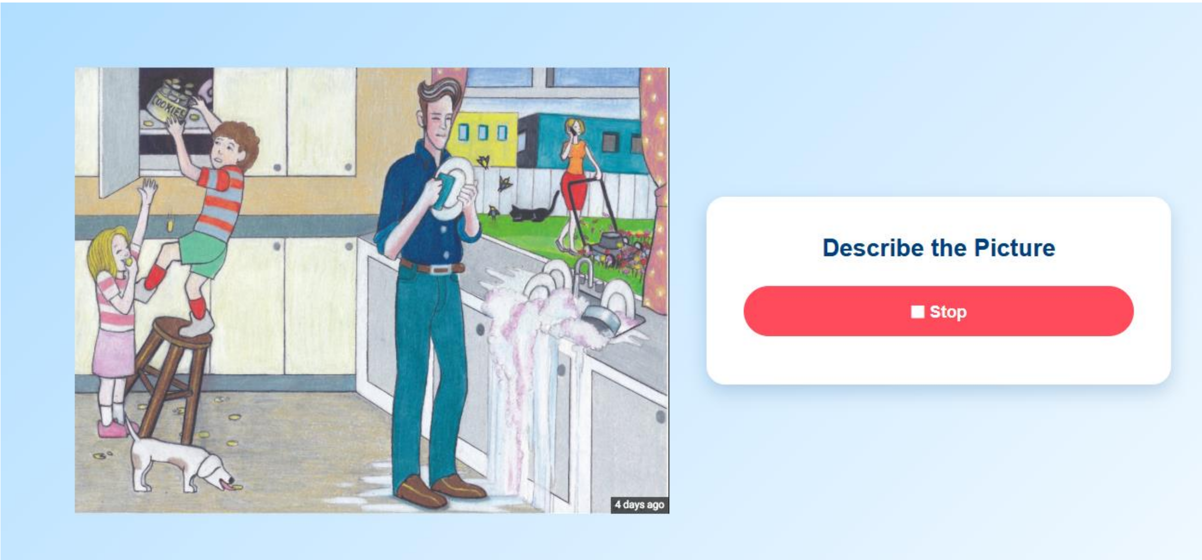


Figure 5. 24: Voice Recording page

The model classifies the recordings into one of two categories: NC (Normal Control) or MCI (Mild Cognitive Impairment). If the prediction indicates MCCI, a message is displayed to seek further medical evaluation, as diagnosing MCI involves multiple factors beyond speech analysis. Figure 5.25 shows the prediction displayed in both cases.

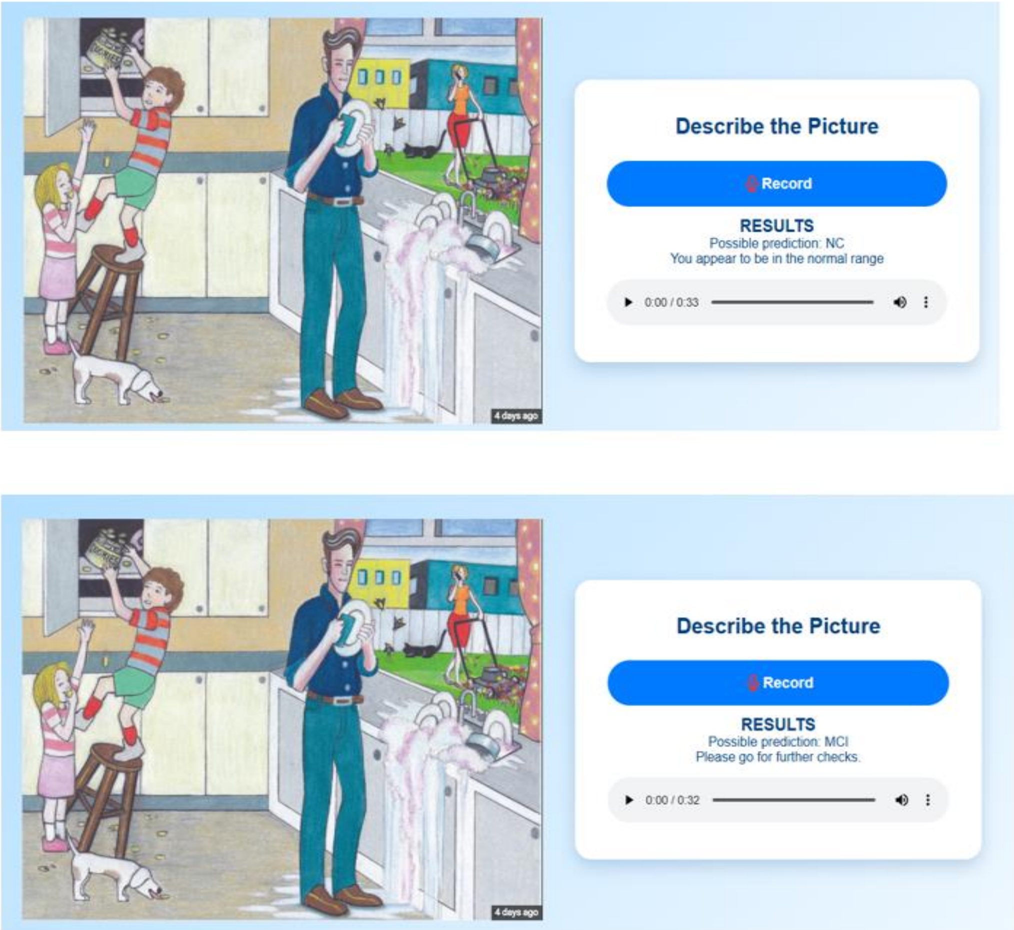


Figure 5. 25: Interface displaying Model's results

Chapter 6: Discussion of Results

6.1 Discussion of Experimental Results

The SVM model, which served as one of the baseline models, initially achieved a relatively high accuracy of 79.38%. However, after performing cross-validation, the accuracy dropped, suggesting that the initial performance may have been overfitting on the training data. While the precision and recall were balanced, averaging around 0.66, the model struggled to generalize beyond the dataset, which became evident after validation. This highlights the limitations of simpler models like SVM when applied to more complex classification problems, especially when dealing with structured audio data.

On the other hand, the CNN-LSTM model, despite its advanced architecture, did not perform as well as expected. It achieved an accuracy of 63.92% on the test set, which is significantly lower than the SVM. One potential reason for this underperformance could be that CNN-LSTMs, while effective for unstructured image or time-series data, are prone to overfitting when the data lacks sufficient diversity and is in tabular format. The high recall for MCI (0.77) suggests that the model could detect many positive cases but at the expense of precision, leading to more false positives. This emphasizes that more complex models are not always better, and they need to be matched with the right type of data. The CNN-LSTM model had low accuracy levels, ranging from 60 to 68 percent even after tweaking its architecture and validating the results with stratified K-fold cross-validation, the accuracy remained moderate. This could be because the data used was structured (consisting of features like MFCCs, jitter, and shimmer), and CNN-based models are more prone to overfitting, especially when dealing with smaller or imbalanced datasets.

On the other hand, the Deep Neural Network (DNN) performed much better, achieving accuracy as high as 84%, especially after addressing class imbalance with data augmentation. This is likely because the DNN architecture suited the structured nature of the input features and extracted meaningful patterns without overcomplicating the learning process. The DNN model performed noticeably better compared to the SVM and CNN-LSTM models, achieving a test accuracy of 84.54%. This higher accuracy shows that DNN handled the structured audio data more effectively and generalized better without overfitting. The model was evaluated using stratifies 5-fold cross validation to ensure balanced class distribution across folds and reduce overfitting. Performance metrics included accuracy, precision, recall, and F1-score.

The precision, recall, and F1-scores are well-balanced, with an average of around 0.85, which means the model consistently classified both MCI and normal control cases. The confusion matrix plotted showed higher false positives than false negatives. This was crucial since in early diagnosis, prioritizing high recall for MCI cases is more important than eliminating false positives. This caused a reasonable balance for early detection.

For class 1 (MCI), the model achieved a precision of 0.88, meaning it reduced false positives, while the recall of 0.79 indicates that it still captured most of the actual positive cases. Class 0 (normal control) had a recall of 0.90, showing that true negatives were classified correctly. The balance between precision and recall for both classes shows that DNN learned important distinguishing features from the dataset better than the other models. This stronger performance may be due to the deeper architecture of the DNN, which could handle the complexity of the data more effectively than simpler models like SVM or CNN-LSTM.

6.2 Dataset Limitation and Model Validity

While the dataset provided a valuable , clinically annotated dataset, it has limitations. The dataset included a relatively small amount of available audio data and the fact that it consists primarily of semi-structured narrative speech tasks. As a result, the model trained on this dataset may affect the overall generalizability and robustness of the model, underscoring the need for further validation using more diverse and representative datasets.

To improve cross-population validity, this study focused on universal acoustic features that are independent of language content and have shown in literature to correlate with cognitive impairment across languages. However, variations in speech rhythm and intonation between speakers of different linguistic and cultural backgrounds may still affect model performance.

6.3 Comparison with Previous Studies

A study conducted by Tóth et al., (2018), that achieved an F1-score of 78.8% using acoustic parameters such as speech tempo, hesitation ration, and pause metrics as features for a model developed using Random Forest algorithm. In contrast, this research achieved a higher F1-score of 85%, using DNN model trained on language-independent acoustic features. This improvement in performance indicates that the use of universal, low level acoustic features may enhance the model’s generalizability and robustness, especially across multilingual datasets, as it avoids dependance on language-specific content The results suggest that deep learning models leverage universal speech characteristics can outperform traditional machine learning approaches in distinguishing MCI from healthy controls, potentially offering a more practical and efficient screening tool.

Bertini et al., (2021), conducted a study which implemented an autoencoder neural network trained in spectrograms of speech signals to classify cognitive impairment. The approach used achieved a notable high accuracy of 90.57%, benefiting from visual audio representation. While the spectrogram-based method benefits from capturing detailed spectotemporal patterns and achieving higher overall accuracy, the feature-based model offers key advantages in interpretability and computational efficiency. These traits are particularly valuable in resource-limited settings, where manually annotated speech data may not be feasible. To detect dementia, a study conducted by Kumar et al., (2022) focused on using a compact set of speech features such as voice quality and cepstral featuress, to distinguish between healthy individuals and those with cognitive impairment. This study achieved a slightly higher accuracy of 87.6%, reinforcing the effectiveness of carefully selected speech features in the early detection of dementia.

6.4 Review of Research Objectives

In Chapter 2.2, the characteristics, signs and symptoms of MCI was explored by reviewing various scholarly articles and medical studies to identify the key indicators of MCI. The studies highlighted that speech impairments are among the indicators of MCI, distinguishing affected individuals from those with normal cognition. Recognizing the significance of these changes in speech, provides the foundation for further research into speech-based screening approaches. The model developed validated these findings as the results demonstrated differences in speech patterns between MCI individuals and NC individuals.

An extensive review of both clinical and self-screening tools was reviewed in Chapter 2.3. The analysis reviewed that while these tools are essential diagnostic aids, they are often susceptible to subjectivity, personal biases and underreporting of symptoms, which can contribute to misdiagnosis cases. One of the most widely tools, the Mini-Mental State Examination, was found to be influenced by an individual’s educational background, leading to disparities in assessment outcomes. Recognizing these limitations, the study sought an alternative approach that could provide a more objective, accessible, and unbiased screening method. The speech-based classification model developed does not rely on subjective selfreports or written assessments, instead it utilizes natural speech patterns to distinguish between individuals with MCI and those with normal control.

The third objective was to develop a speech-based classification model that could distinguish between the two groups. Dataset from DementiaBank was utilized, containing audio recordings from individuals diagnosed with MCI and those with normal cognition. Unlike previous studies discussed in chapter 2.5, the developed model focused exclusively on universal features such as jitter, shimmer, fundamental frequency and Mel-frequency Cepstral Coefficient. These features capture essential aspects of speech production without depending on the language of the speaker. This ensured robustness and generalization of the model. The Deep Neural Network algorithm was used to develop the model, due to its stability and ability to learn complex speech patterns. In addition, the model performed better than the other models tested in the development stage. The model demonstrated a higher classification accuracy than the tested models, reinforcing the potential of speech-based deep learning models as a reliable tool for MCI screening.

To validate the developed model, testing was conducted to assess its accuracy, recall, and overall performance. The model achieved an accuracy of 84 percent, indicating strong ability to classify individuals as either MCI or NC. Additionally, the model demonstrated a high recall, meaning it effectively identified MCI cases while minimizing false negatives. This is particularly crucial in medical screening, where misclassifications can lead to missed diagnoses and delayed interventions. The combination of high accuracy and recall suggests that the model had relatively few misclassifications, making it a reliable and efficient tool for MCU screening.

6.5 Limitations of the Model

Despite the model's promising performance, the developed model has some limitations that must be acknowledged

i. While the model provides a valuable screening tool, it is not intended to serve as a definitive diagnostic tool as diagnosing MCI involves neuropsychological testing, medical history assessment and expert evaluation, rather than relying entirely on speech analysis.

ii. The model's high recall ensured that MCI cases are accurately identified, but it comes at the expense of lower specificity, resulting in a higher number of false positives.

Chapter 7: Conclusion, Recommendations and Future Work

7.1 Conclusions

Speech has been recognized as a biomarker of MCI which is affected at the early stage of the disease. Early detection is crucial in slowing the progression to dementia. The developed model can be used as a self-screening tool that could empower individuals to seek medical attention sooner, since research suggests that one of the major barriers to timely intervention is denial. By providing an accessible and objective screening tool, the model has potential to support early intervention efforts and improve patient outcomes. The model utilized universal features and achieved a favorable accuracy of 84 percent without relying on language dependent features. This approach enhances accessibility across diverse populations. Unlike studies relying on language-specific or transcribed data, this research focused on language-independent acoustic features , which enhanced generalizability and ease of deployment across multilingual settings. The approach assumes that changes in acoustic features directly correspond to cognitive decline, but speech patterns can also be influenced by factors such as cultural speech norms, emotional state, or comorbidities, which were not fully controlled in this study.

While some models in the literature achieved slightly higher accuracy, they often relied on more resource-intensive inputs, whereas the developed solution offers a lightweight approach. The findings of this research affirm that acoustic changes in speech can serve as early indicators of cognitive decline. The simplicity and objectivity of the model make it suitable for self-screening applications, especially important considering that denial and lack of awareness remain barriers to early intervention. By making early screening more accessible, this model supports broader public health goals aimed at slowing progression to dementia through timely medical attention

7.2 Recommendations

Based on the findings of this study, it is recommended to use larger datasets to enhance the model's performance. Larger datasets improve the model's ability to generalize to unseen data, minimizing overfitting, which was evident in this study due to the limited data size. Additionally, exploring advanced data augmentation techniques, such as time-stretching, pitch shifting, and dynamic range adjustments, could diversify the training data and make the models more resilient to variations in input audio.

To further enhance the model’s accuracy and generalizability, I recommend the incorporation of additional universal features such as formants, voice onset time that could provide a deeper understanding into speech production changes associated with MCI. Integrating these features, alongside the existing ones, could improve the model’s accuracy.

7.3 Challenges encountered

During the development of the model some challenges were encountered that affected the model’s performance

i) Overlapping between MMSE scores and clinical diagnoses led to inconsistencies in labeling individuals as MCI or NC. Since MMSE is a widely used screening tool but not a definitive diagnostic tool, some individuals with borderline scores may have been misclassified, affecting the overall evaluation of the model.

7.4 Future Work

To further enhance the model’s real-world applicability, the following improvements can be explored:

i) Increasing the size of the dataset by incorporating more speech samples from varied demographics and languages will improve the model’s robustness and generalizability. A larger dataset might also help address the issues related to data imbalance and overlapping classifications.

ii) Enhancing the feature set by incorporating additional universal features could make the model more sensitive to early-stage MCI detection.

iii) To maximize its impact, the model could be incorporated into digital healthcare platforms for early MCI screening. It can be used as a tool that could assist healthcare providers in early detection and monitoring, facilitating timely medical intervention.

iv) As speech data involves sensitive personal information, future deployment should also consider data privacy frameworks and user content to ensure ethical implementation.

v) Gather feedback from healthcare professionals to refine interpretability and usability of the developed solution

References

Adana Díaz, L., Arango, A., Parra, C., Rodríguez-Lorenzana, A., & Yacelga-Ponce, T. (2021). Impact of Educational Level on Versions (Basic and Complete) of the Montreal Cognitive Assessment. *Dementia and Geriatric Cognitive Disorders*, *50*(4), 341–348. https://doi.org/10.1159/000518747

Al-Sharify, Z. T., Al-Sharify, T. A., Al-Sharify, N. T., & naser, H. Y. (2020). A critical review on medical imaging techniques (CT and PET scans) in the medical field. *IOP Conference Series: Materials Science and Engineering*, *870*(1), 012043. https://doi.org/10.1088/1757-899X/870/1/012043

Anderson, N. D. (2019). State of the science on mild cognitive impairment (MCI). *CNS Spectrums*, *24*(1), 78–87. https://doi.org/10.1017/S1092852918001347

Bai, W., Chen, P., Cai, H., Zhang, Q., Su, Z., Cheung, T., Jackson, T., Sha, S., & Xiang, Y.-T. (2022). Worldwide prevalence of mild cognitive impairment among community dwellers aged 50 years and older: A meta-analysis and systematic review of epidemiology studies. *Age and Ageing*, *51*(8), afac173. https://doi.org/10.1093/ageing/afac173

Boggatz, T. (Ed.). (2020). *Quality of Life and Person-Centered Care for Older People*. Springer.

Breton, A., Casey, D., & Arnaoutoglou, N. A. (2019). Cognitive tests for the detection of mild cognitive impairment (MCI), the prodromal stage of dementia: Meta-analysis of diagnostic accuracy studies. *International Journal of Geriatric Psychiatry*, *34*(2), 233–242. https://doi.org/10.1002/gps.5016

Brown, J. M., Wiggins, J., Dawson, K., Rittman, T., & Rowe, J. B. (2019). Test Your Memory (TYM) and Test Your Memory for Mild Cognitive Impairment (TYM-MCI): A Review and Update Including Results of Using the TYM Test in a General Neurology Clinic and Using a Telephone Version of the TYM Test. *Diagnostics*, *9*(3), 116. https://doi.org/10.3390/diagnostics9030116

Cardoso, S., Barros, R., Marôco, J., de Mendonça, A., & Guerreiro, M. (2024). Different MMSE domains are associated to cognitive decline and education. *Applied Neuropsychology: Adult*, *31*(4), 533–539. https://doi.org/10.1080/23279095.2022.2041018

Carson, N., Leach, L., & Murphy, K. J. (2018). A re-examination of Montreal Cognitive Assessment (MoCA) cutoff scores. *International Journal of Geriatric Psychiatry*, *33*(2), 379–388. https://doi.org/10.1002/gps.4756

Castellano, G., Esposito, A., Lella, E., Montanaro, G., & Vessio, G. (2024). Automated detection of Alzheimer's disease: A multi-modal approach with 3D MRI and amyloid PET. *Scientific Reports*, *14*(1), 5210. https://doi.org/10.1038/s41598-024-56001-9

Chan, C., Lee, T., Chen, L., Wu, J., Tu, J., Liu, Z., Cao, L., He, Y., Huang, J., Tao, J., & Wong, M. N. K. (2022). *Validation of the Efficient Online MCI Screening System (EOmciSS) for Screening of Mild Cognitive Impairment Among Community-dwelling Older Adults (Preprint)*. https://doi.org/10.2196/preprints.40858

Charbuty, B., & Abdulazeez, A. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. *Journal of Applied Science and Technology Trends*, *2*(01), Article 01. https://doi.org/10.38094/jastt20165

Dautzenberg, G., Lijmer, J., & Beekman, A. (2020). Diagnostic accuracy of the Montreal Cognitive Assessment (MoCA) for cognitive screening in old age psychiatry: Determining cutoff scores in clinical practice. Avoiding spectrum bias caused by healthy controls. *International Journal of Geriatric Psychiatry*, *35*(3), 261–269. https://doi.org/10.1002/gps.5227

De Neys, W. (2021). On Dual- and Single-Process Models of Thinking. *Perspectives on Psychological Science*, *16*(6), 1412–1427. https://doi.org/10.1177/1745691620964172

Derbie, A. Y., Dejenie, M. A., & Zegeye, T. G. (2022). Visuospatial representation in patients with mild cognitive impairment: Implication for rehabilitation. *Medicine*, *101*(44), e31462. https://doi.org/10.1097/MD.0000000000031462

Ding, H., Lister, A., Karjadi, C., Au, R., Lin, H., Bischoff, B., & Hwang, P. H. (2024). Detection of Mild Cognitive Impairment From Non-Semantic, Acoustic Voice Features: The Framingham Heart Study. *JMIR Aging*, *7*(1), e55126. https://doi.org/10.2196/55126

Gadhavi, M. (2022, September 28). *Rapid Application Development Guide 2024—Phases, Tools & Benefits*. Radixweb. https://radixweb.com/blog/introduction-to-rapid-application-development

Garrouste, C., & Perdrix, E. (2022). Is there a consensus on the health consequences of retirement? A literature review. *Journal of Economic Surveys*, *36*(4), 841–879. https://doi.org/10.1111/joes.12466

Gould, E. (2023, November 20). *Addressing the Shortage of Dementia Specialists | RTI*. https://www.rti.org/insights/tackling-dementia-specialist-shortage

Huang, L., Yang, H., Che, Y., & Yang, J. (2024). Automatic speech analysis for detecting cognitive decline of older adults. *Frontiers in Public Health*, *12*. https://doi.org/10.3389/fpubh.2024.1417966

Isaacson, R. S., & Saif, N. (2020). A Missed Opportunity for Dementia Prevention? Current Challenges for Early Detection and Modern-Day Solutions. *The Journal of Prevention of Alzheimer's Disease*, *7*(4), 291–293. https://doi.org/10.14283/jpad.2020.23

Islam, N., Hashem, R., Gad, M., Brown, A., Levis, B., Renoux, C., Thombs, B. D., & McInnes, M. D. (2023). Accuracy of the Montreal Cognitive Assessment tool for detecting mild cognitive impairment: A systematic review and meta-analysis. *Alzheimer's & Dementia*, *19*(7), 3235–3243. https://doi.org/10.1002/alz.13040

Jia, X., Wang, Z., Huang, F., Su, C., Du, W., Jiang, H., Wang, H., Wang, J., Wang, F., Su, W., Xiao, H., Wang, Y., & Zhang, B. (2021). A comparison of the Mini-Mental State Examination (MMSE) with the Montreal Cognitive Assessment (MoCA) for mild cognitive impairment screening in Chinese middle-aged and older population: A cross-sectional study. *BMC Psychiatry*, *21*(1), Article 1. https://doi.org/10.1186/s12888-021-03495-6

Jobson, D. D., Hase, Y., Clarkson, A. N., & Kalaria, R. N. (2021). The role of the medial prefrontal cortex in cognition, ageing and dementia. *Brain Communications*, *3*(3), fcab125. https://doi.org/10.1093/braincomms/fcab125

Judge, D., Roberts, J., Khandker, R., Ambegaonkar, B., & Black, C. M. (2019). Physician Perceptions about the Barriers to Prompt Diagnosis of Mild Cognitive Impairment and Alzheimer's Disease. *International Journal of Alzheimer's Disease*, *2019*(1), 3637954. https://doi.org/10.1155/2019/3637954

Lau, S.-Y. J., & Agius, H. (2021). A framework and immersive serious game for mild cognitive impairment. *Multimedia Tools and Applications*, *80*(20), 31183–31237. https://doi.org/10.1007/s11042-021-11042-4

Liu, Y., Jun, H., Becker, A., Wallick, C., & Mattke, S. (2024). Detection Rates of Mild Cognitive Impairment in Primary Care for the United States Medicare Population. *The Journal of Prevention of Alzheimer's Disease*, *11*(1), 7–12. https://doi.org/10.14283/jpad.2023.131

Malek-Ahmadi, M., & Nikkhahmanesh, N. (2024). Meta-analysis of Montreal cognitive assessment diagnostic accuracy in amnestic mild cognitive impairment. *Frontiers in Psychology*, *15*, 1369766. https://doi.org/10.3389/fpsyg.2024.1369766

Mansbach, W., Mace, R., Frangiosa, T., Biggar, V., Comer, M., & Simmons, S. (2020). A Path to Early Diagnosis of MCI and Dementia: Integrating myMemCheck Into the Primary Care Workflow. *Innovation in Aging*, *4*(Suppl 1), 261–262. https://doi.org/10.1093/geroni/igaa057.839

Martínez-Nicolás, I., Llorente, T. E., Martínez-Sánchez, F., & Meilán, J. J. G. (2021). Ten Years of Research on Automatic Voice and Speech Analysis of People With Alzheimer's Disease and Mild Cognitive Impairment: A Systematic Review Article. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.620251

Mattke, S., Jun, H., Chen, E., Liu, Y., Becker, A., & Wallick, C. (2023). Expected and diagnosed rates of mild cognitive impairment and dementia in the U.S. Medicare population: Observational analysis. *Alzheimer's Research & Therapy*, *15*, 128. https://doi.org/10.1186/s13195-023-01272-z

Mueller, K. D., Hermann, B., Mecollari, J., & Turkstra, L. S. (2018). Connected speech and language in mild cognitive impairment and Alzheimer's disease: A review of picture description tasks. *Journal of Clinical and Experimental Neuropsychology*, *40*(9), 917–939. https://doi.org/10.1080/13803395.2018.1446513

Myrberg, K., Hydén, L.-C., & Samuelsson, C. (2020). The mini-mental state examination (MMSE) from a language perspective: An analysis of test interaction. *Clinical Linguistics & Phonetics*, *34*(7), 652–670. https://doi.org/10.1080/02699206.2019.1687757

Pellicer-Espinosa, I., & Díaz-Orueta, U. (2022). Cognitive Screening Instruments for Older Adults with Low Educational and Literacy Levels: A Systematic Review. *Journal of Applied Gerontology*, *41*(4), 1222–1231. https://doi.org/10.1177/07334648211056230

Pillny, M., Krkovic, K., Buck, L., & Lincoln, T. M. (2022). From Memories of Past Experiences to Present Motivation? A Meta-analysis on the Association Between Episodic Memory and Negative Symptoms in People With Psychosis. *Schizophrenia Bulletin*, *48*(2), 307–324. https://doi.org/10.1093/schbul/sbab120

Prats-Sedano, M. A., Savulich, G., Surendranathan, A., Donaghy, P. C., Thomas, A. J., Rowe, J. B., Su, L., & O'Brien, J. T. (2021). The revised Addenbrooke's Cognitive Examination can facilitate differentiation of dementia with Lewy bodies from

Alzheimer's disease. *International Journal of Geriatric Psychiatry*, *36*(6), 831–838. https://doi.org/10.1002/gps.5483

Ramanarayanan, V., Lammert, A. C., Rowe, H. P., Quatieri, T. F., & Green, J. R. (2022). Speech as a Biomarker: Opportunities, Interpretability, and Challenges. *Perspectives of the ASHA Special Interest Groups*, *7*(1), 276–283. https://doi.org/10.1044/2021_PERSP-21-00174

Robin, J., Harrison, J. E., Kaufman, L. D., Rudzicz, F., Simpson, W., & Yancheva, M. (2020). Evaluation of Speech-Based Digital Biomarkers: Review and Recommendations. *Digital Biomarkers*, *4*(3), 99–108. https://doi.org/10.1159/000510820

Sabbagh, M. N., Boada, M., Borson, S., Chilukuri, M., Dubois, B., Ingram, J., Iwata, A., Porsteinsson, A. P., Possin, K. L., Rabinovici, G. D., Vellas, B., Chao, S., Vergallo, A., & Hampel, H. (2020). Early Detection of Mild Cognitive Impairment (MCI) in Primary Care. *The Journal of Prevention of Alzheimer's Disease*, *7*(3), 165–170. https://doi.org/10.14283/jpad.2020.21

Sabbagh, M. N., Boada, M., Borson, S., Doraiswamy, P. M., Dubois, B., Ingram, J., Iwata, A., Porsteinsson, A. P., Possin, K. L., Rabinovici, G. D., Vellas, B., Chao, S., Vergallo, A., & Hampel, H. (2020). Early Detection of Mild Cognitive Impairment (MCI) in an At-Home Setting. *The Journal of Prevention of Alzheimer's Disease*, *7*(3), 171–178. https://doi.org/10.14283/jpad.2020.22

Sanborn, V., Ostrand ,Rachel, Ciesla ,Jeffrey, & and Gunstad, J. (2022). Automated assessment of speech production and prediction of MCI in older adults. *Applied Neuropsychology: Adult*, *29*(5), 1250–1257. https://doi.org/10.1080/23279095.2020.1864733

Sanderson-Cimino, M., Elman, J. A., Tu, X. M., Gross, A. L., Panizzon, M. S., Gustavson, D. E., Bondi, M. W., Edmonds, E. C., Eglit, G. M. L., Eppig, J. S., Franz, C. E., Jak, A. J., Lyons, M. J., Thomas, K. R., Williams, M. E., & Kremen, W. S. (2022). Cognitive practice effects delay diagnosis of MCI: Implications for clinical trials. *Alzheimer's & Dementia : Translational Research & Clinical Interventions*, *8*(1), e12228. https://doi.org/10.1002/trc2.12228

Scharre, D. W., Chang, S. ing, Nagaraja, H. N., Wheeler, N. C., & Kataki, M. (2021). Self-Administered Gerocognitive Examination: Longitudinal cohort testing for the early detection of dementia conversion. *Alzheimer's Research & Therapy*, *13*(1), 192. https://doi.org/10.1186/s13195-021-00930-4

Shen, C., Wang, Z., Chen, H., Bai, Y., Li, X., Liang, D., Liu, X., Zheng, H., Wang, M., Yang, Y., Wang, H., & Sun, T. (2022). Identifying Mild Alzheimer's Disease With First 30-

Min 11C-PiB PET Scan. *Frontiers in Aging Neuroscience*, *14*. https://doi.org/10.3389/fnagi.2022.785495

Siedlecki, S. L. (2020). Understanding Descriptive Research Designs and Methods. *Clinical Nurse Specialist*, *34*(1), 8. https://doi.org/10.1097/NUR.0000000000000493

Tavares-Júnior, J. W. L., Neto, P. B., Bonfadini, J. de C., Bittencourt, L., Lopes, C. H., Mendes, L., Siqueira Neto, J. I., Sousa, V., Amaral, A., Carrilho, C. G., Espindola, J. O., Casali, M. E. A., Veras, A. B., & Alves, G. S. (2021). Clinical characteristics and diagnostic accuracy of the revised Addenbrooke Cognitive Examination (ACE-R) in older adults with a low educational level. *Jornal Brasileiro de Psiquiatria*, *70*, 45–53. https://doi.org/10.1590/0047-2085000000310

United Nations Department of Economic and Social Affairs. (2023). *World Social Report 2023: Leaving No One Behind in an Ageing World*. United Nations. https://doi.org/10.18356/9789210019682

Uysal, S. (n.d.). Diagnosis coding of neurocognitive disorders. *The Clinical Neuropsychologist*, *0*(0), 1–16. https://doi.org/10.1080/13854046.2024.2345925

Vimal, B., Surya, M., Darshan, Sridhar, V. S., & Ashok, A. (2021). MFCC Based Audio Classification Using Machine Learning. *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 1–4. https://doi.org/10.1109/ICCCNT51525.2021.9579881

Wang, H.-L., Tang, R., Ren, R.-J., Dammer, E. B., Guo, Q.-H., Peng, G.-P., Cui, H.-L., Zhang, Y.-M., Wang, J.-T., Xie, X.-Y., Huang, Q., Li, J.-P., Yan, F.-H., Chen, S.-D., He, N.-Y., & Wang, G. (2022). Speech silence character as a diagnostic biomarker of early cognitive decline and its functional mechanism: A multicenter cross-sectional cohort study. *BMC Medicine*, *20*(1), 380. https://doi.org/10.1186/s12916-022-02584-x

Wood, J. L., Weintraub, S., Coventry, C., Xu, J., Zhang, H., Rogalski, E., Mesulam, M.-M., & Gefen, T. (2020). Montreal Cognitive Assessment (MoCA) Performance and Domain-Specific Index Scores in Amnestic versus Aphasic Dementia. *Journal of the International Neuropsychological Society : JINS*, *26*(9), 927–931. https://doi.org/10.1017/S135561772000048X

Xavier, J. F. de S., Feuerstein, S. C., De Moraes, A. C. F., de Oliveira, T. A., da Silva Gomes, E. R., de Almeida Silva, M. I. A., de Oliveira, L. F., de Carvalho, H. B., Marin, K. A., & Nascimento-Ferreira, M. V. (2024). Development of a User-Friendly Self-Screening Tool for Assessing Metabolic Syndrome Risk in Youths from Economically

Challenged Regions. *Journal of Personalized Medicine*, *14*(8), Article 8. https://doi.org/10.3390/jpm14080810

Zaman, K., Sah, M., Direkoglu, C., & Unoki, M. (2023). A Survey of Audio Classification Using Deep Learning. *IEEE Access*, *11*, 106620–106649. IEEE Access. https://doi.org/10.1109/ACCESS.2023.3318015

Zhao, Q., Du, X., Chen, W., Zhang, T., & Xu, Z. (2023). Advances in diagnosing mild cognitive impairment and Alzheimer's disease using 11C-PIB- PET/CT and common neuropsychological tests. *Frontiers in Neuroscience*, *17*. https://doi.org/10.3389/fnins.2023.1216215

Appendices

Appendix A: Similarity Report

A Speech-Based Classification Model for Mild Cognitive Impairment Screening.docx

ORIGINALITY REPORT

15%	15%	14%	12%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	su-plus.strathmore.edu Internet Source	2%
2	Submitted to Strathmore University Student Paper	1%
3	biblioguias.uic.es Internet Source	1%
4	www.mdpi.com Internet Source	1%
5	www.frontiersin.org Internet Source	1%
6	Submitted to University of Newcastle Student Paper	1%
7	Submitted to Flinders University Student Paper	<1%
8	eaapublishing.org Internet Source	<1%
9	Submitted to Saint Joseph's University Student Paper	<1%
10	Submitted to University of Sunderland Student Paper	<1%
11	actacolombianapsicologia.ucatolica.edu.co Internet Source	<1%
12	Submitted to Navitas Professional Institute Pty Ltd	<1%

Student Paper

13	Submitted to The University of Manchester Student Paper	<1%
14	eprints.worc.ac.uk Internet Source	<1%
15	ouci.dntb.gov.ua Internet Source	<1%
16	www.aup.edu.ph Internet Source	<1%
17	www.cambridge.org Internet Source	<1%
18	H L Gururaj, Francesco Flammini, V Ravi Kumar, N S Prema. "Recent Trends in Healthcare Innovation", CRC Press, 2025 Publication	<1%
19	Miguel Ramos-Henderson, Carlos Calderón, Marcos Domic-Siede. "Education bias in typical brief cognitive tests used for the detection of dementia in elderly population with low educational level: a critical review", Applied Neuropsychology: Adult, 2022 Publication	<1%
20	Submitted to University at Buffalo Student Paper	<1%
21	Submitted to University of Huddersfield Student Paper	<1%
22	Submitted to University of Lincoln Student Paper	<1%
23	thescholarship.ecu.edu Internet Source	<1%
24	trepo.tuni.fi Internet Source	<1%

25	Submitted to Ottawa University (Blackboard LTI 1.3 Prod) Student Paper	<1 %
26	Submitted to Vrije Universiteit Amsterdam Student Paper	<1 %
27	doktori.bibl.u-szeged.hu Internet Source	<1 %
28	V. Sharmila, S. Kannadhasan, A. Rajiv Kannan, P. Sivakumar, V. Vennila. "Challenges in Information, Communication and Computing Technology", CRC Press, 2024 Publication	<1 %
29	Submitted to The Scientific & Technological Research Council of Turkey (TUBITAK) Student Paper	<1 %
30	cris.unibo.it Internet Source	<1 %
31	Kieback, Adrian I. C.. "Enhancing Mental Health Diagnosis With Representation Learning: A Transformer Based Approach", San Diego State University, 2024 Publication	<1 %
32	Submitted to University College London Student Paper	<1 %
33	openarchive.ki.se Internet Source	<1 %
34	Submitted to Singapore Institute of Technology Student Paper	<1 %
35	www.wjgnet.com Internet Source	<1 %
36	Submitted to University of New South Wales Student Paper	

		<1%
37	Brielle C Stark, Alyssa Lanzi, Sarah Grace Dalton. "Access to context-specific lexical-semantic information during discourse tasks differentiates speakers with latent aphasia, mild cognitive impairment, and cognitively health adults", Open Science Framework, 2024 Publication	<1%
38	repo.uum.edu.my Internet Source	<1%
39	www.ilo.org Internet Source	<1%
40	Hong, Bryan. "Leveraging Associations to Influence the Organization, Quantity, and Quality of Episodic Recollection", University of Toronto (Canada), 2022 Publication	<1%
41	dergipark.org.tr Internet Source	<1%
42	Fleischmann, Holly. "A Comparison of Inhibitory Control and Episodic Memory Deficits on Functional Variables in Samples of Mild Cognitive Impairment and Matched Healthy Controls", University of Georgia, 2024 Publication	<1%
43	uwspace.uwaterloo.ca Internet Source	<1%
44	Submitted to University of New England Student Paper	<1%
45	ir-library.ku.ac.ke Internet Source	<1%

46	sciendo.com Internet Source	<1 %
47	westminsterresearch.westminster.ac.uk Internet Source	<1 %
48	fastercapital.com Internet Source	<1 %
49	www.parinc.com Internet Source	<1 %

Exclude quotes	Off	Exclude matches	< 25 words
Exclude bibliography	Off		

Appendix B: Ethical Clearance Confirmation

Strathmore UNIVERSITY

6th February 2025

Ms Ahindukha Shisiali Noela,
noela.ahindukha@strathmore.edu

Dear Ms Ahindukha,

RE: Speech-Based Classification Model for Screening Mild Cognitive Impairment

This is to inform you that SU-ISERC has reviewed and **approved** your above **SU-masters** proposal. Your application reference number is **SU-ISERC2609/25.** The approval period is from **6th February 2025 to 5th February 2026.**

This approval is subject to compliance with the following requirements:

i. Only approved documents including (informed consents, study instruments, MTA) will be used.
ii. All changes including (amendments, deviations, and violations) are submitted for review and approval by SU-ISERC.
iii. Death and life-threatening problems and serious adverse events or unexpected adverse events whether related or unrelated to the study must be reported to SU-ISERC within 72 hours of notification.
iv. Any changes anticipated or otherwise that may increase the risks or affected safety or welfare of study participants and others or affect the integrity of the research must be reported to SU-ISERC within 72 hours.
v. Clearance for the export of biological specimens must be obtained from relevant institutions.
vi. Submission of a request for renewal of approval at least 60 days prior to the expiry of the approval period. Attach a comprehensive progress report to support the renewal.
vii. Submission of an executive summary report within 90 days of completion of the study to SU-ISERC.

Before commencing your study, you will be expected to obtain a research license from National Commission for Science, Technology, and Innovation (NACOSTI) https://research-portal.nacosti.go.ke/ and obtain other clearances needed.

Yours sincerely,

Mr Ambrose Rachier,
Chairperson; SU-ISERC

Ole Sangale Rd, Madaraka Estate. PO Box 59857-00200, Nairobi, Kenya. Tel +254 (0)703 034000
Email admissions@strathmore.edu www.strathmore.edu