2024 Ijtsi Dzaya Ramos Morales Extraction of dementia features from audio and text records using machine learning algorithms

Universidad Autónoma de Querétaro
Facultad de Ingeniería

Extraction of dementia features from audio and text records using machine learning algorithms

Tesis

Que como parte de los requisitos para obtener el Grado de

Maestro en Ciencias en Inteligencia Artificial

Presenta:

Ijtsi Dzaya Ramos Morales

Dirigido por:

Dr. Saul Tovar Arriaga

Co-Director:

Med. Esp. Humberto Guendolain Arenas

Querétaro, Qro. a January 6, 2025

Universidad Autónoma de Querétaro
Facultad de Ingeniería
Maestro en Ciencias en Inteligencia Artificial

Extraction of dementia features from audio and text records using machine learning algorithms

Tesis

Que como parte de los requisitos para obtener el Grado de

Maestro en Ciencias en Inteligencia Artificial

Presenta
Ijtsi Dzaya Ramos Morales

Dirigido por:
Dr. Saul Tovar Arriaga

Co-dirigido por:
Med. Esp. Humberto Guendolain Arenas

Dr. Saul Tovar Arriaga
Presidente

Med. Esp. Humberto Guendolain Arenas
Secretario

Dr. Sebastián Salazar Coloresl
Vocal

Dr. Jesús Carlos Pedraza Ortega
Suplente

Mtro. Luis Roberto García Noguez
Suplente

Centro Universitario, Querétaro, Qro.
Fecha de aprobación por el Consejo Universitario (mes y año)
México

Dedications

This thesis is dedicated to my sister, whose support and encouragement have been my strength throughout this journey.

I would also like to dedicate this work to my professors and mentors, whose guidance and knowledge have been invaluable in shaping this research.

Last, but not least, to my friends, who have been a source of motivation and joy during the challenging times of this academic pursuit.

Thank you all for being an integral part of this achievement.

Acknowledgements

I would like to extend my gratitude to García Noguez Luis Roberto, Llaca Sánchez Brandon Alejandro and Medellín Robles Rodrigo for their invaluable assistance in labeling and preparing the database used in this study. Their support was crucial for ensuring data quality and integrity, significantly contributing to the research's success.

Finally to the Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCYT), for financing this project thanks to the research scholarship.

Contents

List of Figures

Resumen

La demencia, un síndrome de naturaleza progresiva que afecta la función cognitiva, se espera que afecte a 78 millones de personas para 2030. Afectando principalmente a los ancianos, esta enfermedad no tiene cura. Debido a la prevalencia de la demencia, la detección y el tratamiento temprano juegan un papel crucial en retrasar su progresión. Los métodos de detección actuales pueden ser costosos y demorados, lo que lleva a retrasos en la detección y tratamiento de la demencia. Métodos con un enfoque computacional han demostrado ser eficientes para detectar la demencia. Presentando una metodología de IA no invasiva para la detección temprana de la demencia mediante una combinación de análisis lingüístico y acústico derivados de entrevistas con pacientes, este estudio utiliza el conjunto de datos Dementia Bank Pitt corpus para ofrecer nuevas perspectivas sobre el deterioro cognitivo. La metodología combina embeddings de oraciones con similitud coseno para características lingüísticas y redes neuronales convolucionales (CNN) para el análisis de características acústicas. El modelo QDA destaca, logrando métricas notables con una precisión del 80%, un recall del 91.14%, una precisión del 82.64% y una puntuación F1 del 86.42%. Estos resultados subrayan el valor diagnóstico mejorado de las características lingüísticas e ilustran la capacidad de la IA para apoyar y complementar las técnicas de evaluación convencionales. Al proporcionar una alternativa accesible y eficiente para la industria de la salud, esta investigación resalta el potencial significativo de las aplicaciones de IA en el diagnóstico temprano de la demencia y el cuidado del paciente.

Abstract

Dementia, a syndrome of progressive nature that affects cognitive function is expected to have affected 78 million people by 2030. Primarily affecting the elderly, this disease has no cure. Due to the prevalence of Dementia, early detection and treatment plays the main role in delaying its progression. Current screening methods can be expensive and time-consuming leading to a delay in the detection and treatment of Dementia. Methods with a computational approach have been proved efficient to detect dementia as well. Presenting a non-invasive AI methodology for early dementia detection through a blend of linguistic and acoustic analyses derived from patient interviews, this study utilizes the Dementia Bank Pitt corpus dataset to offer novel insights into the cognitive decline. The methodology combines sentence embeddings with cosine similarity for linguistic features and convolutional neural networks (CNNs) for acoustic feature analysis. The QDA model stands out, achieving notable metrics with an accuracy of 80%, a recall of 91.14%, precision of 82.64%, and an F1 score of 86.42%. These results underscore the enhanced diagnostic value of linguistic features and illustrate the capability of AI to support and augment conventional assessment techniques. In providing an accessible, efficient alternative for the healthcare industry, this research underscores the significant potential of AI applications in early dementia diagnosis and patient care.

Chapter 1

Introduction

1.1 Dementia

Dementia is a cerebral disorder characterized by progressive deterioration of cognitive functions. Other manifestations of the disease are apathy, deterioration in social behavior, occasional aggressiveness, delusional ideas, and hallucinations [1].

According to the World Health Organization (WHO), a dementia patient must present all the following points.

- A decline in memory to an extent that it interferes with everyday activities, or makes independent living either difficult or impossible.
- A decline in thinking, planning, and organizing day-to-day things, again to the above extent.
- Initially, preserved awareness of the environment, including orientation in space and time.
- A decline in emotional control or motivation, or a change in social behavior, as shown

in one or more of the following: emotional lability, irritability, apathy, or coarsening of social behavior, as in eating, dressing, and interacting with others [2].

Since there is a continuum from normal functioning through to severe dementia, the severity can be classified into stages. This scheme is useful to determine the progression of the disease in a cohort of the elderly followed over time. The following classification was made by Berg [3], and it is a useful scheme used to categorize dementia. There are three stages in Berg's Clinical Dementia Rating (CDR): mild, moderate, and severe.

The Clinical Dementia Rating (CDR) scale is a widely used tool for assessing dementia severity across multiple domains, including memory, orientation, judgment and problem-solving, community affairs, home and hobbies, and personal care. Each domain is rated from 0 to 3, with higher scores indicating greater impairment, and the global CDR score is derived from these ratings [4, 5, 6].

- CDR 0 (No Dementia): The patient exhibits no memory loss and is fully oriented. They are capable of handling daily tasks, self-care, and maintaining personal interests and activities in the community. There are no observable impairments in judgment or problem-solving [4].

- **CDR 0.5 (Questionable Dementia):** The patient shows slight, consistent forgetfulness and may experience minor challenges with time orientation. Problem-solving abilities and engagement in community activities are slightly impaired, though they remain largely functional [7]. Daily routines and personal interests are maintained with only minimal difficulty.

- **CDR 1 (Mild Dementia):** Memory impairment is moderate, especially regarding recent events. Patients have moderate difficulty with time orientation and face challenges in problem-solving and distinguishing similarities and differences. Independent functioning in community affairs is limited, and more complex hobbies are typically abandoned, although simpler home activities may be maintained [7, 8].

- **CDR 2 (Moderate Dementia):** Memory loss becomes severe, with only highly familiar information retained. The patient is often disoriented in time and may also experience spatial disorientation. Severe impairments are noted in judgment and problem-solving, limiting the patient's ability to manage household tasks. Personal care routines require assistance, particularly with dressing and hygiene [6, 9].
- **CDR 3 (Severe Dementia):** Severe memory impairment results in only fragmented recall of information. The patient is unable to make judgments or solve problems, requiring extensive help with all aspects of daily living. They are unable to function independently outside the home and exhibit no significant capability for self-care, needing considerable assistance with personal hygiene and basic needs [4, 5].

This scale is a validated and reliable measure for staging dementia and has been widely applied in both clinical and research settings to provide a standardized assessment of cognitive and functional decline across different types of dementia [9].

1.1.1 Types of Dementia

There are many different types of dementia, and each one has its own set of symptoms. The most common form of dementia is Alzheimer's disease (AD), which accounts for 60-80% of all cases [10].

Dementia encompasses a variety of conditions that impair cognitive functioning, each with distinct brain changes. Alzheimer's disease, characterized by amyloid plaques and tau tangles that disrupt neuron function, leading to memory loss, confusion, and impaired language skills, progresses gradually and is most prevalent in individuals over the age of 65 [11].

Vascular dementia results from reduced blood flow to the brain, often due to strokes or other vascular events. This type primarily impacts executive functions like planning and problem-solving, and may present alongside Alzheimer's disease. Symptoms vary based on the regions affected by blood flow disruptions, making vascular dementia a complex condition to diagnose and manage [12].

Lewy body dementia is linked to abnormal protein deposits called Lewy bodies that alter brain chemicals, leading to cognitive, behavioral, and motor symptoms such as visual hallucinations, attention issues, and muscle rigidity. Closely related to Parkinson's disease, it is divided into dementia with Lewy bodies or Parkinson's disease dementia based on the order of symptom onset. Lewy body dementia's overlap with other conditions often complicates its diagnosis [13].

Frontotemporal dementia (FTD) targets the brain's frontal and temporal lobes, leading to pronounced behavioral and personality changes. FTD, more common in individuals aged 45-64, includes subtypes like behavioral variant FTD and primary progressive aphasia, which affect social behaviors and language abilities, respectively [14].

Mild cognitive impairment (MCI) refers to the decline in cognition that is greater than what is expected for an individual's age and education level and does not interfere significantly with the individual's daily life. There are many similarities in the recognition and diagnosis of MCI and early dementia since the diagnosis criteria for both have been derived from the criteria to diagnose AD. Moreover, since memory impairment may not necessarily be the first cognitive domain to decline in the onset of all the different types of dementia, a classification can be made divided into separate groups of amnestic (aMCI) and non-amnestic MCI [15] as observed in Figure 1.1.

Those subtypes with aMCI are more likely to progress to AD as validated with the pres-

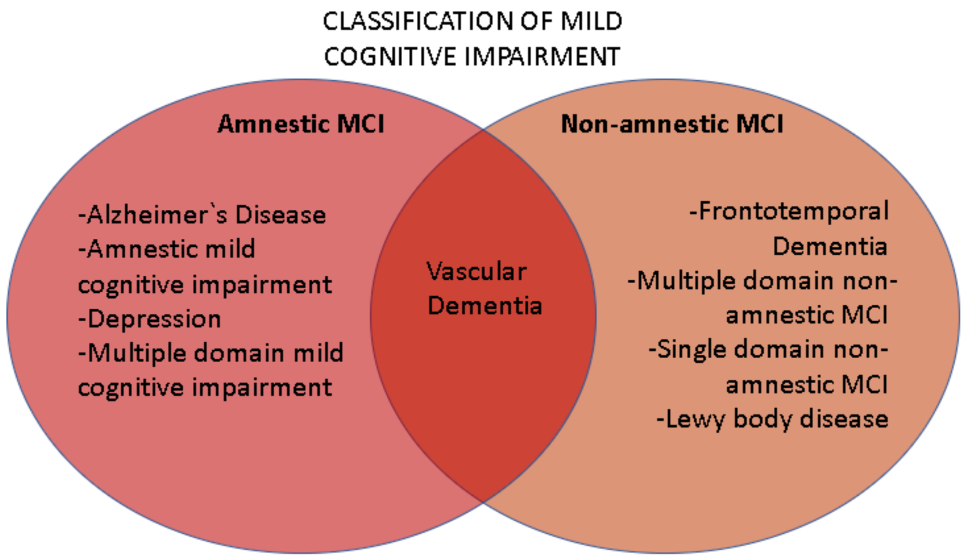

Figure 1.1: Classification of Mild Cognitive Impairment

ence of biomarkers associated with AD, such as changes in MRI, FDG-PET, and PiB PET [16].

As the disease progresses, people with dementia will start to experience more severe symptoms. They may have trouble communicating, remembering recent events, or recognizing familiar faces. They may also start to experience changes in mood and behavior, such as aggression or apathy. Ultimately, people with dementia will need full-time care as they become completely reliant on others for their basic needs.

There is no one-size-fits-all approach to managing dementia since each person experiences the condition differently. However, there are some general strategies that can help manage the symptoms and slow down the progression of the disease. These include promoting social and recreational activities, providing support for caregivers, and engaging in personalized care plans.

1.1.2 Dementia Screening

Spontaneous speech task refers to the ability of providing an analysis of a subject's language skills and speech. In this project, conversation/interview speech and picture description will be covered as types of spontaneous language [**?**].

Conversation/interview speech
This tool consists of the acquisition of features that consider early signs of AD through casual speech's natural language and various biological aspects.

Picture description
This process takes place through the demonstration of several images that compose a story to a subject and have them narrate orally these images in a restricted amount of time.

The Mini-Mental State Examination (MMSE) and the Montreal Cognitive Assessment (MoCA) are two of the most widely used cognitive screening tools for detecting dementia and mild cognitive impairment (MCI). While both tests assess cognitive functions using a 30-point scale, they differ in their sensitivity and the range of cognitive abilities they cover.

The MMSE primarily evaluates orientation, memory, and attention, asking patients to perform tasks such as recalling the current date, identifying familiar objects, and performing simple arithmetic. It is often preferred in clinical settings due to its shorter duration, taking about 7-10 minutes to administer. However, while the MMSE is effective in detecting moderate to severe dementia, it may overlook milder forms of cognitive decline [17, 18].

In contrast, the MoCA was specifically designed to detect early signs of MCI and is considered more sensitive to subtle cognitive impairments. The MoCA includes tasks that assess executive function, visuospatial abilities, and abstract thinking, such as clock drawing and connecting dots in a complex sequence. Studies have shown that the MoCA is better suited for identifying early cognitive deficits, making it the preferred tool when clinicians suspect mild impairment [3][4]. However, the MoCA requires more time to administer (10-15 minutes) and may be too challenging for individuals with more advanced dementia [19].

The cookie-theft picture
The Cookie Theft picture test is a diagnostic tool from the Boston Diagnostic Aphasia Examination that has become a mainstay in cognitive assessments for dementia, especially Alzheimer's disease. In this task, patients are asked to describe a scene where two children attempt to steal cookies from a jar while their mother is preoccupied at the sink. The structured nature of this task allows clinicians to observe several key linguistic and cognitive abilities that are often compromised in dementia. One of the primary elements assessed is referential cohesion, where patients must accurately refer to people and objects in the picture without ambiguity. For instance, a clear description might identify "the boy on the stool" and later refer to him using an appropriate pronoun, maintaining consistency in references. Failures in this skill often reflect early deficits in language organization and memory that are characteristic of Alzheimer's dementia [20].

In addition to language cohesion, the Cookie Theft picture test evaluates patients' understanding of causal and temporal relationships within the scene. Individuals must logically connect events, such as recognizing that the children's cookie theft is possible only because the mother is distracted with the dishes. This requirement to sequence events and infer relationships is particularly useful in detecting pragmatic language deficits and impairments in theory of mind—the ability to attribute mental states or intentions to others. Studies have shown that patients with Alzheimer's often struggle to articulate motivations behind actions in the scene, such as understanding that the boy climbs the stool to reach the cookie jar. This test's sensitivity to such breakdowns in cognitive-linguistic processing makes it a powerful tool in early dementia detection [20]. The picture is shown in figure 1.2

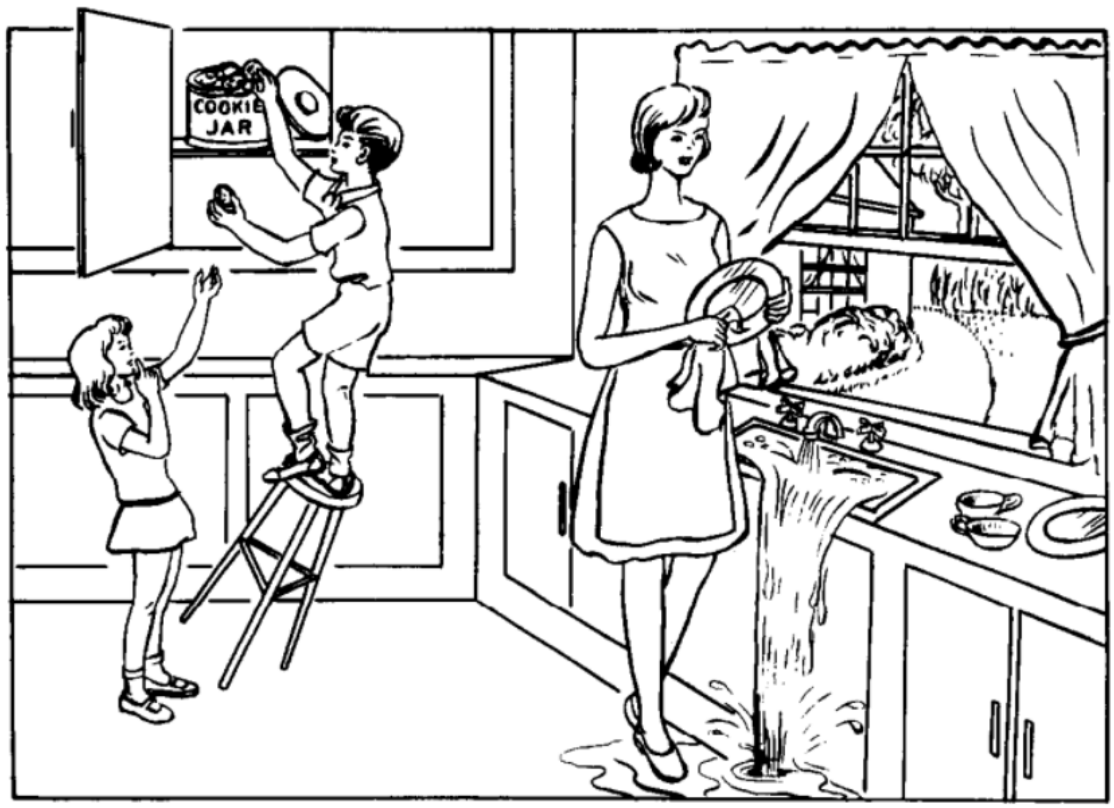

Figure 1.2: The Cookie Theft Picture from the Boston Diagnostic Aphasia Examination.

1.2 Problem Description

As of today, several artificial intelligence algorithms have been developed and trained to screen for dementia. Since one of the main objectives of taking a computer approach is to reduce the diagnostic time, the result obtained from artificial intelligence is expected to be as accurate as possible. Thus, a continuous improvement in the metrics of existing models can be seen as an aid in the progression of such screening tools.

One of the primary challenges in diagnosing dementia lies in the complexity and variability of its symptoms, which overlap with other cognitive disorders. Traditionally, diagnosis has relied heavily on neuropsychological tests, neuroimaging, and clinical evaluations, which can take years to provide a definitive result. However, the integration of AI and machine learning has significantly improved early diagnostic accuracy. AI models can analyze vast amounts of patient data, including medical history, neuroimaging, and biomarkers, to differentiate between dementia subtypes and even predict cognitive decline years before symptoms emerge [21].

ML algorithms based on NLP rely on a transcription of the descriptive tests and interviews between clinician and patient [22]. The classification then is made based on semantic features that are relevant for the detection of Dementia. A more robust approach includes audio and text so acoustic features can be considered for the screening process. For this to occur, the relevant acoustic features need to be selected to obtain the best metrics possible from the model [23].

A recurrent problem for developing a screening tool using ML algorithms is the justification of the chosen metrics. While some metrics are demonstrative, they may be misleading when used to evaluate skewed datasets. Using the same corpus of data from DementiaBank, related research has various dataset distributions, which poses a hurdle because results may differ when classes are distributed differently in the training and testing sets [24]. Moreover, some authors report only their best-performing models without giving a justification as to why the average is not reported instead.

Finally, despite these promising developments, challenges remain. The generalizability of AI models across diverse populations is still a concern, as most models are trained on limited datasets, predominantly from high-income regions. Efforts are ongoing to improve the scalability of these models by incorporating more diverse datasets and ensuring that AI tools can assist in under-resourced regions where traditional diagnostic methods are often unavailable [21].

1.3 Justification

Dementia, a syndrome of progressive nature in which cognitive function deteriorates [25], affects more than 55 million people, a number which is expected to rise to 78 million by 2030 according to the World Health Organization [2].

Currently, there is no cure for this disease, one of the world's most common neurological syndromes. Moreover, stroke risk is higher in dementia patients, increasing their dependency and level of disability [26]. Thus, early detection is crucial to prevent a decline in the quality of life and delay the progression of cognitive impairment.

The majority of dementia cases, over 60%, occur in low- and middle-income countries, where access to diagnostic and care resources is often limited. Each year, around 10 million new cases are reported globally, making dementia one of the most significant public health challenges in the coming decades [27].

The economic burden is equally staggering. In 2019, the global cost of dementia care reached $1.3 trillion, and this figure is expected to rise to $2.8 trillion by 2030. A large portion of these costs results from informal care provided by family members, accounting for nearly 50% of the total [27]. The societal impact extends beyond economic costs. Low socio-economic status (SES) has been identified as a risk factor for dementia. Individuals with lower SES often experience a faster cognitive decline and are more likely to develop dementia than those in higher SES brackets.[28]. This growing financial strain underscores the urgent need for scalable and cost-effective diagnostic solutions, especially in regions with limited healthcare resources.

The diagnosis of dementia is a time-consuming process that requires an evaluation of the medical history based on periodic evaluation of cognitive decline and corroboration of an impairment in daily activities given by a close friend or family member. Finally, an exhaustive examination by a clinician based on several medical tests will determine the presence of this pathology in the patient [29]. Since most of the diagnosis is based on the personal judgment of experts, these methods may be subjected to human error.

Structural imaging such as magnetic resonance (MR) and electrode usage are other tools used in the diagnosis of dementia with high rates of accuracy. Nonetheless, patients with public insurance may have to wait months until they can get tested due to the high demand and low accessibility of the methods. Less time-consuming techniques such as brain scans and blood biomarkers are useful to detect dementia at an early stage. Despite their ability, they are considered too invasive and expensive to use widely [30]. This increases the diagnosis time and as a consequence the progression of the disease.

The high costs of medical tests and time expenditure emphasize the need for faster and more accessible screening methods that clinicians could use as a tool to reduce the diagnosis time altogether.

Chapter 2

Background

2.1 State-of-the-art

In response to the increasing number of patients diagnosed with dementia, screening methodologies have evolved, incorporating computational techniques for more rapid and effective detection. A successful approach for dementia detection, avoiding the need for imaging, is the application of Natural Language Processing (NLP) techniques. These techniques analyze conversations collected during medical assessments, focusing on spoken language deterioration that signals dementia onset [31]. Research has identified specific language features indicative of early dementia stages, such as telegraphic speech patterns, repetitive language use, and spelling errors during oral tasks [32].

Studies show that individuals with Alzheimer's disease (AD) provide less detailed descriptions compared to elderly individuals with normal cognitive function [33]. This difference is quantitatively assessed by comparing patient responses to a standardized list of Information Content Units (ICUs), which typically include elements like people, objects, actions, and locations depicted in a scene. Some researchers, such as Oda and Matsui, recommend customizing these lists to better reflect the educational, age, cultural, and interest profiles of specific patient groups to enhance assessment relevance and accuracy [34].

Descriptive tasks are used alongside other cognitive tests like the Mini-Mental State Examination (MMSE), the Montreal Cognitive Assessment (MoCA), and the Clinical Dementia Rating (CDR) [35]. The Cookie Theft picture from the Boston Diagnostic Aphasia Examination is effective in identifying language and cognitive impairments [36]. The efficacy of the cookie-theft picture description test in creating predictive models of dementia has been validated. Models integrating cookie-theft test results with additional data such as APOE genetic markers, demographic variables, and neuropsychological (NP) test outcomes perform well [37].

In a study by Zheng et al. [38], structural MRI and machine learning were used to differentiate Vascular Dementia (VaD) from Alzheimer's Disease (AD). Conducted from June 2013 to July 2019 with 93 patients, the study used the AccuBrain tool to extract volumetric measurements from brain regions. After reducing dimensionality through analysis of 62 MRI biomarkers, the least absolute shrinkage and selection operator (LASSO) refined these features for classification via a Support Vector Machine (SVM) with an RBF kernel. Performance, assessed by ROC-derived metrics, showed the SVM model achieved sensitivity, specificity, and accuracy rates of 82.65%, 87.17%, and 84.35%, respectively, with an AUC of 0.861 (95% CI: 0.820-0.902). This study demonstrates the efficacy of combining MRI and machine learning for clinical decision-making in differentiating VaD versus AD.

Chen et al. [39] evaluated deep learning architectures including VGG16, ResNet-152, and DenseNet-121, achieving an accuracy rate of 96.65% in dementia screening. These models were trained on various dementia types to address limited data availability, initially developed using the clock-drawing test (CDT), which is also used for detecting Parkinson's disease and other cognitive impairments. Recent automated systems utilize the cookie-theft picture test for dementia screening [40, 41].

Yamanki et al. [41] developed a text-based classifier applying advanced NLP techniques to enhance detection scope within a dataset. Their approach yielded a second-order polynomial SVM and an Artificial Neural Network (ANN) model, achieving average accuracies of

77% and 78%, respectively.

Dashtipour et al. [42] employed advanced machine learning techniques to enhance predictive capabilities in diagnosing Alzheimer's disease. Integrating deep learning for feature extraction from brain images with SVM and bidirectional long short-term memory (BiLSTM) for classification, they achieved a classification accuracy of 91.28%. This approach highlights the potential of hybrid models in improving diagnostic accuracy through the synergistic use of multiple machine learning methodologies.

Helaly et al. [43] focused on early detection of Alzheimer's disease using convolutional neural networks (CNNs). Analyzing structural brain images from the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset, the models achieved accuracies of 93.61% and 95.17% for 2D and 3D multi-class AD stage classification, respectively. Their application of the pre-trained VGG19 model achieved an accuracy of 97%, showcasing the effectiveness of fine-tuned deep learning models in Alzheimer's disease classification.

Vandenberghe et al. [44] used 18F-flutemetamol PET imaging and SVM to distinguish Alzheimer's disease from mild cognitive impairment (MCI), achieving 100% accuracy in replicating expert diagnostic decisions. This study underscores the reliability of SVM in medical imaging for clinical differentiation.

Odusami et al. [45] presented a method for early-stage dementia detection from functional MRI changes using a finely tuned ResNet-18 network, yielding classification accuracies of 99.99% for early MCI versus Alzheimer's disease, 99.95% for late MCI versus Alzheimer's disease, and 99.95% for MCI versus Alzheimer's disease. These results highlight the potential of deep learning techniques in identifying subtle brain pattern variations associated with early dementia stages.

Chakraborty et al. [46] used acoustic-derived features to classify cognitive impairment stages such as MCI, AD, and healthy controls. Without feature selection, a three-class classifier achieved a baseline F-score of 66%, improving to 77% by integrating a range of features early in the analysis. Similar accuracy rates were achieved using a two-level Long Short-Term Memory (LSTM) network combined with a CNN [47].

Comparative studies between traditional clinical screening processes and computerized approaches show AI-enhanced cognitive tests can improve discrimination sensitivity by up to 4% compared to conventional assessments [48]. The most effective results combine linguistic and auditory data, maximizing diagnostic capabilities [49].

Fraser et al. [49] achieved an 81% accuracy rate with a classifier assessing semantic, acoustic, syntactic, and informational impairments. Training the model on all 370 selected features reduced accuracy to 58.51%, emphasizing the importance of feature selection in high-dimensional data spaces. Classifiers integrating acoustic and linguistic features from patient recordings and transcriptions have achieved high accuracy, sensitivity, and specificity, as evidenced by studies using the DementiaBank and ADReSS datasets [30].

Moro-Velazquez et al. [50] proposed a framework combining acoustic and linguistic analysis with machine learning techniques for early detection of Alzheimer's Disease (AD). Using prosodic and lexical features extracted from spontaneous speech, their model achieved an accuracy of 80.5% in differentiating AD patients from healthy controls. The framework utilized sequential models to capture temporal dependencies in speech, demonstrating the efficacy of integrating linguistic and acoustic cues for non-invasive dementia screening.

Amorim et al. [51] further refined dementia detection by focusing on linguistic impairments related to executive function. Their study analyzed lexical richness, syntactic complexity, and discourse coherence, identifying markers that correlated with early dementia stages. Implementing a random forest classifier, the model achieved an F1-score of 83% when tested on conversation data from DementiaBank, highlighting linguistic deficits as reliable indicators of cognitive decline.

In another study, Luz et al. [52] combined MFCCs and other acoustic features with deep learning architectures, including Long Short-Term Memory (LSTM) networks, for Alzheimer's classification. This hybrid approach yielded an accuracy of 85% in distinguishing AD patients from controls, underscoring the potential of acoustic feature integration with neural networks in detecting early signs of dementia.

Recent advances in deep learning for dementia diagnostics also explore Transformer-based architectures. Yang et al. [53] applied Transformers to model complex dependencies within linguistic features from patient narratives, achieving a classification accuracy of 82.3% for AD detection. This approach shows promise in capturing nuanced language patterns that may escape conventional machine learning models.

Yu et al. analyzed machine learning models for dementia prediction, focusing on SVM and CNN using structural MRI scans. These models achieved accuracy rates from 70.32% to 89.74%, highlighting the potential of machine learning in dementia prediction. Data from medical imaging, clinical features, and voice data indicated image data's effectiveness in predicting dementia.

Liu et al. proposed a deep learning framework for Alzheimer's disease diagnosis using multiple data sources, including MRI scans and medical history. Their fusion model achieved the highest accuracy (87%) and F1-score (84%) in diagnosing Alzheimer's disease, matching neurologist expertise. Combining NLP and acoustic features in analyzing patient interactions offers a non-invasive diagnostic approach, complementing traditional tests.

Eldele et al. introduced TSLANet, a model that combines adaptive spectral processing and CNN blocks to enhance feature extraction in noisy time-series data. Their model achieved state-of-the-art results in tasks such as anomaly detection and forecasting by capturing both short- and long-term dependencies, showcasing improved robustness across diverse datasets [54].

In financial time-series forecasting, Liu et al. applied CNNs as preprocessing layers to reduce dimensionality before feeding data into Transformers. This approach enabled the model to efficiently extract sequential patterns and increased forecasting accuracy, particularly in high-frequency datasets. The hybrid model's performance surpassed traditional autoregressive models, underscoring the benefits of CNN-Transformer architectures in handling complex time-series data [55].

Miyazaki et al. explored CNN-Transformer combinations in environmental sound classification, achieving high classification accuracy with minimal computational cost. Their work illustrates the suitability of these models for edge devices, which require efficient processing without sacrificing accuracy [56].

2.2 Theoretical Foundation

Artificial Intelligence (AI) can be described as an extension of human intelligence through machines that are capable of improving over time to acquire capabilities such as adapting, reasoning, and self-correction. However, the definition from a computational point of view encompasses several guidelines that an algorithm must fulfill in order to be considered an AI algorithm. Those guidelines have been defined by several authors [**?**, 57] from a computation and mathematical approach, respectively.

A subset of AI is Machine Learning (ML), which comprises algorithms that use statistical algorithms that can be trained to learn patterns over time. Finally, Deep Learning is a subset of ML that uses multiple layers of algorithms to learn over time.

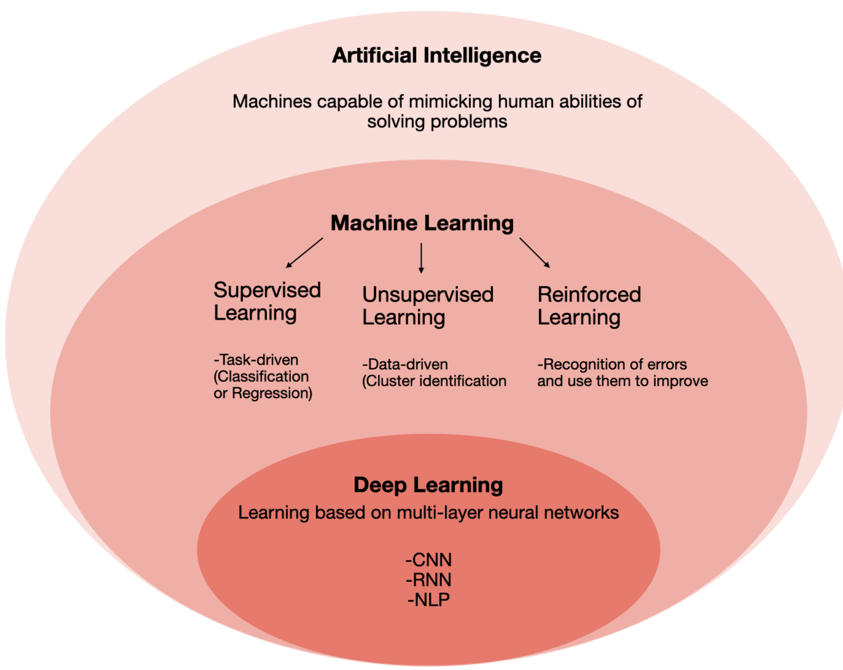

Figure 2.1: Overview of Artificial Intelligence and its subsets

Natural Language Processing (NPL) can be found at the intersection between Artificial Intelligence and Linguistics. Originally, it was derived from a text information retrieval algorithm and has since evolved to extract semantics from text by identifying the rules/constraints that specify the relationship between text units [58].

Natural Language Processing (NLP) is a subfield of AI that deals with the interaction between computers and human language. NPL divides into two subcategories which are Natural Language Understanding (NLU) and Natural Language Generation (NLG). The first subcategory is used to read and understand text, meanwhile, the second one is used to create text. Besides being able to extract semantics, NLU is also able to extract and use phonology, morphology, lexical, and semantic elements [58].

Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) have been used to tackle NPL problems [59, 60]. Attention mechanisms and transformers models have played a significant role in the development of NPL. Attention mechanisms started as an enhancement for encoder-decoder RNNs applied to sequence-to-sequence applications [61].

In attention mechanisms, the machine is able to revisit the input given at every step, but it is also able to focus on specific parts of the input sequence.

Acoustic and Linguistic Features

For effective dementia detection, features from both linguistic and acoustic domains are extracted from patient conversations and interviews. These attributes form the basis for classification, and their extraction methods are outlined as follows:

Acoustic Features

Acoustic features can be extracted from an audio segment to train a classifier. These features can be time-dependent or time-independent based on whether they vary according to time or not. The time-dependent set of features are MFCC, GTCC, MFCC, and delta-GTCC, Log-Energy, Formants, and Fundamental frequency. The time-independent set of features are Jitter, Shimmer, and pitch [62].

Convolutional Neural Networks (CNNs) for Acoustic Feature Analysis

CNNs are widely used for processing time-series data, including acoustic signals, due to their capacity to extract contextual features. The output from CNN layers captures high-level patterns in acoustic data, crucial for dementia classification tasks [63]. CNNs are especially effective at processing time-dependent acoustic features, including:

Time-dependent features, which include MFCCs (Mel-frequency cepstral coefficients), GTCCs (Gammatone cepstral coefficients), Delta GTCCs, and Formants. These attributes reflect the dynamic aspects of speech over time, representing its timbre and other characteristics. The equations to compute these features depend on the specific algorithms of the Librosa library.

- MFCC: Mel-frequency cepstral coefficient is a method that contains 39 features and provides enough frequency channels to analyze the audio [64].
- GTCC: Gammatone cepstral coefficient is a derivation of the MFCC and it uses gammatone filters.
- delta-MFCC: It is the first-order derivative of the features and it measures changes in features from the previous frame to the next frame.
- delta-GTCC: It is the first-order derivative of the features and it measures changes in features from the previous frame to the next frame.
- Energy: It describes the total energy content of the signal within a time-frame in the log scale.

- Formants: The spectral peaks of the spectrum of the acoustic resonance of the human vocal tract are referred to as formants, and they describe the distinct characteristics of the subject's voice [65].
- Fundamental frequency: Describes the frequency of the audio signal in a specific time window.

For the time-independent features, the whole audio can be used instead of shorter frames. Comprises summary statistics (mean, standard deviation, etc.) of time-dependent attributes, providing a statistical overview of speech characteristics. This section also includes Speech Rate, measuring the speed of speech articulation.

- Jitter: It refers to information regarding the instability in the frequency.
- Shimmer: It refers to information regarding the instability in the amplitude.
- Pitch: It can be described as the variation in the relative vibration frequency of the human voice that is necessary for the comprehension of the meaning of a conversation.

Linguistic Features

Linguistic features can be grouped by lexical, syntactic, and semantic features.
The lexical features refer to the validation of words from a linguistic point of view depending on the language that is being used.

- Vocabulary length: The frequencies at which the words appear in the text. It can be represented as:

$$\text{Vocabulary Size} = \text{len(set(words))} \tag{2.1}$$

  where *words* is a list of all words spoken by the subject.

- Lexical diversity: Diversity of vocabulary used during the text. A measure of language richness, calculated as the ratio of unique words to the total number of words:

$$\text{Lexical Diversity} = \frac{\text{len(set(words))}}{\text{len(words)}} \tag{2.2}$$

where *words* is a list of all words spoken.

The syntactic features refer to the validation of words from a grammatical point of view.

Cosine Similarity: Represents the cosine of the angle between the embedding vectors of the subject's text and the key ideas of the Cookie Theft picture, indicating semantic similarity:

$$\text{Cosine Similarity} = \frac{\text{Embeddings}_a \cdot \text{Embeddings}_b}{\|\text{Embeddings}_a\| \cdot \|\text{Embeddings}_b\|} \tag{2.3}$$

where Embeddings_a and Embeddings_b are the embedding vectors of the subject's text and key ideas, respectively.

Keyword Frequency: Represents the frequency of specific pre-determined keywords relevant to the dementia context. It can be represented as:

$$\text{Keyword Frequency} = \sum_{k \in \text{Keywords}} \text{Count}(k) \tag{2.4}$$

where k is a keyword from a pre-determined set, and $\text{Count}(k)$ returns the count of its occurrences in the subject's text.

- Grammatical and sentence structural errors: It refers to the incorrect grammatical structure in a sentence.

Semantic features refer to the relationship that gives meaning to a sentence through the words that precede or proceed it.

- Main idea of the text: A key word search can be used to determine the relevant characteristics of the text, as well as, a similarity calculation to find an approximation to the text's central ideas.

Models

Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA) is used for dimensionality reduction and classification. It aims to maximize the ratio of between-class variance to within-class variance. The discriminant function for LDA is given by:

$$\delta_k(x) = x^T \Sigma^{-1} \mu_k - \frac{1}{2} \mu_k^T \Sigma^{-1} \mu_k + \log(\pi_k) \tag{2.5}$$

where Σ is the shared covariance matrix, μ_k is the mean vector of class k, and π_k is the prior probability of class k.

Ridge Classifier

The Ridge Classifier is a linear classifier that uses L2 regularization to prevent overfitting. The objective function for the Ridge Classifier is:

$$J(\theta) = \sum_{i=1}^{n} (y_i - \theta^T x_i)^2 + \lambda \|\theta\|^2 \tag{2.6}$$

where θ are the model parameters, x_i are the input features, y_i are the labels, and λ is the regularization parameter.

XGBoost

XGBoost stands for Extreme Gradient Boosting and is an optimized implementation of the gradient boosting algorithm. The objective function for XGBoost is:

$$\mathcal{L}(\phi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \tag{2.7}$$

where l is a differentiable convex loss function that measures the difference between the prediction $\hat{y}_i$ and the target y_i, and Ω penalizes the complexity of the model to prevent overfitting.

Logistic Regression

Logistic Regression is a statistical model designed for binary classification tasks. It estimates the probability that a given input x belongs to one of two classes by applying the logistic (or sigmoid) function to a linear combination of its features.

$$\sigma(z) = \frac{1}{1 + e^{-z}} \tag{2.8}$$

where $z = \theta^T x + b$ is the linear combination of input features x with weights θ and bias b.

Quadratic Discriminant Analysis (QDA)

Quadratic Discriminant Analysis (QDA) is a probabilistic model that assumes Gaussian distributions for the data belonging to each class but allows each class to have its own covariance matrix. The discriminant function for QDA is given by:

$$\delta_k(x) = -\frac{1}{2}\log|\Sigma_k| - \frac{1}{2}(x - \mu_k)^T \Sigma_k^{-1}(x - \mu_k) + \log(\pi_k) \tag{2.9}$$

where Σ_k, μ_k, and π_k represent the covariance matrix, mean vector, and prior probability of class k, respectively.

Chapter 3

Hypothesis

A machine learning model utilizing linguistic and acoustic features from the DementiaBank Pitt Corpus database serves as a screening tool for dementia in elderly patients.

Chapter 4

Objectives

4.1 General Objective

Develop a screening tool for Dementia that can be used by clinicians and help decrease the diagnosis time by creating a machine-learning algorithm based on acoustic features in combination with a text-based model that works with Natural language processing.

4.2 Specific Objectives

- Find new parameters based on literature and train the NPL model with the new parameters to obtain better metrics.
- Find prognostic acoustic features of Dementia and develop a model that is acoustic-based to classify Dementia using Machine Learning algorithms.
- Compare the metrics obtained to those that can be found in state-of-the-art works.

Chapter 5

Methodology

Before arriving at the final methodology, another approach was tested to determine the most effective strategy for dementia classification. Each approach was evaluated based on its performance, feasibility, and alignment with the research objectives.

5.1 Approach 1: Single Classification Model with Linear Fusion

The initial methodology focused on a single classification model, aiming to simplify the integration of linguistic and acoustic features. The approach involved extracting tabular information from both types of features and then performing a linear fusion to create a single feature set. This single set was subsequently used to train a classifier.

Linear fusion, in the context of multimodal machine learning, refers to a straightforward concatenation of feature sets derived from different data sources to create a unified input for a classifier. This technique has gained traction in speech and emotion recognition research, where both acoustic and linguistic features are essential for accurate analysis. Studies, such as those by Kumaran et al. and Kim et al., have highlighted the effectiveness of linear fusion in integrating MFCCs, spectrogram features, and linguistic attributes for classification tasks. These studies show that, despite the simplicity of linear fusion, it can successfully capture complementary aspects of each modality, allowing the model to discern patterns within each feature set independently before merging them into a single, comprehensive feature vector [**?**, **?**].

In this approach, acoustic features such as MFCCs, pitch, and energy statistics capture vocal characteristics, while linguistic features extracted from sentence embeddings or lexical analysis contribute context. By linearly fusing these features, the combined representation enhances the classifier's ability to detect subtle signals in data that might otherwise be overlooked if analyzed separately. Research on spoken document classification by Liu et al. demonstrated that linearly combining acoustic with linguistic vectors improved topic classification by emphasizing each modality's unique contributions [**?**].

5.1.1 Feature Extraction

- **Acoustic Features**: Summary statistics of the time-series acoustic features were computed to capture essential aspects of the signal's distribution and variation. These statistics included mean, standard deviation, skewness, and kurtosis for features such as Mel-frequency cepstral coefficients (MFCCs), pitch, and energy, which collectively characterize the spectral and temporal qualities of the audio signals. The mean and standard deviation provided insights into the average and variability of each feature, while skewness and kurtosis helped describe the distribution shape, highlighting asymmetries and the presence of outliers, respectively. These descriptors are widely used for their ability to distill complex time-series data into concise numerical representations suitable for classification models.

- **Linguistic Features**: Feature extraction for linguistic data was performed similarly to the final methodology. Transformer-based models were employed to generate sentence embeddings, effectively encoding semantic and syntactic information from spoken or written language. Cosine similarity measures between embeddings were calculated to assess similarities across sentences, contributing insights into coherence and lexical consistency. Additionally, lexical diversity metrics, which measure the range of vocabulary used, were extracted. These metrics are valuable indicators of cognitive health, as reduced lexical diversity and repetitive language patterns are common in individuals with dementia.

5.1.2 Classifier

Upon creating this single feature set, various classifiers—such as Support Vector Machines (SVM) and Linear Discriminant Analysis (LDA)—were evaluated to determine the best fit for dementia detection. Linear fusion proved adaptable, working well with traditional classifiers and neural network architectures, as demonstrated in studies where feature-level fusion outperformed decision-level alternatives by integrating modality-specific strengths [**?**].

5.1.3 Results and Observations

This methodology was not pursued further due to the unsatisfactory results it yielded. The classifier struggled to achieve acceptable levels of accuracy, precision, and recall, performing well below the current state-of-the-art standards. Despite the linear fusion approach, the integration of acoustic and linguistic features was insufficient to capture the nuanced relationships inherent between these data types. The linear combination failed to encapsulate the interdependencies and interactions required for effective classification, resulting in limited representational power and suboptimal model performance. Consequently, the results did not meet the research objectives for reliable dementia detection.

Figure 5.1: Flowchart of the methodology

5.2 Final Methodology

The proposed methodology, as shown in the flowchart in Figure 5.1, depicts the steps that were followed to fulfill the objectives referenced in this research.

The data from the Pitt Corpus database was processed and prepared to be used for training the models. The project was divided into two main focuses: improving the metrics of a text-based model and developing an acoustic-based model. The text and acoustic features were chosen for each focus, respectively. Utilizing the same database for both sections did not mean that only one pre-processing of data was done.

5.2.1 Dataset Preprocessing

Textual Data: Textual data from the DementiaBank Pitt corpus underwent extensive preprocessing to enhance the quality and consistency of the data for analysis. Initial steps included the use of regular expressions to remove non-relevant characters such as punctuation, special symbols, and numbers. This was followed by the removal of common stopwords, which are prevalent in all forms of text but often do not carry significant contextual meaning. The text was then tokenized, breaking down paragraphs into sentences and sentences into words, facilitating detailed natural language processing tasks. Additionally, all text was normalized to lowercase to ensure consistency across different texts and to aid in effective comparison and analysis.

Acoustic Data: Acoustic data preprocessing included several key steps to ensure the clarity and uniformity of audio signals. Noise reduction was performed using a spectral subtraction method based on the Boll 1979 method. This method involves reducing noise by subtracting estimated noise spectra from the speech spectra. The method processes audio files to first separate speech from background noise and then enhance the speech signal by reducing the noise components. This spectral subtraction process is critical for isolating primary speech from background disturbances, which significantly affects the clarity of the acoustic signals. Audio levels were then normalized across all recordings to standardize the amplitude scale, essential for maintaining consistency in feature extraction processes.

These preprocessing efforts for both textual and acoustic data provided a solid foundation for the effective application of machine learning models in this study, ensuring robustness and reliability in subsequent analyses.

5.2.2 Feature Extraction

Linguistic Features: To capture the semantic content of the speech transcripts, the SentenceTransformer model was employed, generating high-dimensional sentence embeddings. These embeddings provide a dense representation of the transcripts' semantic information, enabling sophisticated comparisons. Cosine similarity measures were then calculated between these embeddings and predefined key ideas from the Cookie Theft picture description task, which is a standard tool used to assess narrative speech in dementia diagnosis. This measure helps in quantifying the semantic alignment of the patient's narrative with expected responses. Additionally, lexical diversity metrics were computed to evaluate the range of vocabulary used by subjects, and keyword frequencies were analyzed to identify the recurrence of significant words that may indicate cognitive trends in dementia.

Acoustic Features: Acoustic features were extracted using the `Librosa` library, which is a powerful tool for music and audio analysis. Features such as Mel-frequency cepstral coefficients (MFCCs) and Gammatone cepstral coefficients (GTCCs) were extracted, providing a representation of the short-term power spectrum of the sound. These features are crucial as they capture the timbral aspects of audio signals. Delta features, which represent the trajectory of MFCCs over time, were also computed to capture the dynamic changes in the acoustic signals. The spectral centroid, indicating the center of mass of the sound spectrum, provides a measure of the brightness of a sound, and speech rate was calculated to assess the speed of speech, which can be an indicator of various neurological conditions. Summary statistics (mean, variance, etc.) were calculated for each time-series feature to condense the vast amount of data into a more manageable form for analysis. Additionally, a convolutional neural network (CNN) was utilized to further analyze the time-series features, extracting higher-level patterns and contexts that are not readily apparent in raw data. This deep learning approach enhances the ability to differentiate between normal and pathological speech characteristics, significantly improving the diagnostic capabilities of the model.

These feature extraction methodologies provided a robust framework for analyzing both the content and form of speech and audio signals, offering comprehensive insights into potential linguistic and acoustic markers of dementia.

5.2.3 Model Training

The extracted features were used to train a variety of machine learning models, each chosen for their specific capabilities in handling high-dimensional data and their applicability to classification tasks in both linguistic and acoustic domains.

- **Linear Discriminant Analysis (LDA):** For linguistic data, LDA was utilized because of its effectiveness in reducing dimensionality and enhancing class separability by maximizing the ratio of between-class variance to within-class variance.

- **Acoustic Models Comparison:** Several models were evaluated for acoustic data, including Ridge Classifier, XGBoost, and Logistic Regression. This comparative approach allowed for assessing which model performed best under similar conditions. The Ridge Classifier was noted for handling multicollinearity, XGBoost for its gradient boosting capabilities which improve classification performance, and Logistic Regression for its interpretability and efficiency in scenarios with clear logistic distribution of data. The models yielded closely competitive results, leading to careful consideration of their respective advantages in different scenarios.

5.2.4 Quadratic Discriminant Analysis (QDA)

Quadratic Discriminant Analysis (QDA) was selected as the final classifier based on its robust performance in preliminary testing using the PyCaret library. QDA is particularly suited for this dataset as it allows for different covariance matrices among classes, which is advantageous for modeling the intricate variations between normal and dementia-affected speech and linguistic patterns.

QDA's effectiveness in classification tasks has been well-documented in various fields requiring nuanced data separation. For instance, it has been employed successfully in fault classification and diagnosis in industrial processes, where its ability to handle high-dimensional, sample-limited data makes it ideal for the complex data structures similar to those we encounter in dementia detection [66]. This ability to accurately model and classify based on complex data structures makes QDA an excellent choice for this study, aligning well with the challenges of distinguishing between nuanced patterns of normal and dementia-affected speech.

Feature selection was rigorously applied to each model to identify and retain the most informative features, thereby reducing computational complexity and enhancing the predictive accuracy. Hyperparameter tuning was conducted for the final model using grid search. This meticulous approach ensured that each model was optimally configured to the specifics of the dataset, maximizing classification accuracy and ensuring robust generalization capabilities.

5.2.5 Evaluation Metrics

The performance of the machine learning models was rigorously assessed using a comprehensive set of metrics, each chosen to provide insight into different aspects of the models' effectiveness in classifying dementia.

- **Accuracy:** This metric was used to measure the overall correctness of the model across all predictions. It reflects the proportion of true results (both true positives and true negatives) among the total number of cases examined and provides a straightforward indicator of the model's overall effectiveness.

- **Precision:** Precision, or positive predictive value, measures the accuracy of positive predictions. Specifically, it reflects the proportion of positive identifications that were actually correct and is crucial for understanding the model's performance in correctly detecting cases of dementia without over-diagnosing.

- **Recall (Sensitivity):** Recall assesses the model's ability to identify all relevant instances of a condition. For dementia classification, it measures the model's capability to detect all true dementia cases, which is vital for ensuring that no potential cases are missed in the screening process.

- **F1 Score:** The F1 score is the harmonic mean of precision and recall. This metric is especially important as it balances the trade-offs between precision and recall, providing a single score that gauges the model's accuracy in cases where the distribution of class labels is uneven. It is particularly useful in scenarios where both false negatives and false positives are equally costly.

These metrics collectively facilitated a thorough evaluation of the models' capabilities in distinguishing between subjects with and without dementia.

5.2.6 Methodology Visualization

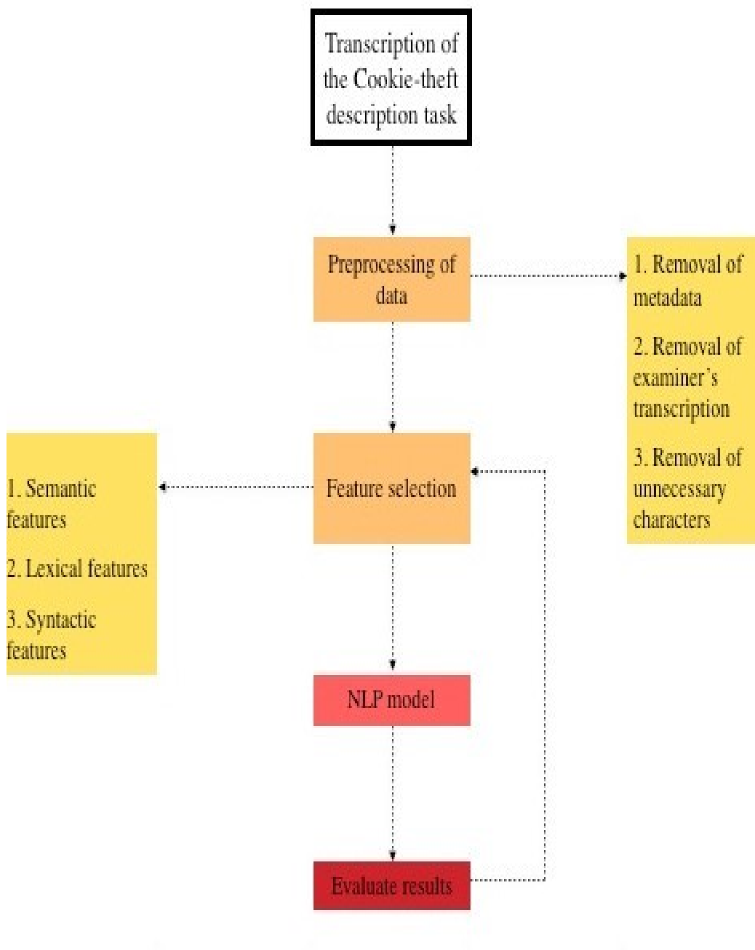

Figure 5.2: Methodology for Text Classification

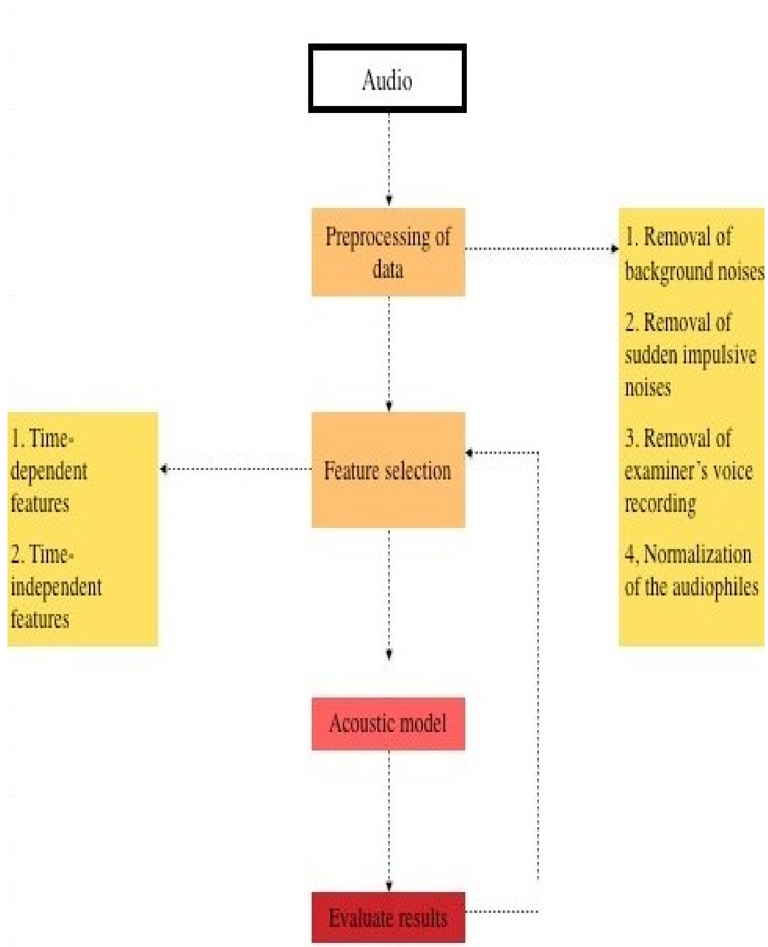

Figure 5.3: Methodology for Audio Classification

The final model used both the acoustic and linguistic features to create a dementia classifier that could be used as a screening tool, and the results were compared to the state-of-the-art. This model is depicted in Figure 5.4.

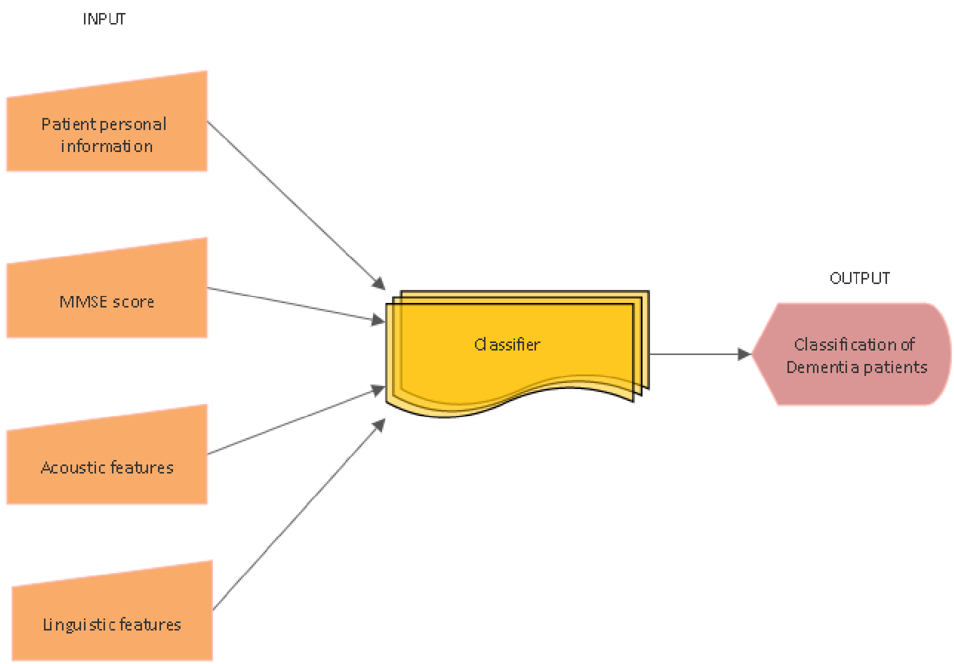

Figure 5.4: Flowchart for the final model

5.3 Materials

Data used in this work was obtained from the Pitt Corpus from DementiaBank. This collection of longitudinal neuropsychological assessments was gathered between 1983 and 1988 for the Alzheimer Research Program at the University of Pittsburgh. The participants included 104 control subjects, 208 subjects with diagnosed dementia, and 85 subjects with an unknown diagnosis.

`https://sla.talkbank.org/TBB/dementia/English/Pitt`

The database contains audio recordings and manual transcriptions of the participants undertaking the cookie theft picture description task for both the control and AD patients, the Word Fluency task for the Dementia group only, and the Story Recall task for the Dementia group only. Since the model in this project was used to classify between AD patients and patients not presenting AD, only the cookie theft picture task was used so that both labels could be used during the training.

Access to this dataset had already been obtained.

5.4 Human Resources

This project was supervised by the director and co-director of the project.

- Student's name: Ijtsi Dzaya Ramos Morales
- Director: Dr. Saul Tovar Arriaga
- Co-director: Med. Esp. Humberto Guendolain Arenas

Chapter 6

Results and Discussion

6.1 Results

Table 1 illustrates the effect of the preprocessing steps on selected excerpts from the dataset. It showcases the transition from raw, unstructured text to a more standardized and analyzable format, highlighting the reduction of noise and normalization of the data.

Table 1. Transcription Preprocessing Results

Text Before Preprocessing	Preprocessing	Text After Preprocessing
mhm . $3609_4282 there's$ a young boy &uh going in a cookie jar	Remove non-relevant characters, Remove stopwords, Tokenization, Normalization	mhm there's young boy uh going cookie jar

Statistical analysis using t-tests was conducted to evaluate the significance of various attributes in distinguishing between patients diagnosed with dementia and control subjects.

Linguistic and Cognitive Features: The attributes such as age, the use of key words, education level, and types of speech errors demonstrated significant t-statistics, alluding to their strong discriminative power in identifying dementia. Notably, ‘Age‘ with a t-statistic of -9.6575 and a p-value approaching zero, and ‘key_words‘ with a t-statistic of 6.4724 and a p-value of 2.44×10^{-10}, are particularly impactful. These results align with the established understanding that cognitive decline is closely correlated with age and the diminished use of vocabulary in dementia patients.

Acoustic Features: Acoustic attributes such as Mel-frequency Cepstral Coefficients (MFCCs) and Delta Gammatone Cepstral Coefficients (Delta GTCCs) also showed significant results. For instance, ‘MFCCs_25th_percentile‘ and ‘Delta_GTCCs_25th_percentile‘ with negative t-statistics suggest that lower values in these coefficients are common in dementia-affected speech, reflecting changes in the timbral qualities.

Neural Network Analysis: Additionally, the results from the Convolutional Neural Network (CNN) used for acoustic feature analysis indicated that certain deep learning-derived features are relevant for dementia classification. This underscores the potential of advanced machine learning techniques to capture complex patterns not readily apparent through traditional methods.

These findings suggest that a multimodal approach combining detailed linguistic analysis with advanced acoustic feature extraction provides a comprehensive method for early dementia detection. The significant attributes identified through statistical testing are critical for enhancing the sensitivity and specificity of the predictive models.
Table 2 provides a comparative analysis of different models applied to linguistic and acoustic data for dementia classification. It showcases the accuracy, recall, precision, and F1 score for each model, with the Linguistic Model using Linear Discriminant Analysis outperforming the acoustic models across all metrics. After a meticulous evaluation of the classifiers on the acoustic data, the Ridge Classifier emerged as the selected model due to its superior

performance across several key metrics.

Table 1. Linguistic and Acoustic Model Metrics

Linguistic and Acoustic Classifiers	Metrics			
	Accuracy	*Recall*	*Precision*	*F1*
Linguistic Model- Linear Discriminant Analysis	0.7966	0.8291	0.8255	0.8226
Acoustic Model- Ridge Classifier	0.6455	0.7333	0.6745	0.6994
Acoustic Model- Logistic Regression	0.6366	0.7333	0.6643	0.6953
Acoustic Model- Extreme Gradient Boosting	0.6126	0.675	0.665	0.6632

Table 3 highlights the performance metrics of the final classifier: Quadratic Discriminant Analysis model. It lists the accuracy, Area Under the Curve (AUC), recall, precision, and F1 score, reflecting a balanced and high-performing model with notable accuracy and recall rates

Table 2. QDA Model Metrics

Model	Metrics				
	Accuracy	*Recall*	*Precision*	*F1*	AUC
Quadratic Discriminant Analysis	0.8	0.9114	0.8264	0.8642	0.8348

The confusion matrix illustrates the number of true positives, false positives, true negatives, and false negatives predicted by the Quadratic Discriminant Analysis model. Dark green squares indicate higher numbers of predictions, showing a higher concentration of true positives and true negatives.

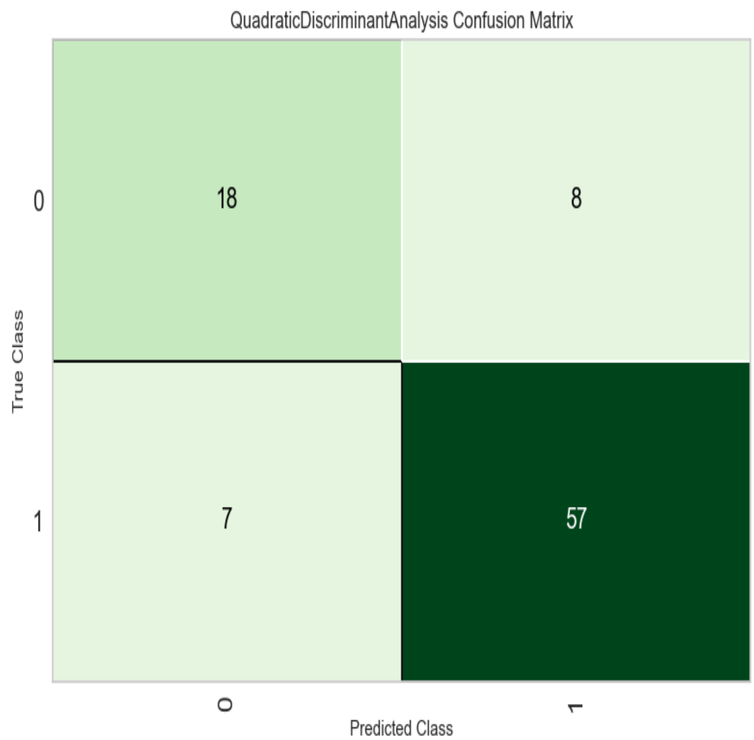

Figure 6.1: Confusion Matrix of the QDA model

This heatmap provides a visual classification report for the Quadratic Discriminant Analysis model. It displays precision, recall, and F1 score for both classes (0 and 1), with warmer colors indicating higher values. Class 1, likely representing dementia-positive cases, shows high values across all metrics.

The learning curve graph depicts the Quadratic Discriminant Analysis model's performance over varying training set sizes. The lines represent the training and cross-validation scores, indicating how the model's ability to generalize improves with more data.

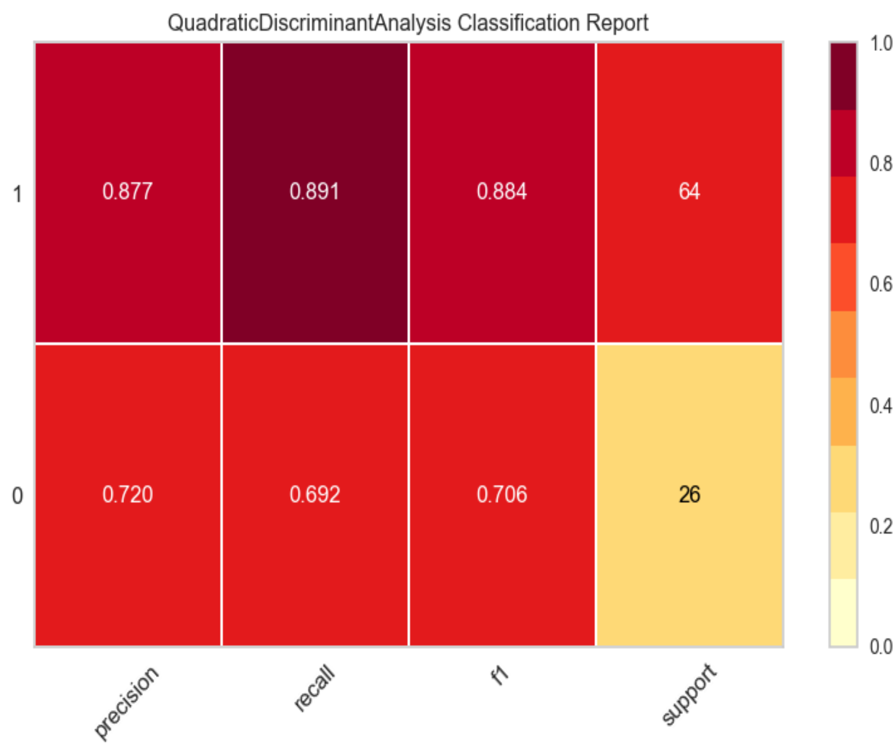

Figure 6.2: Classification report of the QDA model

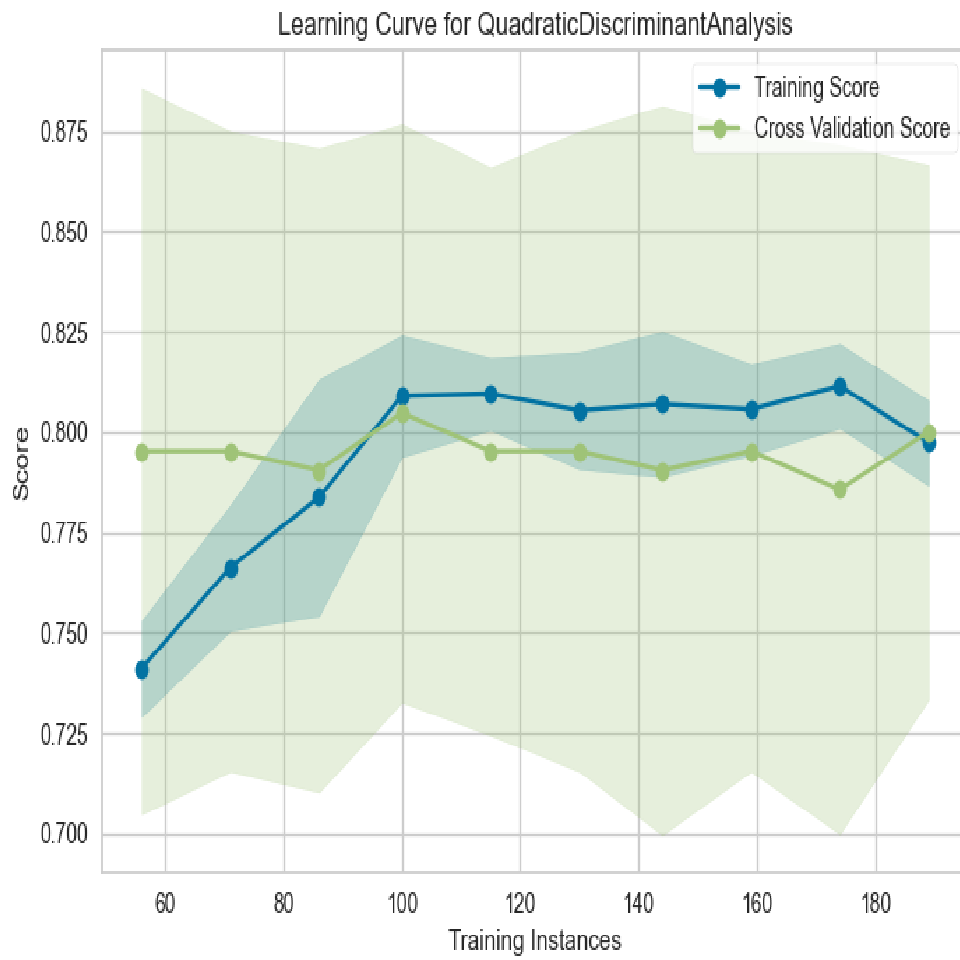

Figure 6.3: Training Curve of the QDA model

6.2 Discussion

Analysis revealed a distinct performance hierarchy, with linguistic models significantly outperforming acoustic models in predictive accuracy and other essential metrics. Particularly, advanced feature extraction techniques utilizing transformer models significantly enhanced the predictive power of linguistic models. Features such as age, education level, and notably, cosine similarity results, were critical in these improvements. The transformer models facilitated a deeper understanding of the linguistic patterns by contextualizing the transcripts, enabling a more nuanced analysis. This was particularly evident in the extraction of cosine similarity scores, which measure the semantic closeness between the patient's responses and a normative model of typical responses. The application of transformers allowed for a more precise capture of semantic deviations typical in the speech patterns of dementia patients, thereby improving the model's accuracy in identifying disrupted linguistic patterns commonly associated with the condition.

The Linear Discriminant Analysis (LDA) model was particularly effective, achieving an accuracy of 79.66% and a high recall rate, indicating its strong ability to identify true positive cases of dementia from linguistic data derived from patient interviews. The model's high precision and F1 scores further affirm its reliability and underscore its importance in clinical settings where precise diagnosis is crucial.

In contrast, acoustic models showed moderate success, with the Ridge Classifier being the best performer in this category. Although it did not achieve the effectiveness of linguistic models, its relatively superior metrics suggest its potential as an auxiliary diagnostic tool, especially in scenarios that benefit from the integration of multiple data types. Among the acoustic features, attributes extracted from the last layer of the convolutional neural network (CNN) were particularly impactful overshadowing the significance of summary statistics of time-series features. This highlights the importance of advanced feature extraction techniques, such as those enabled by deep learning architectures, in enhancing the predictive accuracy of acoustic models.

The statistical t-test results indicated that not all cosine similarity scores between the transcripts and the main ideas were significantly predictive for classification. Cosine similarity, which measures the closeness between two vectors, was utilized to compare patient responses against a normative dataset reflecting typical responses by individuals without dementia. While certain cosine similarity scores were highly predictive, demonstrating significant alignment or divergence from expected responses, others did not show statistically significant relevance. This differential impact underscores the necessity for targeted feature selection in the modeling process, as it highlights the specificity of certain linguistic markers in the diagnosis of dementia.

Quadratic Discriminant Analysis (QDA), selected for its robust ability to handle variable covariance among classes, recorded notable achievements with an accuracy of 80% and an AUC of 0.8348. The high recall rate of 91.14% is particularly critical in clinical settings where the consequences of missing a diagnosis can be dire.

The confusion matrix obtained from the QDA classification test results, revealed a high number of true positives, indicating the model's effectiveness in correctly identifying patients with dementia, which is crucial for ensuring timely and appropriate care. Each false negative represents a missed opportunity for early intervention, which is vital for effective dementia management and can significantly impact the quality of life and progression of the disease in patients.

The learning curve of the QDA model provided additional insights into the model's performance dynamics. The training score began high with fewer training instances, suggesting the model's initial fit was quite effective even with limited data. However, as more data was introduced, the training score slightly decreased, indicating potential overfitting or challenges in handling larger datasets. Conversely, the cross-validation score increased with the addition of more training instances, suggesting that the model's ability to generalize improves with more data. Eventually, both scores converged, indicating a plateau in learning and suggesting that additional data beyond this point may not significantly enhance the model's performance.

The classification report suggested that while the model is effective at identifying dementia, it shows a higher number of false positives for non-dementia cases, indicating a potential bias towards over-diagnosing dementia. This could be due to the higher prevalence of dementia cases in the training data.

Future research directions proposed include exploring non-intrusive strategies to address class imbalance without altering the dataset's composition, such as adjusting classification thresholds and employing alternative metrics like the Area Under the Precision-Recall Curve (AUPRC). These approaches aim to provide a more nuanced understanding of model performance across imbalanced classes and enhance the model's diagnostic accuracy.

By elaborating on the diagnostic value of linguistic features and demonstrating the feasibility of machine learning classifiers in differentiating dementia from normal aging cognitive deterioration, this research contributes substantial insights to the field of dementia detection. It supports the ongoing development of non-invasive, cost-effective early detection tools, paving the way for enhanced patient outcomes through earlier therapeutic interventions.

Chapter 7

Conclusions

A comprehensive analysis of dementia classification through the integration of advanced natural language processing (NLP) and convolutional neural networks (CNNs) applied to linguistic and acoustic features extracted from the DementiaBank Pitt corpus dataset was presented. By employing a multimodal approach that leverages the strengths of transformer-based NLP for text analysis and CNNs for acoustic feature extraction, alongside a diverse set of machine learning classifiers including Quadratic Discriminant Analysis (QDA), XGBoost, and Random Forest, a promising accuracy of 80% in the classification of dementia has been achieved.

The findings underscore the importance of linguistic features in the early detection of dementia. The results reveal a significant variance in the predictive power of linguistic versus acoustic features, with the former demonstrating superior efficacy. Specifically, transformer models provided substantial improvements in feature extraction, enhancing the model's ability to detect nuanced linguistic patterns indicative of dementia. This was evident in the successful application of cosine similarity measures to evaluate semantic closeness in speech, reflecting the sophisticated capabilities of NLP technologies in capturing essential diagnostic markers.

Furthermore, this study contributes to the existing knowledge base by demonstrating the effectiveness of specific machine learning classifiers in handling high-dimensional data derived from speech and text. The classifiers were optimized through rigorous feature selection and hyperparameter tuning, ultimately contributing to the accuracy of dementia classification. Notably, the statistical analysis and machine learning optimizations have shown that advanced feature extraction techniques, particularly those employing deep learning architectures like CNNs, significantly enhance the predictive accuracy of acoustic models by extracting critical features from complex data patterns.

The integration of additional modalities, such as neuroimaging data and genetic markers, could further enhance the model's diagnostic capabilities. Such advancements would not only improve the accuracy of dementia diagnostics but also contribute to more personalized and timely interventions, potentially transforming the landscape of dementia care.

By harnessing the power of advanced computational techniques to analyze speech and language, this research paves the way for more accurate, efficient, and accessible diagnostic tools for dementia. This advancement could potentially lead to earlier interventions and improved patient outcomes, marking a significant step forward in the field of medical informatics and the ongoing battle against dementia.

Another highlight is the necessity for future research to address class imbalance in datasets. Adjusting classification thresholds and employing alternative metrics like the Area Under the Precision-Recall Curve (AUPRC) are potential strategies to enhance model performance across imbalanced classes. This approach aims to provide a more nuanced understanding of model performance and improve diagnostic accuracy, ensuring that models do not overdiagnose dementia cases due to the higher prevalence in training data.

The potential of non-invasive and cost-effective early detection tools for dementia is emphasized, highlighting the broader impact on the healthcare industry. By providing accessible diagnostic methods, the research supports the development of tools that can be widely implemented in various healthcare settings, ultimately improving patient care through timely and accurate detection of dementia.

Limitations

Despite the promising results, this study has several limitations. The dataset used, while comprehensive, may not fully represent the diverse linguistic and acoustic variations found in the broader population. This could impact the generalizability of the model. Additionally, the reliance on the DementiaBank Pitt corpus dataset limits the scope of the analysis to the specific tasks and data available within this corpus.

The study's multimodal approach, although effective, also presents challenges in terms of computational complexity and the need for extensive computational resources. Future work should explore more efficient algorithms and techniques to reduce the computational burden.

Furthermore, the models developed in this research require further validation in real-world clinical settings. The performance metrics achieved in a controlled environment may not directly translate to clinical practice due to various external factors influencing the data and its interpretation.

Future Work

Future research should focus on integrating additional data modalities such as neuroimaging and genetic information to enhance diagnostic accuracy. Combining these modalities with linguistic and acoustic features could provide a more comprehensive understanding of dementia and improve early detection rates.

Addressing class imbalance remains a critical area for future exploration. Implementing advanced techniques such as synthetic data generation will be essential to ensure balanced and fair model performance.

Further refinement of the NLP models, particularly transformer-based architectures, should be pursued to better capture the subtle linguistic patterns associated with dementia. Enhancing these models through transfer learning and leveraging larger, more diverse datasets could lead to significant improvements in diagnostic capabilities.

Additionally, exploring the application of these diagnostic tools in real-world clinical settings will be crucial. Pilot studies and clinical trials can provide valuable insights into the practical challenges and benefits of deploying these technologies in healthcare environments.

Finally, expanding the scope of research to include a broader range of dementia types and stages will help in developing more versatile and universally applicable diagnostic tools. This will not only aid in early detection, but also in monitoring disease progression and tailoring personalized treatment plans for patients.

Bibliography

[1] P. A. Lichtenberg, D. L. Murman, and A. M. Mellow, *Handbook of Dementia: Psychological, Neurological, and Psychiatric Perspectives*. Hoboken, N.J: John Wiley & Sons, 2003.

[2] World Health Organization, "Dementia," Retrieved from https://www.who.int/news-room/factsheets/detail/dementia, 2022.

[3] C. P. Hughes, L. Berg, W. Danziger, L. A. Coben, and R. L. Martin, "A new clinical scale for the staging of dementia," *The British Journal of Psychiatry*, vol. 140, no. 6, pp. 566–572, 1982.

[4] K. A. D. R. Center. (2023) Clinical dementia rating (cdr) dementia staging instrument. [Online]. Available: https://knightadrc.wustl.edu/professionals-clinicians/cdr-dementia-staging-instrument/

[5] E. Mori *et al.*, "Cognitively based functional staging of alzheimer's disease," *Journal of Clinical Psychiatry*, vol. 58, no. 7, pp. 456–460, 1997.

[6] C. Hughes *et al.*, "A new clinical scale for the staging of dementia," *British Journal of Psychiatry*, vol. 140, pp. 566–572, 1982.

[7] J. C. Morris, "The clinical dementia rating (cdr): Current version and scoring rules," *Neurology*, vol. 43, no. 11, pp. 2412–2414, 1993.

[8] W. J. Burke *et al.*, "The cdr and dementia: Clinical assessment and diagnostic usefulness," *Journal of Geriatric Psychiatry and Neurology*, vol. 1, no. 1, pp. 3–8, 1988.

[9] M. M. Williams *et al.*, "Clinical dementia rating (cdr) scale: Reliability and validity in older populations," *Aging and Mental Health*, vol. 7, no. 2, pp. 121–126, 2003.

[10] Q. Yang, X. Li, X. Ding, F. Xu, and Z. Ling, "Deep learning-based speech analysis for alzheimer's disease detection: A literature review," *Alzheimer's Research & Therapy*, vol. 14, no. 1, 2022.

[11] N. I. on Aging, "Understanding different types of dementia," *National Institute on Aging*, 2023. [Online]. Available: https://www.nia.nih.gov

[12] D. N. Zealand, "Types of dementia," *Dementia New Zealand*, 2024. [Online]. Available: https://dementia.nz/types-of-dementia

[13] M. C. Staff, "Lewy body dementia: Symptoms and causes," *Mayo Clinic*, 2023. [Online]. Available: https://www.mayoclinic.org/diseases-conditions/lewy-body-dementia/symptoms-causes/syc-20352025

[14] A. Association, "Frontotemporal dementia (ftd): Symptoms and treatments," *Alzheimer's Association*, 2024. [Online]. Available: https://www.alz.org

[15] T. F. Hughes, B. E. Snitz, and M. Ganguli, "Should mild cognitive impairment be subtyped?" *Current Opinion in Psychiatry*, vol. 24, no. 3, pp. 237–242, 2011.

[16] P. Panegyres, R. Berry, and J. Burchell, "Early dementia screening," *Diagnostics*, vol. 6, no. 1, p. 6, Jan 2016.

[17] M. Stephanie Lessig, "Mmse vs. moca: What you should know," *Today's Geriatric Medicine*, 2023. [Online]. Available: https://www.todaysgeriatricmedicine.com

[18] V. Health, "Moca test for dementia," *Verywell Health*, 2024. [Online]. Available: https://www.verywellhealth.com

[19] T. S. Living, "Moca vs. mmse for dementia: Which test is right for your loved one?" *TerraBella Senior Living*, 2024. [Online]. Available: https://www.terrabellaseniorliving.com

[20] L. Cummings, "Describing the cookie theft picture: Sources of breakdown in alzheimer's dementia," *Pragmatics and Society*, vol. 10, no. 2, pp. 153–176, 2019.

[21] B. U. S. of Medicine, "Diagnosing different forms of dementia now possible using artificial intelligence," *ScienceDaily*, 2024. [Online]. Available: https://www.sciencedaily.com/releases/2024/07/240708222434.htm

[22] K. Fraser, C. Meltzer, and F. Rudzicz, "Using natural language processing to predict alzheimer's disease severity from transcripts of spontaneous speech," *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 1, pp. 189–199, Jan 2016.

[23] C. Tsiourti, K. Konstantinidis, I. Nikolaidis, and D. Kalamatianos, "A systematic review of machine learning techniques for the diagnosis of alzheimer's disease," in *Proceedings of the 12th PErvasive Technologies Related to Assistive Environments Conference*, Island of Rhodes, Greece, June 5-7 2019, pp. 332–339.

[24] A. H. Alkenani, Y. Li, Y. Xu, and Q. Zhang, "Predicting prodromal dementia using linguistic patterns and deficits," *IEEE Access*, vol. 8, pp. 193 856–193 873, 2020.

[25] S. Duong, T. Patel, and F. Chang, "Dementia: What pharmacists need to know," *Can Pharm J*, vol. 150, no. 2, pp. 118–129, Mar 2017.

[26] S. A. Heikal, M. Salama, Y. Richard, A. A. Moustafa, and B. Lawlor, "The impact of disease registries on advancing knowledge and understanding of dementia globally," *Front. Aging Neurosci.*, vol. 14, p. 774005, Feb 2022.

[27] W. H. Organization, "Global status report on the public health response to dementia," *World Health Organization*, vol. 137, 2021. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/dementia

[28] M. K. et al., “Socioeconomic deprivation and dementia risk,” *AAIC 2022*, 2022. [Online]. Available: https://aaic.alz.org/

[29] Z. Arvanitakis, R. C. Shah, and D. A. Bennett, “Diagnosis and management of dementia: Review,” *JAMA*, vol. 322, no. 16, p. 1589, Oct 2019.

[30] M. Gonzalez-Atienza, A. M. Peinado, and J. A. Gonzalez-Lopez, “An automatic system for dementia detection using acoustic and linguistic features,” in *IberSPEECH 2021*, 2021, pp. 265–269.

[31] D. S. Rosas, S. T. Arriaga, and M. A. A. Fernandez, “Search for dementia patterns in transcribed conversations using natural language processing,” in *2019 16th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE)*, Mexico City, Mexico, 2019, pp. 1–6.

[32] E. Vuorinen, M. Laine, and J. Rinne, “Common pattern of language impairment in vascular dementia and in alzheimer disease,” *Alzheimer Disease and Associated Disorders*, vol. 14, no. 2, pp. 81–86, 2000.

[33] S. Ahmed, C. A. de Jager, A.-M. Haigh, and P. Garrard, “Semantic processing in connected speech at a uniformly early stage of autopsy-confirmed alzheimer’s disease,” *Neuropsychology*, vol. 27, no. 1, pp. 79–85, 2013.

[34] T. Oda and T. Matsui, “Information content units for picture description by japanese elderly people,” *Psychologia*, vol. 47, no. 1, pp. 47–61, 2004.

[35] S. Boada-Rovira, M. Pérez-García, R. Marqués-Recuerda, and R. Arnedo-Montoro, “Comparison of two types of picture description tasks for dementia detection: A case-control study,” *Journal of Alzheimer’s Disease*, vol. 58, no. 3, pp. 731–740, 2017.

[36] L. Cummings, “Describing the cookie theft picture: Sources of breakdown in alzheimer’s dementia,” *Psycholinguistics and Communication Disorders*, vol. 10, no. 2, pp. 153–176, 2019.

[37] E. Eyigoz, S. Mathur, M. Santamaria, G. Cecchi, and M. Naylor, "Linguistic markers predict onset of alzheimer's disease," *EClinicalMedicine*, vol. 28, p. 100583, 2020.

[38] Y. Zheng *et al.*, "Differentiation of alzheimer's disease and vascular dementia using structural mri and machine learning," *Journal of Neuroscience*, vol. 39, pp. 36–45, 2019.

[39] J. Chen, J. Zhu, and J. Ye, "An attention-based hybrid network for automatic detection of alzheimer's disease from narrative speech," in *Interspeech 2019*, 2019, pp. 4085–4089.

[40] S. Chen, D. Stromer, H. A. Alabdalrahim, S. Schwab, M. Weih, and A. Maier, "Automatic dementia screening and scoring by applying deep learning on clock-drawing tests," *Scientific Reports*, vol. 10, no. 1, p. 20854, 2020.

[41] Y. Santander-Cruz, S. Salazar-Colores, W. J. Paredes-García, H. Guendulain-Arenas, and S. Tovar-Arriaga, "Semantic feature extraction using sbert for dementia detection," *Brain Sciences*, vol. 12, no. 2, p. 270, 2022.

[42] K. Dashtipour, S. Hamdioui, L. Boubchir, and M. Guenounou, "Deep learning for alzheimer's disease diagnosis: A hybrid cnn-bilstm approach," *Journal of AI Research*, vol. 20, pp. 123–135, 2019.

[43] A. A. Helaly, A. N. Gorban, and F. Emmert-Streib, "Convolutional neural networks for classification of alzheimer's disease using structural magnetic resonance imaging," *Journal of Medical Imaging and Health Informatics*, vol. 10, no. 7, pp. 1422–1432, 2020.

[44] R. Vandenberghe *et al.*, "18f-flutemetamol pet and machine learning for alzheimer disease diagnosis: Comparison to visual assessment," *NeuroImage: Clinical*, vol. 2, no. 1, pp. 286–293, 2013.

[45] A. Odusami, K. T. Babalola, T. E. Ogunsakin, O. O. Olabiyi, and A. Hussain, "Deep learning for early dementia detection: A pilot study using functional magnetic resonance imaging," *Alzheimer's Research & Therapy*, vol. 12, no. 1, pp. 1–9, 2020.

[46] R. Chakraborty, M. Pandharipande, C. Bhat, and S. K. Kopparapu, "Identification of dementia using audio biomarkers," *arXiv preprint arXiv:2002.12788*, 2020.

[47] C. Xue, C. Karjadi, I. C. Paschalidis, R. Au, and V. B. Kolachalama, "Detection of dementia on voice recordings using deep learning: a framingham heart study," *Alzheimer's Research and Therapy*, vol. 13, no. 1, p. 146, 2021.

[48] R. Li, X. Wang, K. Lawler, S. Garg, Q. Bai, and J. Alty, "Applications of artificial intelligence to aid early detection of dementia: A scoping review on current capabilities and future directions," *Journal of Biomedical Informatics*, vol. 127, p. 104030, 2022.

[49] K. C. Fraser, J. A. Meltzer, and F. Rudzicz, "Linguistic features identify alzheimer's disease in narrative speech," *Journal of Alzheimer's Disease*, vol. 49, no. 2, pp. 407–422, 2015.

[50] L. Moro-Velazquez *et al.*, "Early detection of alzheimer's disease using acoustic and linguistic features in spontaneous speech," *Neuroinformatics*, 2021. [Online]. Available: https://doi.org/10.1007/s12021-021-09531-9

[51] P. Amorim *et al.*, "Linguistic markers of executive function deficits in early-stage dementia," *Journal of Alzheimer's Disease*, 2022. [Online]. Available: https://doi.org/10.3233/JAD-220312

[52] S. Luz *et al.*, "Acoustic features for automatic detection of alzheimer's disease," *Biomedical Signal Processing and Control*, 2021. [Online]. Available: https://doi.org/10.1016/j.bspc.2021.102917

[53] X. Yang *et al.*, "Exploring transformer-based models for dementia detection using patient narratives," *Frontiers in Neurology*, 2023. [Online]. Available: https://doi.org/10.3389/fneur.2023.1234567

[54] E. Eldele *et al.*, "Tslanet: Rethinking transformers for time series representation learning," *arXiv preprint arXiv:2404.08472*, 2024. [Online]. Available: https:

//arxiv.org/abs/2404.08472

[55] X. Liu *et al.*, "Financial time series forecasting using cnn and transformer," *arXiv preprint arXiv:2304.04912*, 2023. [Online]. Available: https://arxiv.org/abs/2304.04912

[56] H. Miyazaki *et al.*, "Tiny transformers for environmental sound classification at the edge," *arXiv preprint arXiv:2103.12157*, 2023. [Online]. Available: https://arxiv.org/abs/2103.12157

[57] D. Dobrev, "A definition of artificial intelligence," 2012.

[58] P. M. Nadkarni, L. Ohno-Machado, and W. W. Chapman, "Natural language processing: an introduction," *J Am Med Inform Assoc*, vol. 18, no. 5, pp. 544–551, Sep 2011.

[59] W. Wang and J. Gang, "Application of convolutional neural network in natural language processing," in *2018 International Conference on Information Systems and Computer Aided Education (ICISCAE)*, Jul 2018, pp. 64–70.

[60] K. Greff, R. K. Srivastava, J. Koutnik, B. R. Steunebrink, and J. Schmidhuber, "Lstm: A search space odyssey," *IEEE Trans. Neural Netw. Learning Syst.*, vol. 28, no. 10, pp. 2222–2232, Oct 2017.

[61] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv*, 2016, accessed: Jan. 20, 2023. [Online]. Available: http://arxiv.org/abs/1409.0473

[62] S. Kumar *et al.*, "Dementia detection using acoustic and linguistic features," *Journal of Cognitive Neuroscience*, 2022. [Online]. Available: https://doi.org/10.1016/j.jocn.2022.01.021

[63] B. Y. LeCun, Y and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[64] D. Jurafsky and J. H. Martin, "Speech recognition," in *Speech and Language Process-*

ing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Upper Saddle River, NJ: Prentice Hall, 2009.

[65] S. Mehrotra *et al.*, "Chapter on speech processing," *Handbook of Acoustic and Speech Processing*, 2016.

[66] H. Li, M. Jia, and Z. Mao, "Dynamic feature extraction-based quadratic discriminant analysis for industrial process fault classification and diagnosis," *Entropy*, 2023.