Neuro-Symbolic AI for Mental Health and Well-being

by

Ananda Biswas
21301635
Ananna Paul
20201038
Chinmoy Saha Shourov
21301720
Abdul Munnaf Rafsan
20201008
Alve Ahmed
20201161

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

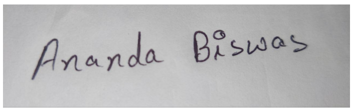
Department of Computer Science and Engineering
Brac University
February 2025

Declaration

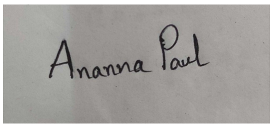
It is hereby declared that

1. The thesis submitted is our own original work while completing the degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material that has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

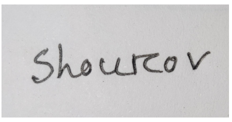
Student's Full Name & Signature:



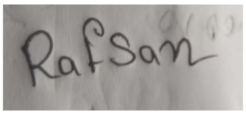
Ananda Biswas
21301635



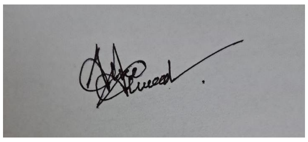
Ananna Paul
20201038



Chinmoy Saha Shourov
21301720



Abdul Munnaf Rafsan
20201008



Alve Ahmed
20201161

Approval

The thesis/project titled "Neuro-Symbolic AI for Mental Health and Well-being" Submitted by:

1. Ananda Biswas
2. Ananna Paul
3. Chinmoy Saha Shourov
4. Abdul Munnaf Rafsan
5. Alve Ahmed

Of Fall, 2024 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on February 09, 2025.

Examining Committee:

Supervisor:
(Member)

Dr. Md. Golam Rabiul Alam
Professor
Department of Computer Science and Engineering
School of Data and Sciences
BRAC University

Program Coordinator:
(Member)

Dr. Md. Golam Rabiul Alam
Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Sadia Hamid Kazi
Associate Professor & Chairperson
Department of Computer Science and Engineering
Brac University

Abstract

Neuro-Symbolic AI (NSAI)- the amalgamation of Neural AI with Symbolic AI leverages the approach of traditional machine learning models where the system of NSAI gives output by learning data's behavior using neural network learning techniques and symbolic reasoning converts the behaviors into human understandable symbols so that the reasoning is transparent. With the advancement of technology, mental illness is one of the inevitable issues that has already suppressed the magnitude of physical sickness. Considering the reputation of artificial intelligence, traditional machine learning techniques are the first contenders for the early detection of mental disorders despite their limitations such as lack of explainability, poor performance on imbalanced data, and a small amount of data. The limitations mislead the results for diagnosing and assisting mental diseases. In addition, Neuro-symbolic AI can be the holistic approach to prevent the lack of precision in existing machine learning models with small amounts of data and detect mental health patients by giving proper explanations as a handcrafted expert does. Our proposed model of Neuro-symbolic AI predicts Alzheimer's disease with 95.45% accuracy which decisively beats the result of traditional machine learning models. Therefore, the paper aims to find a better approach to the concerned people and play the role of assistant to detect the disease earlier and make their lives safe.

Keywords: Neuro-symbolic AI; Neural network; Symbolic reasoning; Traditional machine learning techniques; Mental disorder; Alzheimer's disease; Handcrafted expert.

Acknowledgement

Firstly, we want to be thankful to Almighty God for blessing us with the capability to finish the thesis.

Secondly, we like to express our profound gratitude towards our supervisor Dr. Md. Golam Rabiul Alam who is always with us during the hard times of the thesis period and gives suggestions and encouragement to coordinate the thesis so that the thesis can be leveled up.

Moreover, we are pleased to have a research assistant of the supervisor who helps us in difficult times to complete the thesis.

Additionally, we would like to thank Brac University for providing the opportunity to complete the thesis and study.

Just as importantly, we are grateful to our family for moral support and encouragement throughout the day.

Table of Contents

List of Figures

List of Tables

List of Algorithms

Chapter 1

Introduction

Neuro-symbolic AI is regarded as the outcome of the union of neural networks and symbolic reasoning which are the two separate fields of artificial intelligence. It applies the strengths of knowledge graphs and the learning capabilities of neural networks to solve complex problems. To begin with, the neural network is a subfield of machine learning which forms the base of deep learning. The algorithms of neural networks are inspired by the structure of the human brain or neurons. Neural networks process data by training themselves through multiple layers, recognizing the patterns of the data, and predicting the output. Furthermore, neural networks have shown remarkable performance in various domains such as image recognition, speech processing, and natural language understanding, making them an essential component of modern AI. On the other hand, Symbolic AI is known as classical AI which uses human-understandable symbols, logic, and representations. This method is particularly suitable for solving complex problems with predefined rules and logical inference. Moreover, symbolic AI is valuable in domains like healthcare, where understanding the reasoning behind decisions is critical. The main objective of Neuro-symbolic AI is to address complex problems from a short amount of data and provide an understandable explanation for each action. Integration of knowledge graphs in neural networks unlocks countless opportunities like explainability, performance, robustness, efficiency, interpretability, adaptations, and many more. This hybrid approach aims to overcome the limitations of traditional machine learning models which will offer groundbreaking solutions for complex real-world problems. Mental health has become a global concern. WHO [15] insights that out of every 8 people, 1 person suffers from a mental disorder and it is increasing over time. Approximately, 970 million people suffer from mental disorders right as of 2019's calculation and the number surely jumped after the curse of COVID-19. According to the Johns Hopkins Medicine Organization[33], approximately 9.5% of American adults are suffering from major depression or bipolar disorder. There are nearly 20 types of mental disorders including Neurocognitive disorders (Previously called Dementia). Though Alzheimer's disease is regarded as a mental health problem, it is primarily categorized as a neurodegenerative disease which is a notable example of Neurocognitive disease(Dementia). We are considering Alzheimer's disease for our research resource which is a brain disorder that happens because of the formation of certain proteins in the brain cells. Alzheimer's disease shrinks the brain cells and the brain cells die gradually. Alzheimer's disease is a category of Dementia that has an impact on memory, thinking capability, behavior, and social skills. These changes hamper a person's daily activities. According to a report from the Mayo Clinic (2024) [31], 6.5 million people aged 65 and older suffer from Alzheimer's disease. Among them, more than 70% are 75 years old and older. Approximately 55 million people worldwide live with dementia and every year nearly 10 million new dementia patients are identified, it is estimated that 60% to 70% have Alzheimer's disease (WHO, 2023) [23]. According to an article by the Alzheimer's Association (2024) [27], the number of people aged 65 and older with Alzheimer's disease is expected to rise to 12.7 million by 2050 if early detection and proper medical treatments are not introduced. The symptoms are-

- The early symptom is patients forget the newly learned information.
- Difficult to remember conversations and events.
- Repeat statements and questions again and again.
- There are significant pauses in the middle of a conversation.

Alzheimer's patients require careful observation over time, with regular clinic visits. Yang (2023) [8] experimented that there were only 4 nurses and 1 psychiatrist per 100,000 population which is very low considering the situation of today's world. There is a saying- No Health Without Mental Health. Moreover, there are insufficient resources to accommodate mental health problems. The global shortage of mental health professionals, combined with the rising demand for mental diseases, highlights the urgent need for AI-driven solutions that can assist in diagnosis and treatment planning. Since mental health is a sensitive area, people should be aware of their mental health and follow guidelines from psychiatrists. However, traditional machine learning models do not address mental health problems properly and effectively. Briefly, traditional machine learning models, while powerful in recognizing patterns, often fall short in offering explanations, especially when dealing with complex medical conditions like mental health disorders. Conversations with people having Alzheimer's disease(a subset of mental disorder) are often detectable early, so recording and analyzing the conversation can provide valuable information. So, we combined the neural networks and symbolic reasoning to develop the Neuro-symbolic model to operate the speech dataset(audio) of the diagnosed and undiagnosed patients and predict the outcomes with more precision and proper explanations. This model promises a highly optimistic outlook toward the future of AI research. In conclusion, the emergence of Neuro-symbolic AI in mental health will be a significant factor for mental disorders and deliver positive output to sufferers.

1.1 Research Problem

Machine learning has emerged during the last decade. It grabs the attention of people and influences the research area. It has a huge popularity in the case of classification problems. However, it has limitations when it comes to considering sensitive areas such as mental health issues. Though traditional machine learning models have shown impressive outcomes, a major issue arises whether there is enough labeled data to train the models. Also, collecting and preparing the relevant and necessary data is very challenging. Moreover, people feel insecure to share their health data to the public and it creates a problem of having less number of real dataset which creates a problem to evaluate the model real performance.

A. Bias: Traditional machine learning models often make biased predictions if the model is trained by insufficient data, biased or inconsistent data. Although these biases are generally unintentional, they will create serious issues like wrong medical diagnoses. Amazon's hiring tool is a notable example of how machine learning systems can be affected by biased data. Amazon [16] constructed an AI powered tool that is useful to find the appropriate for their company. However, after implementing the tool, it is found that it preferred resumes from male candidates over female candidates, especially for technical software development roles. After investigation they discovered that since men were more commonly employed in Amazon's technical sectors, the model learned and reflected the implicit gender-bias presented in the training dataset. In the case of mental illness, biased predictions can make the situation worse.

B. Explainability difficulty: The predictions made by the ML models are not always trustworthy because they lack explainability and generate misleading information. In the case of diagnosing a mental illness, it is crucial for doctors to understand properly why this model generates these certain outcomes. Without transparency, it is difficult to trust machine learning systems for diagnosing mental illness. This research paper will propose and apply a Neuro-symbolic AI model to overcome the challenges instead of conventional machine learning techniques.

1.2 Research Contributions

In this work, we develop a Neuro-symbolic model that will be more transparent and effective for diagnosing dementia disease. Our approach integrates the strength of neural networks with rule-based symbolic reasoning and overcomes the limitations of most of the machine learning algorithms, such as the explainability of reaching a decision. The main contributions of this paper are-

1. Neural AI (CNN-LSTM) is a hybrid model which learns the behavioral pattern of data and trains in such a way that it can accurately predict the behavior.
2. The addition of Symbolic AI which relies on human-defined rules, logical reasoning. This well-defined rules will make the system more explainable, reliable and trustworthy.
3. The proposed model of NSAI will aim to distinguish precisely between dementia patients and healthy people by analyzing the features of audio data by ensuring no patients will be affected by the output of the model.
4. We collected our dataset from Dementiabank, where research is conducted on a group of people who ask to read a description of an image.
5. Our model not only focuses specifically on classifying Dementia and Non-dementia individuals; it also provides the explanations why the audio sample is classified with the specific label.

By acquiring these objectives, our research will help to provide the diagnosis of mental illness like dementia (Alzheimer's disease) and overcome the limitation of most of the machine learning algorithms such as explainability of reaching a decision.

1.3 Report Organization

In this section, we will give a brief overview how the report is organized. The report is divided into 5 chapters that are Introduction, Literature Review, Methodology, Performance Analysis, Conclusion. These chapters also subdivided into some categories to present the research in the report.

Introduction: An overview of Neuro-Symbolic AI, Alzheimer's disease, the situation of mental disorders in the world will be given in this chapter. The research problem, motivation will be later discussed in the later part of this chapter.

Literature Review: In this chapter, we discuss existing works in the sector of mental disorder problems. The traditional machine learning models are used to solve the problem but there are some good research paper where they are using innovative and unique models to diagnose mental disorders.

Methodology: Data collection will be the first step of this chapter. This section describes how data has been collected for this research as our primary source. After that, the pre-processing of the data is provided in this chapter. Lastly, the proposed and specification models are highlighted here.

Performance Analysis: The models are evaluated by various performance matrices to legitimize the proposed model.

Conclusion: The research concludes with the summary and findings our research along with the future improvements of this research.

Chapter 2

Literature Review

Ripple Down Rule (RDR) is a rule-based system that permits human experts to transfer their knowledge to a knowledge-based system for decision-making. Unlike traditional machine learning systems, which are often difficult to understand, unable to provide multiple outputs and unsatisfactory performance with imbalance or unbalanced data.RDR offers a solution for diagnosing mental disorders which diagnoses multiple disorders at once. One significant application of the Ripple Down Rule (RDR) is it collects most of the data through surveys that are completed by the participants or their legal guardian's informed consent. The data collection procedure maintains the ethical guidelines granted by the scientific authorities which ensures the validity and reliability of the data. In the case of mental health maintaining privacy is very crucial. Addressing the limitations of current mental disorder screening tests, the RDR method provides a complete solution. Using the algorithm called MCRDR (Multiple Classification Ripple Down Rule), it successfully classifies 17 different mental disorders which many existing models cannot offer. Mental health experts' involvement who designing the rules for diagnosis of mental health increased the credibility and reliability of the RDR system. In addition, RDR is efficient for allowing experts to create rules appropriate to particular populations and environments. RDR also effectively preprocesses the data like grouping ages into different categories, utilizing the available information which provides a more accurate diagnosis. To evaluate the mental disorder accurately RDR system always focuses on relevant high-quality information. To sum up, the Ripple Down Rule (RDR) represents a rule-based, explicable, and morally sound method of diagnosing mental illness. It fills the gaps in current screening tests with a sensitive and comprehensive approach to understanding and diagnosing various mental health conditions. It archives this by integrating the expert-created rules, allowing for customization flexibility, and carefully preparing the data [22](Rahman et al., 2023).

The article "Advances in artificial intelligence for diagnosing Alzheimer's disease through speech" (Abid et al., 2024) [25] discusses a brain disorder called Alzheimer's disease. It usually affects older people and in the United States, this disease is the seventh leading cause of death. More than 6 million people in the United States suffer from Dementia caused by Alzheimer's disease. In its early phases, people face trouble finding words or repeating a sentence multiple times. Later on, it hampers language skills and understanding. Patients face difficulties in communicating with people because of memory loss, and problems with attention and planning. In this article, the researcher mentioned some models that are used to diagnose Alzheimer's disease at an earlier stage by analyzing speech patterns. The first model is the GPT-3 Language Model. This model converts the speech into embeddings, which means one kind of data representation. After that, these embeddings are fed into Support Vector Classifier (SVC) to diagnose AD patients. The accuracy of this model is 80.28% which makes this model more efficient than traditional machine learning models. The researchers make a comparison between Data2Vec and Wav2Vec2 models to determine which model performs better. Internal (ADReSS) and external (DementiaBank Pitt) datasets are used to compare the performance between these two models.The Data2Vec model achieves an AUC (Area Under the Curve) of 0.835 on the external dataset, and 0.846 on the internal dataset which makes the model better than Wav2Vec2 for detecting dementia. Using BERT and ERNIE models, texts that are converted from speech are fine-tuned. These models also analyze the pauses and the gaps in a sentence. The ERNIE model provides an 89.6% accuracy. These models show promising results in diagnosing Alzheimer's disease (Abid et al., 2024).

The paper Robust Cuckoo Search Enabled Fuzzy Neuro Symbolic Reasoning-Based Alzheimer's Disease Prediction at their Earlier Stage (Dhanusha et al., 2023)[18] focuses on a patient who has Alzheimer's disease. It develops a model which is fuzzy neuro-symbolic AI that can detect early Alzheimer's disease. Here they used an algorithm named robust cuckoo search (RCS) which is much optimized and the algorithm has brood parasitism which belongs to cuckoo species. Additionally, this helps to optimize the parameter of the system which helps to increase the performance of the model prediction. Moreover, Fuzzy Neuro Symbolic AI is a hybrid model that provides uncertainty and interpretable results. This hybrid approach can capture complex patterns and relationships in clinical data that are indicative of the early stages of AD. The model would require a dataset with features relevant to AD, such as genetic information, cognitive test results, and possibly neuroimaging data. So, the model they provide gives the perfect detection in the early stage of Alzheimer's Disease compared to the existing model. Important features are selected using techniques like the CS algorithm to improve the model's performance and reduce computational complexity. The traditional Machine learning model cannot predict accurately like this proposed model. The result of model accuracy is much ahead of the existing traditional machine learning model. Finally, their proposed model prediction is much easier to understand and can be actionable in medical centers. So, they used the powerful neuro-symbolic AI for the detection of Alzheimer and they succeeded in predicting the patient who has Alzheimer's disease. This paper would represent a significant advancement in the field of computational neuroscience and medical diagnosis, combining state-of-the-art techniques to tackle a challenging and impactful problem.

Neuroimaging and Machine Learning for Dementia Diagnosis (Ahmed et al., 2019) [5] provides a comprehensive overview of recent advancements and future prospects in utilizing neuroimaging and machine learning techniques for the diagnosis of dementia. The paper focuses on the immense increase of dementia patients in this modern world and recent machine learning techniques, particularly deep learning approaches, for the early detection of dementia. This paper aims to cover a broad range of imaging and machine learning or deep learning technologies for dementia diagnostics. Emphasis is placed on the importance of early detection and prediction of dementia to provide timely treatment and support for patients. Discusses recent neuroimaging modalities such as MRI, PET, and SPECT for dementia diagnosis. With the comparison of SVM (67.01%), Logistics regression(67.04%), and CNN with softmax layer (61%), the proposed method, combined convolutional neural networks, and logistic regression showed better performance (74.93%) by training them on MRI scans of the AD. Highlights the use of imaging technologies for feature extraction and classification methods in diagnosing various forms of dementia. Summarizes the performance of machine learning and deep learning approaches in diagnosing dementia, showcasing the superiority of deep learning methods in analyzing brain images. Indicates the potential of deep learning techniques to outperform traditional machine learning methods in providing accurate and early dementia diagnosis. Identification of challenges and future prospects in dementia research, particularly in improving classification and enabling early detection for early treatment. To sum up, the field is rapidly progressing towards accurate and early diagnosis of dementia through advanced imaging and machine learning techniques.

"Hybrid CNN-LSTM model with efficient hyperparameter tuning for prediction of Parkinson's disease" [21] This paper proposed a CNN-LSTM hybrid model that is designed to analyze and classify data. At first for preprocessing the data, the irrelevant and unnecessary information is removed so that the model can give importance to useful data. After that, the audio data is converted into a visual format called Mel-spectrogram. Then, a pre-trained CNN model ResNet-50 is used. There are several layers in CNN. The first one is the Convolutional Layers that apply filters to extract important features. Each filter is moving around an image and creating a new image which is called feature map. Then, pooling layers are used to simplify the data and lastly, fully connected layers take the data and convert it into a format which can be used for making predictions. The last layer of CNN is the output layer where a mathematical function called softmax is used. This function is used to convert the model result into probabilities. After the feature extraction is done by CNN, the features are passed through the LSTM model.The LSTM layer contains 128 nodes which helps to understand the sequential patterns of the data. The output of the LSTM passed to a fully connected layer . Finally, a softmax matrix is used by the output layer. This matrix helps to classify the input. To get the optimized result, Regularization and dropout probability are tested. This proposed hybrid model provides 93.51% accuracy which is better than other traditional models like CART and SVM & XGBoost.The accuracy of Neural Network, CART, SVM, and XGBoost models is 72.69%, 84.21%, 73.51%, and 90.81% (Lilhore et al., 2023).

The paper "Neuro-Symbolic Speech Understanding in Aircraft Maintenance Metaverse" (Siyaev et al., 2021) [12] suggests a context-oriented speech communication system that is included in aircraft maintenance training and the study of Boeing 737. Neuro-symbolic speech executor (NSSE) is the incorporation of neural network and symbolic reasoning which will processed in the aircraft maintenance metaverse. NSSE is efficient in understanding complex spoken language. NSSE has four steps to do inference which are audio recording, speech-to-text, text to programs, and symbolic programs executor respectively. Firstly, a user needs to wear hololens or smart glass and speak the required command to trigger. After triggered, dynamic length audio recording (DLAR) will handle the audio creation dynamically. It shows the efficiency compared to static approaches and gives a 44.05% improvement. After that, the speech will be extracted from the audio and transcribed into text by using the wav2vec2.0 framework. Furthermore, the transcript will be processed and converted into machine-understandable programs by the deep learning sequence-to-sequence model. Finally, the symbolic programs executor executes the programs and provides audio and visual feedback to the user. The authors experiment with the conventional deep learning approach and the neuro-symbolic AI approach. Before NSSE was applied, the approach misunderstood the complex speech requests and was not efficient for the lack of a context-aware system. After NSSE is applied, the approach achieves 94.7% accuracy for both native and non-native speakers. So, NSSE is an effective model for enhancing speech understanding in context-oriented mechanisms in multiple languages. Therefore, the metaverse gives an affordable alternative for aviation colleges to practice without any physical machines and casualties. The goal of the research paper is to develop a system to control the diverse components in the proposed aircraft maintenance metaverse.

The paper "Interpretable Hierarchical Deep Learning Model for Noninvasive Alzheimer's Disease Diagnosis" [24] researched by Maryam Zokaeinikoo, Pooyan Kazemian, and Prasenjit Mitra represents an innovative approach to Alzheimer's disease diagnosis using noninvasive methods by analyzing the language patterns in the recorded audio of people. This model features a three level hierarchical attention mechanism over self-attention that enhances the model ability to detect key linguistic markers of Alzheimer's from interview recorded audio. Moreover this model good enough accurate to prediction which achieved 96% accuracy rate.They take the Dataset from DementiaBank which is largest healthcare dataset bank.In this model they created the model like that the model can understand the difference between words,sentences and the overall audio interview transcript.This model can identify the linguistic feature which helps to identify the early stage of Alzheimer.Additionally,the model model is not only highly accurate for achieving a 96% accuracy but also excels in interpretability which is a critical factor in clinical applications. With the help of providing clear insights into which aspects of speech most influence the diagnosis, the model builds trust and facilitates the practical use of AI in medical diagnostics.Furthermore,the use of natural speech analysis provides a high amount of accuracy over traditional machine learning model.This model helps in the early detection and management of Alzheimer's disease because of it is more patient-friendly and effective way to identify the disease in its early stagesSo, using natural speech analysis the researchers can gather important information with minimal inconvenience to the patient. This approach leads to better outcomes and helps to get desired outcomes .(Zokaeinikoo et al., 2023).

The paper "On the relevance of logic for AI: misunderstandings in social media, and the promise of neuro-symbolic learning" (Belle, 2023) [17]focuses on the advantages of neural networks and symbolic logic combined in neuro-symbolic AI to produce more complex, adaptable, and understandable AI systems in the context of preventing misunderstanding of social media. Neural networks are excellent at learning from data and identifying patterns, while symbolic logic offers structured, rule-based reasoning, making it useful for jobs needing clarity and precision. Neuro-symbolic AI combines two methods to provide computers the ability to use both adaptive learning and extensive, rule-based knowledge. This leads to improved decision-making processes that combine learned patterns and logical reasoning. To meet the needs of various applications, this integration enables a range of techniques, including applying logical reasoning, integrating logical constraints into neural network training, and extracting rules from learned networks. Gaining a better comprehension of intricate ideas, making models easier to grasp, and managing challenging assignments like multi-agent planning and reasoning are some of the advantages. Furthermore, by combining logical reasoning with learning mechanisms, neuro-symbolic AI makes it easier to use expressive representations that accurately capture domain-specific information, creating more flexible and resilient AI systems that are capable of making complex decisions that are relevant to their context. In the end, this method improves AI capabilities, increasing systems' dependability and credibility and expanding their range of applications.

The paper talks about a Stacked Deep Sense Neural Network model in order to detect Alzheimer's Disease (AD) [14] through using the audio transcripts. All of the transcripts were obtained from the DementiaBank dataset which captures language patterns from the patient's description of a picture. With analyzing these patterns, the paper tries to mark the patients as AD +ve or AD -ve with greater accuracy. In order to conduct this study there are a total three models being developed (Khan, 2022). First of all, the CNN model which uses convolutional and max pooling layers in order to find the features is also followed by the layers which are dense for classification. Secondly, there is CNN+ Bidirectional LSTM which actually adds the CNN's extraction of the features more worthy with Bidirectional LSTM in order to get the dependencies working in sequence for the speech pattern. Finally, the third model was SDDNN which gets the output from CNN , CNN + Bidirectional LSTM and BIdirectional LSTM for attention layers. This architecture uses the layers which are dense to create the system with high accuracy and robustness for the binary classification. There have been used two types of embeddings of word which are randomly initialized and pre-trained GloVe. Moreover, GridSearch for hyperparameter optimization is being used by the researchers, for the parameters like dropout values, learning rates and batch sizes in order to enhance the performance. During the performance the SDDNN model really did better than the other models with an accuracy of 93.31% with GloVe. Whereas the CNN and CNN + Bidirectional LSTM models got 85.05% and 84.89%. It also outperformed in all the other metrics, recall etc. Because of using the attention mechanism they got help for prioritizing critical features and getting a better classification ability. The study shows the importance of deep learning models which are specifically stack architectures which use attention mechanisms for early detection of AD. The writers gave a suggestion about the future work that the SDDNN needs to be expanded for multi class classification in order to find out the several stages of dementia. All these advancements cam give the Doctors with greater medical accessories for testing the diseases (Khan, 2022).

This research [10] analyzes the use of fastText and CNN models to recognize Alzheimer's Disease (AD) from the transcriptions of spontaneous speech. This research explores if n-gram based linguistic information, which can identify speech and language impairments typical in AD, can be useful for detecting AD. The paper uses the ADReSS dataset which consist of audio recordings and transcripts of a “cookie-theft” picture description task performed by individuals with and without AD. From two datasets PAR contains only participant’s utterances and PAR+INV contain both participant and investigator utterances. fastText model is known for its efficiency and ability to capture local word ordering, uses bigrams and trigrams as features, adding them to the input sentences. CNN model is adapted from computer vision which identifies relevant n-grams (bigrams to 5-grams) in the text, potentially signifying AD related language impairments. Both models use pre-trained 100 dimensional Glove word vectors for word embeddings. Bagging is an ensemble technique which is used to combine predictions from 21 models to enhance stability and accuracy. Performance is evaluated using accuracy for AD classification and RMSE for MMSE score prediction. The fastText model with bigrams and trigrams got the highest classification accuracy which is 83.33%. For MMSE score prediction, fastText with bigrams performed best which got an RMSE of 4.28. The idea that n-gram-based features are useful for AD detection is supported by these results. In the classification task, CNN is outperformed by fastText models. Performance is often improved by including investigator utterances, especially for fastText in MMSE prediction. Therefore, the study concludes that n-gram based features, particularly using fastText are promising for AD detection.

The article “Smart Solutions for Detecting, Predicting, Monitoring, and Managing Dementia in the Elderly: A Survey” (Addae et al., 2024) [26] focuses on technologies that assist in detecting predict and managing dementia. Dementia is a brain disorder causing memory loss, and speech disability and can lead to death within 10 years of starting. Different methods are used to collect and analyze data of the elder individuals. These methods include tools like smartphones and cloud platforms for data storage. ML, deep learning, and statistical models are combined for analysis purposes. In wearable technologies cases, researchers collect data from sensors and store it in a custom system. principal component analysis (PCA) is used for data simplification and deep learning classifiers are used to predict the dementia risk. Leave-one-subject-out cross-validation (LOSO) is applied to avoid overfitting means in the training and testing phase data from one individual is eliminated. Logistic regression, support vector machines (SVM), and clustering algorithms like K-means are commonly used in these wearable technologies. In non-wearable studies, data loggers are used to collect data and analyze it through different ML and deep learning models like random forests, neural networks, K-nearest neighbors, and ensemble methods. Both wearable and non- wearable devices have some pros and cons like wearable devices collect real-time data but it has a battery issue and many people forget to wear them. On the other hand, non-wearable devices can monitor people continuously but can miss data because of having limited range. So, this study highlights the combination of wearable and non-wearable technologies to overcome the challenges like small dataset and sensor problems. Emerging technologies like TinyML and blockchain play a vital role in making dementia care safer and accessible to a wide range of people (Addae et al., 2024).

The basic idea of the paper ”Diagnosis of Alzheimer Diseases in Early Step Using

SVM (Support Vector Machine)" [3] is to use Support Vector Machine (SVM) algorithms to find Alzheimer's disease (AD) early on using MRI scans. Changes in these areas show that AD is getting worse, so the authors focus on automated detection methods that use three important brain regions: the cortex, the corpus callosum, and the the hippos.The study investigates how well four distinct SVM-based diagnosis models identify early on Alzheimer's using MRI data. Using the frontal portion and concentrating on the hippocampus area, Model 1 (M1) is 73.33% accurate with 55 true positives. Model 2 (M2) just looks at the sagittal portion (Corpus Callosum), which is just 72% accurate and features 54 true hits. With just 49.33% accuracy and 37 true positives, the axial area (Cortex) under Model 3 (M3) had decreased efficacy. With 90.66% and 68 true positives, Model 4 (M4) comprising the cortex, corpus callosum, and hippocampus areas has the best accuracy. This all-encompassing approach shows that looking at more than one section of the brain at the same time makes diagnosing Alzheimer's disease more accurate than the other models, hence it performs far better.Model 4 did the best, showing that combining views of different parts of the brain makes early Alzheimer's disease detection much more accurate. Instead of looking at parts separately, the article shows that putting together a number of diagnostic traits and places leads to a more accurate assessment. They are going to make their program better in the future and make a longitudinal tracking system so that they can better see how Alzheimer's disease changes over time. This paper stresses how important it is to find Alzheimer's disease early by using cutting-edge machine learning techniques to help with quick and accurate diagnosis.

According to Lee et al. (2023)[20], Comprehensive Integration of Hyperdimensional Computing with Deep Learning towards Neuro-Symbolic AI paper talks about an important topic intro ducing NSHD which refers to a neuro-symbolic method using convolutional neural networks often known as CNN and hyperdimensional computing address as the barrier or limitation of current HD computing which is extracting high data from images. So, the main reason why they used HD computing is for the efficiency of this hyperdimensional computing and to enable the high ability of CNN for extracting large and compound data. Additionally in NSHD, they handle three big problems and those are Human-Interpretable Reasoning, Leveraging Knowledge Beyond Fea ture Extraction. Efficient HD Learning with Feature Reduction. They stated some benefits of using this technology as it has more accuracy than CNN and It needs less memory and space which is a save of resource. Moreover, it can occur certain predictions which help the developer from deep learning. Furthermore, It can han dle large amounts of data set and efficiently predict that dataset. Lastly, NSHD is an efficient AI model combination of CNN and HD computing that helps to predict more accurately in image-based tasks and can work on real-based applications.The paper works on the motivational speech to detect the mental disorder of individual.Mental healthe became one of the major problem in our life.during the conversation deep neural network classifier proposed an accuracy of 89% in the detection of mental health disorders. So,they state that the model they used can perfectly detect mental disorder of human beings.

This paper (Jaimini et al., 2024) [30]presented by incorporating explicit causal representations into already-existing domain knowledge, Causal Neuro-Symbolic AI improves on standard Neuro-Symbolic AI and allows for a deeper comprehension of cause-and-effect linkages within data. Because of this integration, AI systems are better able to carry out difficult causal tasks like causal inference and discovery. These systems can overcome biases and gaps in data by employing causal knowledge, using knowledge graph information to infer missing links, and improving cause discovery. Additionally, this method makes it possible to use heterogeneous data from a variety of sources effectively, which produces results that are stronger. Notably, knowing causality enhances decision-making since AI systems are better able to pinpoint the reasons behind results, which is essential for choosing the right treatments. AI systems are able to capture and analyze causal links more efficiently with the use of mechanisms such as explicit causal representation, causal inference, and integration of domain knowledge. Furthermore, these systems' causal discovery capabilities enhance our understanding of underlying causes by enabling the identification of hitherto undiscovered causal relationships. When causal knowledge is included, models become more reliable and transparent, offering clear justifications for judgments and predictions. This improves interpretability and reliability. All things considered, Causal Neuro-Symbolic AI enhances AI capabilities by offering a thorough comprehension of causation, enhancing decision-making procedures, and guaranteeing more dependable and relevant results across a range of areas.

The study "Predicting Dementia from Spontaneous Speech Using Large Language Models" [13]by Felix Agbavor and Hualou Liang looks into how the GPT-3 language model can be used to analyze natural speech in order to identify Alzheimer's disease (AD). This study shows that GPT-3 text embeddings are better at telling the difference between people with AD and healthy controls and predicting cognitive scores just by looking at speech data.(Agbavor & Liang, 2022). The performance of text embeddings generated from GPT-3 and conventional acoustic characteristics was compared by the authors in their studies. According to the results, GPT-3 embeddings performed noticeably better than acoustic features in terms of accuracy, precision, recall, and F1 score. The Babbage model regularly outperformed the Ada model. In particular, GPT-3 embeddings by themselves outperformed acoustic characteristics and even competed with refined GPT-3 models in AD classification tasks, achieving an accuracy of up to 80.28%. The study also evaluated how well GPT-3 embeddings and acoustic features worked together, but the results showed only a slight improvement in performance. This suggests that although GPT-3 embeddings are very effective, the additional acoustic features do not significantly improve classification accuracy. GPT-3 embeddings outperformed auditory characteristics in predicting Mini-Mental State Examination (MMSE) results, demonstrating their capacity to catch subtle verbal patterns suggestive of cognitive impairment. Overall, this study shows that GPT-3 embeddings could be a non-invasive, low-cost way to diagnose early-stage AD that is better than current methods. It also shows that these methods need more research to be improved and proven to work on bigger datasets(Agbavor & Liang, 2022). The ability of three language models—ChatGPT-3.5, ChatGPT-4, and Bard—to diagnose Alzheimer's disease (AD) using spontaneous speech is assessed in the paper "Performance Assessment of ChatGPT versus Bard in Detecting Alzheimer's Dementia" by Balamurali B.T. and Jer-Ming Chen. The models answered questions about speech transcripts using a zero-shot learning technique, and their accuracy, sensitivity, specificity, precision, and F1 score were assessed.(Balamurali & Chen, 2024).

This research [9]explored two common approaches which are a featured-based model utilizing domain knowledge from speech and a pre-trained Bidirectional Encoder Representation from Transformer (BERT) model employing transfer learning to detect Alzheimer's Disease (AD). Both models used ADReSS dataset, a dataset selected for its balanced representation of age and gender in both AD and non-AD participants. The feature-based model involved extracting 509 linguistic and acoustic features from speech transcripts and audio files. This model is categorized into three types which are Lexico-syntactic features, Acoustic and temporal features and Semantic features. These features were classified using machine learning models including Support vector machine (SVM), neural network, random forest and naive Bayes. The transfer learning with BERT used pre-trained BERT models, specifically the bert base uncased variant, to fine-tune a text classification model using speech transcripts and transcripts were preprocessed to explicitly mark utterance boundaries, considered vital for AD detection. The BERT model got the highest accuracy of 81.8% using 10-fold cross validation on the training set. On the unseen test set, BERT got 83.3% accuracy where the SVM's achieved 81.3% accuracy. Therefore, this research demonstrated that both feature-based and pre-trained language models can effectively detect AD from speech.

The research by Pan et al "Automatic Hierarchical Attention Neural Network" for Detecting AD" [6] tells us that an advanced method in order to find AD by speech generated from picture description. A research "Bidirectional Hierarchical Attention Neural Network " has been developed in order to capture linguistic patterns. By implementing Automatic speech recognition ASR systems and punctuation , transcripts are being transformed by the model into formats which we can analyze for classification, giving good performance through datasets (Pan et al, 2019). The model BHANN used "Bidirectional LSTM" layers in order to understand word level features and "Bidirectional GRU" to understand the sentence level features. For both the levels of the component attention mechanism were applied, like specific words, thematic coherence important for detecting early symptoms of AD. For converting words into vector presentation GloVe embeddings are used which are pre-trained. For BHANN model efficiency validation a couple baseline models were being introduced : a "Bidirectional LSTM" treating the transcripts in single sequence of words and : "Hierarchical Bidirectional Recurrent Neural Network (HBRNN) which removes attention mechanisms (Pan et al, 2019). The DementiaBank dataset used by the models were being tested by the studies. With an IVA dataset of 33 transcripts, showing the result BHANN has performed the baselines with an F score of "73.45%" & "75.68%" which shows the effectiveness of hierarchical structure and attention mechanism. Though there are transcript errors in ASR, BHANN but they tried to keep stable performance which provided robustness. More importantly, IVA data put significance on the model's ability in order to get acquainted with the training sets (Pan et al, 2019). The study tells about an important advancement in the sector Automated dementia diagnosis through the passing of the basic methods in accuracy and robustness. At the time of the current task focusing on the binary classification , future work thought extending the research to multi-class classification, having connection with additional diagnostic classes like "Mild Cognitive Impairment" (MCI). Through addressing the challenges, the model which is proposed certainly has the ability to make a revolution in dementia screening in the medical sites.

Chapter 3

Methodology

The architecture of the system starts with the collection of audio data of dementia and non-dementia people. After done with the pre-processing, the pre-processed data send to two parts of the proposed model. The first part is neural AI that extract the features of the dataset through spectrograms. The spectrogram images handle and feed into the hybrid neural AI model. In contrast, symbolic AI or rule-based system extracts the numerical features and process the data. With the boxplot representation, it generated some rules in terms of confident and uncertain cases. Uncertain cases also feed into the traditional machine learning model and make rule-based system altogether. After adding both parts, we build Neuro-Symbolic AI as our proposed model to level up the accuracy than the traditional machine learning model. In this Neuro-Symbolic Model, the neural AI model integrates with the symbolic model to detect the mental disorder efficiently and it will ignite the opportunity of research. The system then compares the results between conventional machine learning models and the Neuro-Symbolic AI Model (proposed model).

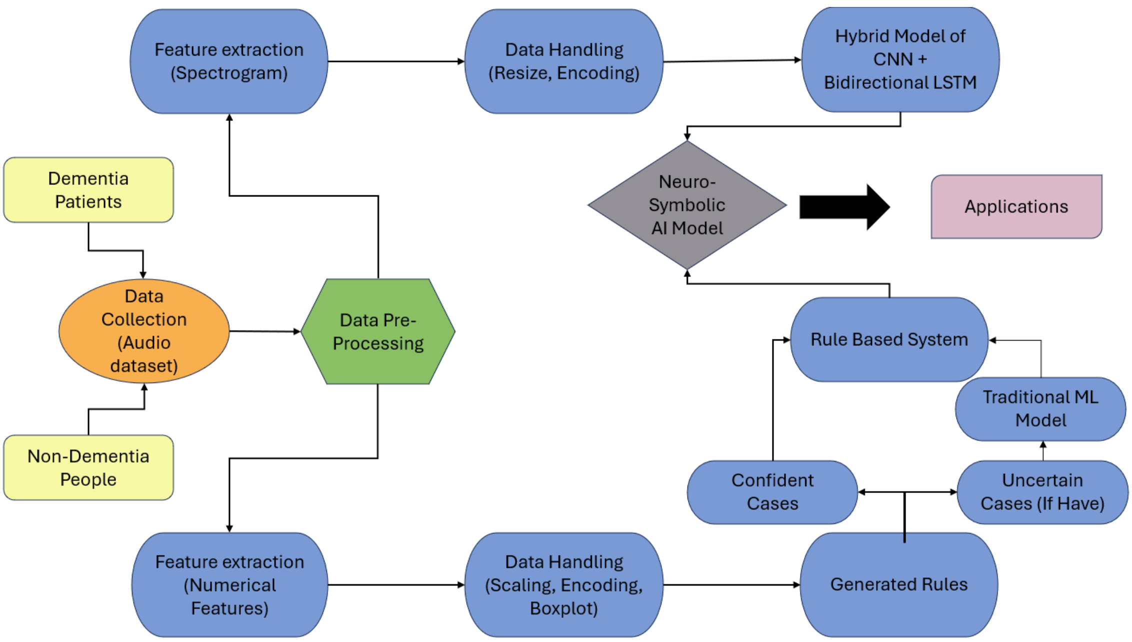


Figure 3.1: Overview of the proposed Neuro-Symbolic AI Model

3.1 Dataset

3.1.1 Dataset Description

TalkBank is a project which consists of fundamental research in the study of human communication. We requested Professor Brian MacWhinney to grant access to the project by ensuring all of the ground rules and proper citations will be followed. We contacted the organization through Dr. MD. Golam Rabiul Alam who is designated as the Professor of Brac University. We collected the audio datasets from the part of the DementiaBank of TalkBank project [1](Becker et al.,1994). Research is conducted where a group of people are asked to read a description of an image. The audio dataset consists of Alzheimer's and healthy people. The research results- 306 people diagnosed with Alzheimer's and 244 healthy people. These audio data are stored in DementiaBank which is a medical domain task. The duration of each audio file is around one to three minutes. This dataset is valuable for research-related works on the classification between dementia and non-dementia people or other parts of the research area. We express our gratitude towards the TalkBank organizer and collaborators for their incredible work in providing datasets.

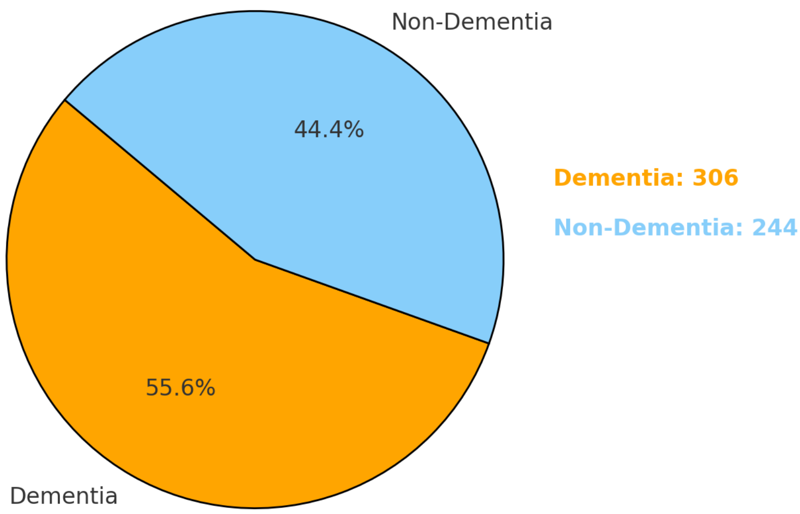


Figure 3.2: Dataset between Dementia and Non-Dementia

3.1.2 Dataset Pre-Processing

The collected audio data must be preprocessed before implementing to the models. This pre-processing includes several operations of cleaning and preparing the data. Firstly, we download the data from the organization from the folders of Dementia and Control. After that, we rename the file names to make consistent. The audio file format is converted from MP3 to WAV. WAV files are uncompressed and offer a higher audio quality compared to MP3. The audio files which duration are less than 5 seconds that are removed. Furthermore, the silence from the beginning and end of the audio data is being trimmed. Additionally, the portion of the audio data is removed when a part of the audio is 30 dB quieter in the audio file, as it is considered background noise.

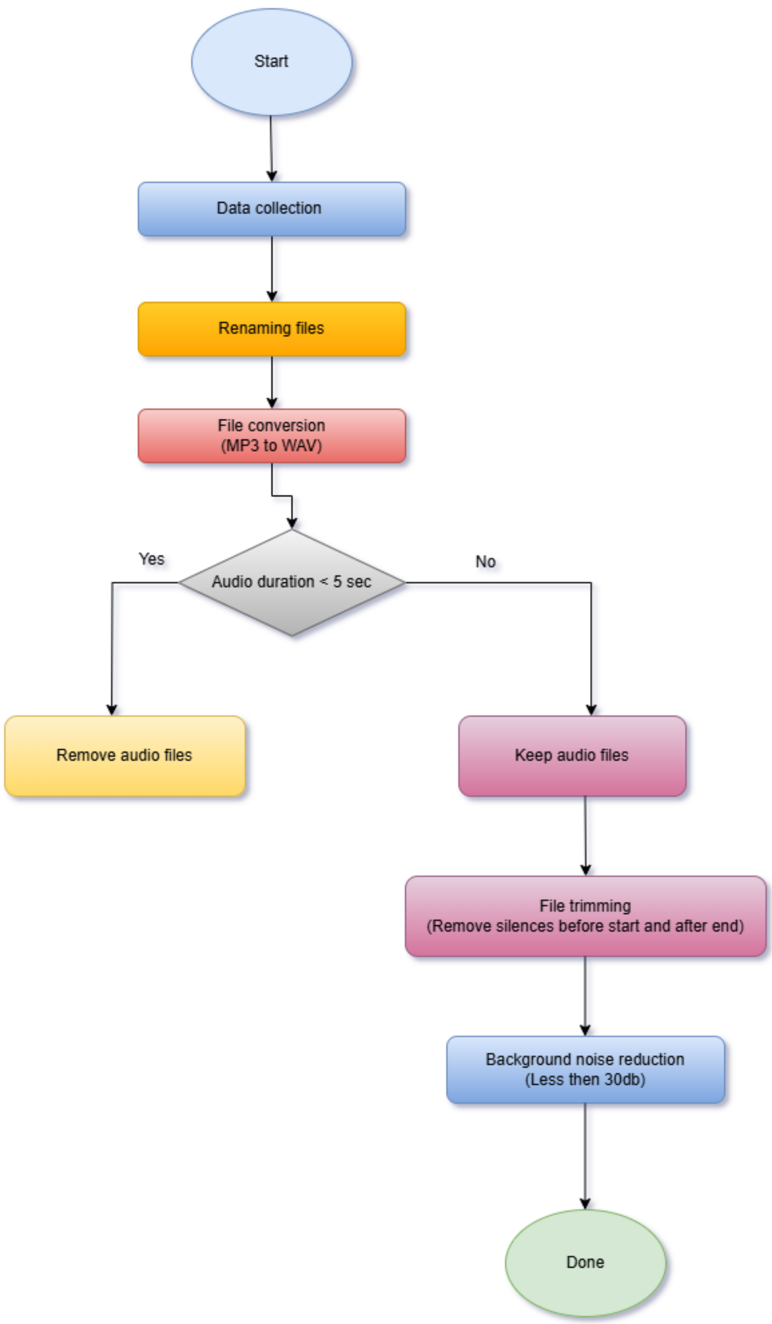


Figure 3.3: Data pre-processing

3.1.3 Feature Extraction of Neural AI model

We converted the audio frequency domain over time and the frequencies converted into mel scale which makes the audio more humanization. Also, the amplitude is converted to decibels to visualize differences in intensity. After that, the spectrograms resize into (224, 224) shape. The pixel values are normalized between 0 and 1 and the labels are encoded. Lastly, we split the data into 80% train data and 20% test data for small dataset. The processed data feed into the CNN-LSTM model. Here are the given some spectrogram images-

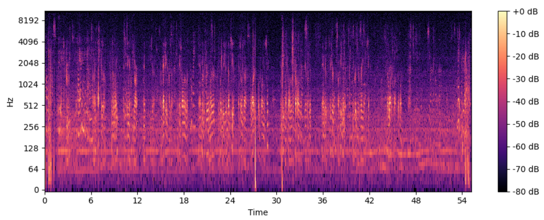


Figure 3.4: Dementia Sample 1

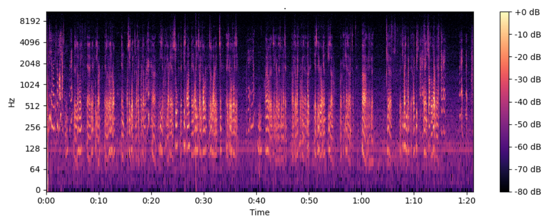


Figure 3.5: Dementia Sample 2

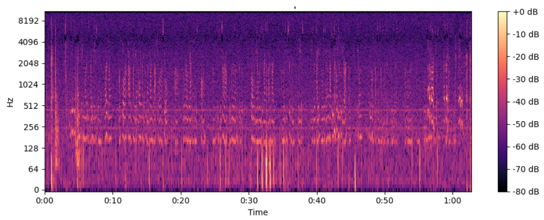


Figure 3.6: Non-Dementia Sample 1

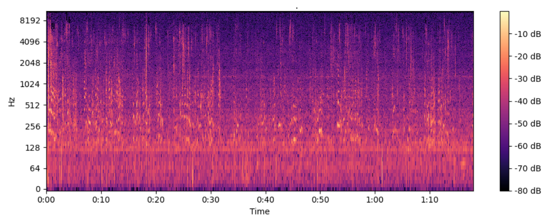


Figure 3.7: Non-Dementia Sample 2

3.1.4 Feature Extraction of Symbolic AI model

Praat (software package) and librosa (python package) are used to analyze different aspects of the sound signals [4]. These carry many features relevant to voice data. Using the Parselmouth (Praat) library, the pitch features like mean and standard deviation of the pitch and intensity features, including mean and maximum are extracted. Librosa extracts features from audio data like pause count, MFCC, Delta-MFCC, Delta-Delta MFCC, zcr, spectral_flatness, spectral_centroid, spectral_bandwidth, and so on. Pause count is measured by detecting the silence portion in the voice data.This is the most significant feature for distinguishing dementia and healthy people. If the silence duration is more than 1 seconds, it is counted as a pause. After that, standard scaling data preprocessing technique is used to ensure the features are properly scaled features of the data and contribute equally during model training. Label encoding converts dementia to 1 and non-dementia to 0.

Data visualization:
We used Boxplot representation to define the states of data in terms of different features. The analysis of Boxplot can be done with Five Number of Summary which is crucial for numerical features (Galarnyk, 2024) [28]. Initially, we figure out the five-number summary and demonstrate the boxplot for each feature of the audio data. This is particularly useful for visualizing the distribution of the data and detecting the skewness, patterns, and possible outliers of the data. The five-number summary includes minimum, first quartile(Q_1), median, third quartile(Q_3),

and maximum. This information helps to compare dementia and non-dementia data for each feature side by side. For each feature boxplot, the x-axis represents the two categories (dementia and non-dementia) being compared, and the y-axis defines the values of the extracted feature from the speech dataset. Here is picture of boxplot and boxplots for various features-

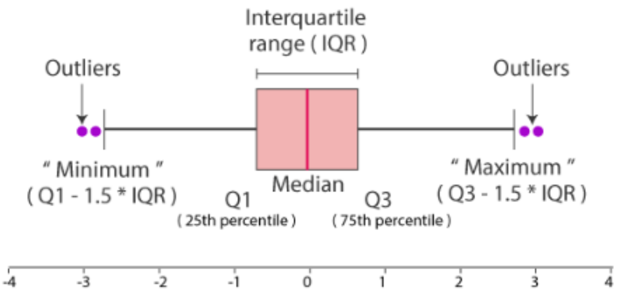


Figure 3.8: Boxplot

(i) Pause Count: The pause count represents the number of pauses are taken during a conversation of Dementia and Non-Dementia people. It is prioritized highly to distinguish between Dementia and Non-Dementia people.

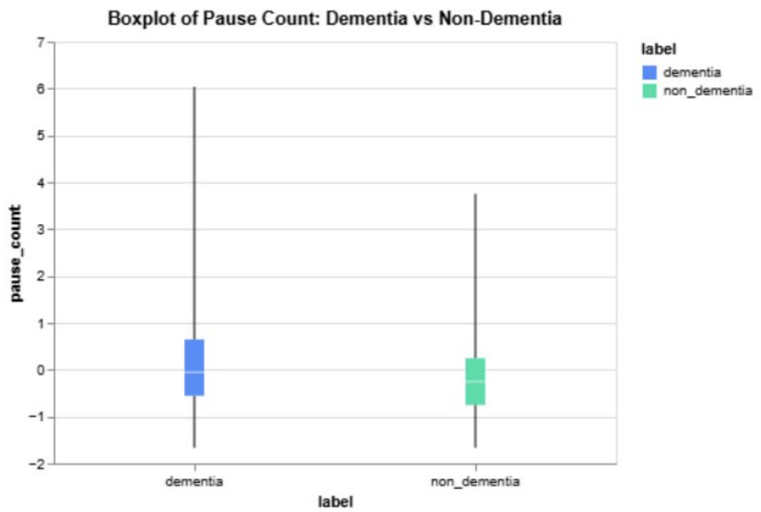


Figure 3.9: Pause Count Feature

(ii) MFCC Mean (1-13): It captures the timbral and spectral features of sound. It contains the MFCC mean 1 to MFCC mean 13 for dementia and non-dementia people. As the MFCC mean values of dementia people and MFCC mean values of non-dementia people mostly overlap, we do not take the five-number summary of MFCC mean for generating the rules.

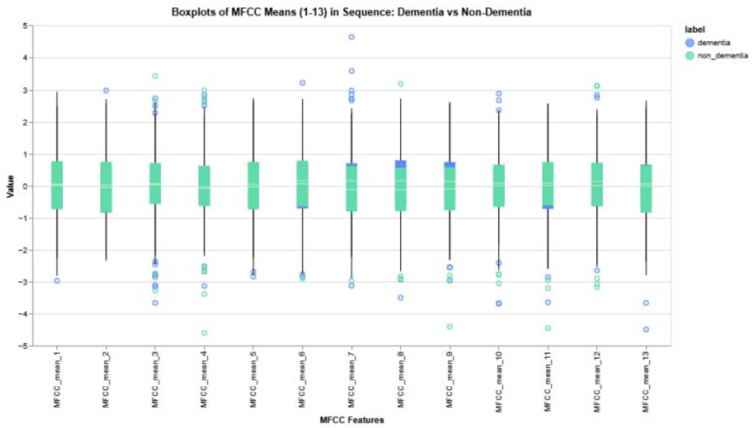


Figure 3.10: MFCC Feature

(iii) Delta MFCC Mean (1-13): It provides the rate of change of original MFCC value. It contains the Delta MFCC mean 1 to Delta MFCC mean 13 for dementia and non-dementia people. As the Delta MFCC mean values of dementia people and Delta MFCC mean values of non-dementia people significantly overlap, we do not take the five-number summary of Delta MFCC mean for generating the rules.

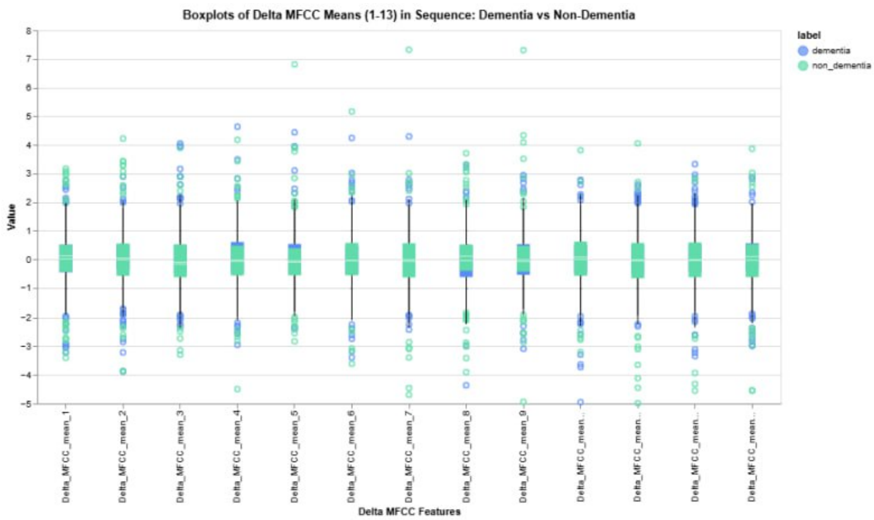


Figure 3.11: Delta MFCC Mean Feature

(iv) Delta Delta MFCC Mean (1-13): It represents the change of the Delta MFCCs. It contains the Delta Delta MFCC mean 1 to Delta Delta MFCC mean 13 for dementia and non-dementia people. As the Delta Delta MFCC mean values of dementia people and Delta Delta MFCC mean values of non-dementia people highly overlap, we do not take the five-number summary of Delta Delta MFCC means for generating the rules.

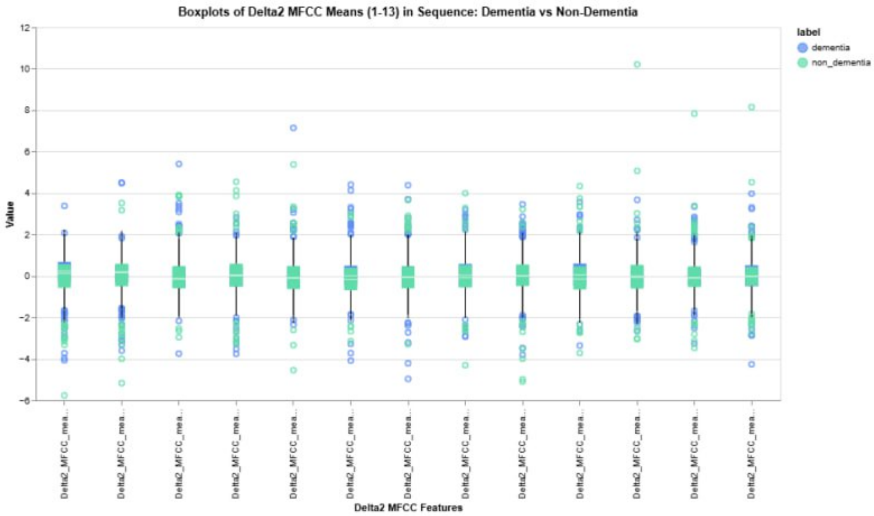


Figure 3.12: Delta Delta MFCC Mean Feature

(v) Chroma Mean (1-12): It captures harmonic properties. It contains the Chroma mean 1 to Chroma mean 12 for dementia and non-dementia people. We do not count this feature since they fall under the same range or overlap.

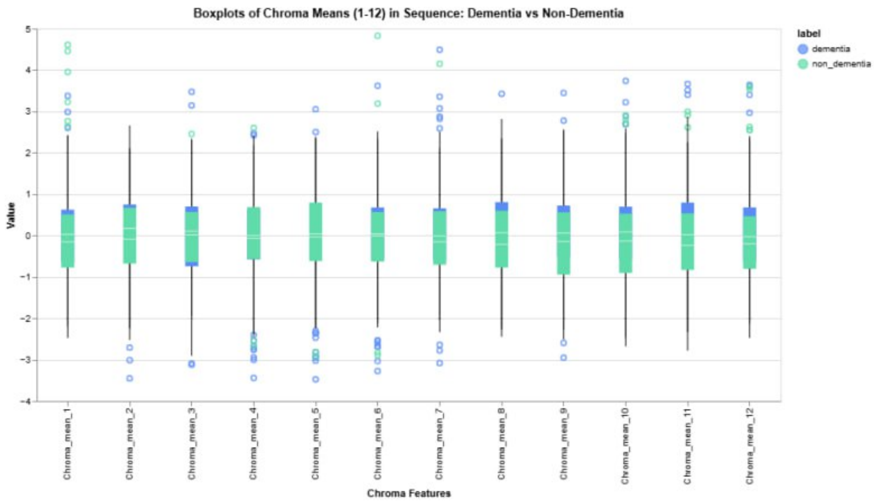


Figure 3.13: Chroma Mean Feature

(vi) ZCR: ZCR reflects the Zero Crossing rate. It detects the noiseness means the changes of audio signals.

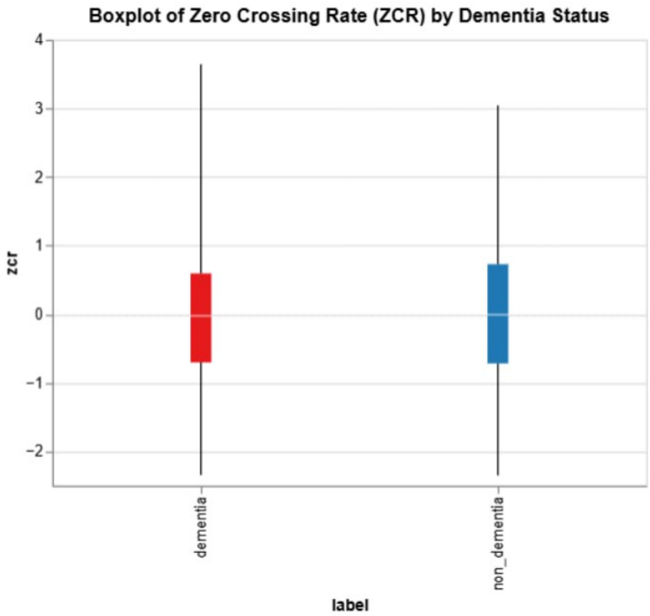


Figure 3.14: ZCR Feature

(vii) RMS: RMS represents as Root Mean Square. It measures the energy or loudness of the voice.

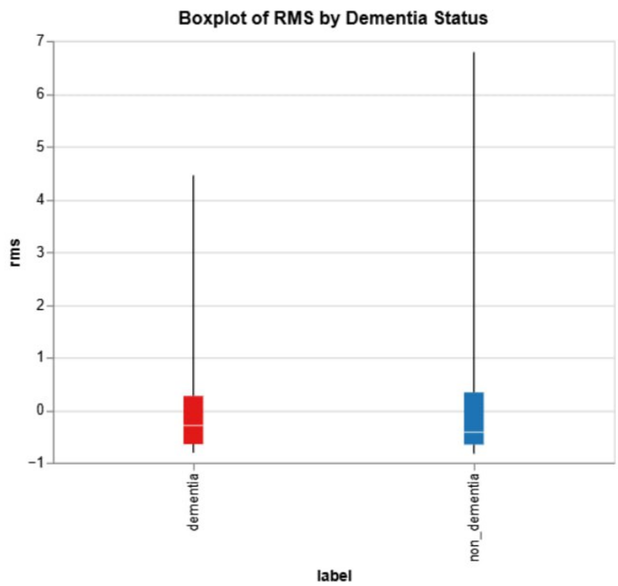


Figure 3.15: RMS Feature

(viii) Spectral Bandwidth: It indicates the ranges of frequencies.

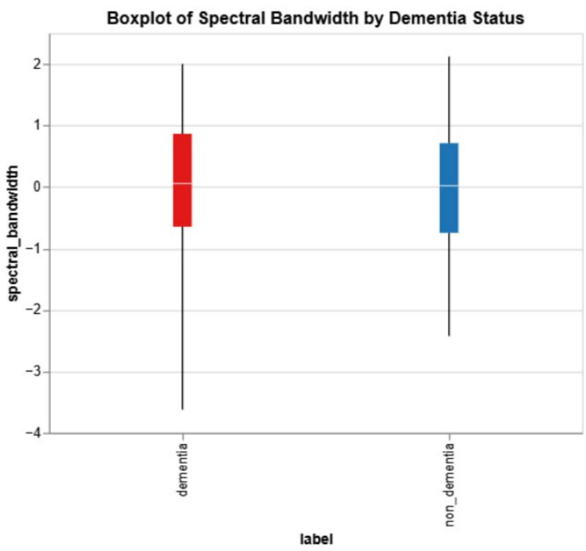


Figure 3.16: Spectral Bandwidth Feature

(ix) Spectral Centroid: It indicates where the center of mass of the spectrum is located.

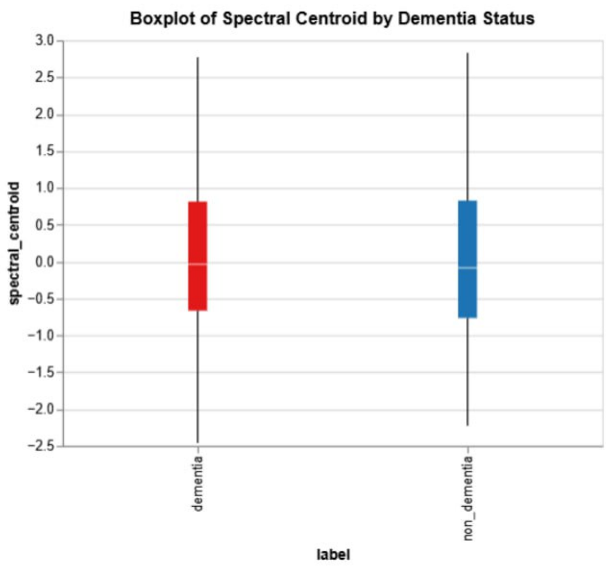


Figure 3.17: Spectral Centroid Feature

(x) Spectral Flatness: It measures how noise-like a sound is compared to being tonal.

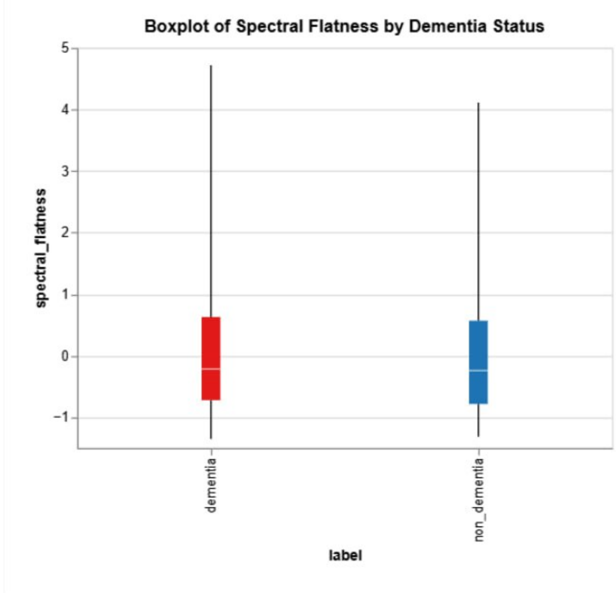


Figure 3.18: Spectral Flatness Feature

(xi) Max Intensity: It measures the maximum loudness of the audio signal.

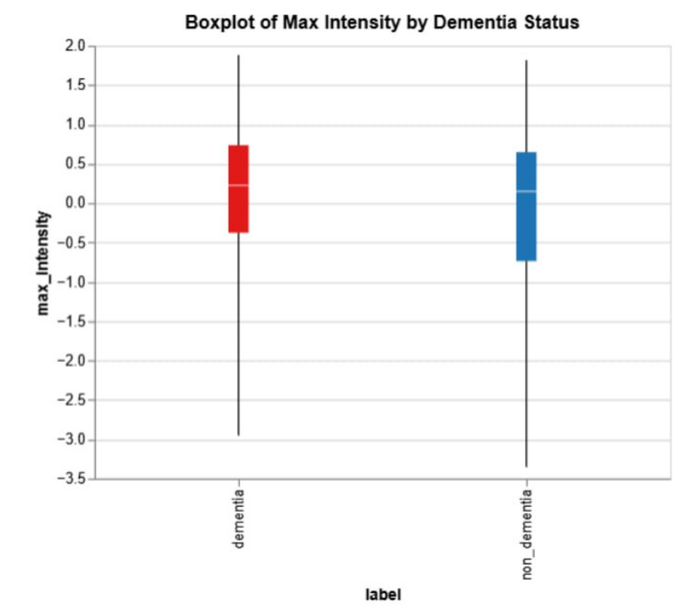


Figure 3.19: Max Intensity Feature

(xii) Mean Intensity: It records the average loudness of audio signal.

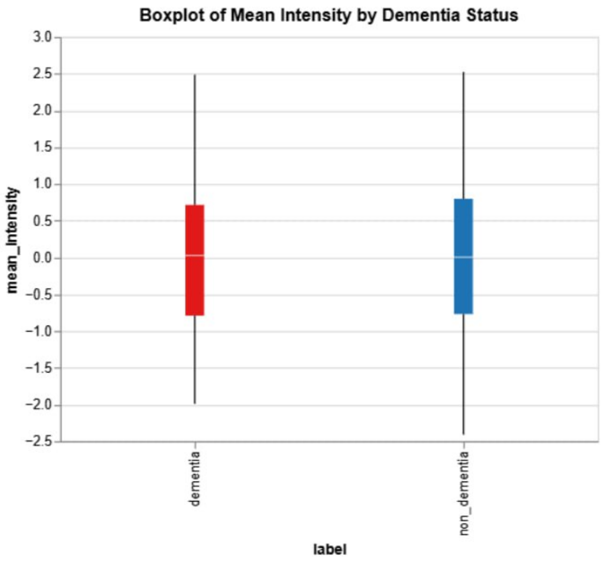


Figure 3.20: Mean Intensity Feature

(xiii) Mean Pitch: It measures the average fundamental frequency of the audio signal.

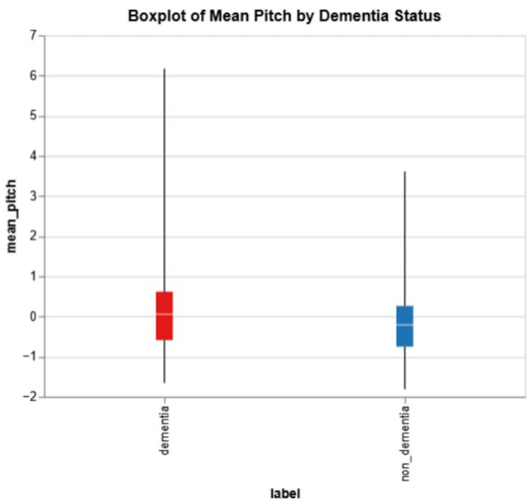


Figure 3.21: Mean Pitch Feature

(xiv) Std Pitch: It measures the variability of pitch in audio signal.

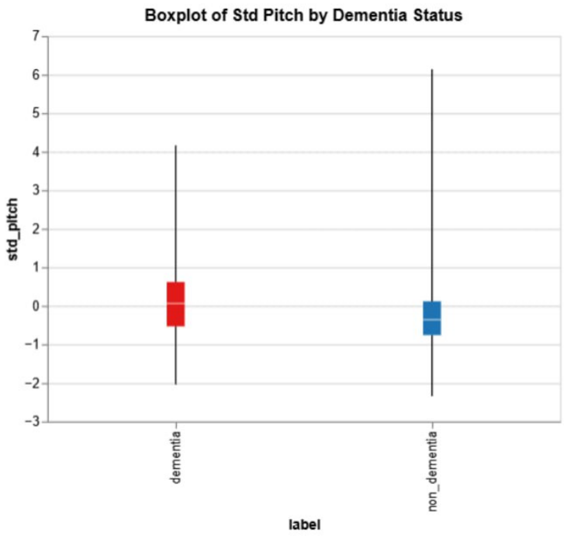


Figure 3.22: Std Pitch Feature

Features	Five-Number Summary of Boxplot	
	Dementia	**Non-Dementia**
Pause Count	Max = 6.03 Q_3 = 0.64, Median = -0.05, Q_1 = -0.55, Min = -1.65	Max = 3.74 Q_3 = 0.24, Median = -0.25, Q_1 = -0.75, Min = -1.65
Chroma Mean 1	Max = 3.37 Q_3 = 0.62, Median = 0.03, Q_1 = -0.63, Min = -2.47	Max = 4.60 Q_3 = 0.51, Median = -0.15, Q_1 = -0.77, Min = -2.18
Chroma Mean 3	Max = 3.47 Q_3 = 0.7, Median = 0.10, Q_1 = -0.74, Min = -3.12	Max = 2.45 Q_3 = 0.56, Median = 0.01, Q_1 = -0.64, Min = -1.94
Chroma Mean 6	Max = 3.61 Q_3 = 0.67, Median = 0.04, Q_1 = -0.57, Min = -3.27	Max = 4.83 Q_3 = 0.569, Median = -0.006, Q_1 = -0.62, Min = -2.89
Chroma Mean 8	Max = 3.42 Q_3 = 0.81, Median = 0.06, Q_1 = -0.62, Min = -2.44	Max = 2.33 Q_3 = 0.60, Median = -0.22, Q_1 = -0.77, Min = -2.26
Chroma Mean 9	Max = 3.44 Q_3 = 0.72, Median = 0.05, Q_1 = -0.51, Min = -2.95	Max = 2.52 Q_3 = 0.56, Median = -0.14, Q_1 = -0.94, Min = -2.51
ZCR	Max = 3.64 Q_3 = 0.59, Median = -0.024, Q_1 = -0.70, Min = -2.34	Max = 3.04 Q_3 = 0.729326, Median = -0.007, Q_1 = -0.71, Min = -2.35
Spectral Bandwidth	Max = 1.99 Q_3 = 0.86, Median = 0.055, Q_1 = -0.65, Min = -3.62	Max = 2.12 Q_3 = 0.712, Median = 0.019, Q_1 = -0.747, Min = -2.421
Spectral Centroid	Max = 2.76 Q_3 = 0.81, Median = -0.034, Q_1 = -0.66, Min = -2.46	Max = 2.829247 Q_3 = 0.83, Median = -0.084, Q_1 = -0.76, Min = -2.23
Spectral Flatness	Max = 4.72 Q_3 = 0.63, Median = -0.214, Q_1 = -0.72, Min = -1.35	Max = 4.11 Q_3 = 0.57, Median = -0.24, Q_1 = -0.78, Min = -1.31
RMS	Max = 4.44 Q_3 = 0.26, Median = -0.29, Q_1 = -0.64, Min = -0.81	Max = 6.78 Q_3 = 0.34, Median = -0.42, Q_1 = -0.65, Min = -0.82
Max Intensity	Max = 1.88 Q_3 = 0.73, Median = 0.22, Q_1 = -0.37, Min = -2.95	Max = 1.81 Q_3 = 0.65, Median = 0.149, Q_1 = -0.73, Min = -3.35
Mean Intensity	Max = 2.48 Q_3 = 0.71, Median = 0.026, Q_1 = -0.78, Min = -1.99	Max = 2.52 Q_3 = 0.79, Median = 0.002, Q_1 = -0.76, Min = -2.41
Mean Pitch	Max = 6.17 Q_3 = 0.61, Median = 0.056, Q_1 = -0.587, Min = -1.65	Max = 3.61 Q_3 = 0.261, Median = -0.211, Q_1 = -0.750, Min = -1.806
Std Pitch	Max = 4.16 Q_3 = 0.61, Median = 0.060, Q_1 = -0.533, Min = -2.030	Max = 6.13 Q_3 = 0.116, Median = -0.360, Q_1 = -0.760, Min = -2.34

Table 3.1: **Five-Number Summary of Dementia vs. Non-Dementia Label**

3.2 Model Specification

We are comparing the several models from our literature review, base line models and proposed model. The proposed model of NSAI is used as our major model of research while others are used for comparison with out model.

3.2.1 Neural AI

The first part of Neuro-symbolic AI is Neural AI which is the hybrid model with the collaboration of CNN and LSTM. To begin the execution of model, we extracted the spectrogram images from the pre-processed data with the STFT. Short-term fourier transform is the most used tool to process speech analysis or audio data. It generates the images along with the frequency over time.

CNN: The first sequential model is two-dimensional (2D) Convolutional neural network (CNN). The motive of the architecture of CNN can do the complex works which can not be done by simple neural networks. CNNs are excellent to capture the spatial patterns and local structures in data. Since the authors extracted the spectrogram from the audio datasets, CNN is highly effective to extract visual patterns in the context of time vs frequency representation. It used to detect the speech irregularities, change of frequency, spectral energy changes for the different labels of data. The CNN model consists of convolutional, max pooling layer and ReLU activation function which represent the equations as follows (Goodfellow, Bengio, & Courville, 2016)[2] -

1. Convolutional Layer Equation

$$Z_{i,j,k} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} X_{i+m,j+n} W_{m,n,k} + b_k \tag{3.1}$$

Where:

- $Z_{i,j,k}$ = Output feature map at position (i, j) for filter k.
- $X_{i+m,j+n}$ = Input spectrogram pixel values.
- $W_{m,n,k}$ = Filter/kernel weights.
- b_k = Bias term.
- M, N = Filter size (e.g., 3×3).
- i, j = Pixel positions in the feature map.

2. ReLU Activation Function

$$A_{i,j,k} = \max(0, Z_{i,j,k}) \tag{3.2}$$

3. Max Pooling Layer

$$P_{i,j,k} = \max_{\substack{m=0,\ldots,M-1 \\ n=0,\ldots,N-1}} A_{i+m,j+n,k} \tag{3.3}$$

Where:

- $P_{i,j,k}$ = Pooled feature map value.
- $A_{i+m,j+n,k}$ = Activation values from the previous layer.
- $M \times N$ = Pooling window size (e.g., 2×2).

Algorithm 1 Convolutional Neural Network (CNN) Model

Start
Input: Training data, epochs, testing data.
conv1 = Conv2D(filters, (kernel_size), activation= "relu", kernel_regularizer)
maxpool1 = MaxPooling2D(pool_size)
conv2 = Conv2D(filters, (kernel_size), activation= "relu", kernel_regularizer)
maxpool2 = MaxPooling2D(pool_size)
flatten()
End

LSTM: In contrast, long short-term memory(LSTM) is capable to learn the long-term dependencies in sequential data. Since the speech data is sequential and dementia related speeches contain long pauses, sluttering, so it can easily captures the sequence of feature maps which generated by CNN. We used bidirectional LSTM instead of regular LSTM to process the data from both forward and backward directions which provided the context of the speech sequence. The standard LSTM hallucinates the speech pattern due to background noises but the bidirectional LSTM helps to improve the speech understanding and prevents the misclassification. The LSTM model process the the extracted features (feature map) of CNN sequentially and process the data. The LSTM model consists of LSTM layer, sigmoid (binary classification) activation function, fully connected layer which represent the equations as follows (Kalita, 2022)**r34-**

1. LSTM Layer Equation

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \tag{3.4}$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \tag{3.5}$$

$$\tilde{C}_t = \tanh(W_C[h_{t-1}, x_t] + b_C) \tag{3.6}$$

$$C_t = f_t C_{t-1} + i_t \tilde{C}_t \tag{3.7}$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \tag{3.8}$$

$$h_t = o_t \tanh(C_t) \tag{3.9}$$

where:

- f_t, i_t, o_t = Forget, input, and output gates.
- C_t = Cell state (memory).
- h_t = Hidden state (output).
- W_f, W_i, W_C, W_o = Weight matrices.
- b_f, b_i, b_C, b_o = Bias terms.
- x_t = CNN-extracted feature input at time t.
- σ = Sigmoid activation function.

2. Fully Connected (Dense) Layer

$$O = W_F F + b_F \tag{3.10}$$

Where:

- O = Output vector (classification probabilities).
- W_F = Weight matrix.
- F = Flattened feature vector from LSTM.

- b_F = Bias term.

Algorithm 2 Long Short-Term Memory (LSTM) Model

1: **Start**
2: bilstm1 = Bidirectional(LSTM(filters, regularizer))
3: dropout1 = Dropout()
4: bilstm2 = Bidirectional(LSTM(filters, regularizer))
5: dropout2 = Dropout()
6: dense1 = Dense(neurons, activation= ”relu”, regularizers)
7: dropout3 = Dropout()
8: dense2 = Dense(neurons, activation= ”sigmoid”)
9: **End**

The Architecture of Hybrid Neural AI Model:

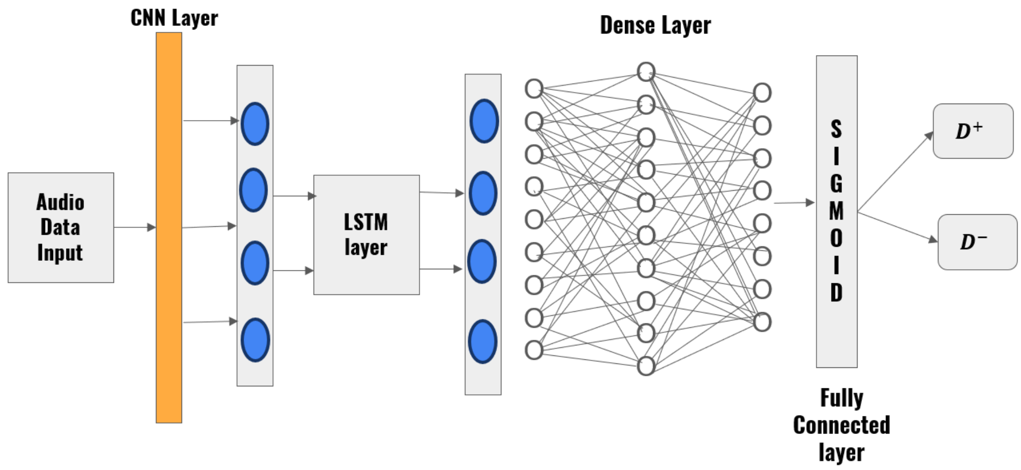


Figure 3.23: Architecture of Hybrid Neural AI Model

The Neural Model Summary Table:

Table 3.2: **Neural Model Summary**

Layer (type)	**Output Shape**	**Param #**
conv2d (Conv2D)	(None, 222, 222, 32)	320
max_pooling2d (MaxPooling2D)	(None, 111, 111, 32)	0
conv2d_1 (Conv2D)	(None, 109, 109, 64)	18,496
max_pooling2d_1 (MaxPooling2D)	(None, 54, 54, 64)	0
flatten (Flatten)	(None, 186624)	0
reshape (Reshape)	(None, 64, 2916)	0
bidirectional (Bidirectional)	(None, 64, 64)	754,944
dropout (Dropout)	(None, 64, 64)	0
bidirectional_1 (Bidirectional)	(None, 128)	66,048
dropout_1 (Dropout)	(None, 128)	0
flatten_1 (Flatten)	(None, 128)	0
dense (Dense)	(None, 128)	16,512
dropout_2 (Dropout)	(None, 128)	0
dense_1 (Dense)	(None, 1)	129

Therefore, the incorporation of CNN-LSTM model strengthens the hybrid model to predict the precise result. CNN extracts the meaningful features from spectrogram images and the feature maps are flattened into sequences and fed into LSTM. LSTM analyzed the data patterns over time and improving the accuracy of the model. We incorporate the two models by extracting feature from spectrogram images via CNN and the extracted features are being reshaped and feed to LSTM layer. Then train the model with 50 epochs to get the expected accuracy.

3.2.2 Symbolic AI

Symbolic AI is the rule-based approach which evaluates any system with the rules or logics or symbols. It make the Neural AI interpretable and describe the black box nature of machine learning models. Symbolic AI is the logical or reasoning part of our proposed model which is crucial to explain our Neural AI model. We develop a symbolic classifier where we generate the human-readable rules based on the statistical thresholds we get from the five-number summary of each feature. This symbolic classifier evaluates whether an audio sample belongs to the dementia or non-dementia class and provides an interpretable and transparent decision. Instead of learning patterns from the machine learning models, this symbolic classifier executes a rule-based approach. The particular features are selected for generating the rules as these features capture significant speech characteristics. The features are pause count, spectral flatness, spectral centroid, spectral bandwidth, zero-crossing

Rate (ZCR), intensity level (mean and max), pitch (mean and standard deviation) and Root Mean Square (RMS) energy.

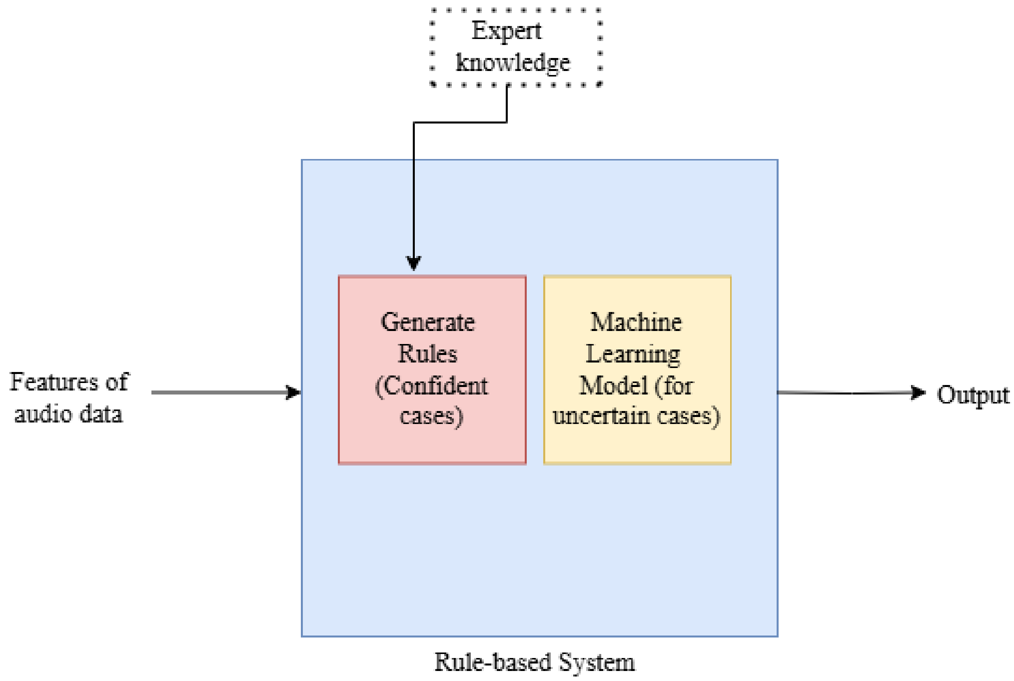


Figure 3.24: Architecture of Symbolic AI Model

Algorithm 3 Symbolic Classification

```
Start
If pause_count > Dementia_Max, Return (1, True)
Else If pause_count < Non-Dementia_Min, Return (0, True)
Else If pause_count is between Non-Dementia_Max and Dementia_Max, Return (1, True)
Else, Return (None, False)
If rms > Non-Dementia_Max, Return (0, True)
Else If rms < Dementia_Min, Return (0, True)
Else If rms is between Dementia_Max and Non-Dementia_Max, Return (0, True)
Else, Return (None, False)
If max_intensity > Dementia_Max, Return (1, True)
Else If max_intensity < Non-Dementia_Min, Return (0, True)
Else If max_intensity is between Dementia_Min and Non-Dementia_Min, Return (0, True)
Else If max_intensity is between Dementia_Max and Non-Dementia_Max, Return (1, True)
Else, Return (None, False)
If spectral_flatness > Dementia_Max, Return (1, True)
Else If spectral_flatness < Non-Dementia_Min, Return (0, True)
Else If spectral_flatness is between Non-Dementia_Max and Dementia_Max, Return (1, True)
Else, Return (None, False)
If spectral_bandwidth > Non-Dementia_Max, Return (0, True)
Else If spectral_bandwidth < Dementia_Min, Return (1, True)
Else If spectral_bandwidth is between Dementia_Max and Non-Dementia_Max, Return (0, True)
Else, Return (None, False)
If std_pitch > Non-Dementia_Max, Return (0, True)
Else If std_pitch < Non-Dementia_Min, Return (0, True)
Else If std_pitch is between Dementia_Max and Non-Dementia_Max, Return (0, True)
Else, Return (None, False)
If mean_pitch > Dementia_Max, Return (1, True)
Else If mean_pitch < Non-Dementia_Min, Return (0, True)
Else If mean_pitch is between Non-Dementia_Max and Dementia_Max, Return (1, True)
Else, Return (None, False)
If zcr > Dementia_Max, Return (1, True)
Else If zcr < Non-Dementia_Min, Return (0, True)
Else If zcr is between Non-Dementia_Max and Dementia_Max, Return (1, True)
Else, Return (None, False)
If spectral_centroid > Non-Dementia_Max, Return (0, True)
Else If spectral_centroid < Dementia_Min, Return (1, True)
Else If spectral_centroid is between Dementia_Max and Non-Dementia_Max, Return (0, True)
Else, Return (None, False)
```

```
If mean_intensity > Non-Dementia_Max, Return (0, True)
Else If mean_intensity < Dementia_Min, Return (1, True)
Else If mean_intensity is between Dementia_Max and Non-Dementia_Max, Return (0, True)
Else, Return (None, False)
Default Return Case (No Decision)
Return (None, False)
End
```

For example, if the pause count of a speech is greater than the maximum pause count observed in dementia patients, the classifier classifies the speech as dementia (1) with confidence. If the pause count is lower than the minimum pause count observed in non-dementia people, it is classified as non-dementia (0) with confidence. If the pause count falls between the max of non-dementia and dementia values, the classifier confidently classifies the speech as dementia (1). The speech sample is classified as uncertain when it falls in an overlapping range. However, the system has a drawback as the significant portion of feature values of dementia and non-dementia categories overlap. As a result, the classifier mostly detects the speech as uncertain (none). To resolve this problem, we use a traditional machine learning model (XGBoost), which analyzes the speech samples when the symbolic classifier is not confident to detect the speech samples. This learns the patterns from the data and handles uncertain cases to define the predicted samples.

3.2.3 Poposed Model(NSAI)

Neuro-symbolic AI bridges the gap between Neural AI and Symbolic AI. We used neuro-symbolic AI as our core architecture of the classification between Alzheimer's or Dementia patients and healthy or Non-Dementia people. The neural AI learns the pattern of data and predicts the result without transparency while symbolic AI provides the handcrafted and expert logics or rules to make neural AI explainable and make it understandable. Neuro-Symbolic AI integrates the Neural AI and Symbolic AI through weighted average approach. Initially, we define the neural model's weight value as 0.6 and symbolic model's weight value as 0.4. Since the symbolic classifier contains misclassified predictions, we give priority more to the Neural AI model. The higher performance of Neural AI model gives the opportunity to gain the major position at Neuro-Symbolic AI while symbolic AI is solely for the explainability. With the addition of expertise, we can elevate the model in terms of performance and consistency. Here is the core architecture of the Neuro-Symbolic AI model:

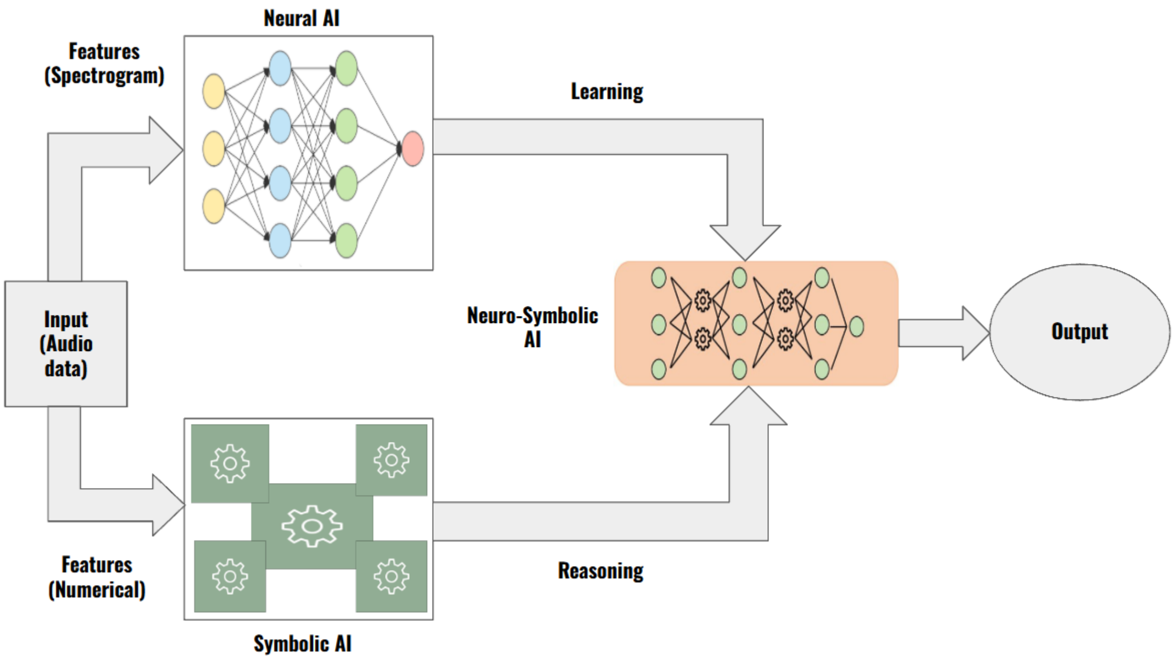


Figure 3.25: Overview of the proposed Neuro-Symbolic AI Model

3.2.4 Base Line Models Specifications

Base line models are like pre-trained models or the simpler models. We are comparing the performance of our proposed model(NSAI) with the other four models as benchmark of our system model which are-

Mobilenet V2: Mobilenet V2 is a pre-trained lightweight deep-learning model efficient in image classification and object detection tasks. This model integrates various key features like a series of convolutional layers to maintain low computational costs and increase accuracy. This model shows its effectiveness in real-time systems like IoT devices rather than classifying dementia (Sharma, 2024)[32].

Inception V3: Inceptionv3 is another image classification pre-trained model. This model consists of AvgPool, Convolution, MaxPool, Concat Layer, Dropout, Fully Connected layer and softmax function. This model provides an optimized result to analyze the medical images (Burkapalli & Patil, 2020) [7].

VGG16: VGG16 is a pre-trained and type of convolutional neural network model that is considered one of the optimal computer vision models. VGG16 consists of convolutional layers, max-pooling layers, dense layers, and softmax layers. This model can classify images very accurately, though are belong to different categories (Rohini, 2021)[11].

Vision Transformer: The Vision Transformer (VIT) model replaces the traditional convolutional neural network. It processes the images by splitting the image into sequences of fixed-size patches and using self-attention mechanisms for extracting features. The VIT model is effective and provides high accuracy for large-scale vision tasks (Inuwa, 2023) [19].

Chapter 4

Performance analysis

In this section, the performance of the proposed neuro-symbolic AI model for Dementia disease detection from speech based features is evaluated. The model combines a neural AI component that takes spectrogram features and a symbolic AI component that uses numerical rules for classification.

4.1 Perfomance Matrices

We assess the model's performance through Confusion matrix, Receiver operating characteristics(ROC Curve), accuracy, loss curve and classification report which includes accuracy, precision, recall, F1-score. The neural model is shown to excel in the detection of Dementia, the symbolic model helps in interpretability, and their combination improves explainability with better interpretation which will be shown with different performance matrices.

Accuracy: Accuracy is a performance measurement that gives a general idea concerning the ability of a constructed model to predict a correct consequence. It measures the ability to index the accurately classified instances of both true positives and true negatives with overall instances of the data set. If the dataset is balanced, then the accuracy alone can define the model's performance. In contrast, there has needed a few parameters to understand the accuracy for the imbalanced data. In this situation, the model predicts the majority class which arises our problem for performance evaluation. Since our collected data is imbalance, so we will use other performance metrics to evaluate the model performance. The accuracy is calculated as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.1}$$

where:

- **TP** (True Positive): Correctly identified positive cases.
- **TN** (True Negative): Correctly identified negative cases.
- **FP** (False Positive): Incorrectly identified positive cases.
- **FN** (False Negative): Inorrectly identified negative cases.

Precision: Precision measures how precisely a model classifies all cases that it deemed to be Dementia cases relative to all actual Dementia cases. This quantifies the reliability of positive predictions by simply indicating how many instances it correctly predicted positive cases. It avoids the false positive of the model and tend to be in true positive always. The higher the precision score, the higher the chances that when the model predicts Dementia, it's actually right. For scenarios where false positives have enormous consequences, precision is very important. If Dementia is detected, a person who is otherwise healthy (false positive) might experience unnecessary stress, medical costs and possibly psychological damage. A precise model ensures most detected Dementia cases are indeed positive, but this may result in a few true patients not being detected (false negatives). The precision is calculated as:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{4.2}$$

where:

- **TP** (True Positive): Correctly identified positive cases.
- **FP** (False Positive): Incorrectly identified positive cases.

Recall: This model is called as sensitivity or true positive ratio. This model generates the true positive predictions of the model. It measures how well the model rates positive cases, so as to classify fewer real patients as non-dementia. It is critical when a miss of the positive case (false negative) is more important than a false positive. The higher the recall, the more actual Dementia cases are identified by the model. This parameter is important in medical domain. For example, a dementia patient does not have dementia but it shows that it analyzes the data and give the false positive prediction. The patient will check to hospital and confirm he or she has the disease or not. If recall is not there, then the dementia patient can be identified as healthy person which cause a bigger problem. The recall is calculated as:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4.3}$$

where:

- **TP** (True Positive): Correctly identified positive cases.
- **FN** (False Negative): Incorrectly identified negative cases.

F_1-Score: F_1-score, harmonic mean of precision, is calculated, making it a balanced measure for model performance when both false positives and false negatives are important. This is especially useful when the data set is imbalanced so that neither precision nor recall dominates evaluation. The F1-score takes into consideration both precision (not making false positives) and recall (not missing negative) accordingly it is a single, comprehensive metric used to assess a classification performance. The high precision and low recall model might miss actual Dementia cases (false negatives), the high recall and low precision model might confuse many healthy people to be Dementia (false positives). At the same time, the F_1-score ensures that the model performs well on both ends of the spectrum. The $F-1$-score is calculated as:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4.4}$$

Confusion Matrix: A confusion matrix is a performance measure that visually shows how precise a classification model does in differentiating out classes. The confusion matrix plays a vital role in diagnosing the errors in the model especially to analyze the misclassified labels. It provides a detailed breakdown of predictions by showing the number of true positives, true negatives, false positives, and false negatives in a structured format:

Actual / Predicted	**Predicted (0)**	**Predicted (1)**
Actual (0)	TP (True Positive)	FN (False Negative)
Actual (1)	FP (False Positive)	TN (True Negative)

Table 4.1: **Confusion Matrix**

where:

- **True Positive (TP)**: Identified label as positive when predicted as positive.
- **True Negative (TN)**: Identified label as negative when predicted as negative.
- **False Positive (FP)**: Identified label as negative when predicted as positive.
- **False Negative (FN)**: Identified label as positive when predicted as negative.

ROC Curve: It is represents as Receiver Operating Characteristics Curve. It is a graphical representation of model's performance in terms of different thresholds. It draws the True Positive Rate (TPR) against the False Positive Rate (FPR). It provides the view of trade-off between recall and specificity. The AUC (Area Under Curve) is computed as the integral of the ROC curve:

$$\text{AUC} = \int_0^1 \text{TPR}(t)\, d(\text{FPR}(t)) \tag{4.5}$$

where:

- **t** is the classification threshold.
- **TPR** is the True Positive Rate.
- **FPR** is the False Positive Rate.
- **AUC = 1** is perfect model (Ideal Scenerio).
- **AUC = 0.5** is Random Classifier (Useless Model).
- **AUC < 0.5** is Worse than Random (Inversed model)

Accuracy Curve: It is also known as training accuracy curve. The accuracy curve represents that how much correct predictions the model makes, given all predictions. It expect training accuracy to increase along with testing accuracy which represents the optimal model. Here is the figure of accuracy curve (Ibrahim, 2024)[29]:

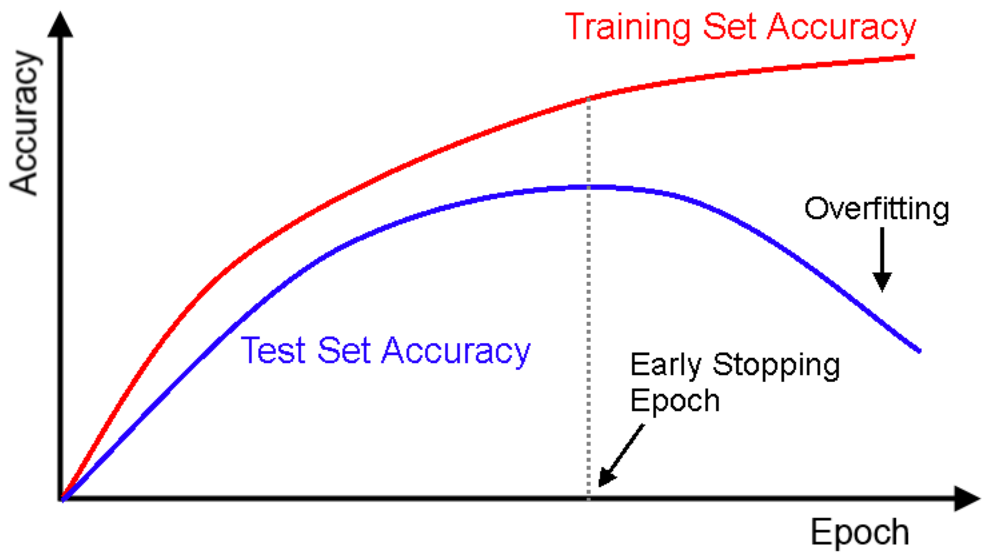


Figure 4.1: Accuracy Curve of a model.

Loss Curve: It is a tool to understand the evaluation of a model. The loss curve represents that the discrepancy between the predictions and true labels of the model. It expects training loss to decrease along with testing loss which represents the model is fitting with the training data. Here is the figure of loss curve (Ibrahim, 2024)[29]:

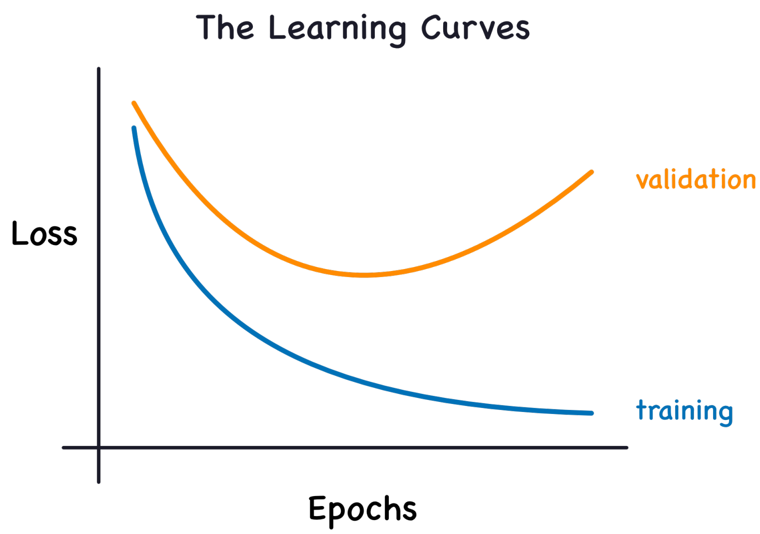


Figure 4.2: Loss Curve of a model.

4.2 Performance Analysis

In this section, we will analyze our results or outcomes of the proposed model. After splitting the dataset, the train dataset contains (80%) of total dataset and (20%) for test dataset. Our dataset contains the labeling also which defined our supervised learning of machine learning models.

4.2.1 Neural AI model

After implementing the Neural AI model, we run the code and find the outcome of the model with the hybrid neural model of CNN and LSTM. The performance will be shown in below with the performance matrices:

Classification report: The model has overall accuracy of 95% which shows promising outcome. The accuracy varies from 90% to 98% when running the code multiple times. The model has a 100% recall for dementia cases in the classification report, meaning that all actual patients are correctly identified. Precisely, its precision for dementia is 0.90, meaning it misclassified some individuals with dementia as healthy. The model achieves perfect precision (1.0) for the non-dementia cases. However, the recall is slightly lower (0.89) meaning it classified a few non-dementia cases as dementia. The model achieves an F1 score of 0.95 in classification report for both dementia and non-dementia cases which is very good as it provides good balance as between precision and recall. The macro average F1-score is 0.95, that is the average F1-score calculated over the classes in an equal manner. This weighted average F1-score of 0.95 (equal to 0.95 weighted and 0.95 unweighted) reflects the class distribution and indicates that the model behaves equivalently for all classes.

Table 4.2: **Classification Report of Neural AI Model**

	Precision	**Recall**	**F1-Score**	**Support**
0	1.00	0.89	0.94	54
1	0.90	1.00	0.95	56
Accuracy			0.95	110
Macro avg	0.95	0.94	0.95	110
Weighted avg	0.95	0.95	0.95	110

Confusion Matrix: These results are further confirmed in the confusion matrix. The result was 6 misclassified dementia (false negatives) out 54 actual non-dementia cases. On the other hand, since all 56 cases of dementia were correctly identified, there were no false positives. The model performs well in the dataset with excellent recall for dementia cases. However, the false positives can be reduced for improvement.

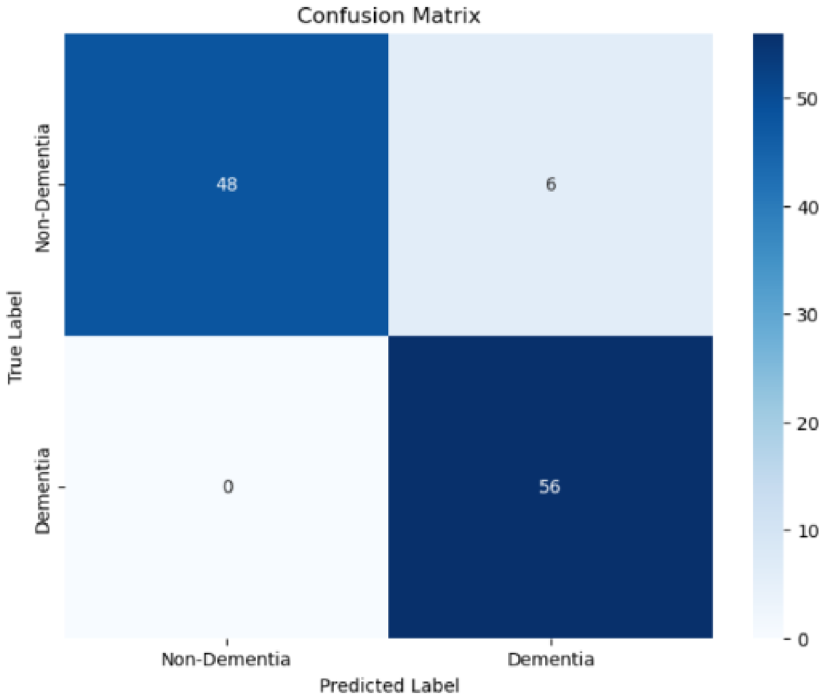


Figure 4.3: Confusion Matrix of neural AI model.

ROC Curve: It represents the performance of a binary classification model to detect dementia and non-dementia cases. The AUC of ROC Curve is 0.96 which is high and ideal for the model. The slight deviation shows a few false positive outcomes.

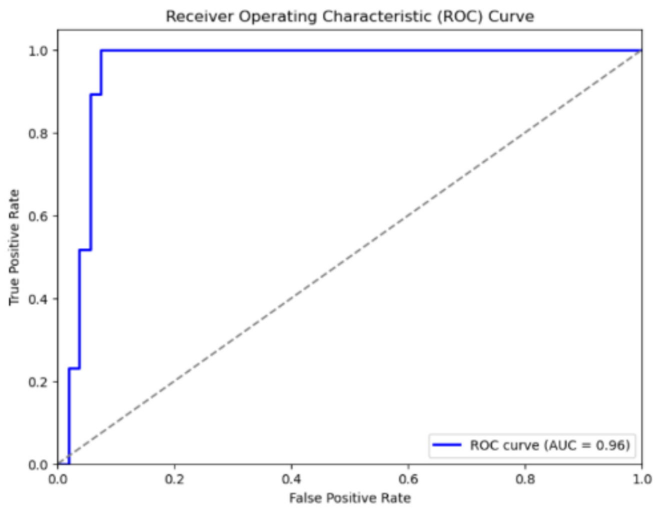


Figure 4.4: ROC Curve of Neural AI model.

Accuracy and Loss Curve: Firstly, the training and testing accuracy curve increases rapidly and the learning is effectively. The training and testing accuracy curves are close that means they have good generalization without major overfitting. High accuracy represents the model is well-trained in 50 epochs. In contrast, the training and testing loss curves gradually decreasing compared to optimizing well. Consistent loss reduction portrays proper learning. The curves shows overall excellent performance.

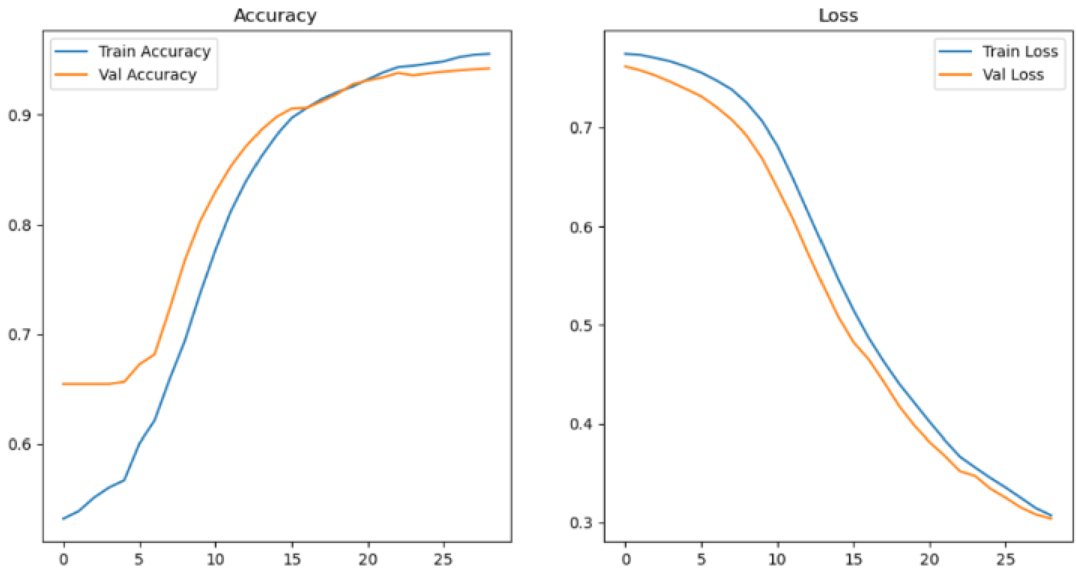


Figure 4.5: Accuracy and Loss Curve of neural AI model.

4.2.2 Symbolic AI model

Symbolic AI model will provide the black-box nature of Neural AI. It will use the generated rules to classify the dementia and non-dementia people which will give the reasons why the people have dementia. It is the new approach to the world wide research to explain the Neural AI model. The performance of the model will be shown in below with the performance matrices-

Classification report: The symbolic model achieves 54% test accuracy which is quite lower. It is struggling to distinguish between labels with rule-based classsification. Since most part of the features overlap with each other, so it is tough to get the correct identification. The precision and recall of classifying dementia are 44% and 41% respectively. Though the low performance of dementia cases, it has moderate performance in non-dementia cases with 60% precision and 62% recall. We need to add more linguistic rules which can refine the logics.

Table 4.3: **Classification Report of Symbolic AI Model**

	Precision	Recall	F1-Score	Support
0	0.60	0.62	0.61	64
1	0.44	0.41	0.43	46
Accuracy			0.54	110
Macro avg	0.52	0.52	0.52	110
Weighted avg	0.53	0.54	0.53	110

Confusion Matrix: These results are further confirmed in the confusion matrix. It has some misclassified labels which make uncertain features. However, it is useful for interpretability and predicted outcomes. The confidence-based thresholding can provide more reliability to diagnose diseases.

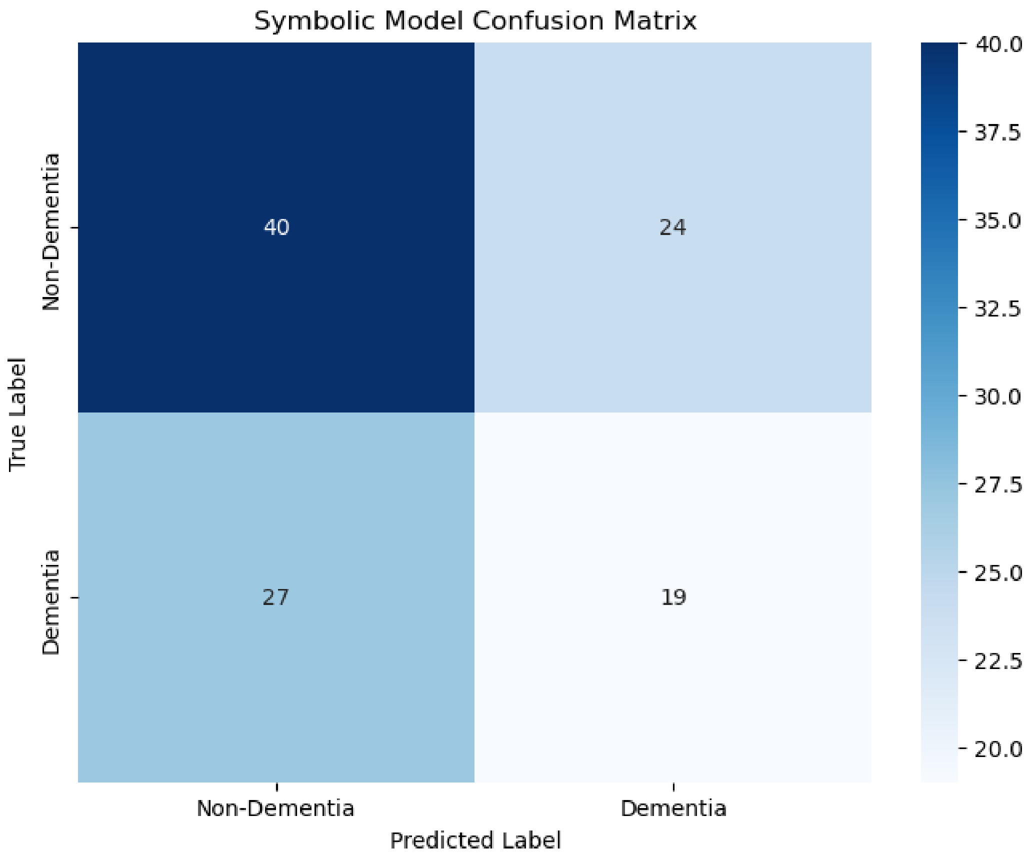


Figure 4.6: Confusion Matrix of symbolic AI model.

4.2.3 Neuro-Symbolic AI model

After implementing the Neural AI model and Symbolic AI model, we add both models to show the final performance of the Neuro-Symbolic AI model. We used weighted average approach to add the models where it get deviated to the performance of Neural AI because of more weight to the Neural AI model. The performance will be shown in below with the performance matrices:

Classification report: The classification report demonstrates the combining strength with Neural and Symbolic AI model. The overall accuracy comes with 95% along with balanced precision and excellent recall and F1-score. It is effectively determines the labels and shows that hybrid approach can be applicable for higher performance.

Table 4.4: **Classification Report of Neuro-Symbolic AI Model**

	Precision	**Recall**	**F1-Score**	**Support**
0	1.00	0.91	0.95	54
1	0.92	1.00	0.95	56
Accuracy			0.95	110
Macro avg	0.96	0.95	0.95	110
Weighted avg	0.96	0.95	0.95	110

Confusion Matrix: These results are further confirmed in the confusion matrix. It has higher classification rate where minor limitations detect 5 false negatives for non-dementia cases. As a result, it ensures that no dementia patients missed from our classification. It balances the predictive labels and true labels.

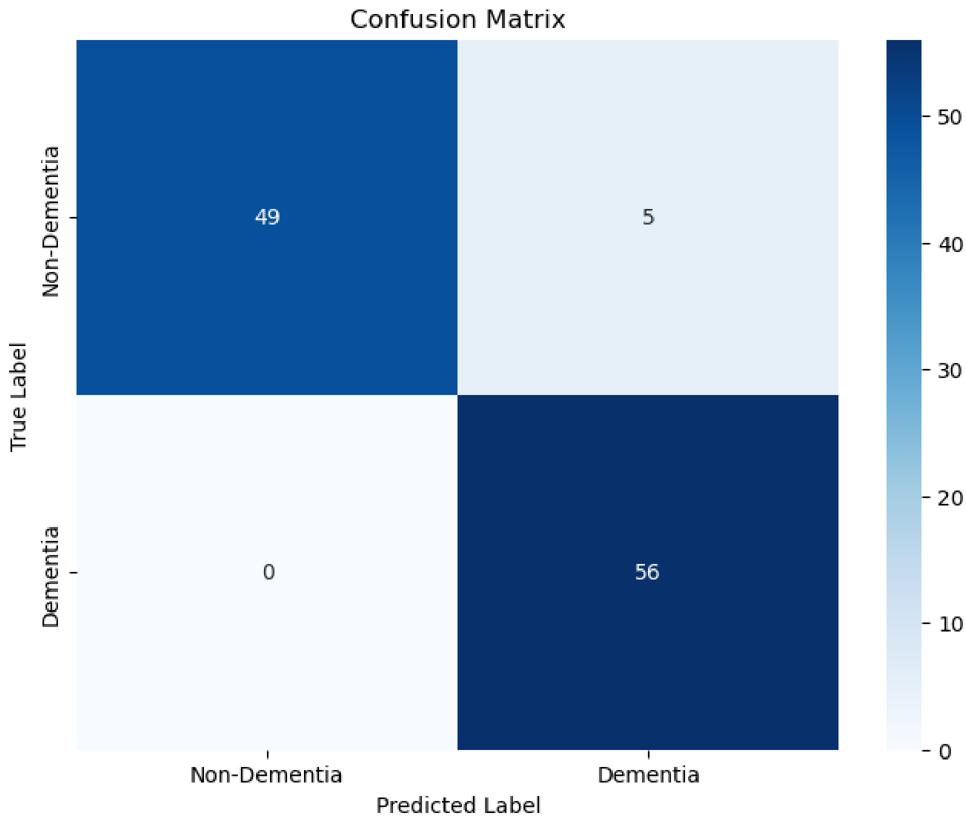


Figure 4.7: Confusion Matrix of Neuro-Symbolic AI model.

4.3 Comparison of Models

We studied some literature review which fall under the same dataset domain. We compare the models with their research. Furthermore, we used some baseline models which implemented by us to compare with our proposed model. Our proposed model of Neuro-Symbolic AI provides the more performance than most of the research and baseline models surely. In terms of accuracy, Neuro-Symbolic AI is the model with the highest accuracy of 95.45% than the most of models. The AoS3 mechanism [24] only slightly higher than NSAI which is 96%. VGG16 of the deep learning models performs the best with the 82.57% accuracy and it is capable to extract features for detection of dementia. However, it is still lacking when compared to the Neuro-Symbolic AI model, which represents the existence of symbolic reasoning as increasing transparency of classification. The accuracy of Inception V3 is 49.54%, which is the lowest one. Therefore we can say that the performance can be increased with the usage of hybrid approach and strengthens the classification. Here is the table of comparing the models with our proposed model:

Table 4.5: **Research Classification Model Performance**

Research	**Classification Model**	**Performance (%)**
(Balagopalan et al., 2021) [9]	Bert	81.8%
(Balagopalan et al., 2021) [9]	Random Forest	73.8%
(Zokaeinikoo, Kazemian, & Mitra, 2023) [24]	A Novel Attention over Self-Attention (AoS3) Mechanism	96%
(Khan et al., 2022) [14]	Stacked Deep Dense Neural Network + NLP (SDDNN) Model	93.31%
Baseline Models	Vision Transformer	59.09%
	Inception V3	49.54%
	MobileNet V2	65.14%
	VGG16	82.57%
Proposed Model	**Neuro-Symbolic AI Model**	**95.45%**

Chapter 5

Conclusion

In this paper, we establish Neuro-symbolic AI, a model designed to diagnose mental disorder like Alzheimer's disease. Neuro-symbolic AI has an important essence in the field of research in this decade. It leverages the potential strengths of statistical learning capabilities of neural networks with the incorporation of the symbolic reasoning techniques of symbolic AI. As we analyzed earlier research works, conventional machine-learning models are unable to provide precise statistics due to the limitations of the models. We aim to overcome the shortcomings of traditional machine learning moles. Alzheimer's disease was diagnosed in this research using the audio data stored in DementiaBank. The proposed Neuro-Symbolic AI model can analyze the data accurately and give output with precision. This model is not only a pivotal solution to the diagnosis of Alzheimer's disease but also plays a key role in further research in related fields. In conclusion, we feel the necessity for the advancement of mental health care as the number of mental disorder patients is growing in short order. Our approach of Neuro-symbolic AI overcomes the failure of supervised learning models, developed in such a way that it can recognize mental disorder (Dementia) and make affordable, accessible to a wide range of people.

5.1 Future Works

Though the proposed neuro-symbolic model provides a positive output, there is still scope for improvement. Since the purpose of this model is to classify dementia and healthy people with proper explainability, expert knowledge can be infused into the symbolic model to handle uncertain and overlapping cases. Expert knowledge improves the interpretability of the model, playing a significant role in the healthcare sector to diagnose diseases accurately. Additionally, the data set we use contains small amounts of data. So, in the future, we aim to research with a larger data set. We can also collect data from hospitals to increase the robustness of the model. Moreover, the model only works for detecting dementia diseases from audio data. In the future, we want to extend the model to diagnose multiple disorders. There will be an opportunity to use this model with text and image data to predict the patients with multiple disorder. Alongside diagnosing diseases, we aim that the model provides expertise solutions simultaneously.

Bibliography

[1] J. Becker, F. Boller, O. Lopez, J. Saxton, and K. McGonigle, "The natural history of alzheimer's disease: Description of study cohort and accuracy of diagnosis," *Archives of Neurology*, vol. 51, no. 6, pp. 585–594, 1994, Please also acknowledge this grant support for the Pitt corpus – NIA AG03705 and AG05133.

[2] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.

[3] A. B. Rabeh, F. Benzarti, and H. Amiri, "Diagnosis of alzheimer diseases in early step using svm (support vector machine)," in *2016 13th International Conference on Computer Graphics, Imaging and Visualization (CGiV)*, 2016, pp. 364–367. DOI: 10.1109/CGiV.2016.76.

[4] Y. Maryn, "Practical acoustics in clinical voice assessment: A praat primer," *Perspectives of the ASHA Special Interest Groups*, vol. 2, no. SIG 3, pp. 14–32, 2017. DOI: 10.1044/persp2.SIG3.14.

[5] M. R. Ahmed, Y. Zhang, Z. Feng, B. Lo, O. T. Inan, and H. Liao, "Neuroimaging and machine learning for dementia diagnosis: Recent advancements and future prospects," *IEEE Reviews in Biomedical Engineering*, vol. 12, pp. 19–33, 2019. DOI: 10.1109/RBME.2018.2886237.

[6] Y. Pan, B. Mirheidari, M. Reuber, A. Venneri, D. Blackburn, and H. Christensen, "Automatic hierarchical attention neural network for detecting ad," in *Proceedings of Interspeech 2019*, International Speech Communication Association (ISCA), 2019, pp. 4105–4109. DOI: 10.21437/Interspeech.2019-1799. [Online]. Available: https://eprints.whiterose.ac.uk/178306/.

[7] V. C. Burkapalli and P. C. Patil, "An efficient food image classification by inception-v3 based cnns," *International Journal of Scientific & Technology Research*, vol. 9, no. 3, pp. 6987–6992, 2020, Accessed: 2024-01-30. [Online]. Available: https://www.researchgate.net/publication/347558848_An_efficient_food_image_classification_by_inception-V3_based_cnns.

[8] J. Yang. "Global mental health workforce rate per 100,000 population by type." Published by Statista on Nov 30, 2023, Statista. (2020), [Online]. Available: URL%20of%20the%20Statista%20page.

[9] A. Balagopalan, B. Eyre, J. Robin, F. Rudzicz, and J. Novikova, "Comparing pre-trained and feature-based models for prediction of alzheimer's disease based on speech," *Frontiers in Aging Neuroscience*, vol. 13, p. 635 945, 2021. DOI: 10.3389/fnagi.2021.635945.

[10] A. Meghanani, C. S. Anoop, and A. G. Ramakrishnan, "Recognition of alzheimer's dementia from the transcriptions of spontaneous speech using fasttext and cnn models," *Frontiers in Computer Science*, vol. 3, p. 624 558, 2021. DOI: 10.3389/fcomp.2021.624558. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fcomp.2021.624558/full.

[11] G. Rohini, *Everything you need to know about vgg16*, Great Learning, Accessed: 2024-01-30, Sep. 2021. [Online]. Available: https://www.mygreatlearning.com/blog/everything-you-need-to-know-about-vgg16/.

[12] A. Siyaev and G.-S. Jo, "Neuro-symbolic speech understanding in aircraft maintenance metaverse," *IEEE Access*, vol. 9, pp. 154 484–154 499, 2021. DOI: 10.1109/ACCESS.2021.3128616.

[13] F. Agbavor and H. Liang, "Predicting dementia from spontaneous speech using large language models," *PLOS Digital Health*, vol. 1, no. 12, e0000168, 2022. DOI: 10.1371/journal.pdig.0000168.

[14] Y. F. Khan, B. Kaushik, M. K. I. Rahmani, and M. E. Ahmed, "Stacked deep dense neural network model to predict alzheimer's dementia using audio transcript data," *IEEE Access*, vol. 10, pp. 32 750–32 765, 2022. DOI: 10.1109/ACCESS.2022.3161749.

[15] W. H. Organization. "Mental disorders." [Online]. (2022), [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/mental-disorders.

[16] Arize AI, *Understanding bias in ml models*, Accessed: 2024-01-30, 2023. [Online]. Available: https://arize.com/blog/understanding-bias-in-ml-models/.

[17] V. Belle, "On the relevance of logic for ai: Misunderstandings in social media, and the promise of neuro-symbolic learning," Jul. 2023. DOI: 10.5281/zenodo.813672210.5281/zenodo.8136721. [Online]. Available: https://gotriple.eu/documents/ftzenodo%3Aoai%3Azenodo.org%3A8136722.

[18] C. Dhanusha, A. V. Senthil Kumar, and V. S. Giridhar Akula, "Robust cuckoo search enabled fuzzy neuro symbolic reasoning-based alzheimer's disease prediction at their earlier stages," in *Computer Networks and Inventive Communication Technologies*, S. Smys, P. Lafata, R. Palanisamy, and K. A. Kamel, Eds., Singapore: Springer Nature Singapore, 2023, pp. 871–886, ISBN: 978-981-19-3035-5.

[19] M. Inuwa, *Introduction to vision transformers (vit)*, Accessed: 2024-01-30, Jun. 2023. [Online]. Available: URL.

[20] H. Lee, J. Kim, H. Chen, *et al.*, "Comprehensive integration of hyperdimensional computing with deep learning towards neuro-symbolic ai," in *2023 60th ACM/IEEE Design Automation Conference (DAC)*, 2023, pp. 1–6. DOI: 10.1109/DAC56929.2023.10248004.

[21] U. K. Lilhore, S. Dalal, N. Faujdar, *et al.*, "Hybrid cnn-lstm model with efficient hyperparameter tuning for prediction of parkinson's disease," *Scientific Reports*, vol. 13, p. 14 605, 2023. DOI: 10.1038/s41598-023-41314-y.

[22] G. A. Rahman, M. G. R. Alam, M. S. Mia, *et al.*, "Rmdd: Ripple down rule-based decision intelligence for mental disorder diagnosis," *IEEE Access*, pp. 10–16, 2023.

[23] World Health Organization, *Dementia*, https://www.who.int/news-room/fact-sheets/detail/dementia, Accessed: 2023-03-15, Mar. 2023.

[24] M. Zokaeinikoo, P. Kazemian, and P. Mitra, "Interpretable hierarchical deep learning model for noninvasive alzheimer's disease diagnosis," *INFORMS Journal on Data Science*, vol. 2, no. 2, pp. 183–196, 2023. DOI: 10.1287/ijds.2020.0005.

[25] M. Abid, M. Asif, Z. Khemane, *et al.*, "Advances in artificial intelligence for diagnosing alzheimer's disease through speech," *Annals of Medicine & Surgery*, vol. 86, pp. 3822–3823, 2024, Published online 20 May 2024. DOI: 10.1097/MS9.0000000000002200.

[26] S. Addae, J. Kim, A. Smith, P. Rajana, and M. Kang, "Smart solutions for detecting, predicting, monitoring, and managing dementia in the elderly: A survey," *IEEE Access*, vol. 12, pp. 100 026–100 056, 2024. DOI: 10.1109/ACCESS.2024.3421966.

[27] Alzheimer's Association, "2024 alzheimer's disease facts and figures," *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, vol. 20, no. 5, pp. 3708–3821, 2024. DOI: 10.1002/alz.13809.

[28] M. Galarnyk. "What is a boxplot? definition and how to use one." Accessed: 2025-01-30. (2024), [Online]. Available: https://builtin.com/data-science/boxplot.

[29] M. Ibrahim, *A deep dive into learning curves in machine learning*, Weights & Biases, Accessed: 2025-01-30, Mar. 2024. [Online]. Available: https://wandb.ai/articles/a-deep-dive-into-learning-curves-in-machine-learning.

[30] U. Jaimini, C. Henson, and A. Sheth, *Causal neuro-symbolic ai: A synergy between causality and neuro-symbolic methods*, Apr. 2024. DOI: 10.13140/RG.2.2.21233.70242.

[31] Mayo Clinic, *Alzheimer's disease: Symptoms and causes*, https://www.mayoclinic.org/diseases-conditions/alzheimers-disease/symptoms-causes/syc-20350447, Accessed: 2024-01-30, 2024.

[32] N. Sharma, *What is mobilenetv2? features, architecture, application and more*, Accessed: 2024-01-30, Nov. 2024. [Online]. Available: https://www.analyticsvidhya.com/blog/2023/12/what-is-mobilenetv2/.

[33] The Johns Hopkins University, The Johns Hopkins Hospital, and Johns Hopkins Health System, *Johns hopkins medicine*, Accessed: 2024-05-25, Copyright © 2024 The Johns Hopkins University, The Johns Hopkins Hospital, and Johns Hopkins Health System. All rights reserved., 2024. [Online]. Available: https://www.hopkinsmedicine.org/.