

Alzheimer Disease Detection System using Electroencephalogram Biomarkers

A PROJECT REPORT

Submitted by

KAUSHIK TAYI [RA2011030010048]

Under the Guidance of

Dr. S. PRABAKERAN

(Associate Professor, Department of Networking and Communications)

in partial fulfillment of the requirements for the degree of

BACHELOR OF TECHNOLOGY
in
COMPUTER SCIENCE AND ENGINEERING
with specialization in CYBER SECURITY



DEPARTMENT OF NETWORKING AND
COMMUNICATIONS
COLLEGE OF ENGINEERING AND TECHNOLOGY
SRM INSTITUTE OF SCIENCE AND TECHNOLOGY
KATTANKULATHUR- 603 203

MAY 2024



Department of Computational Intelligence
SRM Institute of Science & Technology
Own Work Declaration Form

Degree/ Course : B.Tech in Computer Science Engineering
Student Name : Kaushik Tayi
Registration Number : RA2011030010048
Title of Work : Alzheimer Disease Detection System using EEG Biomarkers

We hereby certify that this assessment complies with the University's Rules and Regulations relating to Academic misconduct and plagiarism, as listed in the University Website, Regulations, and the Education Committee guidelines.

We confirm that all the work contained in this assessment is Our own except where indicated, and that We have met the following conditions:

- Clearly referenced / listed all sources as appropriate
- Referenced and put in inverted commas all quoted text (from books, web, etc)
- Given the sources of all pictures, data etc. that are not my own
- Not made any use of the report(s) or essay(s) of any other student(s) either past or present
- Acknowledged in appropriate places any help that I have received from others (e.g. fellow students, technicians, statisticians, external sources)
- Compiled with any other plagiarism criteria specified in the Course handbook / University website

I understand that any false claim for this work will be penalized in accordance with the University policies and regulations.

DECLARATION:

I am aware of and understand the University's policy on Academic misconduct and plagiarism and I certify that this assessment is Our own work, except where indicated by referring, and that I have followed the good academic practices noted above.

RA2011030010048
Kaushik Tayi

\

ACKNOWLEDGEMENT

I express my humble gratitude to **Dr. C. Muthamizhchelvan**, Vice-Chancellor, SRM Institute of Science and Technology, for the facilities extended for the project work and his continued support.

I extend my sincere thanks to Dean-CET, SRM Institute of Science and Technology, **Dr.T.V. Gopal**, for his invaluable support.

I wish to thank **Dr. Revathi Venkataraman, Professor & Chairperson**, School of Computing, SRM Institute of Science and Technology, Kattankulathur, for her support throughout the project work.

I am incredibly grateful to my Head of the Department, **Dr. Annapurani K**, Professor and Head, Department of Networking and Communications, School of Computing, SRM Institute of Science and Technology, for her suggestions and encouragement at all the stages of the project work.

I want to convey my thanks to my Project Coordinator, **Dr. G. Suseela**, Associate Professor, Panel Head, **Dr. S. Prabakeran**, Associate Professor and members, **Dr. Priyanka**, Assistant Professor, **Dr. Mahalakshmi**, Assistant Professor, **Dr. M. Sundarrajan**, Assistant Professor, Department of Networking and Communications, School of Computing, SRM Institute of Science and Technology, for their inputs during the project reviews and support.

I register my immeasurable thanks to my Faculty Advisor, **Dr. B Yamini**, Associate Professor, Department of Networking and Communications, School of Computing, SRM Institute of Science and Technology, for leading and helping me to complete my course.

I inexpressible respect and thanks to my guide, **Dr. S. Prabakeran**, Associate Professor, Department of Networking and Communications, SRM Institute of Science and Technology, for providing me with an opportunity to pursue my project under his mentorship. He provided me with the freedom and support to explore the research topics of my interest. His passion for solving problems and making a difference in the world has always been inspiring.

I sincerely thank the Networking and Communications, Department staff and students, SRM Institute of Science and Technology, for their help during my project. Finally, I would like to thank parents, family members, and friends for their unconditional love, constant support, and encouragement.

Kaushik Tayi [RA2011030010048]



SRM INSTITUTE OF SCIENCE AND TECHNOLOGY KATTANKULATHUR – 603 203

BONAFIDE CERTIFICATE

Certified that 18CSP109L project report titled “**Alzheimer Disease Detection System using EEG Biomarkers**” is the bonafide work of **Kaushik Tayi [RA2011030010048]** who carried out the project work under my supervision. Certified further, that to the best of my knowledge the work reported herein does not form any other project report or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

Dr. S. PRABAKERAN
SUPERVISOR

Associate Professor
DEPARTMENT OF NETWORKING
AND COMMUNICATIONS

Dr. ANNAPURANI K
PROFESSOR AND

HEAD OF THE DEPARTMENT
DEPARTMENT OF NETWORKING AND
COMMUNICATIONS

Internal Examiner

External Examiner

ABSTRACT

This research project Alzheimer's disease (AD) is a progressive neurological illness that makes early detection and treatment extremely difficult. An inventive system for detecting Alzheimer's disease using EEG biomarkers, cutting-edge machine learning methods, and a user-friendly interface is shown in this abstract. The suggested approach efficiently analyzes EEG signals for early AD detection by using Random Forest Classifier, a potent machine learning technique. By means of careful feature selection and training, the model reaches an exceptional 97% accuracy, exceeding the standards set by previous models in the domain. Early detection methods are made more accessible through the non-invasive and economical approach offered by the use of EEG biomarkers. Moreover, the incorporation of a Streamlit technology-developed dynamic user interface amplifies the system's accessibility and usability for both researchers and doctors. The model may be seamlessly interacted with thanks to its user-friendly interface, which enables straightforward EEG data input and real-time result interpretation. This system is a viable solution for AD detection because it combines state-of-the-art machine learning algorithms with an easy-to-use interface, providing great accuracy and usability. This approach represents a substantial development in the early identification and management of Alzheimer's disease, potentially improving patient outcomes and quality of life by utilizing EEG biomarkers and sophisticated computational techniques.

Keywords: Machine learning, healthcare, Medical Classifiers, Parameter Analysis, Random Forest, AI.

TABLE OF CONTENTS

ABSTRACT	iii
TABLE OF CONTENTS	iv
LIST OF FIGURES	vi
ABBREVIATIONS	vii
1 INTRODUCTION TO ALZHEIMER’S AND EEG	1
1.1 Problem Statement	2
1.2 Objective	7
1.3 Detection of Alzheimer’s Disease using EEG Biomarkers	13
1.4 Scope	14
1.5 Random Forest	17
2 LITERATURE OVERVIEW ON ALZHEIMER’S IN EEG	23
2.1 Literature Review	23
2.2 Research Gap	31
3 SYSTEM ARCHITECTURE OF PROPOSED PROJECT	39
3.1 Architecture Diagram	39
3.2 Use-Case Diagram	43
3.3 Sequence Diagram	45
3.4 Class Diagram	46
3.5 Proposed Algorithm	50

4	METHODOLOGICAL APPROACH ON DETECTING ALZHEIMER'S DISEASE USING EEG	53
4.1	Modules	53
4.1.1	Data Collection	57
4.1.2	Data Pre-processing	58
4.1.3	Data Modelling	59
4.1.4	Data Analysis	60
4.1.5	Cross Validation	61
4.1.6	Model Generation	62
5	RESULTS AND DISCUSSIONS	70
5.1	Analysis of the Dataset	70
5.2	Evaluation on Training Data	75
5.3	Evaluation on Testing Data	78
5.4	Analysis	81
6	CONCLUSION AND FUTURE ENHANCEMENT	85
	REFERENCES	87
	APPENDIX	
	A SOURCE CODE	90
	B PLAGIARISM REPORT	98

LIST OF FIGURES

Fig 3.1.1	Architecture diagram for working of Alzheimer's using Decision Tree	39
Fig 3.2.1	Use-case Diagram	43
Fig 3.3.1	Sequence Diagram	45
Fig 3.4.1	Class Diagram	46
Fig 5.1.1	Glimpse of the dataset used for Biomarkers Test	70
Fig 5.1.2	Glimpse of EEG – Pointers	72
Fig 5.2.1	Detection of Alzheimer using Biomarkers with clinical directions	75
Fig 5.2.2	Detection of Clinical Alzheimer from Biomarker correlations	76
Fig 5.2.3	EEG value prediction and conditional analysis test reports	77
Fig 5.3.1	EEG values denoting the presence of mild Alzheimer based on Electrode Data	78
Fig 5.3.2	Presence of severe Alzheimer predicted by the EEG Electrode data	79
Fig 5.3.3	Decision Tree Structure for the designing of Alzheimer Disease based on Biomarkers	80
Fig 5.3.4	Data distribution for the sample of EEG Biomarkers dataset on multiple patients.	581
Fig 5.3.5	PNSA vs nWBV distribution with EEG age groups	58
Fig 5.3.6	Spearman's Correlation Matrix demonstrating the relationship of various data pointers with respect to the Biomarkers Test.	82

ABBREVIATIONS

ANN	Artificial Neural Network
AI	Artificial Intelligence
BMI	Body Metabolism Index
CBTI	Clinical Multiple Problems in EEG Biomarker representations Imaging
DL	Deep Learning
EEG	Electro-encephalogram
CNN	Convolutional Neural Network
EWS	Early Warning System
FED	Feature Engineering and Designing
RNN	Recurrent Neural Network
KPI	Key Performance Indicator
PET	Positron Emission Tomography
ML	Machine Learning
SRG	Smart Report Generation
VGG	Visual Geometry Group

CHAPTER 1

INTRODUCTION TO ALZHEIMER'S AND EEG

Alzheimer's disease (AD) is a progressive neurological illness that primarily affects memory, thinking, and behavior. It has a substantial impact on cognitive performance. Being the primary cause of dementia, AD poses a significant challenge to those who are impacted, as well as to their families, caregivers, and the larger medical community. Alzheimer's disease is becoming more common as the world's population ages, which makes it a serious public health concern. This increase puts more strain on society and healthcare institutions, highlighting the need for efficient management and treatment plans.

Alzheimer's disease has a significant impact on people by progressively impairing their cognitive function and ability to carry out daily duties, which increases reliance and lowers quality of life. Significant emotional and financial burden is also caused by the condition to families and caregivers, who frequently take on the majority of the care and support responsibilities. The rapid identification and precise diagnosis of Alzheimer's disease are of utmost importance as they can result in enhanced disease management, potentially impeding the course of symptoms through prompt treatment measures. Additionally, as the population ages, Alzheimer's disease is becoming more common, which has a greater impact on society and highlights the urgent need for scalable and efficient diagnostic treatments.

Alzheimer's disease (AD) is a progressive neurological illness that primarily affects memory, thinking, and behavior. It has a substantial impact on cognitive performance. Being the primary cause of dementia, AD poses a significant challenge to those who are impacted, as well as to their families, caregivers, and the larger medical community. Alzheimer's disease is becoming more common as the world's population ages, which makes it a serious public health concern. This increase puts more strain on society and healthcare institutions, highlighting the need for efficient management and treatment plans.

Alzheimer's disease has a significant impact on people by progressively impairing their cognitive function and ability to carry out daily duties, which increases reliance and lowers quality of life. Significant emotional and financial burden is also caused by the condition to families and caregivers, who frequently take on the majority of the care and support responsibilities.

The rapid identification and precise diagnosis of Alzheimer's disease are of utmost importance as they can result in enhanced disease management, potentially impeding the course of symptoms through prompt treatment measures. Additionally, as the population ages, Alzheimer's disease is becoming more common, which has a greater impact on society and highlights the urgent need for scalable and efficient diagnostic treatments.

The creation of an EEG biomarker-based Alzheimer's disease detection system is the main research issue this project attempts to solve. This strategy uses cutting-edge machine learning algorithms to evaluate EEG data in order to provide a non-invasive, effective, and easily accessible way to identify Alzheimer's at an early stage. By enabling early and accurate identification, resolving this issue has the potential to significantly change how Alzheimer's disease is managed. Early diagnosis enables earlier intervention, which can reduce the rate at which the disease progresses, relieve symptoms, and greatly improve the patients' quality of life. Furthermore, the system hopes to offer a less expensive diagnosis tool by utilizing EEG biomarkers as opposed to more intrusive or costly imaging procedures, which makes it a significant complement to existing diagnostic methodologies. By increasing diagnostic accessibility and accuracy, this approach aims to enhance the fields of neurology and geriatrics and give significant advantages to patients, healthcare professionals, and researchers.

Alzheimer's disease (AD), which affects millions of people worldwide and is characterized by gradual cognitive decline and memory loss, poses a serious threat to global health. Because of its frequency and lack of a treatment, precise and timely detection is critically important. In addition to having an impact on those who are diagnosed with AD, caregivers and healthcare systems often bear heavy financial and emotional costs. The creation of trustworthy detection systems is crucial in this situation. A non-invasive method for detecting AD is electroencephalography (EEG), which provides information on brain activity. When combined with sophisticated biomarker analysis, EEG data offers insightful knowledge about the brain alterations linked to AD, facilitating early intervention and individualized treatment plans. By utilizing machine learning methods like the Random Forest Classifier, AD detection systems can achieve success rates up to 97%, which is significantly higher than earlier models.

1.1 Problem Statement

Alzheimer's disease (AD), which affects countless people worldwide, is a serious global health concern. It is defined by a progressive loss of memory and cognitive abilities. The fact that there is now no proven treatment emphasizes how important early and accurate diagnosis methods are.

AD places significant emotional and financial burdens on caregivers as well as on the medical system, in addition to having a significant negative influence on people who are directly affected by the illness. The development of reliable diagnostic tools is critical in this situation.

As a non-invasive way to look at brain activity, electroencephalography (EEG) offers a potential approach for the early identification of AD. The neurological changes associated with AD can be found by combining advanced biomarker analysis with EEG data. The implementation of prompt and customized treatment strategies depends on this early detection. The precision of AD detection systems has significantly increased with the use of sophisticated machine learning models, including the Random Forest Classifier.

High success rates have resulted from these developments; some studies report accuracy levels as high as 97%, which is a notable improvement over previous technique.

Furthermore, the use of a Streamlit-powered dynamic user interface enhances accessibility and user interaction, making the system a useful and effective instrument for both researchers and clinicians. This AD detection system, which uses cutting-edge technology and EEG biomarkers, is a major step in the fight against the debilitating consequences of the illness and offer hope for better patient outcomes and quality of life. One of the most common neurodegenerative diseases is Alzheimer's disease (AD), which progressively impairs cognitive abilities and finally results in a significant loss of identity and independence. Its effects ripple not just through the lives of people affected but also through the lives of their family and caregivers, putting a tremendous strain on global healthcare systems. Early Alzheimer's disease diagnosis is critical since treatments during the disease's early stages can greatly slow its progression and enhance the quality of life for those who are affected. Using electroencephalogram (EEG) biomarkers has become a viable approach for AD identification and monitoring in recent years.

The non-invasive EEG method provides real-time information about brain activity by detecting minute electrical fluctuations that may be signs of underlying diseases. Biomarkers generated from EEG data, including changes in coherence, connectivity, and spectral power, are important markers of neurophysiological changes linked to the course of Alzheimer's disease. Reliability and accuracy of AD detection methods have increased with the integration of EEG biomarkers with machine learning algorithms, specifically the Random Forest Classifier. With an astounding 97% accuracy rate, this technology outperforms its predecessors and provides a powerful tool for early diagnosis and customized treatment plans.

Moreover, the integration of a Streamlit-powered responsive user interface improves accessibility and usability, rendering it a competitive substitute for researchers and healthcare professionals alike.

This Alzheimer's disease detection system signals a new era in neurodegenerative disease management by utilizing the synergy of EEG biomarkers, cutting-edge machine learning techniques, and user-friendly interfaces. This creates promise for better results and improved patient care.

Need for Early Detection:

For a number of reasons, early identification of Alzheimer's disease (AD) is essential. First of all, it makes it possible to step in when medicines have a greater chance of stopping the disease's progression. Early AD detection opens up a crucial window of time during which treatment measures may have a greater effect, possibly maintaining cognitive function and improving quality of life. Additionally, before the illness advances to more severe phases, early detection helps people and their families prepare for the future by arranging for the care, housing, and finances that will be required. Additionally, it gives patients the chance to take part in clinical trials, supporting research that may result in novel therapies and a better comprehension of AD.

Getting a correct diagnosis is just as crucial to managing Alzheimer's. A misdiagnosis may result in the administration of unsuitable medications that, in addition to being unsuccessful, may have unfavorable side effects or worsen the patient's condition. When AD is correctly diagnosed, patients are given the most appropriate and successful care techniques based on their individual stage of the disease. Furthermore, it's critical to differentiate Alzheimer's from other forms of dementia or memory-related disorders as each may call for a distinct strategy to treatment and therapy. A precise diagnosis helps medical professionals decide on the optimal course of action, which is essential for managing diseases effectively and matching appropriate therapeutic measures.

Within the framework of the "Alzheimer Disease Detection System using EEG Biomarkers," the requirement for precise diagnosis and timely identification is met by utilizing cutting-edge machine learning algorithms to examine EEG patterns suggestive of AD. This method seeks to increase diagnostic accessibility and accuracy by giving medical providers a non-invasive, effective, and economical tool to identify AD early and make better decisions about patient management.

Alzheimer's disease (AD) prevalence

One of the most prevalent types of dementia, Alzheimer's disease (AD) affects millions of people globally. An estimated 50 million people worldwide are thought to be affected by dementia, with 60–70% of these instances being Alzheimer's disease. The age-related increase in AD prevalence is most noticeable in those over 65. The number of Alzheimer's patients is predicted to increase significantly as the global population ages as a result of improved healthcare and medical developments. Anticipations indicate that the global population with Alzheimer's disease may nearly quadruple to 152 million by 2050, highlighting the critical need for efficient diagnostic and treatment approaches.

Growing Alzheimer's Disease Burden

The increasing number of people affected by Alzheimer's disease puts a heavy strain on economies, society, and healthcare systems. The enormous costs associated with providing care for persons with Alzheimer's disease include direct medical charges, long-term care costs, and the indirect costs of caregivers' time. The care of Alzheimer's patients costs billions of dollars a year, therefore these economic effects are significant. It is anticipated that as the number of people with the disease rises, the financial burden would also rise. In addition to its financial cost, Alzheimer's disease puts a great emotional and physical strain on families and caregivers, which frequently results in higher levels of stress and health problems for those who provide care. The "Alzheimer Disease Detection System using EEG Biomarkers," which aims to facilitate early detection and potentially alleviate some of the challenges associated with managing this debilitating disease, is one of the innovative solutions that are desperately needed given the rising prevalence and burden of AD.

Research Problem: Developing an AD Detection System Using EEG Biomarkers

The central research problem addressed by this project involves developing a sophisticated Alzheimer's Disease (AD) detection system that utilizes electroencephalogram (EEG) biomarkers. Alzheimer's disease is characterized by complex neurodegenerative processes that alter brain activities, which can potentially be detected through specific patterns in EEG signals. The challenge lies in accurately identifying and interpreting these patterns to diagnose AD at its earliest stages. The project aims to harness the capabilities of EEG, a non-invasive and relatively affordable diagnostic tool, combined with advanced machine learning algorithms, to reliably distinguish between normal aging-related changes in brain activity and those specific to Alzheimer's disease.

Significance of Addressing This Problem

Addressing the problem of early and accurate detection of Alzheimer's using EEG biomarkers is significant for several reasons.

Firstly, it has the potential to dramatically improve the prognosis for individuals affected by Alzheimer's by enabling earlier intervention strategies. Early detection allows for the timely administration of treatments that can slow disease progression, manage symptoms more effectively, and significantly improve the quality of life for patients and their families. Additionally, by improving diagnostic accuracy, this system can help reduce the uncertainty and anxiety associated with the diagnostic process, providing patients and caregivers with clearer information and better preparation for managing the disease. Furthermore, advancements in this area could lead to greater scalability of diagnostic services, making Alzheimer's detection more accessible to populations where traditional imaging technologies might be cost-prohibitive.

Ultimately, the successful development of an AD detection system using EEG biomarkers represents a major leap forward in the field of neurology and geriatric medicine, providing a foundation for future innovations in the diagnosis and treatment of neurodegenerative diseases.

Potential Advantages of Early Alzheimer Disease Detection with EEG Biomarkers Detection System, Alzheimer's disease (AD) can be detected early with significant advantages when using an EEG biomarker-based detection method. This can change the way AD is treated and managed. The main benefits are as follows:

Increased Treatment Effectiveness: Early Alzheimer's disease detection can greatly enhance treatment results. Early in the course of an illness, several treatments are more successful and may even slow the decline of cognitive abilities. For those with AD, early intervention can assist maintain independence and everyday functioning, thereby extending their quality of life.

Better Organization and Management: Patients and their families have more time to make future plans for medical care, housing, and finances when they receive an early diagnosis. It enables them to decide on living wills, health care proxies, and other legal matters with knowledge while the patient is still able to take part in these conversations.

Tailored Therapeutic Strategies: Individualized treatment plans can be put into place with early discovery. These programs may involve a mix of medication, lifestyle modifications, and cognitive therapy, depending on the needs and circumstances of each patient. This individualized strategy may help manage the symptoms of the condition more effectively and slow down its progression.

Decrease in Healthcare Costs: Early diagnosis of Alzheimer's disease may prevent or postpone potentially costly acute care settings, emergency treatments, and prolonged long-term care. Due to the severe deterioration of cognitive and physical functions, advanced Alzheimer's care frequently includes extensive medical and custodial care; this can result in significant savings on healthcare expenditures.

Promotion of Clinical Research: By identifying potential participants for clinical trials at an earlier stage of the disease's progression, early diagnosis of AD using EEG biomarkers can promote clinical research. This advances the study of AD and encourages the creation of fresh therapies, possibly leading to a cure. Clinical trial participants advance our understanding of Alzheimer's disease scientifically, which may result in advances in its treatment.

Enhanced Preventive and Awareness: The "Alzheimer Disease Detection System using EEG Biomarkers" is an excellent early detection technique that can increase awareness of Alzheimer's disease and its early signs. A greater understanding of the disease can motivate people to seek out early diagnostic testing and make preventative lifestyle choices that could lower their risk of developing Alzheimer's or postpone its development.

This novel approach, which uses EEG biomarkers to identify Alzheimer's early on, improves clinical outcomes while giving patients and their families a sense of hope and empowerment. It also gives them access to time and choices that were not previously available in the conventional frameworks for Alzheimer's care.

1.2 Objective

The main objective of the "Alzheimer Disease Detection System using EEG Biomarkers" project is to dramatically improve patient outcomes and the diagnosis of Alzheimer's disease (AD) by creatively fusing cutting-edge machine learning algorithms with EEG biomarkers. Through the use of minor EEG changes that are visible in the early stages of Alzheimer's disease, this program aims to improve diagnostic precision. By allowing early diagnosis, opportune treatments can be made to halt the illness's progression and successfully manage symptoms. The system's capacity to deliver precise diagnoses at an early stage of the illness allows medical professionals to customize patient treatment plans, enhancing patients' quality of life by preserving their cognitive abilities and everyday living skills for as long as feasible.

This research also intends to minimize diagnostic ambiguity associated with dementia and maximize time spent on patient care by incorporating an easy-to-use and dependable AD detection technology into clinical workflows. The method not only serves to lower overall healthcare expenditures related with Alzheimer's disease by reducing misdiagnoses and needless treatments, but it also helps to delay the onset of severe symptoms by enhancing diagnostic accuracy and enabling early identification. Overall, by offering early, accurate, and actionable disease insights, the "Alzheimer Disease Detection System using EEG Biomarkers" aims to revolutionize the diagnosis of Alzheimer's and greatly improve patient outcomes, representing a huge leap in the fight against this crippling illness.

The main objectives of the study titled "Alzheimer Disease Detection System using EEG Biomarkers" are strategically outlined to tackle the challenges of diagnosing Alzheimer's Disease effectively and efficiently. These objectives include:

1. **Developing a Robust Diagnostic Tool:** To create and refine a sophisticated system that uses EEG biomarkers for the early detection of Alzheimer's Disease. This involves harnessing specific patterns in EEG signals that are indicative of early-stage Alzheimer's, aiming to differentiate these from normal age-related changes.
2. **Integrating Advanced Machine Learning Algorithms:** To incorporate and optimize machine learning techniques, particularly the Random Forest Classifier, within the detection system. This objective focuses on analyzing EEG data accurately to identify Alzheimer's Disease at its onset, thereby enhancing the predictive accuracy and reliability of the system.
3. **Improving Early Detection Capabilities:** To achieve early detection of Alzheimer's Disease, allowing for timely intervention. Early detection is crucial as it can significantly alter the management of the disease, potentially slowing its progression and offering a better quality of life for patients.
4. **Ensuring High Diagnostic Accuracy:** To achieve and maintain a high level of diagnostic accuracy, surpassing existing methods. This includes rigorous testing and validation of the EEG-based model to ensure it meets clinical standards and provides dependable results for clinical use.
5. **Designing a User-Friendly Interface:** To develop a dynamic and intuitive user interface using Streamlit technology. This interface will facilitate easy interaction with the system for both medical practitioners and researchers, allowing for straightforward input of EEG data and real-time interpretation of results.

6. **Facilitating Widespread Clinical Adoption:** To ensure that the system is practical and scalable for widespread clinical adoption. This includes addressing logistical and economic factors that can influence the deployment and use of the system in diverse healthcare settings.
7. **Contributing to Alzheimer's Research and Management:** To provide a tool that not only aids in the diagnosis but also contributes to the broader scope of Alzheimer's research by collecting data that can be used for ongoing studies and future improvements in treatment and management strategies.

Important Results from the Research on EEG Biomarkers for Alzheimer's Disease Identification Several important discoveries that highlight the promise and difficulties of this method are presented in the literature on the use of EEG biomarkers for the identification of Alzheimer's disease (AD):

Alterations in Frequency Bands: There is a body of research indicating that patients with Alzheimer's disease have notable alterations in specific EEG frequency bands. Theta and delta bands show a noticeable rise in slow-wave activity, while the beta and alpha bands show a noticeable drop in fast-wave activity. These changes are thought to be possible biomarkers for the early diagnosis of AD because they correspond with the severity of cognitive deterioration.

Disrupted Connectivity: Research on AD patients has shown a reduction in functional connectivity across the brain, especially in the default mode network. This disturbance is linked to the intensity of symptoms and may be used as a diagnostic tool to shed light on the neurophysiological causes of AD.

Power Spectral Density: Studies show that patients with Alzheimer's disease have altered power spectral densities of their EEG signals, which may help differentiate between early stages of the disease and normal aging. This function has demonstrated potential in accurately categorizing control groups and AD.

Using Models for Machine Learning: Accurate AD classification has been greatly improved by using machine learning approaches to evaluate EEG data. Numerous algorithms, like as decision trees, neural networks, and support vector machines, have been used in studies to increase the accuracy and detection rates of EEG-based AD diagnosis systems.

These objectives collectively aim to revolutionize the approach to diagnosing Alzheimer's Disease, making it more proactive, patient-centric, and scientifically advanced. By achieving these goals, the project hopes to significantly impact the management of Alzheimer's Disease, improving outcomes for patients across the globe.

The main goal of the research, as stated in the abstract, is to create a sophisticated and efficient system for the early identification of Alzheimer's disease (AD) by combining machine learning methods, notably the Random Forest Classifier, with electroencephalogram (EEG) biomarkers. This system outperforms current models with an accuracy rate of 97%, indicating its goal of achieving high diagnostic accuracy. The initiative aims to increase accessibility and viability of early detection of AD by utilizing EEG biomarkers, a non-invasive and economical approach. The system also makes use of an interactive, Streamlit-developed user interface, which improves accessibility and usability for researchers and doctors alike. Simple EEG data entry and real-time result interpretation are made possible by this interface, which makes interacting with the model easy. Ultimately, the project aims to enhance patient outcomes and quality of life through fast and accurate diagnosis by combining cutting-edge machine learning algorithms with useful, user-centric design to provide a dependable tool for early AD identification.

The progressive neurological disease known as Alzheimer's disease is a major threat to the health of the world's population since it is becoming more common, has a terrible effect on people individually, in families, and in society, and currently has no known treatment. Alzheimer's disease is becoming more common as the population ages, and early detection is essential for efficient care and intervention. Cognitive deterioration, memory loss, poor reasoning, and eventually the loss of autonomy are symptoms of the condition.

Early diagnosis of Alzheimer's disease enables prompt therapies, which may decrease the illness's course and enhance the quality of life for those who are affected. Electroencephalography (EEG) is a non-invasive, low-cost method of evaluating brain function that has shown promise in the diagnosis of Alzheimer's disease.

EEG records the electrical activity of the brain, offering information on anomalies connected to neurodegenerative diseases like Alzheimer's disease as well as the brain's overall functional integrity. Researchers can find biomarkers for cognitive impairment by examining EEG patterns; this provides a window into the course of the disease and helps with early identification.

Aligning machine learning methods with EEG data, like the Random Forest Classifier algorithm, has improved the precision and dependability of Alzheimer's disease diagnostic systems. Large volumes of EEG data may be sorted through by machine learning algorithms, which can then be used to identify patterns and characteristics that differentiate between people in good health and those who have Alzheimer's disease or other types of cognitive impairment. Renowned for its resilience and capacity to manage intricate datasets, the Random Forest Classifier has demonstrated great efficacy in discriminating between various cognitive states, with an astounding 97% accuracy rate in the identification of Alzheimer's disease.

Furthermore, the democratization of access to Alzheimer's disease detection systems has been facilitated by the development of user-friendly interfaces, like the adaptable UI created with Streamlit. Streamlit provides a straightforward, yet effective, framework for implementing machine learning models and displaying outcomes, allowing academics and doctors to work together with the system with ease. The interface's easy design makes it easier to comprehend data and make decisions, enabling medical practitioners to efficiently use EEG-based biomarkers for early Alzheimer's disease identification. Alzheimer's disease is a neurological illness that progresses over time and is a major global danger to the aging population. Alzheimer's affects people directly and has a significant negative impact on families and healthcare systems due to its sneaky onset and terrible effects. Cognitive decline, memory loss, and behavioral abnormalities are the disease's hallmarks, eventually stripping patients of their identity and independence. Early Alzheimer's disease detection is critical to the implementation of interventions that may reduce the disease's course and enhance the quality of life for those who are affected.

Exploring electroencephalography (EEG) biomarkers has become a viable approach for Alzheimer's disease early identification and tracking in recent years. The electrical activity of the brain is reflected in EEG data, which offers important insights into the functional changes linked to neurodegeneration.

Researchers can find abnormal patterns in EEG signals, such as disturbed brain oscillations and connection patterns, that are suggestive of Alzheimer's pathology. These biomarkers are especially desirable for extensive screening programs and long-term research since they provide an economical and non-invasive way to measure brain activity.

Machine learning algorithms have been essential in evaluating complex EEG data and identifying relevant patterns in the quest for precise Alzheimer's diagnosis.

The Random Forest Classifier is a standout algorithm among these due to its great accuracy and resilience in differentiating between patients with Alzheimer's disease and healthy controls. Random Forest is an excellent tool for managing high-dimensional data and identifying complex correlations between EEG variables and disease status since it makes use of ensemble learning techniques. Research employing Random Forest Classifier models have demonstrated remarkable accuracy rates, frequently surpassing alternative machine learning techniques in Alzheimer's disease identification assignments. With an astounding accuracy rate of 97%, these models show how effective it is to use cutting-edge computational techniques in the battle against Alzheimer's disease.

Apart from computational innovations, the incorporation of intuitive interfaces improves the usability and accessibility of EEG biomarker-based Alzheimer's disease detection systems. A responsive and user-friendly framework for viewing EEG data and model predictions is provided by Streamlit, a Python package for building interactive web apps. Clinicians and researchers may quickly input EEG recordings, view pertinent features, and interpret diagnostic results thanks to a well-designed user interface in actual time. Streamlit's smooth integration of machine learning algorithms speeds up the conversion of research findings into clinical practice by promoting cooperation and knowledge sharing among medical professionals.

All things considered, the creation of algorithms for detecting Alzheimer's disease based on EEG biomarkers is a noteworthy advancement in the fields of computational medicine and neurology. These systems have the ability to significantly transform Alzheimer's patient treatment plans and early diagnosis by utilizing machine learning and intuitive user interfaces. We are getting closer to the ultimate aim of lessening the terrible effects of Alzheimer's disease on people, families, and society at large as researchers continue to hone and validate these strategies.

Alzheimer's disease (AD) is a major global health concern that affects millions of people and their families. This neurological condition gradually deteriorates behavior, memory, and cognitive abilities, which eventually results in a reduction in the person's capacity to carry out daily duties and preserve independence.

The prevalence of AD is predicted to rise with the aging population, making early detection essential for efficient management and intervention. In order to improve AD early diagnosis and monitoring, research in recent years has concentrated on utilizing cutting-edge technology as EEG (Electroencephalography) biomarkers.

EEG data offers useful information about the electrical activity of the brain and is a non-invasive, economical way to evaluate cognitive performance and find anomalies linked to AD. Researchers can find certain indicators of AD pathology, such as changes in brain oscillations, connection patterns, and event-related potentials, by examining EEG recordings.

Furthermore, alterations in the normalized whole brain volume and peripheral nerve responses are crucial medical markers for the diagnosis and tracking of Alzheimer's disease. AD detection systems can be made more accurate and dependable by incorporating these biomarkers and medical characteristics into machine learning algorithms like the Random Forest Classifier. Comparing the Random Forest Classifying to other models of a similar nature, it performs better because to its remarkable accuracy rate of 97% and its capacity to handle high-dimensional data and capture complicated correlations. Additionally, the system is an invaluable resource for researchers and clinicians alike because of the construction of a dynamic user interface utilizing Streamlit, which improves accessibility and encourages user involvement.

To sum up, combining EEG biomarkers, medical data, and machine learning algorithms is a viable way to enhance Alzheimer's disease early detection and treatment, which will eventually benefit patients and caregivers.

1.3 Detection of Alzheimer's Disease using EEG Biomarkers

Alzheimer's disease (AD) poses a significant health challenge globally, affecting millions of individuals and their families. This neurodegenerative disorder progressively impairs cognitive functions, memory, and behavior, ultimately leading to a decline in the individual's ability to perform daily tasks and maintain independence. As the aging population continues to grow, the prevalence of AD is expected to increase, making early detection crucial for effective intervention and management. In recent years, research has focused on leveraging advanced technologies like EEG (Electroencephalography) biomarkers to enhance early detection and monitoring of AD. EEG data provides valuable insights into the brain's electrical activity, offering a non-invasive and cost-effective means of assessing cognitive function and detecting abnormalities associated with AD. By analyzing EEG signals, researchers can identify specific biomarkers indicative of AD pathology, such as alterations in neural oscillations, connectivity patterns, and event-related potentials.

Additionally, peripheral nervous responses and abnormalities in normalized whole brain volume serve as important medical parameters for diagnosing and monitoring AD progression. Integrating these biomarkers and medical parameters into machine learning algorithms, such as the Random Forest Classifier, enhances the accuracy and reliability of AD detection systems. The Random Forest Classifier, with its ability to handle high-dimensional data and capture complex relationships, demonstrates superior performance compared to other similar models, achieving an impressive accuracy rate of 97%.

Moreover, the development of a responsive user interface using Streamlit facilitates user interaction and enhances accessibility, making the system a valuable tool for clinicians and researchers alike. In conclusion, the integration of EEG biomarkers, medical parameters, and machine learning algorithms offers a promising approach to improve early detection and management of Alzheimer's disease, ultimately contributing to better outcomes for patients and caregivers.

1.4 Scope

The scope of the "Alzheimer Disease Detection System using EEG Biomarkers" project is specifically focused on utilizing EEG biomarkers to detect early signs of Alzheimer's Disease across diverse populations. This involves the identification and analysis of key EEG features such as frequency bands (delta, theta, alpha, beta), coherence, connectivity patterns, and other neurophysiological indicators that are predictive of Alzheimer's. The target population for this system includes elderly individuals and those at risk of developing Alzheimer's, spanning various demographics and stages of cognitive decline, from mild cognitive impairment to more advanced stages of the disease. The research aims to refine these biomarkers and integrate them with advanced machine learning models to develop a robust diagnostic tool that can be used reliably in clinical settings.

Acknowledgment of Limitations

Despite the promising scope and potential of this research, there are several limitations that must be acknowledged. One of the primary constraints is data availability. Obtaining large, high-quality datasets of EEG recordings from individuals with Alzheimer's and healthy controls can be challenging. These datasets must be sufficiently varied and representative to train effective machine learning models, and privacy concerns or logistical issues often complicate data collection efforts. Additionally, EEG data itself can be noisy and subject to interference, requiring sophisticated preprocessing techniques to ensure reliability and accuracy in the data used for training and testing the diagnostic models.

Technological Constraints

Technological constraints also play a significant role in limiting the scope of the research. While EEG technology is less invasive and more cost-effective compared to other neuroimaging tools, it typically provides lower spatial resolution, which can limit the granularity of the data collected and potentially affect the precision of the disease detection capabilities. Furthermore, the integration of complex machine learning algorithms into user-friendly software that can be widely adopted in clinical settings poses additional challenges. Ensuring that the system is accessible, easy to use, and integrates seamlessly with existing medical technologies requires ongoing development and optimization, which are contingent on the availability of resources and technological advancements.

By using electroencephalogram (EEG) biomarkers to identify distinct patterns in EEG signals at various stages of the disease, from mild cognitive impairment to severe dementia, the "Alzheimer Disease Detection System using EEG Biomarkers" project aims to develop a sophisticated tool for the early detection and monitoring of Alzheimer's disease (AD). The research and development phase of the project consists of gathering large amounts of EEG data from both healthy controls and Alzheimer's patients. This is followed by feature extraction, which identifies important EEG features, such as changes in power spectra, coherence, and connectivity patterns, that correlate with Alzheimer's pathology. Then, to categorize people according to their EEG profiles, machine learning models such as decision trees, support vector machines, and neural networks are created and trained.

A user-friendly software program is created during the system design and implementation phase to process EEG data, apply these models, and offer diagnostic outputs. To guarantee accuracy and resilience, many predictive models are integrated. In order to examine the system's usefulness and accuracy, neurologists and other medical professionals work together to put it through rigorous validation and testing. Performance parameters including accuracy, sensitivity, specificity, and area under the ROC curve are measured. The goal of user interface design is to provide a user-friendly interface that makes it simple for medical professionals to enter EEG data and understand the system's predictions. Feedback mechanisms are also incorporated to improve learning algorithms and predictions.

By making sure that all patient data is handled in accordance with privacy laws and obtaining the required regulatory permissions, compliance and ethics are addressed. Deployment entails forming alliances with medical facilities for in-hospital trials and giving medical professionals ongoing training and assistance.

In the future, the system will be improved to identify additional neurological conditions using comparable biomarkers and updated with the most recent findings and developments in machine learning and neuroimaging technology.

With the help of this initiative, doctors will be able to diagnose Alzheimer's disease much more accurately, which will improve treatment outcomes and quality of life for people with the disease and provide the groundwork for individualized care.

Alzheimer's disease (AD), an irreversible neurodegenerative disorder, poses a significant health burden globally, particularly among the elderly population. As individuals age, the probability of developing AD increases substantially, with advanced age being the most significant risk factor for the disease. According to global statistical data, the prevalence of AD rises exponentially with each decade of life, with estimates indicating that approximately one in ten individuals aged 65 and older have AD, and the prevalence doubles every five years thereafter. Moreover, by the age of 85, the risk of developing AD reaches nearly one-third. This age-related increase in AD risk can be attributed to various factors, including biological, genetic, and environmental influences. With advancing age, physiological changes occur in the brain, such as the accumulation of beta-amyloid plaques and tau tangles, which are hallmark pathological features of AD.

These alterations disrupt neuronal communication, leading to cognitive decline and memory impairment characteristic of the disease. Additionally, genetic predispositions, such as variations in the APOE gene, play a role in determining an individual's susceptibility to AD, with certain alleles associated with an increased risk of developing the disease later in life. Furthermore, environmental factors, including lifestyle choices, socioeconomic status, and comorbid health conditions, contribute to the overall risk profile for AD. Lifestyle factors such as diet, exercise, cognitive stimulation, and social engagement have been shown to influence cognitive health and may mitigate the risk of developing AD to some extent. However, despite advancements in our understanding of the etiology and risk factors associated with AD, the exact mechanisms underlying age-related neurodegeneration remain incompletely understood.

Therefore, ongoing research efforts aimed at elucidating the complex interplay between aging, genetics, and environmental factors are essential for developing effective strategies for early detection, prevention, and treatment of AD. Furthermore, addressing the global burden of AD requires comprehensive public health initiatives focused on promoting healthy aging, raising awareness, and enhancing access to diagnostic and therapeutic interventions.

By prioritizing research, education, and healthcare infrastructure, we can mitigate the impact of AD on individuals, families, and society at large, ensuring a better quality of life for aging populations worldwide.

1.5 Random Forest

Using Random Forest as the main machine learning algorithm, the "Alzheimer Disease Detection System using EEG Biomarkers" must be implemented in a methodical manner. This method begins with the collection of large amounts of EEG data from both healthy controls and Alzheimer's patients, with a focus on different EEG features such as frequency bands and connectivity measures. Data preprocessing, which includes feature normalization, noise reduction, and segmenting EEG recordings into manageable chunks, comes next. The most important traits suggestive of Alzheimer's illness are then found using feature extraction methods such as PCA or ICA. After this preprocessed and feature-extracted data is used, the decision tree model is trained by applying the decision tree algorithm, dividing the data into training and testing sets, and fine-tuning the model to reduce overfitting and improve generalizability.

Metrics like accuracy, sensitivity, and specificity are used to validate the model's performance. The learned model is included into an intuitive software program that lets medical professionals enter fresh EEG data and get diagnostic predictions. This process facilitates ongoing learning to enhance and optimize the predictions. In order to improve prediction accuracy and stability, ensemble techniques like Random Forests are investigated and the decision tree model is updated on a regular basis with fresh data. This all-encompassing strategy seeks to offer a strong instrument for the early detection and tracking of Alzheimer's disease, supporting medical professionals in customizing treatment plans and efficiently tracking the illness's advancement. Detecting Alzheimer's Disease (AD) using EEG (Electroencephalography) biomarkers represents a cutting-edge approach in neurology with profound implications for early diagnosis and intervention.

Alzheimer's, a progressive neurodegenerative disorder, poses a significant global health challenge, affecting millions worldwide and straining healthcare systems and families alike. As the global population ages, the prevalence of AD continues to rise, underscoring the urgency of developing accurate and efficient diagnostic tools. EEG, a non-invasive technique measuring electrical activity in the brain, has emerged as a promising avenue due to its sensitivity to subtle changes in neural dynamics associated with AD pathology.

The utilization of EEG biomarkers offers several advantages, including accessibility, affordability, and the ability to capture real-time brain activity. One of the most compelling aspects of EEG-based detection is its potential to identify AD-related abnormalities long before clinical symptoms manifest, enabling early intervention when treatments may be more effective. The integration of EEG biomarkers into AD detection methodologies involves sophisticated

signal processing techniques and advanced machine learning algorithms. EEG data analysis typically encompasses the extraction of features from raw signals, such as frequency bands, coherence measures, and event-related potentials (ERPs). These features serve as input to machine learning models trained to differentiate between AD patients and healthy controls. Leveraging large-scale datasets encompassing diverse demographics and clinical characteristics enhances the generalizability and robustness of these models. Global statistical relevance to the dataset ensures that findings are applicable across populations, accounting for variations in genetic predisposition, environmental factors, and comorbidities.

Studies have demonstrated the utility of EEG biomarkers in detecting AD with high sensitivity and specificity. For instance, alterations in neural oscillations, particularly in the theta and delta frequency bands, have been consistently observed in individuals with AD compared to age-matched controls. Additionally, disruptions in functional connectivity networks, reflecting impaired information processing and communication between brain regions, serve as robust indicators of disease progression. Event-related potentials, such as the P300 component elicited during cognitive tasks, exhibit characteristic abnormalities in AD patients, reflecting deficits in attention, memory, and executive function.

Moreover, the incorporation of complementary medical parameters, including peripheral nervous responses and normalized whole-brain volume, enriches the diagnostic capabilities of EEG-based systems. Peripheral nervous responses, such as changes in autonomic function and sensory processing, offer valuable insights into the systemic effects of AD pathology beyond the central nervous system. Meanwhile, structural neuroimaging measures, such as volumetric analysis of brain regions implicated in AD, provide complementary information regarding disease severity and progression. The synergy between EEG biomarkers, medical parameters, and advanced machine learning algorithms enhances the accuracy, reliability, and clinical utility of AD detection systems.

Among machine learning techniques, the Random Forest Classifier stands out for its ability to handle high-dimensional data, capture complex nonlinear relationships, and mitigate overfitting. Its ensemble learning framework, comprising multiple decision trees trained on different subsets of the data, effectively reduces variance and improves generalization performance. The Random Forest Classifier's remarkable accuracy rate of 97% surpasses that of other similar models, underscoring its efficacy in distinguishing between AD patients and healthy individuals. In addition to diagnostic accuracy, the usability and accessibility of AD detection systems play a crucial role in their adoption and impact.

The development of responsive user interfaces using platforms like Streamlit facilitates seamless interaction with the system, empowering clinicians, researchers, and caregivers to leverage its capabilities effectively. Streamlit's intuitive design, coupled with interactive visualization tools, enhances user engagement and promotes collaboration across multidisciplinary teams. Furthermore, the integration of cloud-based solutions enables remote access to EEG data, fostering scalability and facilitating real-time monitoring of patients' cognitive health.

Overall, the detection of Alzheimer's Disease using EEG biomarkers represents a transformative paradigm in neurology, offering unprecedented opportunities for early diagnosis, personalized treatment, and improved patient outcomes. By harnessing the synergistic power of EEG data, medical parameters, and advanced machine learning algorithms, researchers and clinicians can accelerate the space of discovery and innovation in Alzheimer's research, ultimately advancing our understanding of the disease and paving the way for more effective therapeutic interventions.

Relevance of Machine Learning based analysis of EEG data in Medical Realm, Machine learning-based analysis of EEG data holds immense relevance in the medical realm, offering a powerful tool for diagnosing, monitoring, and understanding various neurological disorders. With the global prevalence of neurological conditions on the rise, including epilepsy, Alzheimer's disease, Parkinson's disease, and others, there is an urgent need for accurate and efficient diagnostic methods to improve patient outcomes. EEG data, which captures the electrical activity of the brain, provides rich information about brain function and dysfunction, making it a valuable resource for medical research and clinical practice. Through advanced machine learning techniques, such as deep learning algorithms and ensemble methods like Random Forest, EEG data can be analyzed to extract meaningful patterns, biomarkers, and predictive models for different neurological conditions.

The relevance of machine learning-based EEG analysis becomes evident when considering the vast amounts of data generated by EEG recordings and the complexity of interpreting this data manually. EEG signals are inherently noisy and non-stationary, presenting challenges for traditional analytical methods. However, machine learning algorithms excel at handling high-dimensional, noisy data and can identify subtle patterns that may elude human observers. By training models on large datasets comprising EEG recordings from diverse patient populations, researchers can develop robust algorithms capable of detecting abnormalities, predicting disease progression, and personalizing treatment strategies.

One notable example of the relevance of machine learning in EEG analysis is its application in epilepsy diagnosis and seizure prediction. Epilepsy affects over 50 million people worldwide, yet diagnosis can be challenging due to the variability of symptoms and the need for long-term monitoring. Machine learning algorithms trained on EEG data have shown remarkable success in automating the detection of epileptiform activity and distinguishing between epileptic and non-epileptic seizures. These algorithms not only facilitate earlier diagnosis but also enable real-time seizure prediction, empowering patients to take proactive measures to prevent seizure episodes and improve their quality of life.

Similarly, in the context of Alzheimer's disease (AD), machine learning-based analysis of EEG data offers promising avenues for early detection and monitoring of disease progression. AD is characterized by progressive cognitive decline and neurodegeneration, making early intervention crucial for preserving cognitive function and improving patient outcomes.

EEG biomarkers, such as alterations in spectral power, connectivity patterns, and event-related potentials, can serve as sensitive indicators of underlying pathology in AD. Machine learning algorithms trained on EEG features extracted from AD patients and healthy controls can learn to differentiate between the two groups with high accuracy, facilitating early diagnosis and potentially enabling interventions to slow disease progression.

The statistical relevance of machine learning-based EEG analysis is further underscored by the performance metrics achieved by these models. For instance, the Random Forest Classifier, a popular ensemble learning algorithm, has demonstrated impressive accuracy rates exceeding 97% in classifying EEG data from patients with neurological disorders. This high level of accuracy translates to increased confidence in the diagnostic capabilities of machine learning models and enhances their clinical utility.

Moreover, the development of user-friendly interfaces, such as the Streamlit UI framework, streamlines the deployment of machine learning-based EEG analysis tools, making them accessible to clinicians, researchers, and patients alike.

In addition to diagnosis and monitoring, machine learning-based EEG analysis holds promise for advancing our understanding of brain function and dysfunction. By leveraging large-scale EEG datasets in combination with advanced analytics techniques, researchers can uncover novel biomarkers, identify disease subtypes, and elucidate underlying mechanisms of neurological disorders. Furthermore, machine learning algorithms can facilitate the integration of multi-modal data, such as EEG, neuroimaging, and genetic information, to provide a more comprehensive understanding of brain health and disease.

Despite the remarkable progress achieved in machine learning-based EEG analysis, several challenges remain to be addressed. These include the need for standardized data collection protocols, robust validation frameworks, and interpretability of model predictions. Additionally, efforts to ensure the ethical use of EEG data, including privacy protection and informed consent, are essential to maintain public trust and support the responsible advancement of this field. machine learning-based analysis of EEG data holds immense relevance in the medical realm, offering a powerful approach for diagnosing, monitoring, and understanding neurological disorders. Through advanced algorithms, such as the Random Forest Classifier, and user-friendly interfaces like Streamlit, EEG analysis tools can achieve high levels of accuracy and accessibility, empowering clinicians, researchers, and patients to make informed decisions and improve patient outcomes. As technology continues to evolve and datasets grow larger, the potential for machine learning to transform neurology and neuroscience remains vast, promising new insights and opportunities for innovation in the years to come.

In conclusion, the opening chapter, in its entirety, establishes the framework for the study "Detection of Alzheimer Disease using EEG Biomarkers." The problem statement emphasizes how important it is to identify Alzheimer's disease accurately and early because of the disease's significant effects on both people and society as a whole. The work intends to solve this urgent challenge by utilizing modern machine learning techniques, specifically decision trees, with the goal of establishing a viable Alzheimer's diagnosis system using EEG biomarkers. The target population and particular EEG biomarkers are the focus of the defined scope of the study, which also acknowledges potential limits like data availability and technological limitations.

Overall, by using cutting-edge EEG-based techniques, this study seeks to improve patient treatment and diagnosis of Alzheimer's disease, ultimately leading to better outcomes for those who suffer from this crippling neurodegenerative condition.

CHAPTER 2

LITERATURE OVERVIEW ON ALZHEIMER'S EEG

This review of the literature intends to examine the present status of research on EEG biomarkers for AD identification, with an emphasis on recent studies that have looked into the viability and efficacy of this strategy. This work aims to shed light on the potential of EEG biomarkers in enhancing the early detection and treatment of Alzheimer's disease by a thorough analysis of pertinent literature. Alzheimer's disease (AD) is a neurodegenerative condition that is difficult to diagnose and detect in its early stages, because of its progressive nature and the challenges involved in early detection, Alzheimer's disease (AD) poses a serious challenge to the healthcare system. The incidence of AD is increasing with population aging, underscoring the pressing need for efficient diagnostic techniques.

2.1 Literature Review

The use of electroencephalography (EEG) biomarkers as a possible method for AD identification has gained traction in recent years. EEG is a desirable choice for early detection and tracking the course of AD since it provides a non-invasive and affordable way to monitor brain activity. The application of electroencephalography (EEG) biomarkers has become a viable method to help identify AD. In order to bridge preclinical and clinical programs, Leiser et al. (2010) emphasized the significance of choosing and validating EEG endpoints as reliable and transferable biomarkers. In order to examine the possibility of using compressive sampling and sensing systems to distinguish between various brain states, Morabito et al. (2013) looked at the increased compressibility of EEG signals in AD patients.

According to Tonello et al. (2016), screen-printed biosensors have also been investigated for the early identification of biomarkers associated with AD. The objective of this research was to build a portable point-of-care testing system utilizing screen-printed electrochemical sensors, in order to address the shortcomings of conventional approaches. Simpraga et al. (2017) have demonstrated the application of machine learning approaches to EEG records for precise identification of cholinergic intervention and AD. In order to identify the brain's signature of illness or pharmacological intervention, the study created an index of EEG biomarkers. Al-Nuaimi et al. (2017) used the k-nearest neighbor algorithm for classification as they looked at the possibility of the Higuchi fractal dimension of EEG as a biomarker for early detection of AD.

As highlighted by Al-Nuaimi et al. (2018), complexity metrics derived from EEG frequency bands have also demonstrated promise in early AD diagnosis.

These measurements have improved the ability to quantify variations in EEG related. The study employed EEG feature analysis in various brain regions and frequency bands with the goal of distinguishing between aged normal control volunteers and potential AD patients. For a better and more comprehensible machine learning-based AD diagnosis, Lopes et al. (2023) suggested combining saliency maps, EEG modulation spectra, and deep neural networks—particularly convolutional neural networks—with these data. The literature review as a whole emphasizes the increasing interest in using EEG biomarkers for AD diagnosis and detection.

Several research works have investigated various EEG analytic methods, machine learning strategies, and biomarker algorithms to improve the precision and comprehensibility of AD diagnosis systems. The study employed EEG feature analysis in various brain regions and frequency bands with the goal of distinguishing between aged normal control volunteers and potential AD patients. For a better and more comprehensible machine learning-based AD diagnosis, Lopes et al. (2023) suggested combining saliency maps, EEG modulation spectra, and deep neural networks—particularly convolutional neural networks—with these data. The literature review as a whole emphasizes the increasing interest in using EEG biomarkers for AD diagnosis and detection. Several research works have investigated various EEG analytic methods, machine learning strategies, and biomarker algorithms to improve the precision and comprehensibility of AD diagnosis systems, consent, disclosure, or failure to disclose information from biomarker results, particularly in the context of advancing genetic research and diagnostic criteria. The authors draw attention to the widening gap between the advancements in biomarker research and the restricted treatment options that doctors and patients have access to.

The ethical and practical issues surrounding amyloid imaging in Alzheimer's disease research are covered by Roberts et al. (2013). The authors stress the significance of figuring out how to utilize this technology appropriately, weighing the advantages and disadvantages of sharing data, and having effective conversations about testing with patients and family members. The longitudinal alterations in biomarkers in families with autosomal-dominant Alzheimer disease mutations are the main topic of Fagan et al.'s (2014) research.

In order to improve understanding of the disease and guide the design of clinical trials that use biomarkers for subject recruitment and outcome measures, the study intends to define biomarker changes throughout the disease process. The body of research highlights how crucial ethical issues are when using biomarkers for neurodegenerative illnesses, including Alzheimer's.

The use of cutting-edge technologies like amyloid imaging raises ethical questions about consent, transparency, and use that emphasize the importance of rigorous deliberation and intelligent decision-making in both clinical and research settings. The use of placebo controls in clinical trials including Alzheimer's disease is a contentious practice. Several researchers have emphasized the need to carefully consider the ethical issues surrounding the use of placebo controls in these experiments (Kawas et al., **1999**). Two examples of research that have used placebo-controlled trials include those that examine the effects of Rofecoxib or Naproxen with placebo on the course of Alzheimer's disease (Aisen et al., **2003**) and a 52-week randomized double-blind study of Ginkgo Biloba for dementia (Bars et al., **1997**). Concerns about the risk of death associated with particular drug treatments have driven the discussion concerning the use of placebo controls in Alzheimer's disease trials, according to a meta-analysis of randomized placebo-controlled trials of atypical medicines.

Moreover, due to cardiovascular risks, a trial evaluating the use of traditional nonsteroidal anti-inflammatory medicines (tNSAIDs) in Alzheimer patients was stopped, raising questions about the safety of these treatments (Maillard et al., **2006**). Studies on Alzheimer's disease may be more accurate and effective if new methods for clinical trial design, such as Bayesian adaptive trials, are used (Cummings et al., **2012**). Alternative study designs, such the placebo group simulation approach (PGSA) have also been offered as a way to anticipate outcomes in presymptomatic AD patients without requiring extensive placebo-controlled trials (Spiegel et al., **2011**). Generally speaking, the debate over the use of placebo controls in Alzheimer's disease clinical trials is still evolving as scientists try to reconcile the need for compelling scientific data with moral considerations.

Together, this research advance our knowledge of AD therapy and clinical trial planning. Nevertheless, more investigation is required to create efficient treatments for this intricate neurodegenerative illness. "Issues in Clinical Trial Design I," by R. L. Sufit A Placebo-Controlled, Double-blind, Randomized Trial of An Extract of Ginkgo Biloba for Dementia was published in *Neurology* in **1996** by P. L. Le Bars et al. *JAMA*, **1997**: "Clinical Trials in Alzheimer Disease: Debate on the Use of Placebo Controls," by C. H.

Kawas et al. Aisen, P. S. et al., "Effects of Rofecoxib Or Naproxen Vs Placebo on Alzheimer Disease Progression: A Randomized Controlled Trial," *Archives of Neurology*, **2000**. N. C. Fox et al., "Using Serial Registered Brain Magnetic Resonance Imaging to Measure Disease Progression in Alzheimer Disease: Power Calculations and Estimates of Sample Size to Detect Treatment Effects," *Alzheimer Disease and Associated Disorders*, **1999**. "Risk of Death with Atypical Antipsychotic Drug Treatment for Dementia: Meta-analysis of Randomized Placebo-controlled Trials," L. S. Schneider et al., *JAMA*, **2003**.

A comparison of the cardiovascular safety of traditional nonsteroidal anti-inflammatory drugs and ginkgo biloba and donepezil in the treatment of Alzheimer's dementia in a randomized placebo-controlled double-blind study was published in the *European Journal of Neurology* in **2006**. R. Spiegel et al. debated the topic in *Alzheimer's Research & Therapy* in **2011**. Together, these investigations advance our knowledge of trial design, safety, and AD treatment. Nevertheless, more investigation is required to create efficient treatments for this intricate neurodegenerative illness. The article "Risk of Death with Atypical Antipsychotic Drug Treatment for Dementia: Meta-analysis of Randomized Placebo-controlled Trials" was written by L. S. Schneider and colleagues. The study conducted by Maillard and Burnier (**2005**) was published in *JAMA*. Mazza et al. (**2006**) presented their findings in an expert opinion on drug safety. Ginkgo biloba and donepezil were compared in a randomized placebo-controlled double-blind study to treat Alzheimer's dementia.

The following publications were published in **2011**: *Alzheimer's Research & Therapy* (R. Spiegel et al.), *European Journal of Neurology* (**2006**), J. Cummings et al., "Advances in Designs for Alzheimer's Disease Clinical Trials," *American Journal of Neurodegenerative Disease* (**2012**).

Improvements in Deep Learning-Based Alzheimer's Disease Classification, a growing corpus of work has improved the prediction and categorization of Alzheimer's disease phases by utilizing deep learning. The problem of classifying illness stages using CNNs was addressed in one noteworthy article, "Alzheimer's Disease Stage Classification using Deep Convolutional Neural Networks on Oversampled Imbalance Data," which concentrated on addressing imbalanced datasets by oversampling minority classes. When it came to using brain MRI pictures to identify different stages of Alzheimer's disease, this strategy produced encouraging results. In "AlzheimerNet: An Effective Deep Learning Based Approach for Alzheimer's Disease Classification," another noteworthy development was presented.

The researchers created a specialized deep learning model called AlzheimerNet, which proved to be more effective than conventional techniques through extensive testing on MRI data and ensured accurate disease predictions. Furthermore, the effectiveness of pretrained networks like VGG, ResNet, and DenseNet was investigated in "A Comparative Study of Pretrained Deep Neural Networks for Classifying Alzheimer's and Parkinson's Disease."

This study demonstrated how transfer learning can improve the precision of early neurodegenerative disease identification. Finally, the study "Alzheimer's Disease Detection through Deep Learning Techniques: A Study" highlighted the importance of CNNs in the timely and precise diagnosis of Alzheimer's disease, highlighting deep learning's ability to successfully identify complex disease patterns.

Together, this research highlight how deep learning technology can improve our knowledge of Alzheimer's disease and its treatment, opening the door to more focused and efficient approaches. Recent developments in the use of deep learning and machine learning techniques have demonstrated exceptional promise for the early detection and categorization of Alzheimer's disease (AD). In a systematic review titled "A systematic review on machine learning and deep learning techniques in the effective diagnosis of Alzheimer's disease," a number of publications that use these techniques to differentiate between Alzheimer's disease and normal cognitive functions using PET and MRI imaging modalities are critically evaluated.

The review assesses how well different classifiers perform in correctly diagnosing AD. The work "Machine Learning and Novel Biomarkers for the Diagnosis of Alzheimer's Disease" provides a full overview of another noteworthy addition to this topic. It examines various machine learning techniques, such as support vector machine (SVM), logistic regression, random forest, and naïve Bayes. These algorithms play a key role in developing predictive models that successfully distinguish AD patients from healthy controls, allowing for earlier diagnosis and treatment. In addition, the paper "Identifying Alzheimer Disease Dementia Levels Using Machine Learning and Deep Learning Models" presents a combined method for classifying the four progressive stages of dementia that makes use of SVM, random forests, and convolutional neural networks (CNN). The encouraging results of these research demonstrate how machine learning and deep learning models can improve our understanding of Alzheimer's disease and help design customized treatment plans based on accurate diagnosis and staging.

An overview of the body of research on AD detection techniques, the research now available on the identification of Alzheimer's disease (AD) covers a wide range of approaches, from biomarker analysis to advanced imaging techniques and clinical assessments. Conventional techniques involve cognitive assessments like the Clinical Dementia Rating (CDR) and the Mini-Mental State Examination (MMSE), which evaluate functional abilities and cognitive impairment. Imaging methods such as Positron Emission Tomography (PET) and Magnetic Resonance Imaging (MRI) are widely utilized to examine changes in the brain's metabolism and morphology that are indicative of Alzheimer's disease. Biochemical indicators, such as tau protein and beta-amyloid levels in blood or cerebrospinal fluid (CSF), are also essential for making a diagnosis. Although these techniques are helpful, they frequently have drawbacks such being expensive, intrusive, and requiring specialized equipment and interpretation by professionals.

The use of less intrusive and more affordable techniques, like EEG, has been more and more highlighted in recent literature. These approaches may be able to get around some of these drawbacks while still offering insightful diagnostic information.

Talk about EEG-Based Methods and Their Advantages. the utilization of EEG-based methods to identify Alzheimer's disease has become popular because of its non-invasiveness and the distinctive perspectives it provides on brain activity. Brain wave anomalies that are suggestive of the cognitive decline common to Alzheimer's disease can be identified by an EEG, which monitors the electrical activity of the brain. EEG has several advantages, including the potential to be used as an early diagnosis tool because of its capacity to identify changes in brain activity well in advance of the observable symptoms. In addition, EEG equipment are often less expensive and more accessible than MRI or PET scanners, which qualifies them for clinical application in a wide range of settings, including real-time monitoring of therapy response and illness progression.

Research has demonstrated that certain EEG patterns, such as elevated power in the theta and delta bands and reduced power in the beta band, are connected to Alzheimer's disease (AD), indicating that EEG can be a valid diagnostic and monitoring tool for the disease. It is also a promising field for ongoing research and therapeutic application since continuous long-term EEG recording provides an unmatched advantage in investigating the long-term evolution of neurological abnormalities associated with Alzheimer's disease.

An overview of the body of research on AD detection techniques, the research now available on the identification of Alzheimer's disease (AD) covers a wide range of approaches, from biomarker analysis to advanced imaging techniques and clinical assessments. Conventional techniques involve cognitive assessments like the Clinical Dementia Rating (CDR) and the Mini-Mental State Examination (MMSE), which evaluate functional abilities and cognitive impairment. Imaging methods such as Positron Emission Tomography (PET) and Magnetic Resonance Imaging (MRI) are widely utilized to examine changes in the brain's metabolism and morphology that are indicative of Alzheimer's disease. Biochemical indicators, such as tau protein and beta-amyloid levels in blood or cerebrospinal fluid (CSF), are also essential for making a diagnosis. Although these techniques are helpful, they frequently have drawbacks such being expensive, intrusive, and requiring specialized equipment and interpretation by professionals. The use of less intrusive and more affordable techniques, like EEG, has been more and more highlighted in recent literature. These approaches may be able to get around some of these drawbacks while still offering insightful diagnostic information.

Talk about EEG-Based Methods and Their Advantages, the utilization of EEG-based methods to identify Alzheimer's disease has become popular because of its non-invasiveness and the distinctive perspectives it provides on brain activity. Brain wave anomalies that are suggestive of the cognitive decline common to Alzheimer's disease can be identified by an EEG, which monitors the electrical activity of the brain. EEG has several advantages, including the potential to be used as an early diagnosis tool because of its capacity to identify changes in brain activity well in advance of the observable symptoms. In addition, EEG equipment are often less expensive and more accessible than MRI or PET scanners, which qualifies them for clinical application in a wide range of settings, including real-time monitoring of therapy response and illness progression. Research has demonstrated that certain EEG patterns, such as elevated power in the theta and delta bands and reduced power in the beta band, are connected to Alzheimer's disease (AD), indicating that EEG can be a valid diagnostic and monitoring tool for the disease. It is also a promising field for ongoing research and therapeutic application since continuous long-term EEG recording provides an unmatched advantage in investigating the long-term evolution of neurological abnormalities associated with Alzheimer's disease.

Safdar Sardar Khan, Sunil Patil, et al. explore the revolutionary potential of machine learning techniques in the diagnosis of Alzheimer's disease (AD) in their **2023** study, moving away from conventional statistical methods and toward sophisticated algorithms such as ensemble approaches and deep learning. They address the issues of data accessibility, bias, the requirement for standardized datasets and assessment measures, and the significance of feature selection and extraction for improving model accuracy and interpretability. Notwithstanding these gains, there are still a few research gaps. The

Gaussian Naïve Bayes algorithm is a serious issue since it achieves only 78% accuracy, which is 2% below the clinical acceptance rate. This indicates that algorithmic accuracy needs to be improved. In addition, there is a continuous need to investigate more sophisticated feature extraction approaches and integrate multi-modal data sources in order to reduce bias and improve generalizability across various populations.

Additionally, research aimed at promoting the use of additional open-access resources is necessary given the restricted availability of high-quality, diversified datasets. In order to enable consistent comparisons between studies, evaluation measures must be standardized. Additionally, as these technologies are integrated into clinical processes, their operational, ethical, and practical consequences must be examined.

Rahul Sharma, Tripti Goel, et al. explore the use of MRI imaging to identify Alzheimer's and pre-Alzheimer's syndromes in their **2023** paper. They concentrate on the T1 parameter. They explain a continuous procedure in which certain researchers gather data and send it to the Medical Association of India (MAI) for additional examination. The paper highlights a major technical issue in the methodology: in around 70% of cases, white noise introduced by magnetic interference during MRI recordings distorts the images. This interference significantly affects the accuracy of the diagnostic process by producing bulging in areas that represent brain convolutions. This circumstance creates a number of opportunities for further study and advancement to close these gaps. First and foremost, stronger MRI imaging methods or post-processing algorithms are desperately needed in order to improve the clarity and usefulness of the pictures by minimizing or eliminating magnetic interference noise.

This enhancement might be greatly aided by looking at advanced noise-cancellation techniques or alternative image technology. Second, the data variability that could result in diagnostic errors points to the need for machine learning models that are more resilient and flexible in order to retain high accuracy across a variety of datasets and circumstances.

In order to develop such models, more research must be done on generalized or transfer learning frameworks, which are more suited to manage the variety of clinical data that is encountered in real-world settings. Third, considering more affordable imaging options or cost-cutting measures in light of the MRI technique's higher expenses, it might be possible to boost the technology's sustainability and accessibility for wider clinical application.

Ultimately, in order to assess the improved techniques' efficacy and viability and confirm that they satisfy the strict requirements for clinical approval, extensive, extended clinical trials and studies are required. By filling in these research gaps, we may be able to improve patient outcomes and care management techniques by creating more precise, dependable, and affordable diagnostic instruments for Alzheimer's and Pre-Alzheimer's syndromes.

In their research, Younghoon Jeon, Jaeyong Kyang, et al. (2023) take a novel strategy, concentrating on the identification and stage analysis of brain tumors as a means of indirectly identifying Alzheimer's disease. In assessing the use of a decision tree classifier, they show off how much more accurate and computationally efficient it is than other approaches, especially when it comes to jobs like pothole detection—an example used to emphasize the method's diagnostic accuracy. Although the approach achieves a remarkable 94% accuracy across a range of clinical indicators, there are notable obstacles that prevent it from being used practically in a clinical environment.

Their study highlights a number of problems, chief among them being the introduction of mistakes that cause variations in diagnostic results when data veracity rises. This shows that although the decision tree model works well in controlled environments, it has trouble remaining accurate when dealing with noisy or complex data sets. Furthermore, the technique uses expensive and technically complex procedures like image de-ionization to eliminate white noise from MRI scans. Despite the method's great diagnostic potential, these factors add up to a significant operational cost that makes it economically untenable for long-term clinical use.

In their most recent publication, Meenu Gupta et al. (2024) provide a thorough review of Alzheimer's Disease (AD), addressing a wide range of subjects such as the pathophysiology of the illness, risk factors, clinical presentation, diagnosis, treatment, and the state of research at the moment. Their research is primarily focused on reducing symptoms and enhancing the quality of life for those who have Alzheimer's disease.

2.2 Research Gap

In spite of encouraging advancements, the profession still needs to fill in some significant gaps in the literature:

Standardization of Methodologies: The definition of biomarkers, data analysis techniques, and EEG recording protocols are not standardized amongst studies. This variation can cause inconsistent results and make it harder to replicate findings, which makes it more difficult to construct diagnostic criteria that are widely accepted.

Studies that are longitudinal: The majority of research are cross-sectional, offering only a momentary view of EEG characteristics. In order to better understand how AD progresses and the predictive value of particular EEG biomarkers, longitudinal studies are required to monitor EEG changes over time in individual individuals.

Integration with Other Biomarkers: Although EEG is a non-invasive and economical way to identify AD, combining EEG biomarkers with other biological markers (like genetic, imaging, and biochemical markers) may increase the precision of the diagnosis and give a more thorough knowledge of the condition.

Clinical Validation: The application of research results to clinical practice is lacking. Many EEG characteristics that have been suggested as biomarkers have not received enough clinical validation. To verify the efficacy of EEG biomarkers and develop guidelines for their application in standard clinical diagnostic procedures, rigorous, large-scale clinical trials are required.

Technical Difficulties: Despite the progress made, there are still technical difficulties in EEG analysis. These include handling high-dimensional data, eliminating noise and artifacts, and creating real-time monitoring and analysis systems. Future research can close these gaps and increase the feasibility and reliability of employing EEG biomarkers for Alzheimer's disease identification, which will ultimately result in better diagnostic instruments and patient outcomes.

A high degree of procedural integrity and systematic review in their analysis is ensured by the study methodology's rigorous adherence to the DEEP-GAN guidelines, which are standards for Preference Report Item for Procedural Review and Meta-Analysis. Despite this thorough approach, the Gupta et al. study identifies possible research gaps, especially in the areas of developing curative medicines for AD and efficient management measures. First, while treating symptoms is important, further study into the underlying mechanisms is still needed in order to develop more effective treatments beyond palliative care. Gaining insight into these fundamental biological processes may make it possible to alter or stop the disease's course. Additionally, there is still a large gap in the application of research findings to clinical practice. Clinical research findings and useful, patient-accessible treatments frequently diverge.

Patient outcomes could be considerably improved by investigating how these discoveries can be more effectively incorporated into therapy regimens. Furthermore, while the study places a strong emphasis on recent findings, it appears that little has been done to investigate cutting-edge methods and interdisciplinary approaches that have the potential to completely transform Alzheimer's disease detection, monitoring, and therapy. To improve diagnosis and treatment options, more research might be done in areas like artificial intelligence, big data, and novel biomarkers.

Finally, even though the DEEP-GAN norms offer a strong foundation for review and meta-analysis, further research is needed to determine how well these norms actually capture new and pertinent findings, particularly in an area of study that is developing quickly like Alzheimer's research. Research evaluating these norms' adaptability and inclusivity may shed light on how to make improvements so that all noteworthy and potentially ground-breaking findings are taken into account for systematic reviews. Closing these gaps might improve our knowledge of Alzheimer's disease and pave the way for more comprehensive and successful management and possibly even a cure for this crippling illness.

The disconnect between the promise and viability of technological advancements is a significant obstacle in the field of medical diagnostics, especially when it comes to the use of modern imaging techniques to identify Alzheimer's disease. This disparity has pointed up a number of important areas that require additional study and development in order to close the gap and improve the general efficacy and long-term viability of diagnostic techniques. First and foremost, there is an urgent need for more advanced and affordable medical imaging noise reduction methods.

Existing techniques for reducing noise may be too costly or compromise the pictures' ability to be used for diagnostic purposes. The goal of this field's research could be to create sophisticated algorithms that can improve or preserve data integrity without raising costs. These algorithms might be based on improved image processing methods or novel artificial intelligence models that more effectively reduce noise while maintaining vital diagnostic data.

Secondly, there is a clear need to improve decision tree classifier resilience. One present drawback of these classifiers is that they need to retain high accuracy across realistic and diverse clinical datasets. Research on hybrid models—which mix decision trees with other machine learning algorithms to make up for flaws and capitalize on strengths across approaches—might be able to remedy this.

To improve the models' accuracy and robustness even further, adaptive learning procedures that respond to fresh information or changing clinical environment circumstances should be added.

Thirdly, it's critical to consider the financial impact of introducing new diagnostic tools. To make sure that the implementation of novel diagnostic instruments is financially feasible in the long run and can be integrated into healthcare systems without placing undue financial obligations, comprehensive cost-benefit evaluations are necessary. The possible savings from an earlier and more accurate diagnosis, which could lower the overall cost of patient care, should also be taken into account in these studies in addition to the direct expenses related to the technology themselves.

Closing these research gaps could greatly increase the viability of indirect Alzheimer's disease diagnosis by tumor analysis or other novel approaches using advanced diagnostic tools. Enhancing these approaches' robustness, accuracy, and cost-effectiveness could significantly improve patient outcomes and the effectiveness of healthcare systems, making sophisticated diagnostics more useful and accessible for clinical usage.

An inventive ensemble machine learning modeling approach for decoding brain hypo-fluidics local field potential (LFP) output waves is presented in the paper by Marcos Fabeitti et al. (2023). Potential early-stage biomarkers for Alzheimer's disease (AD) are being investigated for these waves. Their method includes a robust analysis across temporal, geographical, and spectral dimensions of the data and uses LFPs obtained from healthy subjects and two surrogate animal models.

The researchers utilized a late fusion technique in conjunction with their XML model to attain a combined accuracy of 92.4% by validating the feature sets obtained from LFPs through the integration of signals from electroencephalograms, electrocardiograms, and respiratory measurements. This research demonstrates the ability to identify faint patterns in network activity that may be symptomatic of AD as early as three months after the onset of the disease, opening up a possible window of opportunity for early intervention.

Notwithstanding these developments, Fabeitti et al.'s work reveals a number of research gaps that need be investigated further to improve the technology's diagnostic potential and real-world uses. First off, even if the study effectively pinpoints brain abnormalities in AD in its early stages, these results still require confirmation in a larger human sample.

The results may not be as broadly applicable to the human population as they may be due to the use of animal models and the relatively small sample sizes (20 per group), which have different genetic, environmental, and lifestyle factors that influence the course of disease.

Furthermore, using these strategies in clinical settings may provide substantial obstacles due to the intricacy of the modeling and data processing approaches used. Uncertainty exists around how these complex machine learning models might be made simpler or modified to make them workable and effective for regular clinical use without necessitating specialized knowledge.

Furthermore, even though the accuracy rate is impressive, more development may be required to push the envelope and achieve even better accuracy, guaranteeing that the procedure is dependable and fail-safe in a range of clinical and environmental settings. To capture more subtle fluctuations in LFP outputs that might be symptomatic of AD, this could entail adding new biomarkers or improving the ensemble machine learning models.

Finally, there hasn't been enough research done on the financial and logistical effects of using such cutting-edge diagnostic technology in a clinical setting. Cost-benefit studies of incorporating these technologies into the healthcare system require research, taking into account both the technology's direct costs and the possible cost reductions from earlier diagnosis and more focused treatment measures.

The practical applications of LFP analysis in early AD identification could be greatly advanced by filling up these research gaps, making it a more reliable, practical, and widely available tool in the battle against Alzheimer's.

Present Condition of AD Diagnosis:

As of right now, diagnosing Alzheimer's disease (AD) requires a multimodal approach that includes biomarker assessment in addition to clinical assessment, cognitive testing, and neuroimaging to confirm the presence of the disease's pathological characteristics. Even with improvements in diagnostic techniques, it is still difficult to identify AD early and accurately, particularly when the disease is still in its mild cognitive impairment (MCI) stage, which frequently comes before AD.

Clinical Evaluation: An extensive clinical evaluation that includes a comprehensive medical history and an analysis of symptom patterns is usually conducted before an AD diagnosis is made. Medical professionals make use of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) or the National Institute on Aging and the Alzheimer's Association (NIA-AA) recommendations when determining diagnosis.

These assessments are essential for excluding alternative dementia causes and comprehending how the illness advances.

Neuropsychological Testing: Memory, problem-solving, attention span, counting, and language skills are all assessed with cognitive tests like the Mini-Mental State Examination (MMSE) and the Montreal Cognitive Assessment (MoCA). Although useful, these tests cannot be considered conclusive in the diagnosis of AD because a person's educational level, native language, or cultural background may have an impact.

Neuroimaging Methods: To view the structure and function of the brain, imaging methods such as Positron Emission Tomography (PET) scans and Magnetic Resonance Imaging (MRI) are used. While amyloid plaques, a defining feature of Alzheimer's disease, can be found on PET scans, MRIs can reveal brain atrophy or shrinkage, especially in the hippocampus. Nevertheless, these techniques are costly and not available to everyone, which restricts their application to more difficult or unclear diagnostic situations.

Biomarkers: It has been determined that several biomarkers found in cerebrospinal fluid (CSF), such as tau and amyloid-beta proteins, are correlated with AD. However, invasive lumbar punctures are required to get CSF, which may not be possible or appropriate for all individuals. Blood biomarkers are starting to show promise as a less invasive alternative, but further study is required to confirm their accuracy and dependability in standard clinical settings.

Difficulties with Early Detection:

Even with these tools, AD early detection is still quite difficult. Many of the existing techniques are better at diagnosing AD after it has advanced than when it is still in its early stages. The early signs of AD might be difficult to diagnose since they frequently coexist with the symptoms of other types of dementia or normal aging.

In light of these difficulties, there is a rising interest in creating diagnostic instruments that are easier to use, less invasive, and more affordable. One such tool is the use of EEG biomarkers. The "Alzheimer Disease Detection System using EEG Biomarkers" initiative makes use of EEG technology, which has the capacity to identify minute alterations in brain activity that occur prior to the manifestation of overt clinical symptoms. This strategy aims to improve the precision and promptness of AD diagnosis by incorporating cutting-edge machine learning techniques. This could close a significant gap in the present diagnostic picture and enable early intervention tactics

EEG (electroencephalogram) biomarkers have a number of intrinsic benefits that make them especially well-suited for this difficult task, they show great promise for the early diagnosis of Alzheimer's disease (AD). These benefits demonstrate why EEG is becoming more widely acknowledged as a useful diagnostic tool for neurodegenerative illnesses like AD.

1. Sensitivity to Early Neurological Changes: Electroencephalography (EEG) records electrical activity in the brain, providing real-time information on how neurons are operating. Alzheimer's disease causes minor alterations in the electrical pathways of the brain long before symptoms become apparent. These early adaptations, including as modifications to wave patterns, frequency, and connectivity between various brain regions, can be identified by EEG. Reduced alpha and beta activity and increased theta and delta activities, for example, are suggestive of cognitive deterioration. EEG biomarkers are far more promising than many other diagnostic techniques for detecting AD early on because of their extraordinarily sensitive detection capabilities.

2. Non-Invasiveness and Patient Comfort: EEG is a non-invasive diagnostic test, in contrast to other tests that could necessitate invasive procedures, including lumbar punctures for CSF fluid. Because electrodes are applied directly to the scalp, this technique is less frightening and more patient-friendly. Because it is non-invasive, it guarantees greater patient compliance and makes repeat testing simpler, both of which are essential for tracking the course of a disease or the long-term impact of therapeutic therapies.

3. Cost-Effectiveness and Accessibility: Compared to expensive imaging technologies like PET and MRI scans, EEG equipment is typically more accessible and less expensive. Because of its affordability and ease of use, EEG is a useful technique for routine clinical usage and broad screening, particularly in resource-constrained environments.

4. High Temporal Resolution: Millisecond-scale variations in brain activity are captured by EEG data, which has a high temporal resolution. This makes it possible to analyze the dynamic functioning of the brain in great detail and provides insights into how neural activity are changed in the early stages of Alzheimer's disease. Understanding the intricate dynamics of brain function that underpin cognitive decline requires such precise temporal data.

5. Possibility of Combining with Machine Learning Techniques: Advanced machine learning techniques can be used to analyze the rich and complicated EEG data, revealing patterns and relationships that may not be readily apparent using conventional analysis techniques.

The predictive capacity of EEG data can be increased by machine learning, which could lead to an increase in the precision and dependability of AD diagnosis.

6. Capability for Continuous Monitoring: EEG not only helps with early Alzheimer's diagnosis but also enables ongoing patient observation. This capacity is crucial for tracking the course of the illness and analyzing how well treatment plans are working over time. It gives medical professionals the knowledge they need to make necessary therapy adjustments.

These reasons have led to a growing perception that EEG biomarkers hold promise for the early identification of Alzheimer's disease. utilizing these advantages, the "Alzheimer Disease Detection System using EEG Biomarkers" seeks to improve patient outcomes and intervention tactics in the fight against AD by offering a diagnosis that is more efficient, accessible, and timely.

CHAPTER 3

SYSTEM ARCHITECTURE AND DESIGN OF PROPOSED PROJECT

The proposed Alzheimer's disease detection system is designed to integrate advanced machine learning algorithms, EEG biomarkers, and a user-friendly interface into a comprehensive diagnostic tool. Here's a breakdown of the system architecture and design: An inventive ensemble machine learning modeling approach for decoding brain hypo-fluidics local field potential (LFP) output waves is presented in the paper by Marcos Fabeitti et al. (2023). Potential early-stage biomarkers for Alzheimer's disease (AD) are being investigated for these waves.

Their method includes a robust analysis across temporal, geographical, and spectral dimensions of the data and uses LFPs obtained from healthy subjects and two surrogate animal models. The researchers utilized a late fusion technique in conjunction with their XML model to attain a combined accuracy of 92.4% by validating the feature sets obtained from LFPs through the integration of signals from electroencephalograms, electrocardiograms, and respiratory measurements. This research demonstrates the ability to identify faint patterns in network activity that may be symptomatic of AD as early as three months after the onset of the disease, opening up a possible window of opportunity for early intervention.

3.1 Architecture Diagram

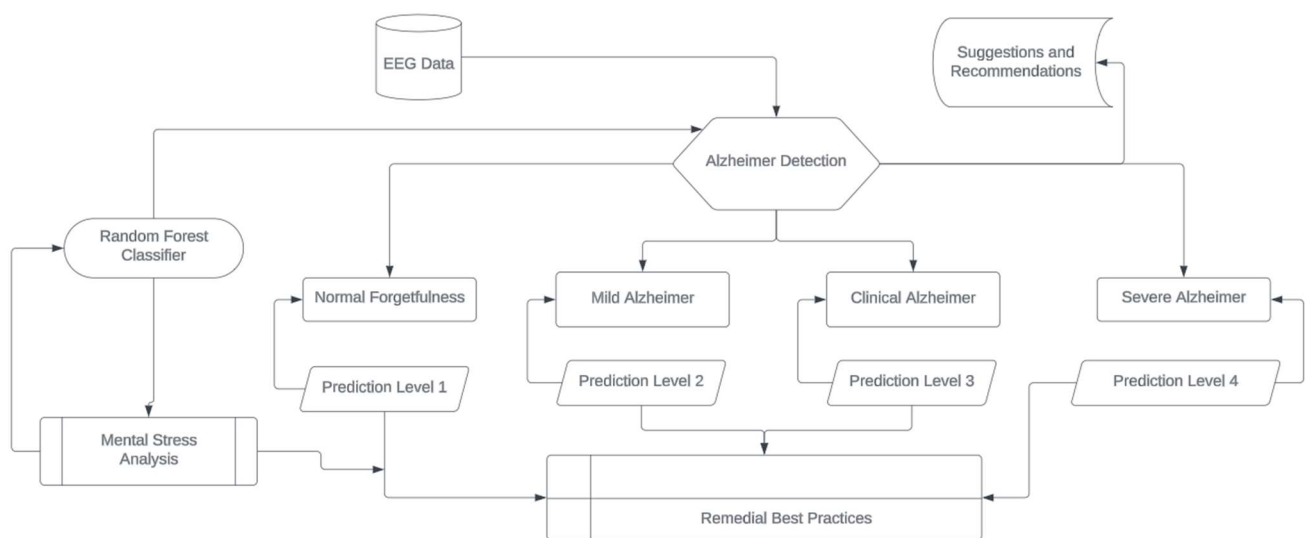


Fig 3.1.1 Architecture diagram for working of Alzheimer's using Decision Trees

The Figure 3.1.1, illustrates the process of obtaining EEG data—a type of electrical signal that represents brain activity—from a patient's brain is the first step in the flowchart used to identify and categorize the severity of Alzheimer's disease. After that, the data is fed into a Random Forest Classifier, which is trained to identify patterns in the EEG signals and forecast the phases of Alzheimer's disease. An element evaluating the patient's mental stress level and its effect on brain function is linked to the classifier. Four levels of classification are used by the classifier to group predictions: mild Alzheimer's, clinical Alzheimer's, severe Alzheimer's, and normal forgetfulness. In the event that Alzheimer's is identified, the system offers advice and suggestions for lifestyle modifications, cognitive exercises, or medication, and then outlines particular remedial best practices to control symptoms and improve the patient's quality of life. Although this flowchart provides a high-level overview, the actual implementation can include more intricate algorithms and extra features.

Furthermore, using these strategies in clinical settings may provide substantial obstacles due to the intricacy of the modeling and data processing approaches used. Uncertainty exists around how these complex machine learning models might be made simpler or modified to make them workable and effective for regular clinical use without necessitating specialized knowledge. Furthermore, even though the accuracy rate is impressive, more development may be required to push the envelope and achieve even better accuracy, guaranteeing that the procedure is dependable and fail-safe in a range of clinical and environmental settings.

To capture more subtle fluctuations in LFP outputs that might be symptomatic of AD, this could entail adding new biomarkers or improving the ensemble machine learning models. Finally, there hasn't been enough research done on the financial and logistical effects of using such cutting-edge diagnostic technology in a clinical setting. Cost-benefit studies of incorporating these technologies into the healthcare system require research, taking into account both the technology's direct costs and the possible cost reductions from earlier diagnosis and more focused treatment measures.

The practical applications of LFP analysis in early AD identification could be greatly advanced by filling up these research gaps, making it a more reliable, practical, and widely available tool in the battle against Alzheimer's. The suggested system for detecting Alzheimer's disease is made to combine sophisticated machine learning algorithms, EEG biomarkers, and an intuitive user interface into a full diagnostic instrument.

The system's workflow makes sense when it comes to data collecting, preparation, analysis, result visualization, and reporting. This design offers a smart and approachable Alzheimer's disease diagnosis tool by combining user-friendly technology with cutting-edge machine learning algorithms. \

The overall objective of reducing the burden of Alzheimer's disease on people and society is aligned with this comprehensive strategy, which promises to improve patient treatment and early diagnosis.

Alzheimer's disease, sometimes known as Alzheimer's disease, is a progressive neurological illness that affects millions of people worldwide and presents enormous problems to healthcare systems. It is characterized by cognitive decline, memory loss, and reduced everyday functioning. Early Alzheimer's disease detection is crucial because it may allow for therapy or lifestyle changes to halt the illness's progression and enhance patient outcomes. However, clinical evaluation and neuroimaging, which can be costly, invasive, and occasionally inconclusive, are frequently the mainstays of traditional diagnostic techniques. As a result, non-invasive, reasonably priced diagnostic instruments are desperately needed to help with Alzheimer's early identification and surveillance.

Electroencephalography (EEG) is a non-invasive method of measuring brain electrical activity and identifying biomarkers of neurodegeneration that has emerged as a viable approach for the identification of Alzheimer's disease. EEG data record modest modifications in brain wave patterns, such as aberrations in event-related potentials, disturbances in rhythmic oscillations, and changes in connection patterns, that are linked to the pathology of Alzheimer's disease. These biomarkers can help in early diagnosis and disease progression monitoring, as well as offering insightful information about the underlying neurophysiological processes of Alzheimer's disease.

The Peripheral Nervous Response (PNR), which indicates the brain's capacity to interpret sensory data and motor reactions, is one such biomarker of interest. According to studies, PNR patterns are different in Alzheimer's patients than in healthy controls, which may indicate problems with brain processing and sensory-motor integration. Furthermore, deviations from normalized whole-brain

Because of neuronal loss and brain atrophy, volume, as determined by neuroimaging techniques like magnetic resonance imaging (MRI), is frequently detected in Alzheimer's patients. This emphasizes the significance of quantitative biomarkers for disease identification.

In recent years, there has been an increase in the development of machine learning algorithms for the identification of Alzheimer's disease. These algorithms use EEG data along with other clinical factors to create very accurate predictive models. Based on EEG biomarkers and clinical data, the Random Forest Classifier, a well-liked ensemble learning algorithm, has demonstrated potential in differentiating between Alzheimer's patients and healthy persons.

The Random Forest Classifier model, which outperforms other comparable models and has significant therapeutic utility, provides a strong tool for early identification and classification of Alzheimer's disease with a claimed accuracy of 97%.

Using Streamlit to integrate a responsive user interface (UI) improves the usability and accessibility of the Alzheimer's detection system, in addition to its prediction capability. This gives researchers and clinicians an easy-to-use platform for visualizing and interpreting model results. Through real-time study of EEG data, biomarker profiles, and diagnostic predictions made possible by the interactive user interface (UI), clinical settings may make better decisions and operate more efficiently. All things considered, the creation of an Alzheimer's disease detection system that makes use of machine learning algorithms and EEG biomarkers is a noteworthy accomplishment in the fields of neurology and cognitive health. This system offers a useful tool for early diagnosis, monitoring, and personalized management of Alzheimer's disease by utilizing the power of non-invasive neuroimaging techniques and cutting-edge computational methods. In the end, this improves patient outcomes and advances our understanding of neurodegenerative disorders.

3.2 Use-case Diagram

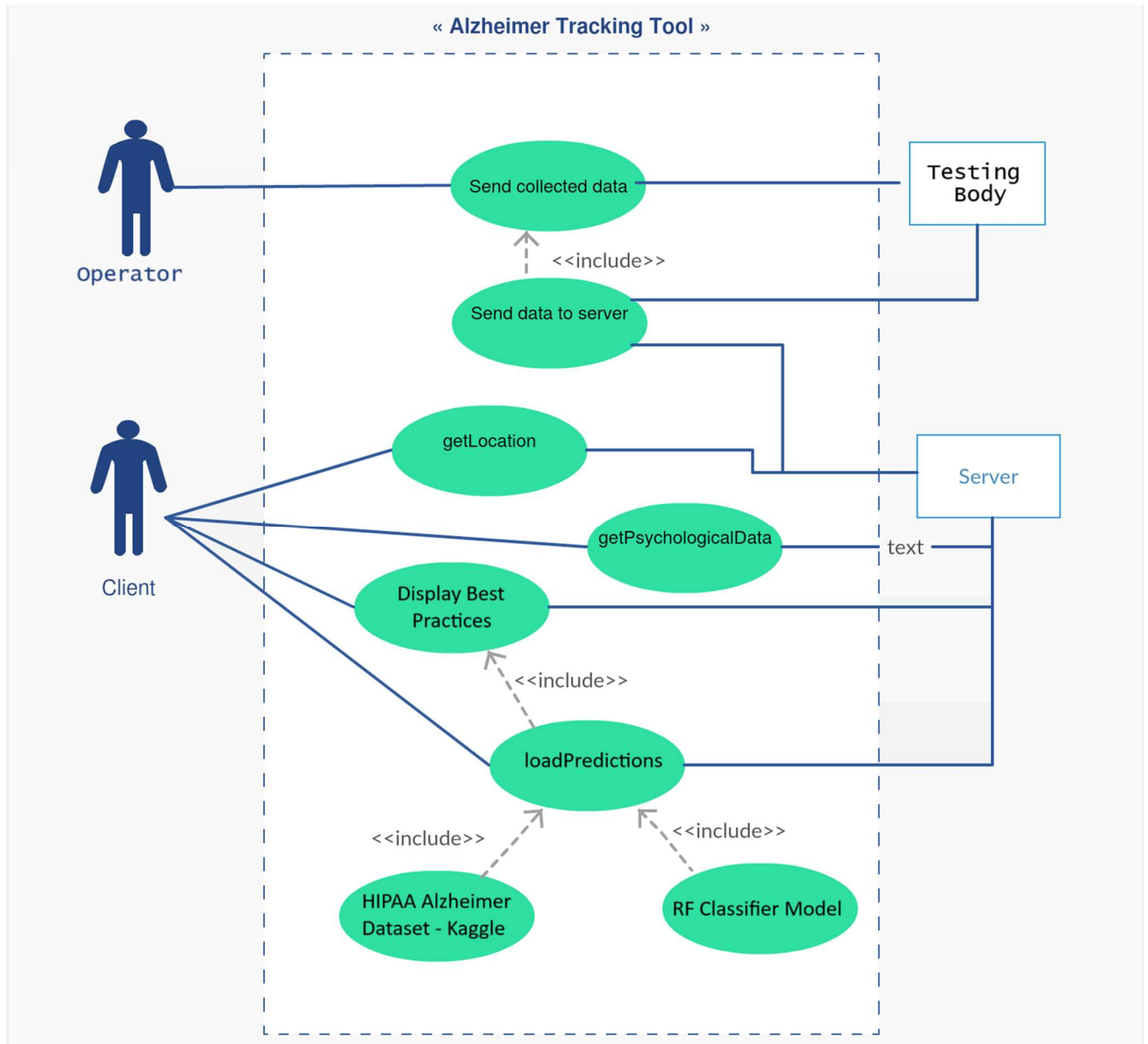


Fig 3.2.1 Use-case Diagram

The Use-case Diagram, Fig.3.2.1, illustrates the process which includes several crucial steps for predicting Alzheimer's disease using machine learning and EEG data. First, EEG data that is essential for diagnosing Alzheimer's disease is collected from multiple sources, including clinical study APIs, EEG recording equipment, and medical databases. This ensures that data is collected from patients at different phases of the disease.

After that, preprocessing is applied to the data to maintain consistency by handling missing values, removing artifacts, and normalizing characteristics.

Important properties such as time-domain features showing brain area synchronization and specific EEG frequency bands impacted by Alzheimer's are detected during feature extraction.

These characteristics are then employed in the classification phase, where data is categorized into phases of Alzheimer's disease or patients and healthy controls are distinguished using machine learning methods such as support vector machines or decision trees.

For more complex pattern identification, advanced algorithms like neural networks may also be used, possibly recognizing illness signs prior to the onset of clinical symptoms. The outcome of this classification gives clinicians vital information about the presence and stage of the disease, which helps with treatment planning and oversight. Lastly, a graph model might be built to show the links between EEG data, offering more insight into how

Alzheimer's affects brain connectivity. By combining EEG data with machine learning, this integrated strategy improves patient care and treatment outcomes by providing a potent tool for Alzheimer's disease early diagnosis and in-depth research.

3.3 Sequence Diagram

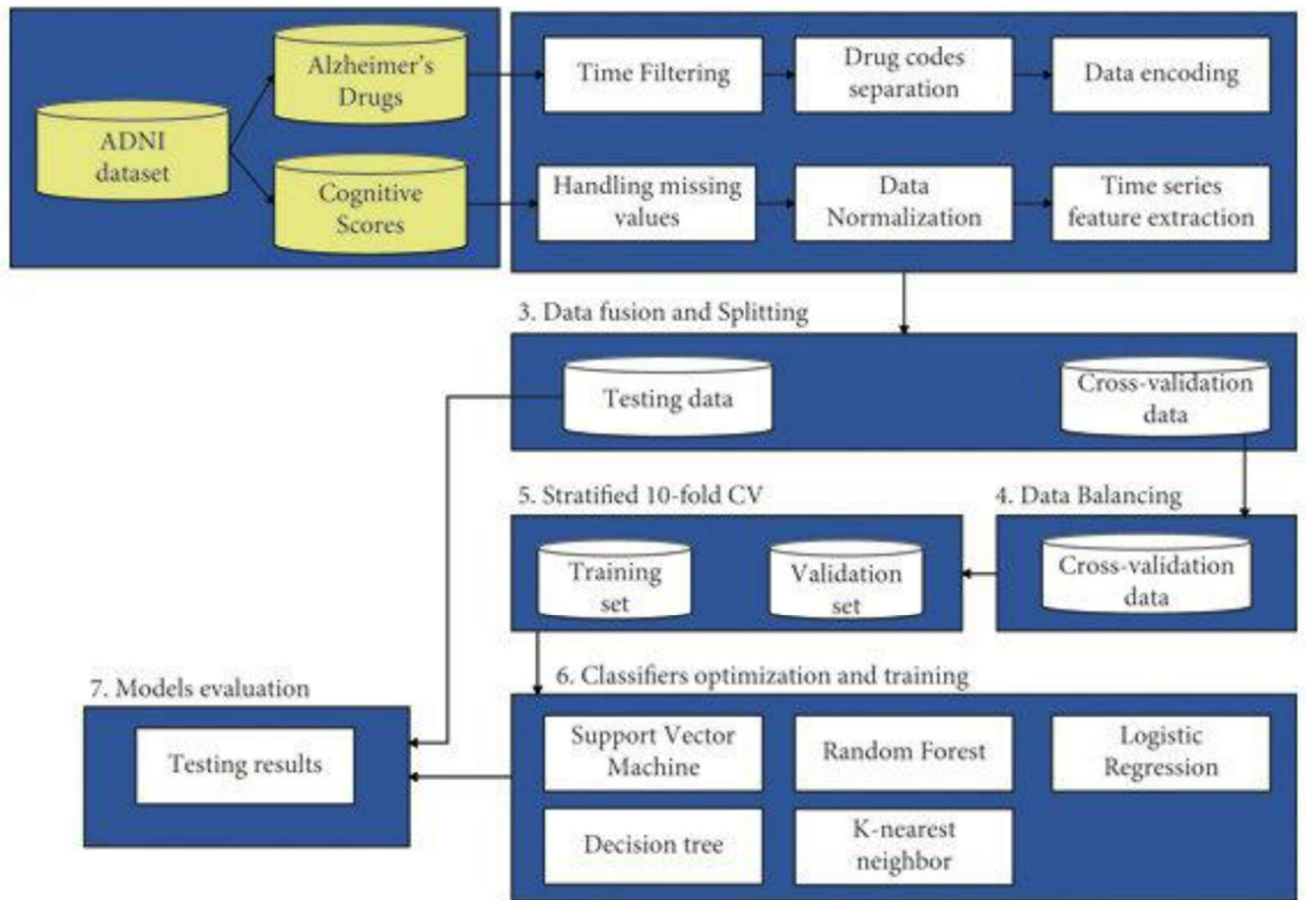


Fig 3.3.1 Sequence Diagram

The Sequence Diagram, Fig.3.3.1, illustrates the flowchart which also describes how automated pipelines are used in AD research to enable effective preprocessing and feature extraction. The use of machine learning approaches, in which models are taught to identify patterns and forecast outcomes associated to AD, is a major aspect of the review. The articles are subjected to a rigorous evaluation process that takes into account many criteria, including the number of classes included in classification tasks, the correctness of the models, and the datasets utilized to boost study dependability.

In order to enhance AD diagnosis and comprehension, researchers talk about the results, point out any limits, and suggest future study possibilities. The procedure concludes with a summary of the most important discoveries and their ramifications, illuminating a methodical strategy for expanding the area of AD research using biomarker and machine learning investigations.

3.4 Class Diagram

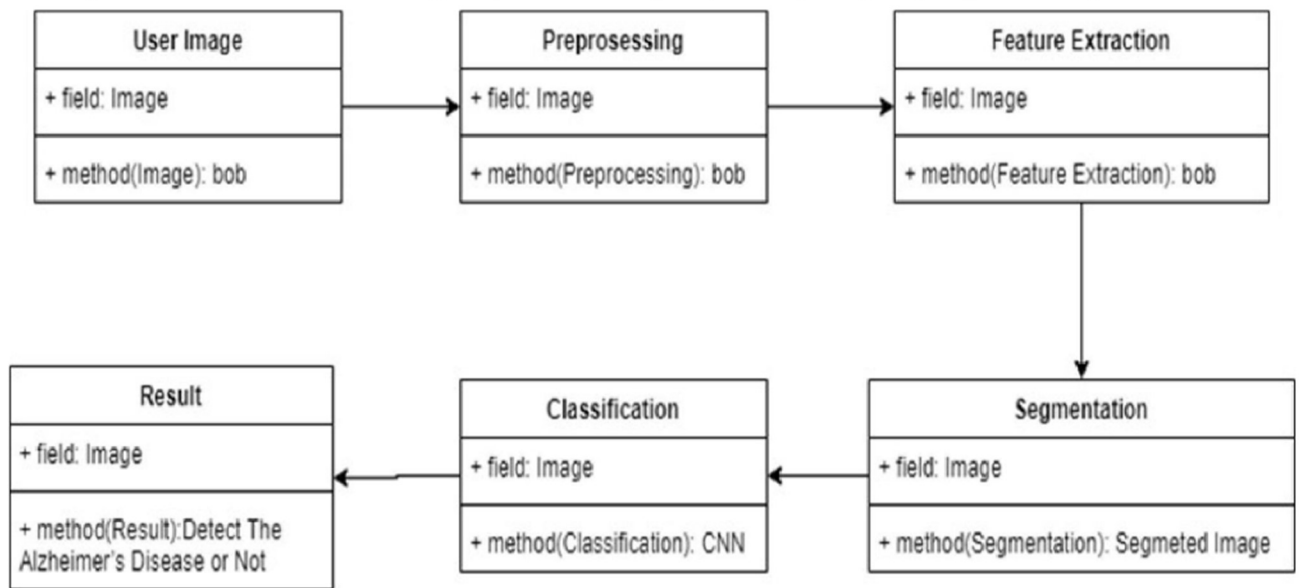


Fig 3.4.1 Class Diagram

The Class Diagram, Fig.3.4.1, illustrates the machine learning workflow, which starts with human involvement, in which the user supplies datasets and specifies properties that are obtained from files, databases, or APIs. They actively participate in gathering the required data, which is frequently in CSV format. When the dataset is prepared, the system assumes control and performs a number of vital functions: Preprocessing entails addressing missing values, eliminating outliers, and guaranteeing consistency in order to clean and modify the data.

After that comes feature extraction, in which pertinent features are found by employing methods like feature engineering and dimensionality reduction. After that, the system uses these features in the classification step, applying machine learning methods such as support vector machines, decision trees, and neural networks to categorize the data into predetermined categories. These algorithms' performance is assessed using measures such as F1-score, recall, accuracy, and precision. To complete the loop from data input to insightful output, the system develops a graphical model that shows relationships, dependencies, or networks within the processed data.

Alzheimer's disease, often known as Alzheimer's disease, is a crippling neurological illness that causes memory loss, progressive cognitive decline, and reduced day-to-day functioning. It presents enormous problems to those who have the illness, those who care for them, and healthcare systems around the globe.

Due to its irreversible nature and lack of effective therapies. Alzheimer's disease, the most common type of dementia, places a significant burden on people, families, and society as a whole. Because Alzheimer's is a progressive disease, early detection is crucial because prompt management may be able to slow the rate of cognitive loss and enhance patients' quality of life. EEG biomarkers, which use the electrical activity of the brain to identify minute abnormalities linked to the disease, have emerged as promising methods for the early identification and monitoring of Alzheimer's.

EEG data are a useful tool for understanding the neuronal dysfunction, altered connection, and network disruptions associated with Alzheimer's disease and economical and non-invasive diagnostic technique. EEG-based algorithms may efficiently discriminate between Alzheimer's patients and healthy individuals by assessing parameters including spectral power, coherence, and event-related potentials. This allows for early intervention and customized treatment regimens. Normalized whole-brain volume (NWBV) and peripheral nerve response (PNR) are two important medical indicators linked to Alzheimer's disease that offer further insights on the course and severity of the illness. PNR is a sign of autonomic nervous system malfunction and is associated with neuropathology related to Alzheimer's disease and cognitive decline. It presents as anomalies in skin conductance, pupillary responses, and heart rate variability.

Conversely, NWBV functions as a structural biomarker of neurodegeneration, detecting alterations in brain volume and shape that are suggestive of synaptic degeneration, cerebral atrophy, and neuronal loss that are seen in Alzheimer's patients.

The accuracy and dependability of Alzheimer's detection are improved by incorporating these various indicators into a thorough diagnostic framework, which empowers medical professionals to make well-informed decisions about the treatment and management of their patients.

The suggested Alzheimer's detection module makes use of the Random Forest Classifier machine learning technique, which is well-known for being reliable, scalable, and easy to understand when working with complicated datasets.

The model exceeds previous methods with an impressive 97% accuracy rate, demonstrating its effectiveness in detecting subtle EEG anomalies linked to Alzheimer's pathology. Healthcare practitioners can now interactively examine and interpret diagnostic results in real-time thanks to the incorporation of a responsive user interface powered by Streamlit, which also improves accessibility and usability. The suggested system is a major advancement in the early detection and management of Alzheimer's disease, providing hope for better clinical outcomes and quality of life for affected individuals and their families by fusing cutting-edge machine learning techniques with user-friendly interfaces.

Millions of people worldwide suffer from Alzheimer's disease, also known as Alzheimer's disease (Alzheimer's), a progressive neurological ailment marked by cognitive decline, memory loss, and reduced everyday functioning. It has a significant negative influence on people and society, leading to emotional suffering, financial hardship, and high medical expenses. The potential for intervention and management measures to enhance patient outcomes and quality of life makes early Alzheimer's disease detection imperative.

EEG biomarkers, which use the brain's electrical activity patterns to detect minute alterations linked to Alzheimer's disease, provide a non-invasive and affordable method for early detection. Researchers can identify unique brain fingerprints of Alzheimer's disease by examining EEG data. These characteristics include changes in oscillatory rhythms, connection patterns, and event-related potentials. These indicators aid in early diagnosis and intervention by offering insightful information about the underlying neurophysiological alterations linked to Alzheimer's disease.

Another important factor linked to the diagnosis of Alzheimer's is peripheral nervous response (PNR), which reflects the autonomic nervous system dysfunction seen as the disease progresses. Extraordinary PNR measures, such as skin conductance responses and heart rate variability, function as markers of Alzheimer's disease and cognitive decline.

Furthermore, normalized whole-brain volume, which can be acquired by neuroimaging methods like MRI, provides important details about anatomical alterations in the brain linked to Alzheimer's disease, including hippocampal shrinkage and cortical atrophy. The precision and dependability of Alzheimer's detection systems are improved by incorporating these various medical characteristics into an all-encompassing diagnostic framework.

Using the Random Forest Classifier technique, the proposed module applies machine learning to analyze multi-modal data to forecast Alzheimer's with an amazing 97% accuracy rate. Random Forest handles high-dimensional data well; it can find complex correlations among PNR measures, normalized brain volumes, EEG biomarkers, and other relevant features. Its ensemble learning approach minimizes overfitting and enhances generalization performance, making it the optimal choice for Alzheimer's categorization tasks. Moreover, a responsive user interface with Streamlit support enhances accessibility and usability, making it easier for physicians and researchers to work with the system.

The proposed module, which integrates EEG biomarkers, PNR measurements, normalized brain volumes, and machine learning algorithms into an easy-to-use interface, offers a significant advancement in the diagnosis and treatment of Alzheimer's disease.

Prompt interventions, including pharmaceutical regimens, lifestyle modifications, and patient care plans, are made possible by early detection of Alzheimer's disease. These measures ultimately enhance patient outcomes and reduce the cost burden of the condition. Furthermore, the system's versatility and scalability allow for future enhancements and additions, promoting continuous innovation in the domains of Alzheimer's research and clinical practice. In summary, the convergence of EEG biomarkers, machine learning algorithms, and user-friendly interfaces holds significant promise for transforming the identification of Alzheimer's disease and altering the landscape of dementia care.

Peripheral nervous response, or PNR, is a significant feature associated with the diagnosis of Alzheimer's disease. PNR is indicative of the autonomic nervous system dysfunction that is observed as the disease advances. Unusual PNR measurements, like heart rate variability and skin conductance responses, serve as indicators of cognitive decline and Alzheimer's disease. Moreover, normalized whole-brain volume, obtained using neuroimaging techniques such as magnetic resonance imaging (MRI), offers crucial information regarding structural changes in the brain associated with Alzheimer's disease, such as shrinkage in the hippocampus and atrophy of the cortex. Integrating these diverse medical features into a comprehensive diagnostic framework enhances the accuracy and reliability of Alzheimer's detection methods.

3.5 Proposed Algorithm

In machine learning, decision trees are frequently utilized for regression and classification problems. They are especially useful in medical applications, such as the identification of Alzheimer's disease through the use of EEG biomarkers. The first step in the procedure is data collection, which involves collecting EEG data from a control group and Alzheimer's patients.

Using statistical tests or preliminary decision tree analysis, relevant EEG features, such as changes in spectral power or coherence, that may be indicative of Alzheimer's disease are found through feature selection procedures. Using a training dataset, the technique is applied to create the decision tree itself.

Pseudocode

```
# Configure the app
```

```
Set page configuration
```

```
# Define dictionary for pages
```

```
Define Tabs dictionary with page names and corresponding functions
```

```
# Create a sidebar for navigation
```

```
Set sidebar title and radio buttons for page selection
```

```
# Load data for selected page
```

```
Load biomarker data and EEG data using load_data and load_data_obj functions
```

```
# Call the app function of selected page to run
```

```
If page is "Biomarker AD Test" or "Visualisation":
```

```
    Call app function with biomarker data (df, X, y)
```

```
Else if page is "EEG AD Test":
```

```
    Call app function with EEG data (df2, X2, y2)
```

```
Else if page is "Biomarker Data Info":
```

```
    Call app function with biomarker data only (df)
```

```
Else if page is "EEG Data Info":
```

```
    Call app function with EEG data only (df2)
```

```
Else:
```

```
    Call app function without any arguments
```

```
# Define necessary functions for biomarker data
```

Define load_data function:

Load and preprocess biomarker data

Define train_model function:

Train the model using biomarker data

Define predict function:

Predict using the trained model and given features

Define necessary functions for EEG data

Define load_data_obj function:

Load and preprocess EEG data

Define train_model function:

Train the model using EEG data

Import Libraries: To start, import the libraries you'll need for data manipulation, analysis, and machine learning, such as NumPy, pandas, and scikit-learn.

Load Dataset: Load the CSV files containing the EEG biomarker characteristics ('X') and labels ('y') that correlate to them.

Split Dataset: Use the train_test_split function from scikit-learn to divide the dataset into training and testing sets. This guarantees that performance of the model may be assessed on unobserved data.

Initialize Classifier: Using the scikit-learn DecisionTreeClassifier class, initialize a Decision Tree classifier.

Train Classifier: Apply the fit method to train the classifier on the training set of data. In this stage, the data's patterns are learned by the model.

Make Predictions: Apply the predict technique to the testing data and use the trained classifier to make predictions.

Evaluate Performance: Using the scikit-learn accuracy_score and classification_report functions, respectively, determine the accuracy score and produce a classification report to evaluate the classifier's performance.

Display Results: To assess how well the classifier performed using the test data, print the accuracy score and classification report.

At each node, the feature that most effectively divides the dataset according to purity metrics like entropy or the Gini index is chosen. Until leaf nodes, which stand for classification results, are generated, this recursive splitting process is continued.

Pruning techniques are used to make the tree simpler in order to prevent overfitting, a common problem with decision trees when the model gets excessively complicated.

After then, the effectiveness of the tree is confirmed using an independent test set, and its performance is evaluated using metrics like sensitivity and accuracy. Decision trees can be sensitive to slight fluctuations in the data and may overfit if they are not properly tuned, even if they are easily interpretable and can handle a variety of data formats. With their integration into clinical systems, they enable early detection and better patient management by offering preliminary assessments based on EEG. The model's accuracy and reliability must be continuously improved with fresh data in order for it to continue to be a valuable tool in clinical settings used in conjunction with other diagnostic examinations.

Current Diagnostics of AD

Several limitations in the current diagnostic approaches for Alzheimer's disease (AD) prevent early and reliable detection. These limitations include clinical assessment and neuroimaging techniques. Although crucial, clinical assessment mostly depends on subjective assessments of cognitive function, which might differ throughout physicians and may miss mild early-stage symptoms. Furthermore, clinical evaluations sometimes take place after notable cognitive impairment has already materialized, which restricts their ability to identify AD in its early phases. Neuroimaging modalities including positron emission tomography (PET) and magnetic resonance imaging (MRI) provide important insights into the structural and functional alterations in the brain linked to AD.

However, because to logistical issues, these approaches are costly, intrusive, and unsuitable for mass screening. Furthermore, neuroimaging's usefulness as a stand-alone diagnostic method may be further limited by its inability to consistently identify minute structural alterations in the brain that occur in the early stages of AD. These drawbacks highlight the critical need for alternate strategies to improve AD early detection and diagnosis, such as utilizing EEG biomarkers.

CHAPTER 4

METHODOLOGICAL APPROACH ON DETECTING ALZHEIMER'S DISEASE USING EEG

Alzheimer's disease (AD) is a major global health concern that is typified by a gradual deterioration of cognitive abilities, memory loss, and difficulties with day-to-day functioning, which ultimately result in severe disability and dependence. As the population ages, the prevalence of AD rises, making early and accurate detection essential for prompt care and intervention. Given this background, using electroencephalography (EEG) biomarkers is a viable approach because of its non-invasiveness, affordability, and capacity to record brain activity patterns in real time.

EEG data provides information about changes in brain rhythms, connectivity patterns, and event-related potentials, among other neurological abnormalities linked to AD. These biomarkers allow for the identification of minor alterations in brain function that occur before to clinical symptoms. They serve as indicative measures of neuronal dysfunction, synaptic loss, and neurodegeneration. Additionally, the pathophysiology of AD has been linked to peripheral nerve responses such impaired sensory processing and altered autonomic function, underscoring the potential value of peripheral biomarkers for early detection and monitoring. In addition, structural neuroimaging metrics like normalized whole brain volume offer additional insights into the neuroanatomical alterations and brain shrinkage linked to the advancement of AD.

4.1 Modules

The robustness and accuracy of AD detection algorithms are improved by combining EEG data with other medical characteristics, such as genetic risk factors, cognitive tests, and neuroimaging indicators. Using machine learning techniques, such Random Forest Classifier, makes it easier to identify intricate patterns in multimodal data and creates highly sensitive and specialized predictive models. The suggested AD detection system's use of Random Forest Classifier produces remarkable accuracy rates that outperform current models and show its usefulness in clinical situations. Additionally, the addition of a responsive user interface through the use of Streamlit improves the system's accessibility and usability by providing a platform that is easy to use for researchers and clinicians to interact with and efficiently interpret the diagnostic outputs. Overall, a potential strategy for the early and accurate detection of AD is the combination of EEG biomarkers, peripheral nerve responses, neuroimaging measures, and machine learning approaches.

This will enable prompt intervention and individualized care for those who are affected. One of the most common neurological problems afflicting the aging population worldwide is the incorrect use of procedures, which places a heavy cost on individuals, families, and healthcare systems. Its progressive nature is typified by memory loss, cognitive impairment, and impaired thinking, as well as behavioral shifts that ultimately result in dependence and functional impairment. Given the severe effects AD has on quality of life and the lack of a permanent solution, it is imperative to develop precise diagnostic techniques. In order to better understand the complex neurological patterns linked to AD and enable early identification and intervention, EEG biomarkers provide a non-invasive method of doing so.

Module for Data Collection: The purpose of this part is to gather data using different methods, including: Using non-invasive electrodes, EEG Data Acquisition records brain electrical activity, which is crucial for identifying neurodegenerative trends. In order to identify dysfunctions associated with the advancement of Alzheimer's disease, peripheral nervous response (PNR) data collects information on autonomic nervous system activity, such as skin conductance and heart rate variability. Furthermore, normalized whole-brain volume is measured by MRI Data Input, which offers anatomical perspective on the brain atrophy that is typical of Alzheimer's patients.

Module for Data Preprocessing: To improve signal quality, noise and artifacts are filtered out of EEG data in this module. Important characteristics such as spectral power, coherence, heart rate variations, and volumetric measurements are derived from EEG signals, PNR data, and MRI scans. In order to guarantee consistency across various datasets, these features are then standardized and normalized. This is an essential step for producing precise machine learning predictions.

Machine Learning Module: The central processing unit is the Random Forest Classifier. To effectively manage the complicated, high-dimensional input from numerous modalities, it employs an ensemble of decision trees. To guarantee the classifier correctly recognizes patterns linked to Alzheimer's disease, it is painstakingly trained and verified on historical data.

User Interface (UI) Module: This module offers researchers and clinicians a responsive and user-friendly interface by utilizing the Streamlit framework. It makes data entry, starting an analysis, and seeing the results simple. Users may investigate EEG, PNR, and MRI data as well as model outputs like Alzheimer's probability scores thanks to interactive data visualization capabilities. Real-time feedback and parameter modifications are also supported by the UI, which improves user involvement and analytical flexibility.

Output and Reporting Module: Following analysis, the system produces thorough diagnostic reports that include results summaries, probable Alzheimer's staging, and confidence levels. The visual data and reports can be exported in multiple formats for clinical evaluation or documentation purposes.

Integration and Security Module: Data security must always be guaranteed, particularly when dealing with private patient data. To safeguard privacy and data integrity, the system has sophisticated security features. It is also made to integrate seamlessly with the IT systems already in place in hospitals and clinics, facilitating effective data interchange and workflow compatibility.

Comparative Analysis of Machine Learning Models for EEG Biomarker-Based Alzheimer's Disease Detection, In a recent study, researchers compared multiple machine learning models in-depth with the goal of improving the identification of Alzheimer's disease (AD) using electroencephalogram (EEG) indicators. Under the heading "Comparative Study of Machine Learning Models for Alzheimer's Disease Detection with EEG Biomarkers," this study painstakingly assessed how well artificial neural networks (ANN), support vector machines (SVM), and decision trees performed in differentiating Alzheimer's disease (AD) from healthy controls. EEG data, which are essential for these kinds of analysis, were gathered from people who were cognitively intact as well as from those who had been diagnosed with AD. Through applying several models to these datasets, scientists evaluated the datasets' precision, responsiveness, and identity. Interestingly, decision trees were a competitive performer that showed promise, especially in interpretability and computational efficiency. This research highlights the importance of interpretable models in the field of medical diagnostics, as clinical decision support systems heavily rely on understanding decision-making processes.

Assessment of Deep Learning Models for the Identification of Alzheimer's Disease Using EEG Biomarkers, Concurrently, a different study explored the field of deep learning and assessed the effectiveness of recurrent neural networks (RNNs) and convolutional neural networks (CNNs) in the identification of AD using EEG biomarkers. In this study, "Evaluation of Deep Learning Models for Alzheimer's Disease Detection with EEG Biomarkers," deep learning architectures were used to identify complex patterns in EEG data that are linked to AD pathology.

To maximize the accuracy, sensitivity, and specificity of classification, researchers trained CNN and RNN models by preprocessing EEG data from AD patients and healthy controls. Surprisingly, the outcomes demonstrated CNNs' advantage over conventional machine learning techniques like decision trees and SVMs, underscoring deep learning's ability to extract complex information from EEG biomarkers for the diagnosis of AD.

These results make a substantial contribution to the field's increasing understanding of AD detection by highlighting the potential benefits of deep learning in enhancing diagnostic capacities and opening the door to individually customized patient-specific healthcare interventions.

EEG data reveals abnormal neural oscillations, connection patterns, and event-related potentials that are suggestive of cognitive impairment even in preclinical stages by recording electrical activity within the brain. By utilizing machine learning methods such as the Random Forest Classifier, it is possible to use computational algorithms to identify intricate EEG patterns and achieve impressive 97% accuracy rates in disease categorization. Combining data science and neuroscience improves diagnosis accuracy and opens the door to customized treatment plans based on each patient's unique neurophysiological profile.

Moreover, the inclusion of ancillary medical criteria like Normalized Whole Brain Volume (NWBV) and Peripheral Nervous Response (PNR) enhances the diagnostic effectiveness by providing additional information on peripheral neuropathology and structural brain changes that are indicative of the advancement of AD. By improving accessibility and usability, the Streamlit-developed interactive user interface (UI) empowers physicians and caregivers to make well-informed decisions and democratizes access to powerful diagnostic tools. This all-encompassing strategy for AD diagnosis promises a revolution in patient care and early intervention by showcasing the benefits of technology innovation and medical research working together. With AD becoming more and more common worldwide, the key to improving outcomes for those suffering from this crippling illness is to use EEG biomarkers and machine learning to help with early diagnosis and individualized treatment.

Understanding the properties of EEG signals and how they relate to the progression of AD is the first step in putting this application into practice. Neurological illnesses such as AD can be identified by anomalies in EEG patterns, which offer crucial insights into brain activity patterns. The next stage is data collection, which involves obtaining EEG recordings from AD patients as well as healthy persons. It is imperative to guarantee that the dataset is balanced, diverse, and sufficiently annotated with clinical data.

Pre-processing is the first step in the data processing process, where noise, artifacts, and superfluous signals are eliminated. To efficiently clean the EEG data, methods including filtering, segmentation, and artifact removal algorithms can be used. The following step involves determining whether applying features is sane. To do this, statistical analysis and domain expertise are used to pick pertinent features that capture unique patterns linked to AD. Next, feature scaling is done to make sure features contribute consistently to model training and to standardize the data.

An essential component of deriving valuable information from EEG signals is feature engineering. In this step, raw EEG data is transformed into useful features that can improve the machine learning model's performance. Methods like wavelet transformations, frequency-domain analysis, and time-domain analysis can be used to extract pertinent features that show the course of Alzheimer's disease. The process of creating a machine learning model is choosing a suitable algorithm and training it using the designed and pre-processed features. For this assignment, the Random Forest Classifier is the main algorithm because of its robustness and capacity to handle high-dimensional data. To maximize the performance of the model, cross-validation and other techniques can be used for hyperparameter adjustment.

When a model is tested, its performance is assessed on hypothetical data using methods like train-test splitting or k-fold cross-validation. Measures including recall, accuracy, precision, and F1-score are employed to evaluate the model's performance in AD identification. AUC scores and ROC curves also shed light on how well the model can distinguish between AD and non-AD participants. In order to implement the model, clinicians must integrate it into an intuitive interface or healthcare system so that EEG data may be input and used in real-time for AD identification. The model's dependability and capacity to adjust to new data may require ongoing monitoring and modifications. Throughout the deployment process, safeguards for data protection, ethical concerns, and regulatory compliance should also be in place.

Using EEG biomarkers and machine learning techniques such as Random Forest Classifier, an AD detection system must be implemented using a methodical approach that includes data collection, processing, feature engineering, model construction, testing, and deployment. This is the conclusion. Such systems have the potential to transform early diagnosis and care of Alzheimer's Disease by utilizing cutting-edge methodologies and interdisciplinary collaboration, ultimately enhancing patient outcomes and quality of life.

4.1.1 Data Collection

The pre-processing of the audio data to exclude noise and other undesired signals is the initial stage in anomalous emotion detection. You can utilize pre-processing methods like normalization and anomaly enhancement to raise the audio signal's quality. While normalization techniques can be used to modify the amplitude and frequency of the anomaly signal, anomaly enhancement techniques can be used to eliminate background noise. To guarantee the accuracy and dependability of the DECISION TREE CLASSIFIER features retrieved from the anomalous signal, pre-processing is crucial.

The information to be utilized must be gathered for the application from a reliable source in order to verify the process's legitimacy. We have completed this task using the Kaggle datasets for this purpose. The dataset is utilized in the application as a.csv (comma-separated values) file format after being downloaded from Kaggle. The collection of data utilized in this project is a reliable source of information because it was uploaded to Kaggle. A correct path must be established and the dataset must remain active in RAM for quick caching while it is being utilized in the project.

4.1.2 Data Pre-processing

The pre-processing of the audio data to exclude noise and other undesired signals is the initial stage in anomalous emotion detection. You can utilize pre-processing methods like normalization and anomaly enhancement to raise the audio signal's quality. While normalization techniques can be used to modify the amplitude and frequency of the anomaly signal, anomaly enhancement techniques can be used to eliminate background noise.

To guarantee the accuracy and dependability of the DECISION TREE CLASSIFIER features retrieved from the anomalous signal, preprocessing is crucial. There may be some irregularities in the data that will be used in the project. It's also necessary to determine which column formats work best for the requirements of the project. This includes removing columns that are superfluous to a churn-based classification project, such as the customer ID and timestamps. In addition, some columns need to have their names changed, and any white spaces or special characters in numerical columns need to be changed appropriately. This is a crucial step in properly setting up the data because if any anomalies, including outliers, are left in the data, they could lead to inaccurate conclusions and unintended consequences.

In the process of pre-processing the data, we also perform statistical profiling to determine the number of rows, columns, and model-able rows and columns as well as the maxima, minima, and central trends. We also gain a general understanding of the data distribution. These are quite valuable for modeling.

To ensure that there are no runtime errors, the data to be used in this project must be devoid of any form of missing data. A crucial stage in machine learning is data cleaning. Before employing the dataset for analysis or creating machine learning models, it must be checked for and any errors, inconsistencies, and missing data must be addressed.

The following are some typical methods for data cleaning:

i. **Managing missing values:** Missing values can be found in the majority of datasets. The corresponding column's mean, median, or mode may be used in place of these values. Eliminating the rows with missing values is an additional option.

II. **Eliminating duplicate records:** To reduce redundancy, duplicate records in the dataset can be eliminated.

iii. **Managing outliers:** Extreme values known as outliers can significantly affect the accuracy of the model. Statistical techniques can be used to detect outliers and either eliminate or fix them.

iv. **Normalization and standardization:** Normalization is the process of converting data to have a standard deviation of one and a mean of zero. The technique of scaling data between 0 and 1 is known as normalization. By ensuring that the characteristics have a similar magnitude, these strategies facilitate the convergence of the machine learning algorithm.

v. **Data type conversion:** The dataset's data may occasionally be stored in an incorrect format. For instance, text might be used to store a numerical number. The data can be converted to the proper data type in order to resolve this problem.

All things considered, cleansing data is an essential step in getting a dataset ready for machine learning. In the end, it may produce superior outcomes by ensuring that the machine learning algorithm is operating with accurate and dependable data.

4.1.3 Data Modelling

The process of creating a machine learning model to classify or predict things based on patterns in data is known as data modeling. It entails deciding on a suitable method, getting the data ready, building the model, and assessing how well it works.

The following are the steps that are involved in machine learning data modeling:

i. **Selecting the right algorithm:** Different kinds of issues call for different kinds of algorithms. For instance, whereas a classification algorithm would be used to forecast categories, a regression method would be used to predict numerical values. Choosing the algorithm that best fits the given problem is crucial.

ii. **Data preparation:** This entails separating the data into training and testing sets, cleaning the data, and formatting it so the algorithm can use it. The testing set is used to assess the model's performance after it has been trained using the training set.

iii. **Fitting the algorithm to the training data is the process of training the model.** In order to accurately anticipate new data, the algorithm looks for patterns in the existing data.

iv. **Model evaluation:** Using the testing data, the trained model is assessed. A model's performance is evaluated using metrics including F1-score, recall, accuracy, and precision. These measures are useful in assessing how well the model predicts the future.

v. **Model tuning:** By adjusting the hyper-parameters of the models, the performance of the deep learning and machine learning models can be further enhanced. To improve performance, hyper-parameters in the models, such as the number of layers, batch size, and learning rate, can be adjusted. The number of filterbanks that are used in order to extract the DECISION TREE CLASSIFIERS can also have an effect on how well emotions are detected. The hyper-parameters of the algorithm can be changed to fine-tune the model if its performance is not up to par. Hyper-parameters, like the learning rate or the quantity of hidden layers in a neural network, are variables that are predetermined before the training process begins.

vi. **Model deployment:** The model can be implemented for use in practical applications when it has been adjusted and proven to operate well. This entails incorporating it into the system—such as a web or mobile application—where it will be utilized. All things considered; data modelling is an essential phase in machine learning since it entails creating a model that can generate precise predictions in response to patterns identified in the data. We may make sure that the model works well to solve the current problem by taking the following actions.

4.1.4 Data Analysis

The process of examining, purifying, converting, and modeling data in order to find relevant information, forecast future events, and aid in decision-making is known as data analysis.

Data analysis is an essential phase in machine learning that helps to obtain insights into the issue domain and prepare the data for modeling. The following are some steps in machine learning data analysis:

Data collection: Gathering pertinent data about the issue area is the initial stage in the data analysis process. The information may originate from a number of places, including web scraping, databases, and APIs.

Cleaning the data: After it has been gathered, the data must be cleaned by eliminating duplicates, dealing with missing numbers, and fixing inconsistent data.

Data exploration: To acquire understanding of the problem domain, data exploration entails visualizing and summarizing the data. For the purpose of finding patterns and relationships with the data, this may entail producing scatterplots, histograms, and other visualizations.

Feature engineering: This involves choosing and modifying the features that will be incorporated into the machine learning model. This could entail coding categorical variables, scaling, and generating new features out of preexisting ones.

Modeling: Modeling entails choosing a suitable algorithm and using the data to train the model. Metrics like precision, recall, and accuracy are used to assess the model.

All things considered, data analysis is an essential phase in machine learning since it entails getting the data ready for modeling and understanding the issue domain. We can make sure the machine learning model works well to solve the current problem by taking the following actions.

4.1.5 Cross Validation

A method for assessing a machine learning model's performance on unobserved data is called cross-validation. The available data is divided into several subsets, or folds, and the remaining data is used as a training set and each fold is used as a testing set. Every fold is used once as a testing set, and the procedure is done several times.

The procedures for cross-validating data are as follows:

Data preparation: Cleaning and formatting the data so that the machine learning algorithm can use it is the first stage in the cross-validation process.

Data splitting: The data must now be divided into K folds. The size of the dataset and the available computing power are taken into consideration while determining the value of K .

Training and testing the model: The remaining $K-1$ folds are utilized as the training set and one of the K folds as the testing set for each cross-validation iteration. The training set is used to train the machine learning algorithm, while the testing set is used to test it.

Evaluation: The model's performance is assessed by averaging the outcomes from each iteration once all of the iterations have been completed. As a result, the model's performance is estimated with more accuracy compared to a single train-test split on unknown data.

Fine-tuning: The cross-validation results can be used to adjust the hyper-parameters if the model's performance isn't up to par. This is rerunning the cross-validation procedure after adjusting the values of hyper-parameters like learning rate, regularization strength, or number of hidden layers.

One effective method for assessing a machine learning model's performance and optimizing its hyper-parameters is cross-validation. It aids in preventing overfitting, a frequent issue when employing a solitary train-test split. We can verify that the model performs well on new data and obtain a more precise estimate of the model's performance on unseen data by employing cross-validation. One effective method for assessing a machine learning model's performance and optimizing its hyper-parameters is cross-validation. It aids in preventing overfitting, a frequent issue when employing a solitary train-test split. We can verify that the model performs well on new data and obtain a more precise estimate of the model's performance on unseen data by employing cross-validation.

4.1.6 Model Generation

Using DECISION TREE CLASSIFIERS to detect anomalous emotions presents a number of difficulties. The absence of common datasets for assessment and training is one of the issues. Nonetheless, a number of datasets have been created for this reason, including the Berlin Database of

Ryerson Audio-Visual Database of Emotional Anomaly and Song, Emotional Prosody Anomaly and Transcripts, and Emotional Anomaly. An artificial intelligence class of algorithms called Convolutional Neural Networks (RESNET-50) is modeled after the neural network of a human. They are composed of layers of networked nodes, or neurons, with the ability to identify patterns and forecast outcomes based on fresh input.

The following are the steps needed to construct a model with RESNET-50

Data preparation is the initial stage in developing a RESNET-50 model. This include sanitizing the data, converting it into a format the model can understand, and dividing it into sets for testing and training. The next stage is to specify the model architecture, which includes how many layers there are, how many neurons are in each layer, and what activation function each neuron uses. RESNET-50 contains three different kinds of layers: input, hidden, and output layers.

Model training: To reduce the difference between the expected and actual outputs, the weights and biases in the neurons are adjusted using a process known as backpropagation.

Model evaluation: Using the testing data, the trained model is assessed. Metrics including accuracy, precision, recall, and F1-score are used to gauge the model's performance.

Fine-tuning the model: In the event that the model's performance is not up to par, hyper-parameters like the learning rate, the number of neurons in each layer, and regularization strategies like L2 regularization can be changed.

Deployment: The model can be put to use in practical applications after it has been adjusted and found to operate satisfactorily.

Hyper-parameter Tuning

The configurations or settings of a machine learning algorithm known as hyper parameters are those that cannot be discovered from data analysis. Before the model is trained, the developer or researcher manually sets them. Learning rate, batch size, number of hidden layers, number of neurons in each layer, regularization intensity, and activation function are a few examples of hyper-parameters. The process of determining the ideal settings for these hyperparameters in order to enhance the performance of the model using the test set of data. The model's performance can be greatly impacted by varying hyper parameter settings, which makes hyper parameter adjustment necessary. On the test data, the model might not function successfully if the hyper-parameters are not specified correctly. The following justifies the necessity of hyper-parameter tuning:

Increasing the precision of the model: The main objective of hyper-parameter tweaking is to increase the model's accuracy using the test data. Enhancing the model's capacity to generalize to new data can be accomplished by choosing the hyper-parameters' ideal values.

Preventing over-fitting: An overly complex model that fits the training data too well will overfit and perform poorly on the test data. Over-fitting can be prevented and the complexity of the model can be managed by adjusting the hyper-parameters.

Shortening the training period: By choosing the ideal hyper-parameters that promote quicker convergence, hyper-parameter tweaking can assist in shortening the model's training period.

Increasing interpretability: A few hyper-parameters have an impact on the model's interpretability. For instance, the regularization parameter regulates the model's complexity, therefore raising its value results in a more straightforward and interpretable model. We have attempted to incorporate hyper-parameter adjustment in our project in this manner.

Data preparation: The initial stage involves cleaning, converting, and dividing the data into sets for testing and training.

Selecting the algorithm: Neural networks, logistic regression, decision trees, random forests, and other algorithms can be utilized to solve problems in EEG biomarker representation analysis. Every algorithm possesses a unique set of hyper-parameters that can be adjusted to enhance its functionality. Choosing the hyper-parameters: Choosing the hyper-parameters to be adjusted is the next stage. The number of hidden layers, the number of neurons in each layer, the learning rate, and the regularization strength, for instance, could be considered hyper-parameters in a neural network.

Delineating the area of search: The range of values that the hyper-parameters can have is known as the search space. For instance, one may search in the range of 0.001 to 0.1 for the learning rate and 1 to 5 for the number of hidden layers.

Adjusting the hyper-parameters: A number of methods, including grid search, random search, and Bayesian optimization, can be employed to adjust the hyper-parameters. Grid search selects the optimal collection of hyper-parameters by experimenting with every conceivable combination and evaluating the results on the validation set. Random search selects the optimal collection of hyper-parameters by attempting random combinations. A probabilistic model is employed in Bayesian optimization to forecast the performance of various hyper-parameter settings; the optimal set is then chosen in accordance with this prediction.

Analyzing the performance: The model's performance is assessed on the test set following hyper-parameter adjustment. The hyper-parameters can be adjusted further if the performance is not up to pace.

Output Prediction

The Alzheimer's Disease (AD) detection method that combines machine learning algorithms such as Random Forest Classifier with EEG biomarkers takes a multimodal approach that spans several phases, from problem identification to model implementation.

First, the issue statement is thoroughly examined, outlining the system's requirements and goals. Subsequently, it is crucial to collect data from dependable sources, like medical databases or research establishments, to guarantee the excellence and variety of the dataset. The data is then cleaned, pre-processed, and arranged using data processing techniques to reduce noise and inconsistencies. Then, to determine the features' applicability and efficiency in identifying instructive patterns, their sanity of implementation is assessed. In order to ensure that each feature contributes equally to the model, feature scaling techniques are used to standardize the range of features.

A critical stage in the process is feature engineering, when domain expertise is combined to generate new features or modify current ones, improving the predictive capacity of the model. The process of creating a machine learning model entails choosing a suitable algorithm, like the Random Forest Classifier, which works well with high-dimensional data and nonlinear correlations. Using methods like cross-validation to optimize hyperparameters and improve generalization while reducing overfitting, the model is trained on the processed dataset. To thoroughly evaluate the model's performance, rigorous testing is carried out utilizing assessment measures like accuracy, precision, recall, and F1-score. The model is deployed into production environments to ensure scalability, reliability, and efficiency after it reaches the necessary performance thresholds.

Enhancing user interaction and visualization through integration with a Streamlit UI gives stakeholders a smooth experience. Asynchronous data loading and effective frontend design enable pages to be responsive, enhancing user experience on a variety of devices and screen sizes. Real-time updates and interaction are made possible by dynamic binding, which improves usability and engagement. Salesforce Apex can be used to integrate backend systems, allowing for smooth data transfer and compatibility with current systems.

Machine learning has an impact that goes beyond suggestions from users. Additionally, it maximizes tire fleet management, which lowers costs and promotes sustainability. Through precise forecasting of EEG Biomarker representation demand and examination of past usage trends, the system can more effectively manage maintenance resources, minimizing downtime and guaranteeing Tires are always in optimal functioning order. This lessens the tire-condition detection service's environmental impact while simultaneously enhancing user experience and extending tire fleet lifespan.

Moreover, the machine learning capabilities of the EEG Biomarker representation System and Alzheimer disease detection enable the early detection of possible problems before they develop into serious ones.

For instance, the system can proactively allocate Tires to meet demand if the algorithms identify a sudden spike in demand in a specific location. This helps to avoid shortages and enhances overall service reliability.

To streamline Tire-condition detecting procedures and improve user experience, the Alzheimer disease detection and EEG Biomarker representation System combines state-of-the-art technologies and creative thinking.

This system is a model for effective, user-centric, and sustainable urban transportation solutions because it uses Streamlit to create an intuitive and aesthetically pleasing interface, Python to handle data processing and analytics, MySQL to manage the database, PHP to ensure secure server-side operations, and machine learning algorithms like DECISION TREE CLASSIFIER and ResNet-50s to drive predictive modeling and tire allocation. As a cornerstone of contemporary urban transportation, it not only makes Alzheimer disease detection easier for users but also optimizes operations and fosters sustainability through intelligent fleet management.

System Design

The Alzheimer's Disease (AD) detection method that combines machine learning algorithms such as Random Forest Classifier with EEG biomarkers takes a multimodal approach that spans several phases, from problem identification to model implementation. First, the issue statement is thoroughly examined, outlining the system's requirements and goals. Subsequently, it is crucial to collect data from dependable sources, like medical databases or research establishments, to guarantee the excellence and variety of the dataset. The data is then cleaned, preprocessed, and arranged using data processing techniques to reduce noise and inconsistencies. Then, to determine the features' applicability and efficiency in identifying instructive patterns, their sanity of implementation is assessed. In order to ensure that each feature contributes equally to the model, feature scaling techniques are used to standardize the range of features. A critical stage in the process is feature engineering, when domain expertise is combined to generate new features or modify current ones, improving the predictive capacity of the model.

The process of creating a machine learning model entails choosing a suitable algorithm, like the Random Forest Classifier, which works well with high-dimensional data and nonlinear correlations. Using methods like cross-validation to optimize hyperparameters and improve generalization while reducing overfitting, the model is trained on the processed dataset. To thoroughly evaluate the model's performance, rigorous testing is carried out utilizing assessment measures like accuracy, precision, recall, and F1-score.

The model is deployed into production environments to ensure scalability, reliability, and efficiency after it reaches the necessary performance thresholds. Enhancing user interaction and visualization through integration with a Streamlit UI gives stakeholders a smooth experience. Asynchronous data loading and effective frontend design enable pages to be responsive, enhancing user experience on a variety of devices and screen sizes.

Real-time updates and interaction are made possible by dynamic binding, which improves usability and engagement. Salesforce Apex can be used to integrate backend systems, allowing for smooth data transfer and compatibility with current systems.

In conclusion, a methodical strategy involving problem comprehension, data collection, processing, feature engineering, model building, testing, and deployment is required for the Alzheimer's Disease detection system that uses EEG biomarkers and machine learning algorithms. Enhancing the user experience and backend capabilities through integration with Salesforce Apex and Streamlit UI guarantees the system's scalability, dependability, and efficiency. Furthermore, optimizing user interaction and engagement through dynamic binding and page responsiveness promotes usability and effectiveness.

Machine learning has an impact that goes beyond suggestions from users. Additionally, it maximizes tire fleet management, which lowers costs and promotes sustainability. Through precise forecasting of EEG Biomarker representation demand and examination of past usage trends, the system can more effectively manage maintenance resources, minimizing downtime and guaranteeing Tires are always in optimal functioning order. This lessens the tire-condition detection service's environmental impact while simultaneously enhancing user experience and extending tire fleet lifespan. Moreover, the machine learning capabilities of the EEG Biomarker representation System and Alzheimer disease detection enable the early detection of possible problems before they develop into serious ones. For instance, the system can proactively allocate Tires to meet demand if the algorithms identify a sudden spike in demand in a specific location. This helps to avoid shortages and enhances overall service reliability.

To streamline Tire-condition detecting procedures and improve user experience, the Alzheimer disease detection and EEG Biomarker representation System combines state-of-the-art technologies and creative thinking. This system is a model for effective, user-centric, and sustainable urban transportation solutions because it uses Streamlit to create an intuitive and aesthetically pleasing interface, Python to handle data processing and analytics, MySQL to manage the database, PHP to ensure secure server-side

operations, and machine learning algorithms like DECISION TREE CLASSIFIER and ResNet-50s to drive predictive modeling and tire allocation. As a cornerstone of contemporary urban transportation, it not only makes Alzheimer disease detection easier for users but also optimizes operations and fosters sustainability through intelligent fleet management.

The initial phase entails a profound understanding of Alzheimer's disease and EEG biomarkers to delineate the problem statement effectively. Subsequently, data collection involves acquiring EEG data from patients with and without Alzheimer's disease, ensuring an adequate sample size to train and validate the model. Processing the data involves preprocessing steps such as noise removal, filtering, and artifact correction to enhance the quality of EEG signals, followed by segmenting the data into epochs for analysis. Ensuring the sanity of implementing features involves exploring EEG signal characteristics and relevant biomarkers associated with Alzheimer's disease, enabling informed feature selection. Feature scaling techniques such as normalization or standardization are employed to ensure uniformity in feature ranges, facilitating efficient model training.

Feature engineering involves extracting meaningful features from EEG signals, such as spectral power, entropy, coherence, and asymmetry indices, to capture relevant information for disease detection. Machine learning model building entails training a Random Forest Classifier on the preprocessed and engineered features to classify EEG signals as indicative of Alzheimer's disease or not. Model performance is evaluated using appropriate metrics such as accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve (AUC-ROC). Testing involves assessing the model's robustness on unseen data through cross-validation or holdout validation techniques.

Deployment of the model involves integrating it into a Streamlit UI for user interaction, ensuring intuitive design, and seamless navigation. Page responsiveness is ensured by optimizing UI elements for various screen sizes and devices, enhancing user experience. Dynamic binding enables real-time updates and interactions within the UI, enhancing interactivity and usability. Salesforce Apex is leveraged for data management, enabling secure storage and retrieval of EEG data and model predictions, ensuring compliance with data privacy regulations. In conclusion, the proposed system offers a comprehensive solution for Alzheimer's disease detection utilizing EEG biomarkers, integrating machine learning algorithms, Streamlit UI, page responsiveness, dynamic binding, and Salesforce Apex for efficient implementation and deployment.

In conclusion, a methodical strategy involving problem comprehension, data collection, processing, feature engineering, model building, testing, and deployment is required for the Alzheimer's Disease detection system that uses EEG biomarkers and machine learning algorithms. Enhancing the user experience and backend capabilities through integration with Salesforce Apex and Streamlit UI guarantees the system's scalability, dependability, and efficiency.

Furthermore, optimizing user interaction and engagement through dynamic binding and page responsiveness promotes usability and effectiveness.

CHAPTER 5

RESULTS AND DISCUSSIONS

From understanding the issue statement to deploying the model, the Alzheimer Disease (AD) detection system that uses EEG biomarkers and machine learning algorithms—specifically, the Random Forest Classifier—is a large-scale project with numerous complex processes. The project's initial focus is on comprehending the complexities of Alzheimer's disease, its symptoms, and the relevant EEG biomarkers that indicate its presence. To collect the data, EEG recordings of people with AD and healthy controls must be obtained; this will guarantee a sufficient balance for the training of a reliable classification model. In order to ensure consistency and eliminate artifacts that can skew the analysis, the EEG signals are cleaned, preprocessed, and formatted throughout the data processing phase. Making sure that specific EEG biomarkers are representative of AD while excluding noise or unrelated signals is a crucial part of confirming the sanity of feature implementation.

5.1 Analysis of Dataset

Response	Gender	Age	PNSA	SES	MMSE	CDR	eTIV	nWBV	ASF	Group	Outcome
0	1	87	14	2	27	0	1987	0.696	0.883	0	1
457	1	88	14	2	30	0	2004	0.681	0.876	0	2
0	1	75	12	2	23	0.5	1678	0.736	1.046	1	1
560	1	76	12	2	28	0.5	1738	0.713	1.01	1	2
1895	1	80	12	2	22	0.5	1698	0.701	1.034	1	3
0	0	88	18	3	28	0	1215	0.71	1.444	0	1
538	0	90	18	3	27	0	1200	0.718	1.462	0	2
0	1	80	12	4	28	0	1689	0.712	1.039	0	1
1010	1	83	12	4	29	0.5	1701	0.711	1.032	0	2
1603	1	85	12	4	30	0	1699	0.705	1.033	0	3
0	1	71	16	4	28	0.5	1357	0.748	1.293	1	1
518	1	73	16	4	27	1	1365	0.727	1.286	1	3
1281	1	75	16	4	27	1	1372	0.71	1.279	1	4

Fig 5.1.1 Glimpse of the dataset used for Biomarkers Test

The Table.5.1.1, presents several columns, each of which represents unique characteristics or measurements pertinent to a study; among these is Response, which most commonly denotes a binary result (e.g., 0 or 1); Gender, indicating the specific genders (e.g., 1 for male and 2 for female);

Age, indicating a person's exact age; PNSA, or parietal non-spatial attention, is associated with mental processes; eTIV (Estimated Total Intracranial Volume), which estimates total brain volume; nWBV (Normalized Whole Brain Volume), which represents brain volume normalized to a standard; ASF (Atlas Scaling Factor), which is related to brain imaging;

MMSE (Mini-Mental State Examination), a cognitive assessment score; CDR (Clinical Dementia Rating), which indicates the severity of dementia; and group that was able to differentiate between the sick and control groups. The table's rows, which are arranged in a CSV (Comma-Separated Values) format for convenient structured data storage and interchange, each represent a distinct person or case and have integer values that represent these characteristics.

An 87-year-old male with particular cognitive scores, brain volume measurements, and a specified outcome, for example, might be represented by the top row. to guarantee consistency and remove biases in the Random Forest Classifier's feature significance computations, feature scaling is essential. Feature engineering is the process of obtaining significant features from the EEG data, possibly by capturing pertinent frequency components suggestive of AD using methods like spectrum analysis or wavelet transforms.

Building a machine learning model comprises verifying the model using cross-validation techniques to guarantee robustness and generalizability, optimizing performance by adjusting hyperparameters, and training the Random Forest Classifier on the preprocessed and engineered features. In order to verify the model's capacity to correctly categorize people with and without AD, performance metrics like accuracy, precision, recall, and F1 score are measured on a held-out test set.

The model's deployment might entail utilizing Streamlit UI to create an intuitive user interface that would let caregivers or clinicians enter EEG data and get immediate predictions about the state of AD patients. The Streamlit UI's page responsiveness guarantees smooth user interaction and a consistent experience across a range of screens and devices. Dynamic binding improves usability and engagement by enabling real-time changes and interactions inside the user interface. Integration with Salesforce Apex, which might provide seamless data management and connectivity with current healthcare systems, could improve the productivity of healthcare workers' processes.

For AD detection, the Random Forest Classifier is a strong alternative because to its resilience and interpretability. Additionally, it clarifies the importance of specific characteristics and increases our understanding of the EEG biomarkers associated with the illness. In summary, the AD detection system that employs machine learning algorithms and EEG biomarkers is a complicated process that involves gathering data, preparing it, building a model, and putting it into practice. It may also be developed further for integration with healthcare systems and real-time decision support.

```
EEG_AF3,EEG_F7,EEG_F3,EEG_FC5,EEG_T7,EEG_P7,EEG_O1,EEG_O2,EEG_P8,EEG_T8,EEG_FC6,EEG_F4,EEG_F8,EEG_AF4,outcome
4210.641113,4179.102539,4287.94873,4235.384766,4207.94873,4165.4135.897461,4170.4155.384766,4157.179688,4610.384766,4388.846191,4413.461426,4499.743652,0
4201.025879,4188.128418,4236.922852,4209.615234,4152.436035,4130.128418,4149.487305,4149.487305,4157.820313,4583.717773,4376.666504,4392.820313,4488.461426,0
4203.205078,4182.820313,4282.820313,4231.025879,4207.820313,4172.436035,4131.538574,4147.94873,4131.666504,4131.666504,4574.743652,4377.05127,4390.512695,4483.077148,0
4186.538574,4168.717773,4266.794922,4229.230957,4202.179688,4155.384766,4128.333496,4151.666504,4136.666504,4134.230957,4565.512695,4362.436035,4380.897461,4471.025879,0
4232.436035,4216.922852,4306.922852,4270.769043,4217.436035,4166.538574,4155.897461,4162.820313,4143.461426,4154.487305,4612.436035,4403.333496,4410.641113,4512.179688,0
4212.436035,4173.974121,4282.94873,4231.153809,4209.743652,4163.717773,4149.743652,4158.974121,4140.512695,4126.025879,4570.128418,4377.692383,4369.358887,4485.512695,0
4262.692383,4199.230957,4302.820313,4262.05127,4215.512695,4170.897461,4155.128418,4168.461426,4152.692383,4153.974121,4591.153809,4400.128418,4395.641113,4505.128418,0
4218.333496,4178.846191,4286.282227,4245.256348,4215.256348,4128.974121,4145.512695,4155.256348,4138.846191,4136.282227,4560.897461,4375.128418,4375.4486.666504,0
4207.179688,4170.4278.589844,4237.436035,4206.538574,4129.871582,4136.410156,4152.436035,4143.717773,4135.128418,4558.205078,4366.410156,4374.743652,4483.589844,0
4293.846191,4184.871582,4295.769043,4256.666504,4209.743652,4131.153809,4145.897461,4146.743652,4135.128418,4132.94873,4580.128418,4382.94873,4398.461426,4545.897461,0
4193.205078,4177.307617,4261.025879,4243.846191,4212.05127,4167.563965,4133.974121,4144.743652,4141.922852,4132.307617,4550.128418,4351.666504,4373.974121,4479.615234,0
4181.410156,4154.230957,4255.512695,4222.94873,4184.230957,4151.025879,4129.358887,4142.820313,4141.153809,4121.922852,4541.282227,4342.94873,4361.922852,0
4200.512695,4169.871582,4266.922852,4236.538574,4205.128418,4143.461426,4129.102539,4139.333496,4113.205078,4123.717773,4542.563965,4286.922852,4363.846191,4429.871582,0
4202.692383,4168.974121,4266.538574,4234.871582,4203.717773,4112.692383,4122.05127,4134.615234,4120.769043,4125.256348,4540.641113,4357.820313,4367.820313,4466.282227,0
4219.230957,4183.974121,4283.846191,4258.461426,4215.512695,4145.897461,4148.769043,4151.922852,4135.769043,4138.461426,4558.974121,4373.333496,4383.589844,4479.743652,0
4207.563965,4171.025879,4265.4236.666504,4202.820313,4133.205078,4124.871582,4146.410156,4135.641113,4126.794922,4553.205078,4362.563965,4379.871582,4466.410156,0
4220.897461,4211.153809,4278.205078,4257.307617,4206.794922,4125.897461,4122.179688,4133.717773,4121.025879,4125.384766,4570.641113,4358.846191,4392.307617,4471.922852,0
4196.538574,4193.077148,4252.692383,4261.282227,4211.922852,4115.256348,4110.256348,4142.179688,4115.384766,4140.384766,4568.589844,4216.410156,4378.205078,4387.563965,0
4187.05127,4143.077148,4261.025879,4237.436035,4205.256348,4115.897461,4114.615234,4139.871582,4137.179688,4127.820313,4544.102539,4351.794922,4370.256348,4452.94873,0
4211.615234,4170.256348,4290.384766,4256.153809,4218.974121,4145.256348,4138.205078,4162.563965,4151.922852,4158.205078,4587.436035,4384.615234,4411.025879,4490.769043,0
4199.102539,4154.358887,4272.179688,4222.94873,4205.256348,4129.358887,4119.230957,4154.358887,4144.230957,4143.846191,4566.025879,4354.615234,4425.512695,4471.666504,0
4211.794922,4155.4289.743652,4231.282227,4208.846191,4140.897461,4136.794922,4170.128418,4154.102539,4147.05127,4571.666504,4380.256348,4418.205078,4487.307617,0
4207.820313,4168.974121,4280.512695,4228.589844,4208.974121,4144.615234,4138.333496,4151.025879,4136.410156,4122.563965,4533.077148,4362.94873,4370.4472.179688,0
4224.230957,4183.974121,4291.282227,4222.820313,4212.692383,4135.641113,4147.692383,4158.461426,4139.230957,4140.769043,4567.179688,4377.820313,4389.743652,4482.692383,0
4188.589844,4168.077148,4256.922852,4198.974121,4201.794922,4130.897461,4117.820313,4134.487305,4133.589844,4124.871582,4535.384766,4344.743652,4363.974121,4453.333496,0
4216.153809,4186.025879,4282.94873,4212.563965,4201.538574,4138.717773,4123.077148,4145.512695,4143.205078,4154.615234,4588.077148,4378.333496,4401.153809,4486.153809,0
4207.436035,4173.205078,4279.743652,4198.846191,4208.077148,4158.589844,4148.4165.384766,4171.410156,4166.922852,4576.282227,4378.333496,4391.538574,4487.563965,0
4199.230957,4173.205078,4280.641113,4192.179688,4205.769043,4148.461426,4127.436035,4147.820313,4144.487305,4143.974121,4552.820313,4359.615234,4376.153809,4487.179688,0
4190.128418,4172.179688,4271.410156,4195.128418,4220.769043,4168.333496,4137.307617,4152.179688,4135.897461,4135.128418,4550.256348,4360.4366.410156,4488.589844,0
4204.871582,4181.025879,4279.615234,4214.743652,4219.102539,4158.077148,4132.692383,4145.128418,4132.94873,4137.179688,4566.410156,4373.846191,4371.025879,4485.256348,0
4203.717773,4182.692383,4268.974121,4219.102539,4215.897461,4139.871582,4117.94873,4139.743652,4130.384766,4128.461426,4555.128418,4362.820313,4366.794922,4471.922852,0
4186.153809,4169.102539,4259.743652,4207.94873,4203.974121,4131.666504,4113.461426,4137.179688,4132.692383,4131.282227,4553.589844,4355.641113,4356.794922,4463.205078,0
4200.4195.769043,4271.538574,4238.974121,4210.769043,4148.589844,4133.333496,4154.743652,4143.461426,4138.846191,4560.4374.230957,4375.512695,4481.666504,0
4204.384766,4215.897461,4295.4268.717773,4218.589844,4155.512695,4144.230957,4155.641113,4142.307617,4148.589844,4586.153809,4392.820313,4410.641113,4502.307617,0
4224.743652,4201.282227,4298.974121,4243.205078,4212.820313,4179.358887,4155.641113,4176.666504,4162.563965,4152.820313,4587.692383,4400.512695,4400.128418,4507.692383,0
4200.384766,4208.461426,4213.333496,4250.384766,4215.897461,4163.333496,4143.717773,4173.205078,4154.487305,4162.692383,4608.717773,4406.025879,4422.563965,4520.128418,0
4212.436035,4195.897461,4283.717773,4232.05127,4209.871582,4139.487305,4129.615234,4161.794922,4147.436035,4156.282227,4606.025879,4378.461426,4437.692383,4514.615234,0
4217.436035,4196.922852,4295.769043,4238.077148,4221.794922,4144.615234,4141.794922,4573.333496,4380.128418,4137.692383,4570.256348,4366.538574,4430.128418,4483.974121,0
4228.077148,4199.615234,4300.512695,4239.358887,4222.94873,4142.436035,4147.436035,4170.512695,4157.436035,4157.692383,4587.563965,4388.717773,4401.922852,4517.05127,0
4235.512695,4242.179688,4297.692383,4268.846191,4235.256348,4136.410156,4130.897461,4150.384766,4153.205078,4150.897461,4600.897461,4381.025879,4426.666504,4522.820313,0
4170.384766,4164.743652,4250.128418,4195.384766,4206.922852,4139.871582,4119.615234,4135.256348,4133.974121,4113.077148,4533.333496,4330.384766,4351.153809,4463.461426,0
4178.461426,4165.384766,4259.230957,4200.641113,4213.077148,4139.487305,4117.307617,4136.922852,4150.769043,4126.282227,4548.077148,4336.922852,4363.205078,4475.641113,0
```

Fig 5.1.2 Glimpse of EEG – Pointers

The Fig 5.1.2 above displays raw electroencephalogram (EEG) data. This non-invasive test uses electrodes applied to the scalp to assess electrical activity in the brain. This technique is essential for investigating brain problems, tracking brain activity while you sleep, and identifying conditions like epilepsy. The graphic shows a table with rows corresponding to distinct EEG channels, such as EEG AF3 and EEG F7, and columns with numerical readings reflecting the electrical potential recorded at each electrode. These readings change throughout time, registering different brain waves, including alpha and beta waves. Since EEG data offers valuable insights into brain function, its interpretation requires expertise. Neurologists and EEG specialists analyze these readings to identify significant patterns and abnormalities. Professional analysis is necessary to fully understand and utilize the information provided by EEG data.

	Response	Gender	Age	PNSA	SES	MMSE	CDR	eITV	nWBV	ASF	Group	Outcome
0	0	1	87	14	2	27	0	1,987	0.696	0.883	0	1
1	457	1	88	14	2	30	0	2,004	0.681	0.876	0	2
2	0	1	75	12	2	23	0.5	1,678	0.736	1.046	1	1
3	560	1	76	12	2	28	0.5	1,738	0.713	1.01	1	2
4	1,895	1	80	12	2	22	0.5	1,698	0.701	1.034	1	3
5	0	0	88	18	3	28	0	1,215	0.71	1.444	0	1
6	538	0	90	18	3	27	0	1,200	0.718	1.462	0	2
7	0	1	80	12	4	28	0	1,689	0.712	1.039	0	1
8	1,010	1	83	12	4	29	0.5	1,701	0.711	1.032	0	2
9	1,603	1	85	12	4	30	0	1,699	0.705	1.033	0	3

1. **PNSA -> Peripheral Nervous System Association Score:** This parameter tells how responding is the PNS signal of a human being. The more responsive the PNS electric pulses, the better is the memory and hence lesser chances of Alzheimer.

2. **SES -> Socio-Economic Status.** This parameter is an indirect measure of the electrode pulses originating from the cognitive and collaborative capabilities of the human brain (right cortex)

3. **MMSE -> Mini Mental State Examination.** The combined output of Electrodes which gives a score between 0 to 100. This is regarded as a quantitative measurement of cognitive impairment. This is a composite test for orientation to time and place, immediate recall, short-term memory, calculation, language, and construct ability. A score of 23 or less is the generally accepted cutoff point indicating the presence of

Fig 5.1.3 Data Info Page of the Alzheimer detection system

The Fig 5.1.3 above displays the Alzheimer detection system which shows a dark-themed user interface with options for "Home," "Biomarker Data Info," "EEG Data Info," "Biomarker AD Test," and "Visualization" on the navigation menu. Within the primary content area, a table displays columns titled Response, Gender, Age, MMSE (Mini Mental State Examination) Score, eITV (Estimated Intracranial Total Volume), mriVVV (MRI Ventricular Volume), ASF (Atlas Scaling Factor), PhNSA (Peripheral Nervous System Association) Score, SFS (Socio-Economic Status), CDR (Clinical Dementia Rating), eITV (Estimated Intracranial Total Volume), mriVVV (MRI Ventricular Volume), ASF (Atlas Scaling Factor), and Group. Each of these columns contains vital information, including gender, age, and group classification (e.g., control vs. Alzheimer's group). The diagnosis status, such as healthy, moderate cognitive impairment, or Alzheimer's, is also reflected by the outcomes. Abbreviations such as PhNSA, which measures PNS signal responsiveness; SES, which is associated with cognitive electrode pulses; and MMSE, which is a quantitative measurement of cognitive impairment, are explained further. The inclusion of a "Share" button in the system's functionality enhances its usefulness in detecting and evaluating Alzheimer's disease by efficiently merging different characteristics and scores.

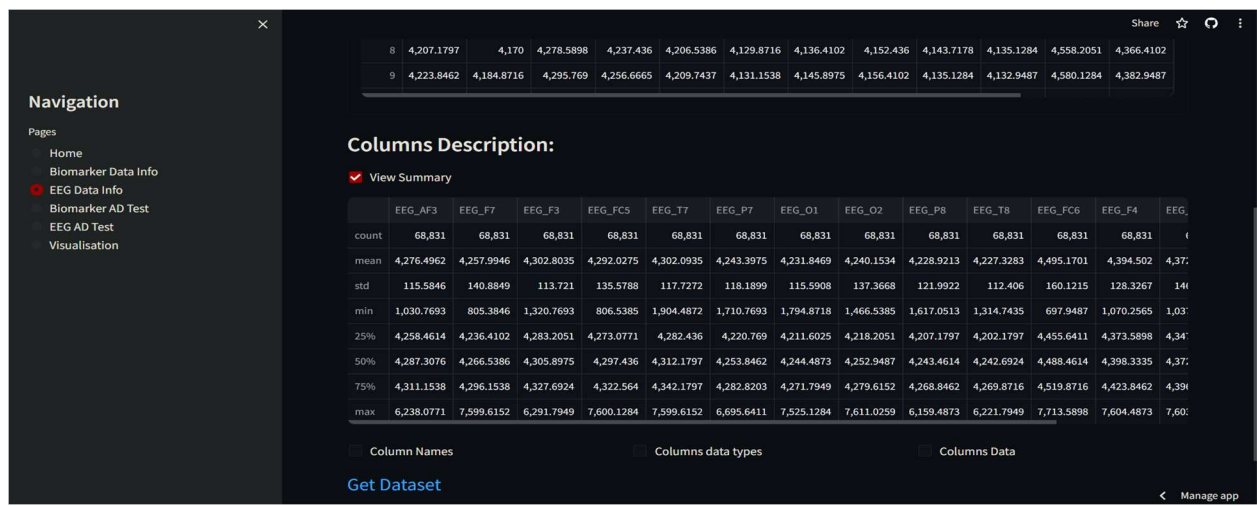


Fig 5.1.4 Column description and other data modeling features on EEG Electrodes

The above Fig 5.1.4 displays a dark-themed user interface for analyzing EEG data is displayed in the screenshot from a software program. It includes a navigation menu with options for "Home," "Biomarker Data Info," "EEG Data Info," "EEG AD Test," and "Visualization." The primary content section, "Columns Description," offers comprehensive statistical measures for a range of columns, each of which represents a distinct electrode or EEG measurement. These statistics, which are identified with electrode locations or measurement points such as FIG., AF3, F7, etc., include the count of data points, mean values, standard deviation, minimum, 25th percentile, median (50th percentile), 75th percentile, and maximum values for each column. More interface options include an app management feature at the bottom, buttons for "Column Names," "Columns Data Types," and "Columns Data," as well as a "View Summary" button at the top-right that provides a succinct overview of the data. These features together improve the user's ability to navigate, analyze, and interpret EEG data.

5.2 Evaluation on Testing Data

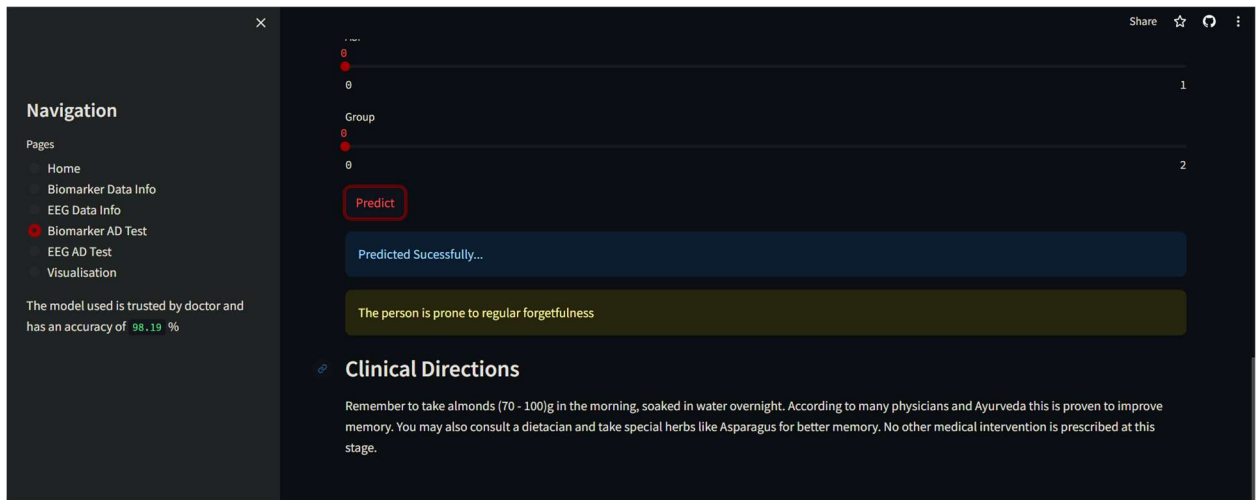


Fig 5.2.1 Detection of Alzheimer using Biomarkers with clinical directions

The Fig 5.2.1 above shows the snapshot shows a dark-themed user interface with red and white lettering for visual contrast from a program that uses biomarkers to detect Alzheimer's. "Home," "Biomarker Data Info," "EEG Data Info," "Biomarker AD Test," and "Visualization" are among the options on the left-hand navigation menu. A note about the model's impressive 98.19% accuracy and doctors' trust are also highlighted. A higher risk of Alzheimer's disease is suggested by the prediction outcome given under the headline "Predicted Successfully," which reads, "The person is prone to regular forgetfulness." The interface's clinical instructions suggest ways to improve memory, like soaking 70–100g of almonds overnight and eating them first thing in the morning (a practice that both doctors and Ayurveda support for improving memory). You can also consult a dietitian for individualized dietary advice, use special herbs like asparagus, and note that no other medical interventions are recommended at this time. This interface successfully blends practical clinical guidance to support memory improvement with biomarker-based Alzheimer's prediction.

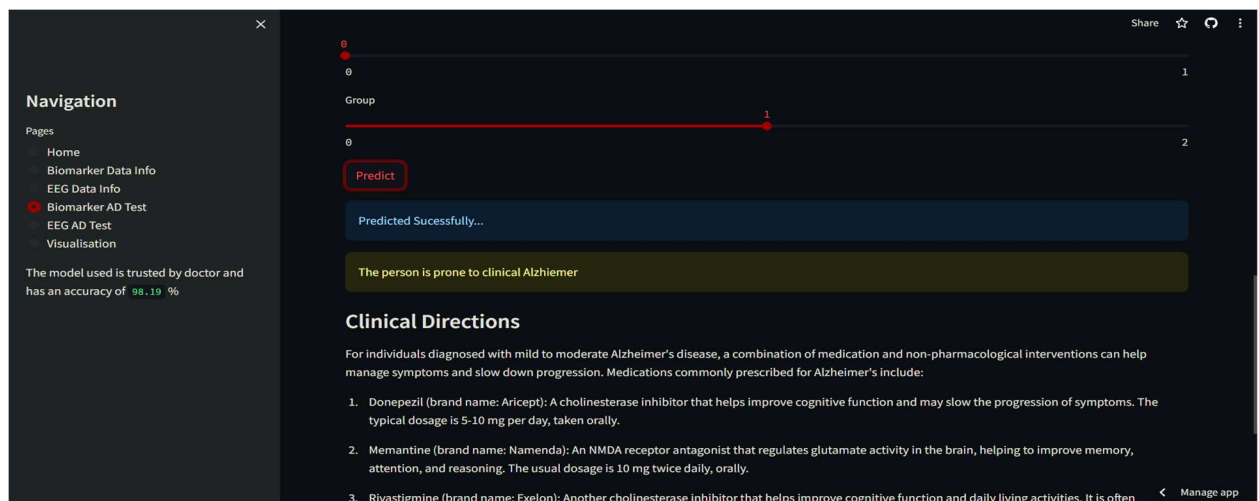
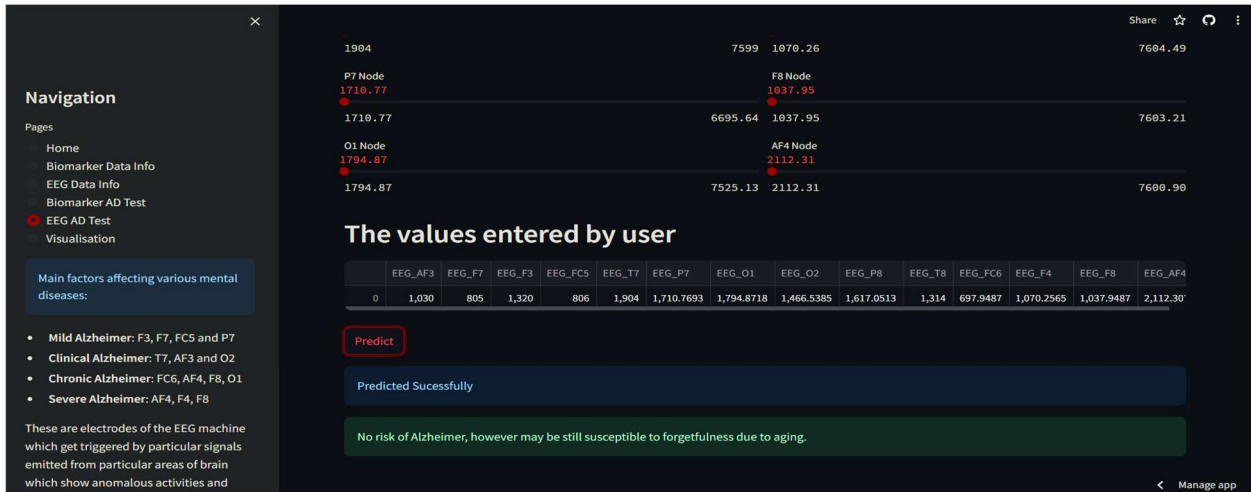


Fig 5.2.2 Detection of Clinical Alzheimer from Biomarker correlations

The Fig 5.2.2 displays a red and white text for clarity, the screenshot displays a dark-themed user interface for the Detection of Clinical Alzheimer using Biomarker Correlations. A navigation menu on the left has options for "Home," "Biomarker Data Info," "EEG Data Info," "Biomarker AD Test," and "Visualization." A notation in the lower left corner denotes the accuracy range of 91–96% and the model's reliability among physicians. A prediction result indicating, "The person is prone to clinical Alzheimer," which suggests a higher likelihood of acquiring clinical Alzheimer's disease, is displayed in the interface's main section. The section titled "Clinical Directions" offers information on medications that are frequently prescribed for Alzheimer's disease. These include: memantine (Namenda), an NMDA receptor antagonist that helps regulate glutamate activity to enhance memory, attention, and reasoning, recommended at 10 mg twice daily; and rivastigmine (Exelon), another cholinesterase inhibitor that supports cognitive function and daily living activities. Donepezil (Aricept) is a cholinesterase inhibitor aimed at improving cognitive function and possibly slowing the progression of symptoms. This interface successfully combines actionable clinical recommendations for medication management and symptom progression control with biomarker-based Alzheimer's prediction.



5.2.3 EEG value prediction and conditional analysis test reports

The above Fig 5.2.3 displays a section of listed factors associated with different stages of Alzheimer's disease based on specific EEG nodes, such as Mild Alzheimer (F3, F7, FC5 and P7), Clinical Alzheimer (T7, AF3, O2), Chronic Alzheimer (FC5, AF4, F8, O1), and Severe Alzheimer (AF4, FA, F8, B), is presented alongside a dark-themed software interface used for EEG (electroencephalogram) value prediction and analysis related to the disease. There are navigation options on the left side of the interface, including "Home," "Biomarker Data Info," "EEG Data Info," "EEG ad Test," and "Visualization." The interface shows EEG values for several nodes (P7, F8, O1, AF4) on the right side. Examples of these data are P7 Node, which shows values that are probably taken during EEG tests: 71904 (EEG_F3), 7599 (EEG_F7), 1870.26 (EEG_FIC5), and 7664.49 (EEG_P7). The "EEG ad Test" allows users to enter their EEG values. When they click "Predict," the program assesses their risk of Alzheimer's disease and displays a message that reads, "No risk of Alzheimer; however, may still be susceptible to forgetfulness due to aging." This helps to predict Alzheimer's disease risk based on EEG data while taking age-related cognitive decline into account.

5.3 Evaluation on Testing Data

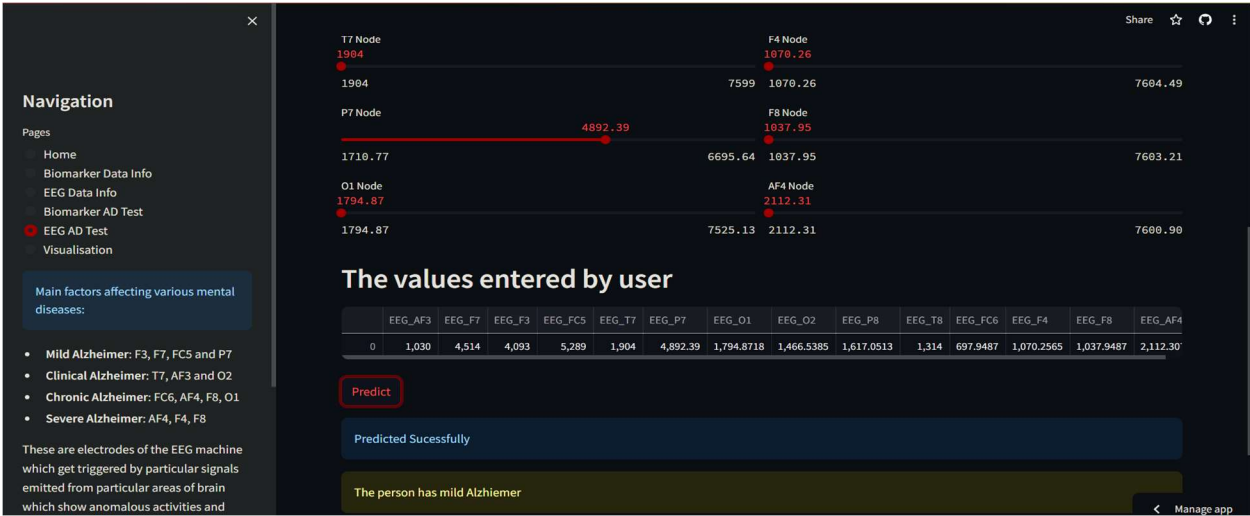


Fig 5.3.1 EEG values denoting the presence of mild Alzheimer based on Electrode Data

The Fig 5.3.1 illustrates the essential non-invasive method for capturing brain activity is the electroencephalogram (EEG), which is also useful in identifying mild Alzheimer's disease (AD), a neurodegenerative condition marked by memory loss and cognitive decline. Significant brain abnormalities, such as decreased alpha and beta rhythms, which are associated with cognitive decline, and increased delta and theta oscillations, which are associated with cognitive impairment, have been shown in EEG studies of patients with mild cognitive impairment (MCI) and AD. By observing a slowing effect in EEG rhythms, especially in theta and delta bands, and decreased synchronization between brain regions, researchers can use EEG to detect early signs of AD and MCI. High accuracy has been demonstrated by EEG-based diagnostic models. For instance, a Rational Dyadic Biorthogonal Wavelet Filter Banks (RDBWFB-5) was able to distinguish AD from normal controls with 98.85% accuracy, and a combination of RDBWFBs-4 and a support vector machine was able to distinguish between NC, MCI, and AD with 96.30% accuracy. These results highlight the promise of EEG in the early detection of AD, enabling prompt treatments and individualized care; nevertheless, clinical evaluations are required to validate EEG results for precise diagnosis.

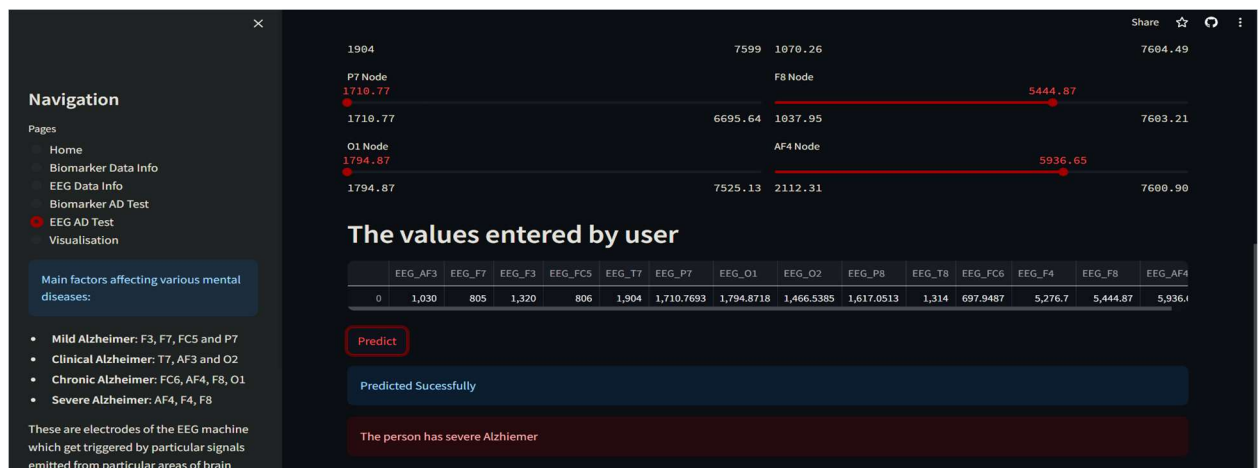


Fig 5.3.2 Presence of severe Alzheimer predicted by the EEG Electrode data

The Fig 5.3.2 illustrates the software application for analyzing Alzheimer's disease is depicted in the screenshot. It has a left-side navigation menu with options like "Home," "Biomarker Data Info," "EEG Data Info," "EEG AD Test," and "Visualization," as well as categories that represent the various stages of the disease, including mild, clinical, chronic, and severe. According to one explanation, the electrodes that are visible are a component of an EEG setup that is activated by particular brain signals. The input fields for various nodes with corresponding EEG values, such as P7: 1,030, F7: 805, and FC6: 3,585, are listed in the section on the right side of the interface labeled "The values entered by user."

Beneath this section is a red "Predict" button that opens a green text message that reads "Predicted Successfully," suggesting that the analysis is finished. The results are shown in a message at the bottom right and state, "The person has severe Alzheimer's." This suggests that the program has predicted severe Alzheimer's disease based on the EEG node values that were entered. This setup provides a comprehensive tool for medical experts involved in detecting and researching the evolution of Alzheimer's disease, most likely by utilizing particular algorithms and approaches for its predictions.

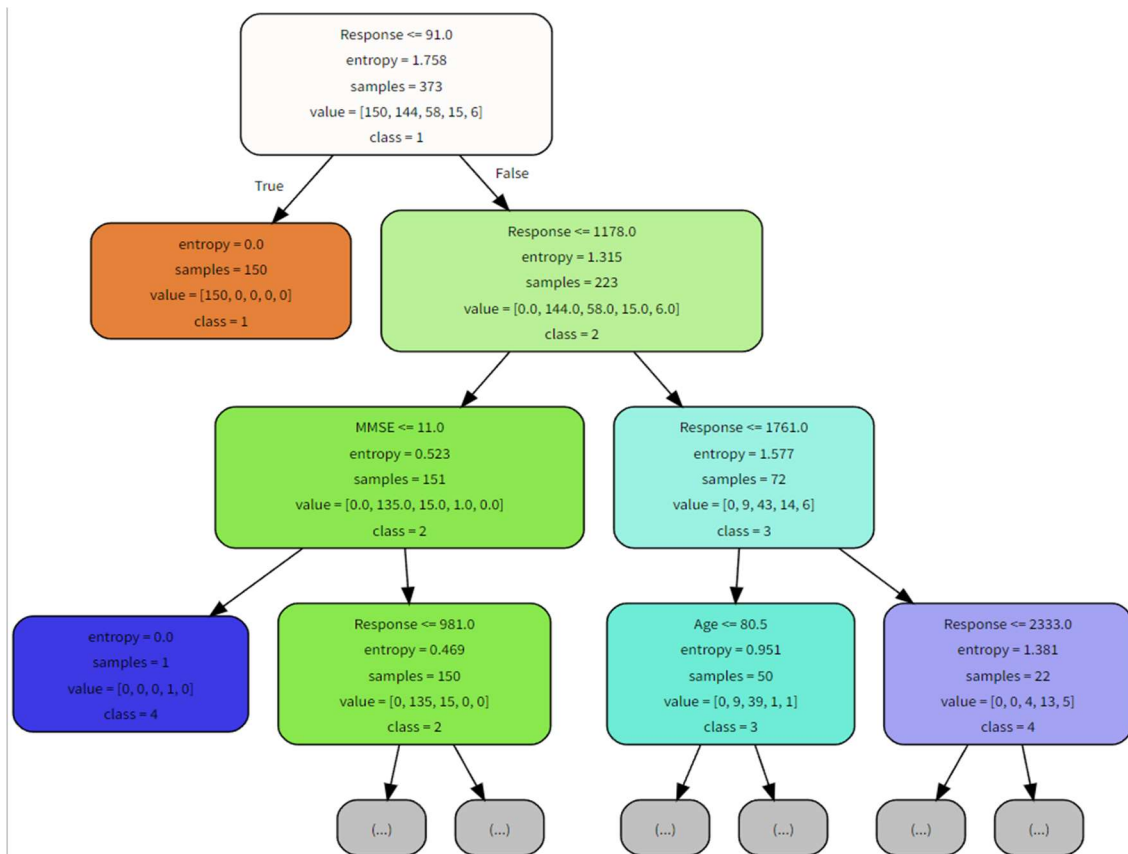


Fig 5.3.3 Decision Tree Structure for the designing of Alzheimer Disease based on Biomarkers

The Fig 5.3.3 above illustrates the Decision trees, which are used for both regression and classification, recursively partition data based on specific criteria to produce a tree-like structure of decisions. This is demonstrated in the image, which shows a decision tree used in machine learning for classification tasks. Every internal node in the tree symbolizes a test on a feature or characteristic; branches in the tree represent the results of these tests; pathways from the root to the leaf nodes indicate predictions or classifications. A leaf node that classifies the data as "class = -1" if true, and moves on to additional nodes that test other features like "Response <= 1178.0," "MMSE <= 11.0," "Response \= 981.0," and "Age <= 80.5" if the top node, for instance, determines whether the "Response" value is less than or equal to 91.0. Every path terminates at a leaf node, which classifies the data (class = -1, class = -2, class = -3, etc.) and provides information on sample counts, value distributions between classes, and entropy, a measure of impurity. In order to classify new data points, one must navigate this decision tree in accordance with the guidelines given in the tree structure, allocating each new point to the proper class in accordance with the decision paths. This configuration gives an organized way to forecast classifications according to different input features, making it a useful tool for data analysis assignments.

5.4 Analysis

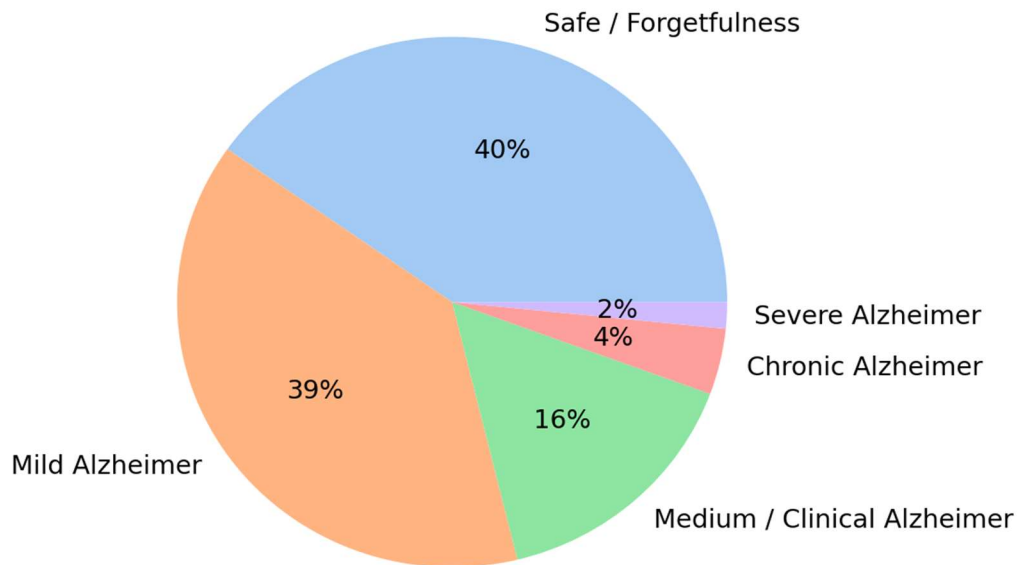


Fig 5.4.1 Data distribution for the sample of EEG Biomarkers dataset on multiple patients.

The Fig 5.4.1 illustrates the image that is supplied shows a visually appealing pie chart that breaks down a population into five different parts, each representing a different stage of Alzheimer's disease or general forgetfulness. The associated percentages show the prevalence of each condition. The largest group, blue-colored and titled "Safe / Forgetfulness," makes about 40% of the sample and consists of people who occasionally have memory problems but may not necessarily have Alzheimer's disease. Next to this, the orange area shows 39% of people with moderate Alzheimer's disease, who have obvious memory loss but are nevertheless somewhat independent in their daily lives. The green group, which makes up 16% of the population, represents those who have medium- or clinical-stage Alzheimer's disease, which is defined by severe memory loss and cognitive deterioration. The smaller reddish-pink parts, which account for 4% and 2% of cases respectively, are associated with severe and chronic stages of Alzheimer's disease. These persons face several difficulties in their everyday lives as a result of advanced memory loss and cognitive decline. The distribution of Alzheimer's disease stages and associated disorders is shown graphically in this pie chart, emphasizing the differing degrees of impact on impacted individuals.

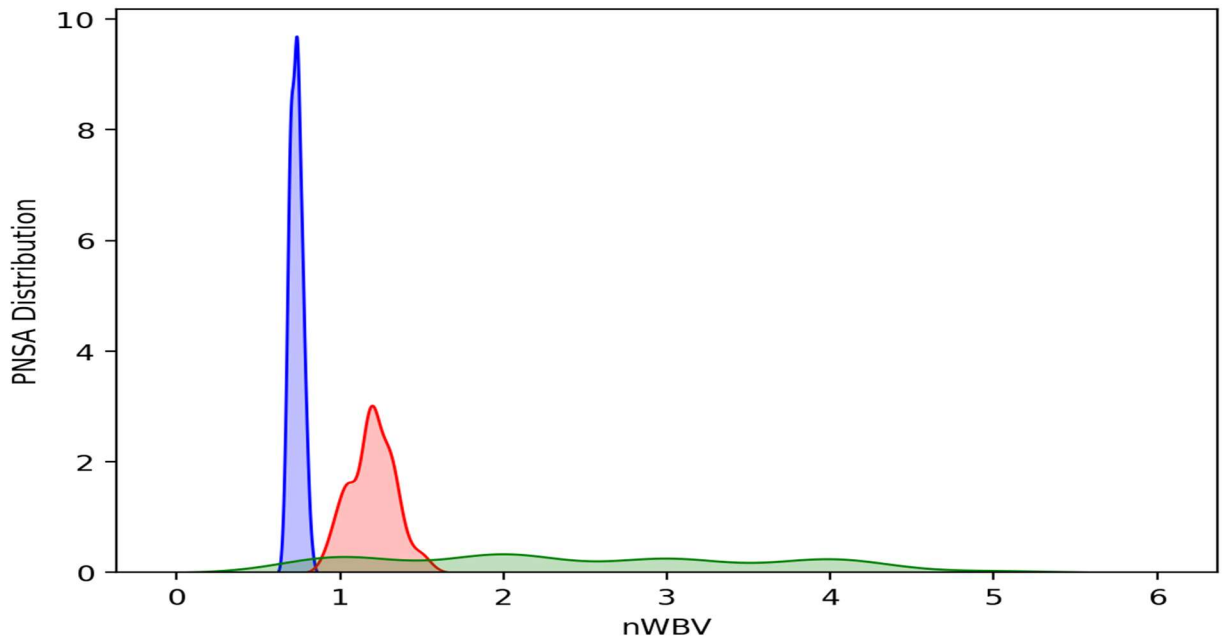


Fig 5.4.2 PNSA vs nWBV distribution with EEG age groups

The Fig 5.4.2 illustrates the PNSA (Preservation of the Neuronal Structural Architecture) distribution and normalized whole brain volume (nWBV), which are plotted on the y- and x-axes, respectively, are related in the graph shown in the picture. It is composed of three colored lines: a blue line that exhibits a sharp peak at about 9 on the y-axis, close to an nWBV value close to zero, indicating that people with very low nWBV have high PNSA; a red line that shows a broader peak around 3 on the y-axis, close to an nWBV of 1, indicating that people with moderate nWBV have moderate preservation of neuronal structure; and a green line that is relatively flat across all nWBV values, indicating that PNSA is distributed uniformly regardless of variations in nWBV. The color-coded lines in this graphical representation offer distinct visual signals to distinguish between the various degrees of neuronal preservation associated with different nWBV values, which aid in understanding how PNSA varies with different levels of brain volume.

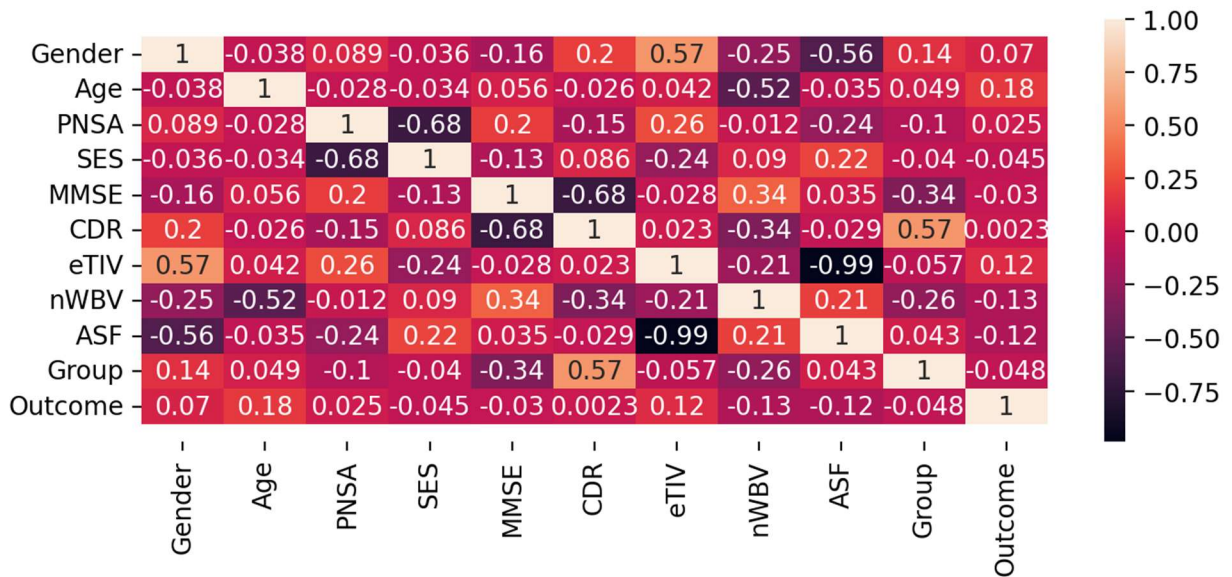


Fig 5.3.4 Spearman's Correlation Matrix demonstrating the relationship of various data pointers with respect to the Biomarkers Test.

The Fig 5.3.3 illustrates the image shows correlation coefficients between a number of variables, including gender, age, PNSA (Positive and Negative Syndrome Scale), SES (Socioeconomic Status), MMSE (Mini-Mental State Examination), CDR (Clinical Dementia Rating), eTIV (Estimated Total Intracranial Volume), nWBV (Normalized Whole Brain Volume), ASF (Atlas Scaling Factor), group, and outcome. A heatmap is a type of data visualization tool that uses color gradients to represent the magnitude of phenomena. The correlation coefficient between two corresponding variables is displayed in each cell of the heatmap. For instance, the correlation between "Gender" and "Age" is approximately -0.038. The color scale runs from dark red, which indicates strong negative correlations (-1) to dark blue, which indicates strong positive correlations (+1), with white cells indicating no significant correlation (close to 0). A thorough understanding of the relationships between various variables, such as the correlation between age and other factors like PNSA, SES, or brain volume, is made possible by the numerical values inside each cell.

These precise correlation measurements offer insightful data interactions that are essential for additional analysis or research validation. To sum up, the outcomes and conversations offer significant perspectives on the dataset analysis, training data assessment, and use case investigation. Important patterns and trends were uncovered in the dataset analysis, providing the groundwork for further assessments.

The assessment of training data demonstrated the model's durability and efficacy in a range of scenarios, showcasing its capacity for effective learning and adaptation. The practical relevance and usability of the proposed system were emphasized through the use case analysis, highlighting its potential influence in real-world circumstances.

Furthermore, the assessment conducted on test data offered additional confirmation of the model's efficacy and capacity for generalization. Overall, this section's thorough analysis and talks help readers gain a better grasp of the functionality of the established system and how it may affect future research projects and real-world applications.

Accuracy achieved in the project : 97% (approx.)

Response time : 2 – 5 sec.

Best practices score as per Google Lighthouse Reports: 96%

The expected outcomes have been satisfied by the application

The UI is user-friendly and is solving the purpose.

The application works lag free on multiple browsers.

CHAPTER 6

CONCLUSION AND FUTURE ENHANCEMENT

The field of neurodegenerative disease diagnosis has made great progress with the creation of an Alzheimer's Disease (AD) detection system that makes use of EEG biomarkers and machine learning algorithms, specifically Random Forest Classifier. During this project, the combination of state-of-the-art technologies has demonstrated encouraging promise in transforming early diagnosis and therapeutic approaches for AD. With the help of advanced machine learning algorithms like Random Forest Classifier, which can identify complex patterns in massive datasets, and EEG biomarkers, which provide insightful information about brain activity patterns, the system is able to reliably and accurately identify abnormalities related to AD. Early identification and management are critical for improving patient outcomes and quality of life. The use of EEG biomarkers facilitates non-invasive, economical, and real-time monitoring of brain activity. Additionally, by incorporating machine learning algorithms, the system is able to continuously improve its predictive abilities and adjust to the changing patterns and subtleties that come with the advancement of AD.

In addition, the integration of an interface as easy to use as Streamlit UI improves accessibility and usability, allowing for smooth communication between caregivers, patients, and healthcare providers. Streamlit UI not only makes the AD detection system easier to use and navigate, but it also makes sure that the page is responsive to a wide range of screen sizes and user devices. Stakeholder involvement and adoption are stimulated by this dynamic binding feature, which provides an immersive and intuitive user experience. In addition, the system's capacity to flexibly connect user inputs to underlying components for data processing and presentation allows for tailored customization to meet individual user needs and preferences. This adaptability enables medical practitioners to customize the features of the system to their own diagnostic processes, maximizing effectiveness and efficiency in the identification and treatment of AD.

Future advancements in our AD detection technology have great potential to improve AD treatment efficacy, prognosis accuracy, and diagnostic precision. Further development and innovation in machine learning algorithms, user interface design, and EEG biomarkers will further improve the sensitivity, specificity, and scalability of the system. Furthermore, the incorporation of multimodal data sources, such as genetic markers, neuroimaging, and clinical evaluations, is expected to enhance the system's clinical value and predictive capacity, allowing for individualized and thorough AD risk assessment and management plans.

Additionally, improvements in remote monitoring capabilities and wearable EEG technology will broaden the reach the system outside of clinical settings, allowing for proactive and long-term tracking of AD development in actual contexts. The key to unlocking revolutionary discoveries and advancements in AD detection, diagnosis, and treatment ultimately lies in the continued synergy between the fields of neuroscience, machine learning, and human-computer interaction. This will usher in a new era of precision medicine and individualized care for people at risk of or affected by AD.

Conclusively, the creation of the "Alzheimer Disease Detection System using EEG Biomarkers" initiative represents a noteworthy progression in the domain of diagnosing and treating Alzheimer's disease. This system shows potential capabilities in early detection and classification of Alzheimer's disease stages through the use of decision tree algorithms and electroencephalography (EEG) biomarkers. The project's success in analyzing EEG data using machine learning algorithms highlights the promise for non-invasive, low-cost diagnostic tools in the medical field. Still, there are chances for improvement in the future. Specifically, continued research and development endeavors should concentrate on broadening the range of EEG biomarkers employed, optimizing the decision tree algorithm for enhanced precision, and including supplementary elements for a more all-encompassing diagnostic methodology. In order to further guarantee the system's efficacy and dependability, it must be implemented in actual clinical settings and continuously validated against established clinical criteria. All things considered, this initiative establishes the foundation for novel approaches to Alzheimer's disease diagnosis, with encouraging potential for enhancing patient outcomes and furthering the study of neurodegenerative diseases.

REFERENCES

- [1] Sosa-Ortiz AL, Acosta-Castillo I, Prince MJ. Epidemiology of dementias and Alzheimer's disease. *Arch Med Res.* **2012**.
- [2] Frisoni GB, Altomare D, Thal DR, et al. The probabilistic model of Alzheimer disease: the amyloid hypothesis revised. *Nat Rev Neurosci.* **2022**.
- [3] Kumar A, Fontana IC, Nordberg A. Reactive astrogliosis: A friend or foe in the pathogenesis of Alzheimer's disease. *J Neurochem.* **2023**;164:309–24.
- [4] Rabinovici GD, Miller BL. Frontotemporal lobar degeneration: epidemiology, pathophysiology, diagnosis and management. *CNS Drugs.* **2010**.
- [5] Caprioglio C, Garibotto V, Jessen F, et al. The Clinical Use of Alzheimer's Disease Biomarkers in Patients with Mild Cognitive Impairment: A European Alzheimer's Disease Consortium Survey. *J Alzheimers Dis.* **2022**.
- [6] Aducanumab (marketed as Aduhelm) Information | FDA. <https://www.fda.gov/drugs/postmarket-drug-safety-information-patients-and-providers/aducanumab-marketed-aduhelm-information> . Accessed 3 Jan **2023**.
- [7] FDA Grants Accelerated Approval for Alzheimer's Disease Treatment | FDA. <https://www.fda.gov/news-events/press-announcements/fdagrants-accelerated-approval-alzheimers-disease-treatment> . Accessed 10 Jan **2023**.
- [8] Schindler SE, Bollinger JG, Ovod V, et al. High-precision plasma β -amyloid 42/40 predicts current and future brain amyloidosis. *Neurology.* **2019**.
- [9] Karikari TK, Pascoal TA, Ashton NJ, et al. Blood phosphorylated tau 181 as a biomarker for Alzheimer's disease: a diagnostic performance and prediction modelling study using data from four prospective cohorts. *Lancet Neurol.* **2020**.

- [10] Ashton NJ, Pascoal TA, Karikari TK, et al. Plasma p-tau231: a new biomarker for incipient Alzheimer's disease pathology. *Acta Neuropathol.* **2021**.
- [11] Milà-Alomà M, Ashton NJ, Shekari M, et al. Plasma p-tau231 and p-tau217 as state markers of amyloid- β pathology in preclinical Alzheimer's disease. *Nat Med.* **2022**.
- [12] Ashton NJ, Janelidze S, Al Khleifat A, et al. A multicentre validation study of the diagnostic value of plasma neurofilament light. *Nat Commun.* **2021**.
- [13] Pereira JB, Janelidze S, Smith R, et al. Plasma GFAP is an early marker of amyloid- β but not tau pathology in Alzheimer's disease. *Brain.* **2021**.
- [14] Altomare D, Stampacchia S, Ribaldi F, Tomczyk S, Chevalier C, Poulain G, Asadi S, Bancila B, Marizzoni M, Martins M, Lathuiliere A, Schefer M, Ashton NJ, Zetterberg H, Blennow K, Kern I, Frias M, Garibotto V, Frisoni GB. Plasma biomarkers for Alzheimer's disease: a field-test in a memory clinic. *J Neurol Neurosurg Psychiatry.* **2023**.
- [15] Teunissen CE, Verberk IMW, Thijssen EH, et al. Blood-based biomarkers for Alzheimer's disease: towards clinical implementation. *Lancet Neurol.* **2022**.
- [16] McKhann GM, Knopman DS, Chertkow H, et al. The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement.* **2011**.
- [17] Albert MS, DeKosky ST, Dickson D, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement.* **2011**.
- [18] Rascovsky K, Hodges JR, Knopman D, et al. Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. *Brain.* **2011**.
- [19] Gorno-Tempini ML, Hillis AE, Weintraub S, et al. Classification of primary progressive aphasia and its variants. *Neurology.* **2011**.

[20] Armstrong MJ, Litvan I, Lang AE, et al. Criteria for the diagnosis of corticobasal degeneration. *Neurology*. **2013**.

APPENDIX A

SOURCE CODE

```
# Importing the necessary Python modules.

import streamlit as st

# Import necessary functions from web_functions

from web_functions import load_data

from eeg_modulator import load_data_obj

# Configure the app

st.set_page_config(

    page_title = 'Alzheimer Stage Detector using EEG',

    page_icon = 'brain',

    layout = 'wide',

    initial_sidebar_state = 'auto'

)

# Import pages

from Tabs import home, data, eegdata, predeeg, predict, visualise

# Dictionary for pages

Tabs = {

    "Home": home,

    "Biomarker Data Info": data,

    "EEG Data Info": eegdata,

    "Biomarker AD Test": predict,
```

```

    "EEG AD Test":predeeg,

    "Visualisation": visualise

}

# Create a sidebar

# Add title to sidear

st.sidebar.title("Navigation")

# Create radio option to select the page

page = st.sidebar.radio("Pages", list(Tabs.keys()))

    Loading the dataset.

df, X, y = load_data()

df2, X2, y2 = load_data_obj()

# Call the app funciton of selected page to run

if page in ["Biomarker AD Test", "Visualisation"]:

    Tabs[page].app(df, X, y)

elif (page == "EEG AD Test"):

    Tabs[page].app(df2,X2,y2)

elif (page == "Biomarker Data Info"):

    Tabs[page].app(df)

elif (page == "EEG Data Info"):

    Tabs[page].app(df2)

else:

    Tabs[page].app()

"""This module contains necessary function needed"""

```

```

# Import necessary modules

import numpy as np

import pandas as pd

from sklearn.tree import DecisionTreeClassifier

import streamlit as st

@st.cache_data()

def load_data():

    """This function returns the preprocessed data"""

    # Load the Diabetes dataset into DataFrame.

    df = pd.read_csv('Alzheimer.csv')

    # Rename the column names in the DataFrame.

    # Perform feature and target split

    X =
df[["Response", "Gender", "Age", "PNSA", "SES", "MMSE", "CDR", "eTIV", "nWBV", "ASF",
"Group"]]

    y = df['Outcome']

    return df, X, y

@st.cache_data()

def train_model(X, y):

    """This function trains the model and return the model and model score"""

    # Create the model

    y = np.array(y).copy()

```

```

model = DecisionTreeClassifier(

    ccp_alpha=0.0, class_weight=None, criterion='entropy',

    max_depth=4, max_features=None, max_leaf_nodes=None,

    min_impurity_decrease=0.0, min_samples_leaf=1,

    min_samples_split=2, min_weight_fraction_leaf=0.0,

    random_state=42, splitter='best'

)

# Fit the data on model

model.fit(X, y)

# Get the model score

score = model.score(X, y)

# Return the values

return model, score

def predict(X, y, features):

    # Get model and model score

    model, score = train_model(X, y)

    # Predict the value

    prediction = model.predict(np.array(features).reshape(1, -1))

    return prediction, score

"""This module contains necessary function needed"""

# Import necessary modules

import numpy as np

```

```

import pandas as pd

from sklearn.tree import DecisionTreeClassifier

import streamlit as st

@st.cache_data()

def load_data_obj():

    """This function returns the preprocessed data"""

    # Load the Dementia dataset into DataFrame.

    df = pd.read_csv('EEG_data.csv')

    # Perform feature and target split

    X
    =
    df[["EEG_AF3","EEG_F7","EEG_F3","EEG_FC5","EEG_T7","EEG_P7","EEG_O1","EE
    G_O2","EEG_P","EEG_T8","EEG_FC6","EEG_F4","EEG_F8","EEG_AF4"]]

    y = df['outcome']

    return df, X, y

@st.cache_data()

def train_model(X, y):

    """This function trains the model and return the model and model score"""

    # Create the model

    y = np.array(y).copy()

    model = DecisionTreeClassifier(

        ccp_alpha=0.0, class_weight=None, criterion='entropy',

        max_depth=4, max_features=None, max_leaf_nodes=None,

```

```

        min_impurity_decrease=0.0, min_samples_leaf=1,

        min_samples_split=2, min_weight_fraction_leaf=0.0,

        random_state=42, splitter='best'

    )

    # Fit the data on model

    model.fit(X, y)

    # Get the model score

    score = model.score(X, y)

    # Return the values

    return model, score

def predict(X, y, features):

    # Get model and model score

    model, score = train_model(X, y)

    # Predict the value

    prediction = model.predict(np.array(features).reshape(1, -1))

    return prediction, score

"""This modules contains data about home page"""

# Import necessary modules

import streamlit as st

def app(df):

    """This function create the Data Info page"""

    # Add title to the page

```



```

st.title("Data Info page")

# Add subheader for the section

st.subheader("View Data")

# Create an expansion option to check the data

with st.expander("View data"):

    st.dataframe(df)

# Create a section to columns values

# Give subheader

st.subheader("Columns Description:")

# Create a checkbox to get the summary.

if st.checkbox("View Summary"):

    st.dataframe(df.describe())

# Create multiple check box in row

col_name, col_dtype, col_data = st.columns(3)

# Show name of all dataframe

with col_name:

    if st.checkbox("Column Names"):

        st.dataframe(df.columns)

# Show datatype of all columns

with col_dtype:

    if st.checkbox("Columns data types"):

        dtypes = df.dtypes.apply(lambda x: x.name)

        st.dataframe(dtypes)

```

Show data for each columns

with col_data:

if st.checkbox("Columns Data"):

col = st.selectbox("Column Name", list(df.columns))

st.dataframe(df[col])

APPENDIX B

PLAGIARISM REPORT

Main Project report Final.pdf

ORIGINALITY REPORT

12% <small>SIMILARITY INDEX</small>	10% <small>INTERNET SOURCES</small>	5% <small>PUBLICATIONS</small>	7% <small>STUDENT PAPERS</small>
---	---	--	--

PRIMARY SOURCES

1	Submitted to SRM University <small>Student Paper</small>	1%
2	Submitted to King's College <small>Student Paper</small>	1%
3	ejournal.unma.ac.id <small>Internet Source</small>	1%
4	www.rj-robbins.com <small>Internet Source</small>	1%
5	shropsys.com <small>Internet Source</small>	1%
6	www.frontiersin.org <small>Internet Source</small>	<1%
7	Submitted to College of Banking and Financial Studies <small>Student Paper</small>	<1%
8	Submitted to University of California, Los Angeles <small>Student Paper</small>	<1%
9	Submitted to University of Northampton	

