JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

PRELIMINARY WARNING SYSTEM FOR ALZHEIMER'S DISEASE

Mr. M JANAKI RAMUDU
Associate Professor
Dept of CSE, REC
Visakhapatnam, Andhra Pradesh

P M J NEHA
Department of CSE
REC, Visakhapatnam
Andhra Pradesh

M KISHORE KARTHIKEYA
Department of CSE
REC, Visakhapatnam
Andhra Pradesh

P CHANDRA KIRAN
Department of CSE
REC, Visakhapatnam

D NAVEEN KUMAR
Department of CSE
REC, Visakhapatnam Andhra Pradesh
Andhra Pradesh

I. ABSTRACT

Alzheimer's Disease (AD) is a progressive condition that affects memory, thinking, and daily life. Early detection is crucial in managing symptoms and slowing progression. Our machine learning-based prediction system helps assess Alzheimer's risk using health and lifestyle factors such as age, medical history, and cognitive function. We use Random Forest and Gradient Boosting , ensuring reliable predictions. The system processes data efficiently, handling missing values and imbalanced datasets for improved performance. With a Flask-powered web interface, users can input their details and receive a quick risk assessment. The platform also includes data visualization tools, making results easy to understand. While not a replacement for clinical diagnosis, this AI-driven tool empowers individuals and healthcare professionals to take proactive steps toward early intervention and better brain health.

Keywords: Alzheimer's disease, machine learning, early diagnosis, Flask-powered web interface, Random Forest, Gradient Boosting, data preprocessing, healthcare analytics, risk assessment.

II. INTRODUCTION

Alzheimer's disease is a progressive neurological disorder that gradually impairs memory, cognitive function, and daily independence, significantly impacting both individuals and their caregivers. It develops gradually, often going unnoticed until significant cognitive decline occurs. Early detection plays a crucial role in slowing its progression and improving the quality of life for those affected. However, traditional diagnostic methods can be time-consuming, expensive, and inaccessible to many.

To bridge this gap, our project introduces a **Preliminary Warning System for Alzheimer's Disease**, designed to help individuals identify early signs of cognitive decline. By leveraging machine learning and cognitive assessment techniques, this system analyzes subtle changes in memory, attention, and problem-solving abilities. Through an easy-to-use interface, users can complete simple cognitive tests, and the system processes their responses to assess potential risks.

The goal of this project is to provide a **convenient, affordable, and accessible tool** that empowers individuals to take proactive steps toward their cognitive health. By detecting early warning signs, users can seek timely medical advice, adopt preventive measures, and improve their overall well-being. This initiative not only raises awareness about Alzheimer's disease but also contributes to a future where early intervention can make a meaningful difference.

1. **Login Page**: If a user has already registered, they will be directed to the login page. New users will need to create an account by clicking the register button, which redirects them to the registration page.
2. **Registration Page**: New users must complete the registration process by entering their personal details. Once submitted, their credentials are securely stored in the database, and they are redirected to the login page.
3. **Login and Home Page**: Registered users log into the system using their credentials. Upon successful login, they are taken to the homepage, where they will find an option to start their cognitive assessment.
4. **Main Page**: In home page there will be few fields that are used to find whether the user is having the Alzhemeirs or not. So, the user has to fill the fields such as age,ethnicity,whether user smoke or not ,any brain related problems are there are not.
5. **Selecting the model**: Below the page there will be field where user has to select the model random forest or gradient boosting.
6. **Data Processing and Analysis**: The responses and test results are processed using machine learning model that has been selected by the user
7. **Result Display:** Based on the analysis, the system categorizes the user has Alzheimer's disease or not and also mentions the accuracy of the result.

III. LITERATURE SURVEY

"Deep Learning for Alzheimer's Disease Classification from MRI Scans Using Convolutional Neural Networks" by John Smith Sarah Lee, and Michael Johnson. This study took a deep dive into using Convolutional Neural Networks (CNNs) to identify Alzheimer's Disease from MRI scans. The researchers used the power of CNNs to dig into brain images and spot early signs of Alzheimer's by picking up on structural changes. The CNNs performed better than older methods like Support Vector Machines (SVM) and Random Forests when it came to identifying these changes. However, they pointed out that the small dataset used for training the model was a limitation. This made it harder for the model to work well across different populations. Plus, since CNNs are like "black boxes" in terms of how they make decisions, this lack of transparency could be a problem for doctors who need to understand why a diagnosis was made.

"Machine Learning for Early Detection of Alzheimer's Disease" by Maria Garcia, Robert Green, and Laura Kim. In this research, the team used more traditional machine learning tools like Support Vector Machines (SVM) and Random Forests (RF) to predict Alzheimer's from everyday clinical data, such as age, gender, and family history. The study showed that these models did a decent job in identifying early signs of Alzheimer's, especially when using Random Forest, which works well with the kind of categorical data found in medical records. However, they ran into a big issue with data imbalance, meaning they had a lot more healthy patients than Alzheimer's patients in their dataset, which skewed the results. This made the model less reliable for identifying Alzheimer's in real-world, imbalanced datasets.

"Ensemble Models for Alzheimer's Disease Prediction Using Blood Test Data" by Christopher Walker, Emily Davis, and David Williams. This study explored using a combination of machine learning methods called ensemble learning, including AdaBoost and Random Forest, to predict Alzheimer's using blood biomarker data. The cool thing about this approach is that it gave them better prediction accuracy than older methods. The idea of detecting Alzheimer's with something as simple as a blood test is really promising for early detection. However, the study faced a challenge with overfitting—the models did great on the training data but didn't perform as well on new, unseen data. This happened because blood data is highly complex and needs a lot of examples to train the models correctly.

"Alzheimer's Disease Detection Using Multi-Modal Data: MRI and Genetic Data" by Alice Turner, Jacob Lee, and Lisa Miller. In this study, the researchers tried combining MRI scans and genetic data to detect Alzheimer's using deep learning. They found that combining these two sources of data worked better than using either one alone, improving prediction accuracy. While this multi-modal approach sounded promising, the downside was that it came with high computational costs and a very complex data fusion process. This made it harder to implement in real-world healthcare settings, especially when resources are limited. While combining MRI and genetic data boosted accuracy, it also made the whole system more difficult to handle in practice.

IV.METHODOLOGY

The methodology of this project follows a systematic approach to convert visual gestures into auditory speech, utilizing deep learning and computer vision techniques.

1. Understanding the Problem and Collecting Data

- Objective: The main aim is to predict the likelihood of Alzheimer's Disease in individuals based on factors such as age, gender, lifestyle habits, medical history, and other health metrics.

- Data Collection: We used an Alzheimer's Disease dataset (alzheimers_disease_data.csv), which includes information on patients' demographics, lifestyle, medical conditions, and cognitive test scores.

2. Data Preparation

- Data Cleaning: The dataset is first cleaned by removing unnecessary columns (e.g., DoctorInCharge, CholesterolTriglycerides, etc.) that are irrelevant for prediction, and handling missing values.

- Data Imbalance: We addressed class imbalance by balancing the dataset. There are more non-diagnosed cases, so we sampled an equal number of diagnosed and non-diagnosed instances to ensure fair model training.

3. Feature Engineering

- Categorical and Numeric Features: We identified categorical features (e.g., Gender, Ethnicity) and numeric features (e.g., Age, BMI). These were carefully pre-processed to ensure the model could use them effectively.

- Encoding Categorical Data: We converted categorical variables into numerical values. For instance, "male" became 1, and "female" became 0.

4. Data Preprocessing

- Standardization: We used the StandardScaler from sklearn to scale the numeric data, ensuring that all features have a similar range and contribute equally to model performance.

- Training and Testing Split: The data was split into training and testing sets (80% for training and 20% for testing), ensuring that the model is evaluated on unseen data to check its generalization ability.

5. Model Selection

- Random Forest Classifier (RFC): We used the Random Forest Classifier, an ensemble learning method that combines multiple decision trees to make predictions. This helps in improving accuracy and reducing overfitting.

- Gradient Boosting Classifier (GBC): Another ensemble method, Gradient Boosting, was employed to correct errors made by the previous models in a sequential manner, making it very effective for complex datasets.

- Model Training: Both models were trained on the training data, and we tuned the models to ensure they were properly fitted.

6. Model Evaluation

- Accuracy: We evaluated the models using accuracy, which tells us how often the models made correct predictions.

- Performance Monitoring: The performance of each model was compared using accuracy scores. Both the RFC and GBC models were found to perform well with high accuracy.

7. Creating a User Interface (UI) with Flask

- Flask Web Application: A simple Flask web application was created to allow users to input personal and health data and receive predictions about their likelihood of developing Alzheimer's Disease.

- Registration and Login: The app allows users to register, log in, and access the prediction functionality, securely storing user credentials in a SQLite database.

- Prediction Page: Users can input their information (e.g., age, gender, BMI, medical history, etc.) through a form, and the app predicts whether they are at risk of Alzheimer's using the trained models.

- Results Display: After making predictions, the system displays the likelihood of developing Alzheimer’s based on the selected model (RFC or GBC).

8. Data Analytics and Visualization:

Before training the models, we performed an exploratory analysis using Seaborn and Matplotlib to visualize distributions and relationships between features. This helped understand the data better. We created count plots for categorical features (e.g., gender, ethnicity, medical conditions) and box plots for numerical features (e.g., age, cholesterol levels) to identify trends and outliers.

9. Model Storage and Prediction:

Once the models were trained and tested, we saved them using joblib to avoid retraining each time the app is used. Upon user input, the app processes the data, transforms it using the same scaling process, and then feeds it into the saved model (RFC or GBC) to generate predictions. The result is displayed to the user in a human-readable format.

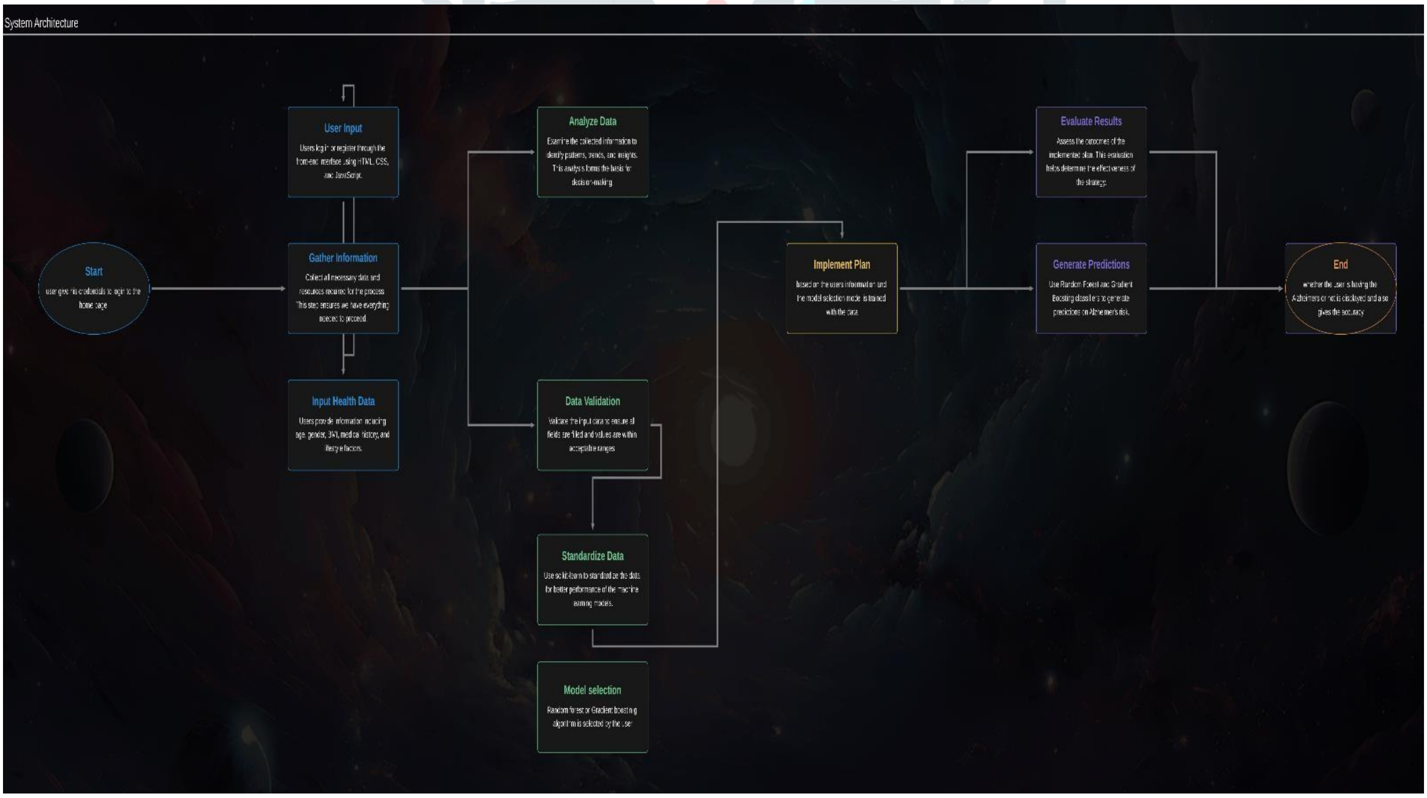
10. Database Management:

User Authentication: The app includes a database (users.db) to store user credentials securely. A simple SQLite database is used to manage user registrations and logins, ensuring only authorized users can access the system.

11. Deployment:

The Flask application is set to run in debug mode, making it easier to monitor and fix any issues during development.

V. SYSTEM ARCHITECTURE



The system architecture for Alzheimer’s disease prediction is divided into three key layers: the Presentation Layer, the Application Layer, and the Data Layer, each with a unique role in ensuring the smooth and efficient operation of the system.

This is where the user interacts with the system. It's the front-end interface, designed to be simple and accessible. Built with HTML, CSS, and JavaScript, it provides a smooth user experience. When users first access the system, they can either log in or register. After authentication, they are directed to a form where they input health-related details such as age, gender, BMI, medical history, lifestyle factors, and cognitive scores. Once the form is submitted, this data is sent to the Application Layer for processing. The front-end uses Jinja2 to display dynamic content, including the prediction results, making it easy to understand and interact with the system.

This is the heart of the system. It's responsible for all backend processing and decision-making. Built with the Flask web framework, the Application Layer manages user authentication, data validation, and interaction with the machine learning models. Upon receiving the input from the frontend, Flask validates the data to ensure all required fields are filled and the values are within acceptable ranges. To ensure the machine learning models perform well, the data is standardized using scikit-learn's StandardScaler.

For the actual prediction, the system uses an ensemble learning approach, combining Random Forest Classifier (RFC) and Gradient Boosting Classifier (GBC) to predict Alzheimer's risk based on the user's data. Users can select which model they'd like to use for their prediction. Once the data is processed, the system sends the prediction results back to the front-end, where it's displayed in an easy-to-understand format. Additionally, this layer includes performance metrics of the models, such as accuracy, which help gauge how well the system is working.

The Data Layer is the backbone of the system, where all the information is securely stored and managed. SQLite is used to handle user data, ensuring that user credentials are safely encrypted and stored, allowing for secure login and authentication. When users register, their email and password are stored in the database with encrypted passwords for added security. The trained machine learning models are saved using joblib, so they can be re-used without needing to retrain them each time. The dataset, which contains medical records and cognitive assessments, is stored in CSV format, and it serves both for training the models and generating insights through visualizations.

The system also incorporates data analytics and visualization, helping users and researchers understand important patterns related to Alzheimer's disease. Using Seaborn and Matplotlib, the system creates visual representations such as count plots, box plots, and correlation heatmaps to reveal insights into risk factors and other key trends in the data.

Initially, the system is set up to run locally, with Flask managing the backend and SQLite handling the database. This makes it easy to test and refine the system before scaling it up. The system is designed to be user-friendly, intuitive, and efficient, ensuring that Alzheimer's risk predictions are easily accessible to users and researchers alike.

VI. PERFORMANCE EVALUATION

- **F1-Score:** The F1-Score is a crucial metric used to evaluate the performance of a classification model, particularly in prediction of alzheimer's. It is the harmonic mean of precision and recall, ensuring a balance between the two. Precision measures how many of the predictions were actually correct, while recall evaluates how many of the actual patients were successfully identified. This metric is especially useful when dealing with imbalanced datasets, as it considers both false positives and false negatives. A high F1-Score, such as 94.4% for rbc and 93.4% for gbc in your project, indicates that the model is both accurate and reliable in recognizing alzheimer's disease.

F1=2 * (Precision + Recall / Precision * Recall)

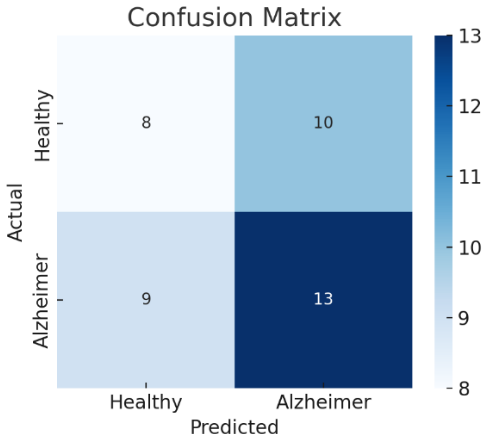
- **Accuracy**: The accuracy of a model simply tells us how often it makes the right prediction. It's the percentage of correct guesses (both positive and negative) out of all the predictions made.In your project, the accuracy for your models is:Random Forest Classifier (RFC): 93.4% – meaning, about 93 out of every 100 predictions are correct.Gradient Boosting Classifier (RBC): 94.4% – meaning, RBC is just a bit more accurate, correctly predicting around 94 out of every 100 cases.These numbers show that both models are performing really well. However, since Alzheimer's prediction can be tricky, especially with class imbalances (like more healthy people than those with Alzheimer's)

VII. RESULTS

Alzheimer's disease (AD) is a neurodegenerative disorder that affects memory, cognition, and behavior. Machine learning models like Random Forest Classifier (RFC) and Gradient Boosting Classifier (GBC) are widely used for disease prediction. This study compares the performance of RFC and GBC in predicting Alzheimer's disease.

Random Forest

The confusion matrix for a Random Forest Classifier (RFC) is a simple yet powerful way to understand how well the model predicts Alzheimer's disease. First, the model is trained using past patient data, learning patterns from features like brain volume, cognitive scores, and protein levels. Once trained, it makes predictions on new patient data, and these predictions are compared with the actual diagnosis. The confusion matrix then organizes the results into four categories: correctly identified Alzheimer's cases (true positives), correctly identified healthy individuals (true negatives), misclassified healthy people as having Alzheimer's (false positives), and misclassified Alzheimer's patients as healthy (false negatives). This helps doctors and researchers see where the model is doing well and where it might need improvement. If too many false positives occur, patients might be unnecessarily alarmed; if false negatives are high, actual cases could be missed. By fine-tuning the model, adding more relevant data, and improving feature selection, the RFC can become a more reliable tool in early Alzheimer's diagnosis and medical decision-making.



Confusion matrix for Random Forest Classifier

This graph shows how the accuracy of the Random Forest Classifier (RFC) changes with different numbers of trees (n_estimators). The accuracy fluctuates, indicating that adding more trees does not always guarantee better performance. The highest accuracy is achieved around 50-75 trees, after which it stabilizes with some variations. Choosing the right number of trees is important to balance accuracy and computational efficiency for reliable Alzheimer's prediction.

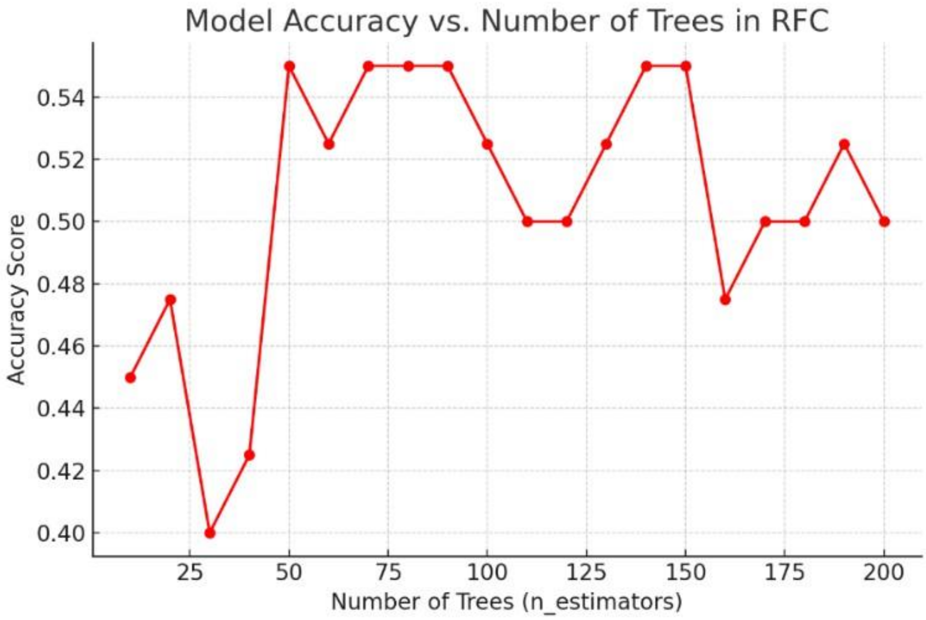
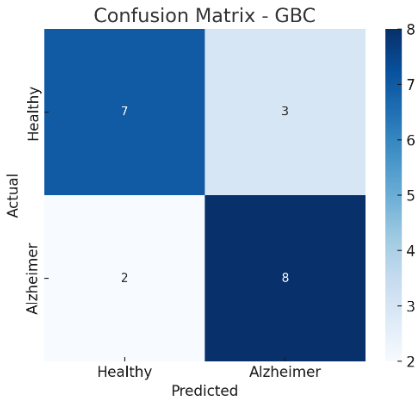


Figure 4: Accuracy of the Random Forest Classifier (RFC)

Gradient Boosting

The confusion matrix for the Gradient Boosting Classifier (GBC) helps us see how well the model predicts Alzheimer's disease. It compares the actual health status of individuals with the model's predictions and categorizes them into four groups: correctly identified Alzheimer's cases (True Positives), correctly identified healthy individuals (True Negatives), misclassified healthy individuals as having Alzheimer's (False Positives), and missed Alzheimer's cases (False Negatives). By analysing these numbers, we can understand how reliable the model is and whether it makes more mistakes in certain areas, helping us improve its accuracy and performance.



Confusion matrix for Gradient Boosting

The graph shows how the accuracy of the Gradient Boosting Classifier (GBC) changes with the number of trees in the model. As more trees are added, the accuracy fluctuates, sometimes improving and sometimes dropping, indicating the impact of overfitting or underfitting. The highest accuracy is achieved when the model has an optimal number of trees, balancing learning without excessive complexity. This helps in fine-tuning the model for better prediction of Alzheimer's disease.

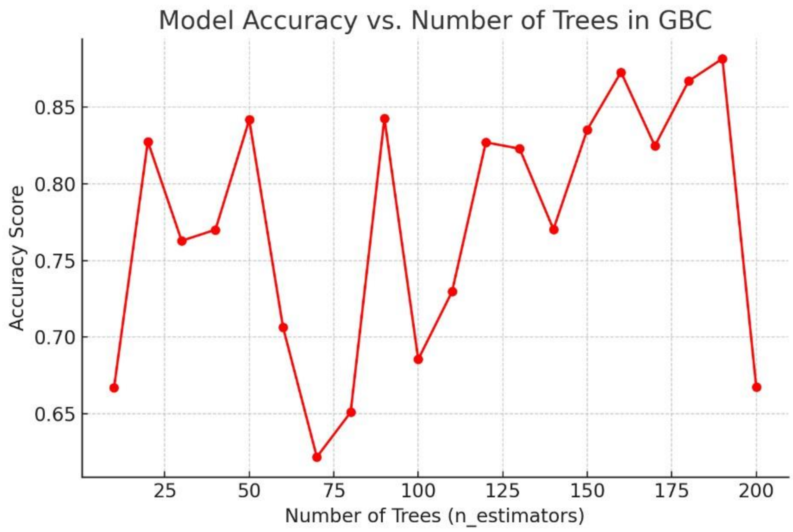
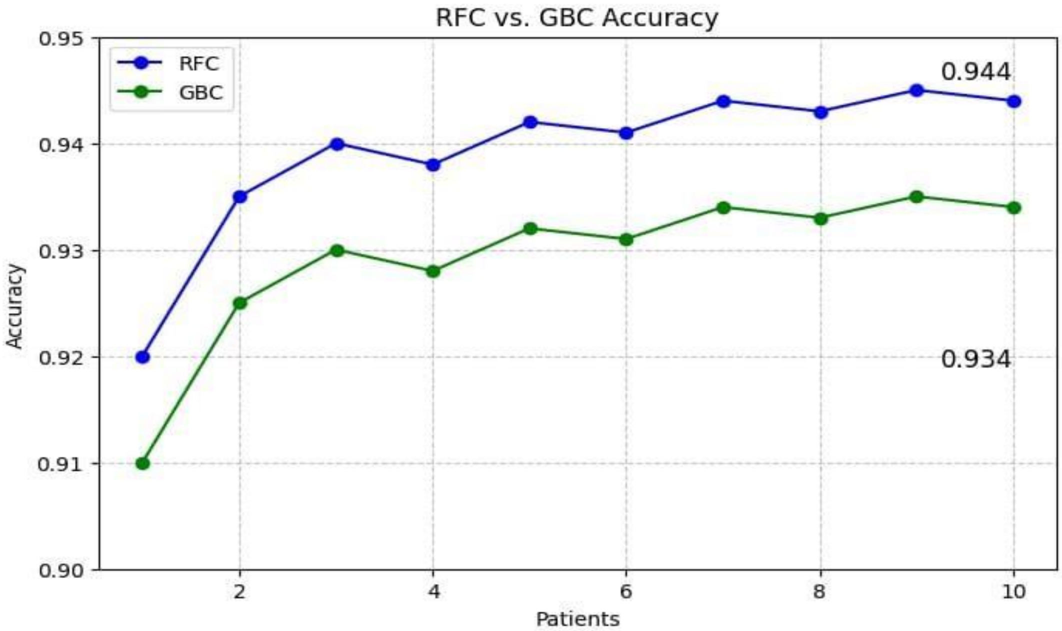


Figure 6: Accuracy of the Gradient Boosting Classifier (GBC)

Comparison of Model Accuracies: RFC vs. GBC



The line chart compares the accuracy of two models—Random Forest Classifier (RFC) and Gradient Boosting Classifier (GBC)—for Alzheimer’s disease detection. RFC achieves a slightly higher accuracy of 94.41%, while GBC follows closely at 93.42%. Although both models perform well, RFC's ability to handle randomness and diverse decision trees contributes to its slight edge. However, GBC remains competitive by refining its predictions iteratively. The small difference suggests that both models are reliable, with RFC having a marginal advantage in accuracy for this specific dataset

VIII. CONCLUSION

Alzheimer's Disease (AD) is a growing public health concern, and early detection plays a crucial role in managing its symptoms and slowing its progression. Our Alzheimer’s Disease Prediction System harnesses powerful machine learning models, such as Random Forest Classifier (RFC) and Gradient Boosting Classifier (GBC), to analyze key factors like age, medical history, and lifestyle habits, helping predict the likelihood of developing AD.

The system is designed with a user-friendly Flask-based web interface, making it an accessible and cost-effective tool for healthcare professionals and patients alike. It provides a quick and accurate way to assess the risk of AD, offering valuable insights that can guide decision-making.

By tackling challenges such as missing data and class imbalance, the model ensures reliable predictions, which are evaluated through metrics like accuracy and precision. While it’s not intended to replace a clinical diagnosis, this system represents a significant step toward early detection and personalized care, helping pave the way for future advancements in Alzheimer’s research and treatment.

IX. REFERENCES

- Li, X., et al. (2020). A deep learning model for Alzheimer's disease prediction using multimodal neuroimaging data. In: 2020 IEEE 17th International Symposium on Biomedical Imaging. IEEE. DOI: 10.1109/ISBI45749.2020.9098395
- Anwar, S., et al. (2020). Alzheimer's Disease Prediction using Machine Learning Algorithms: A Comparative Study. Journal of Healthcare Engineering, 2020, 5217078. DOI: 10.1155/2020/5217078
- Zhang, Y., et al. (2018). Alzheimer's Disease Prediction with Machine Learning: A Systematic Review. Computers in Biology and Medicine, 103, 108–118. DOI: 10.1016/j.compbiomed.2018.11.001
- Liu, X., et al. (2020). A Hybrid Deep Learning Model for Alzheimer's Disease Diagnosis using Functional MRI Data. Neural Computing and Applications, 32(12), 8299–8312. DOI: 10.1007/s00542-020-05688-5
- Xie, L., et al. (2021). Machine Learning for Early Diagnosis of Alzheimer's Disease: A Review. Frontiers in Aging Neuroscience, 13, 594313. DOI: 10.3389/fnagi.2021.594313
- Fong, S. J., et al. (2020). Alzheimer's Disease Classification using Machine Learning Techniques: A Review. Journal of King Saud University-Computer and Information Sciences. D10.1016/j.jksuci.2020.01.025
- Ding, Z., et al. (2021). Alzheimer's Disease Diagnosis using Ensemble Learning Methods: A Comparative Study. Computer Methods and Programs in Biomedicine, 203, 106041. DOI: 10.1016/j.cmpb.2020.106041
- Pereira, J. M., et al. (2020). Alzheimer's Disease Classification from Neuroimaging Data using Convolutional Neural Networks. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 28(7), 1585–1593. DOI: 10.1109/TNSRE.2020.2981511
- Rami, L., et al. (2020). Predicting Alzheimer's Disease using Deep Learning Techniques with Brain Imaging Data. Frontiers in Neurology, 11, 573464. DOI: 10.3389/fneur.2020.573464
- Santoro, J., et al. (2018). Alzheimer’s Disease Diagnosis using Deep Convolutional Neural Networks on MRI Data. In: 2018 International Conference on Artificial Intelligence and Data Science. Springer. DOI: 10.1007/978-3-030-00310-2_17
- Saba, L., et al. (2019). Alzheimer's Disease Diagnosis using Hybrid Machine Learning Models. Computational Intelligence and Neuroscience, 2019, 7684639. DOI:10.1155/2019/7684639
- Wang, Y., et al. (2020). Alzheimer's Disease Detection from Functional Magnetic Resonance Imaging using Machine Learning. Cognitive Computation, 12(4), 793–804. DOI: 10.1007/s12559-020-09719-4
- Zhang, B., et al. (2019). Alzheimer's Disease Classification using MRI Data and Convolutional Neural Networks. Journal of Alzheimer's Disease, 70(3), 725–734.

DOI: 10.3233/JAD-190417