2023 2nd International Conference for Innovation in Technology (INOCON)
Bangalore, India. Mar 3-5, 2023

Early Detection of Alzheimer's Disease Using Svm, Random Forest & FNN Algorithms

B. Shanmukha Sai Varma
B. *Tech, Information Technology*
V R Siddhartha Engineering College,
Vijayawada
varma.borukati@gmail.com

Dr. G. Kalyani
Ph. D, Information Technology
V R Siddhartha Engineering College,
Vijayawada
gkalyani@vrsiddhartha.ac.in

Koyagura Asish
B. Tech, Information Technology
V R Siddhartha Engineering College,
Vijayawada
asishwork999@gmail.com

M. Indumathi Bai
B. Tech, Information Technology
V R Siddhartha Engineering College,
Vijayawada
indusimharaidu@gmail.com

2023 2nd International Conference for Innovation in Technology (INOCON) | 979-8-3503-2092-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/INOCON57975.2023.10101298

Abstract **- Alzheimer's disease is a brain disorder in which the central nervous system or the nerve tissues in the brain stop working or die, causing tissue damage. Symptoms include dementia, difficulty reading, writing, and performing various other tasks independently. Alzheimer's disease, also known as AD, is an incurable disease at the current age. Though various medical procedures and tests, such as MMSE, CDR, eTIV, nWBV, etc., help in identifying the disease, there is no permanent cure for this disease. Hence, a comprehensive model for detecting the disease in an individual in the early stages of the disease is needed. To address this task, this paper uses a few machine learning algorithms to train the model to later classify the patient's data to perform a medical data analysis and predict the patient's status, whether the patient has AD, doesn't have AD, or has the possibility of getting AD in the future. This model benefits the medical field by increasing the efficiency of results and decreasing the processing time of medical reports.**

Keywords: Alzheimer's disease (AD), machine learning model, dementia, medical data analysis.

I. INTRODUCTION

Alzheimer's disease is a neurological disorder of the brain caused by damage to brain cells caused by a variety of factors such as old age, accidents, certain medical conditions, and so on.This disease damages the brain by causing damage or death of the cells in the brain tissues. The patients suffering from this disease suffer from various cognitive, behavioral, and psychological disorders. Cognitive symptoms include mental decline, difficulty thinking and understanding, and memory loss. Repetition of own words, irritability, and difficulty taking care of themselves are examples of behavioral symptoms. Psychological symptoms include depression, hallucinations, etc.

While this is a very dangerous disease, an optimal and effective treatment method to cure it has yet to be found. This is due to the fact that the brain is one of the most sensitive parts of the human body, which limits the treatment methods with our current technology. Hence, there is a need to detect the disease in its early stages to avoid being too late to prevent its growth in a person. Machine learning is a field of study in which the user makes the machine in such a way that the machine learns some methods and processes to perform actions independently without the need for human intervention. Using this technology, machines can complete various data analytic processes using machine learning algorithms, saving users from having to repeat the processes themselves. Therefore, we want to combine the data from the previous batches of patients tested for Alzheimer's disease with machine learning algorithms to build a trained model that can determine the status of a patient tested for Alzheimer's disease, which makes it easier and more accurate compared to manual prediction by the working doctors.

The scheme in this paper uses structured data (i.e., a csv file format) to train the machine learning model to find the stage of disease in the patient. Algorithms such as support vector machines (SVM), random forests (RF), and feed-forward neural networks (FNN are used in this paper.

II. LITERATURE REVIEW

We find a vast number of machine learning algorithms using various methods for achieving the objective of training the machine. The various algorithms use different approaches to give out the most optimal solution. SVM became one of the most popular machine learning algorithms due to the kernel concept [1, 6]. The use of decision trees and random forests made various types of complex data easier to manage and understand, which made prediction easier [1]. The neural network concept brought the capabilities of the learning in a machine to a new level [2, 5]. MRI scanning and prediction also helped in understanding more of the brain condition and improving accuracy. Many papers used SVM for building the machine to predict the disease, as it gives relatively good accuracy for the results. But all the algorithms used for training the model have their own advantages, such as performing various data analytics to broaden the scope of research. Also, the latest techniques, such as neural networks, have many advantages. All these different kinds of methods give us a very broad scope of research in the concept of machine learning. Methods are to be tried and tested to understand the concepts and apply them properly. Hence, for this problem, we choose 3 algorithms SVM, Random Forests, and Feed-Forward Neural Network—to work on the creation of the required model.

Authorized licensed use limited to: University of West Florida. Downloaded on December 12,2024 at 22:21:04 UTC from IEEE Xplore. Restrictions apply.

III. DESIGN METHODOLOGY

A. Design Flow

To achieve our ultimate goal of developing a model for Alzheimer's disease prediction, we first collect a dataset of Alzheimer's patients' past test results. This dataset was downloaded from kaggle.com, a dataset website. Now, the dataset is pre-processed to remove any null or unwanted values and convert the data into a structured form. We also convert the data labels Demented, Non-Demented, and Converted from strings to integer values 1, 0, and -1. Now, we divide the data into training and testing data, with about 70% training and 30% testing data. Now, we find and extract the most important feature that is influencing the results of the tests in the dataset: feature labels. This provides us with the most optimal solution model with the best accuracy rate. Now, train the data using machine learning algorithms like SVM, Random Forest, and FNN. Finally, the trained model can be tested using test data for accuracy percentage and other measures. If the accuracy of the model is not satisfactory, we train the model again. After that, we finally complete the construction of a trained model for early detection of Alzheimer's disease.

Figure 1 shows the workflow design of our trained machine used for the detection of Alzheimer's disease. This design is made to give three outcomes in the results: 1) demented, 2) non-demented, and 3) converted. These results are the state of the patient's brain condition, which in turn describes the possibility of Alzheimer's disease.

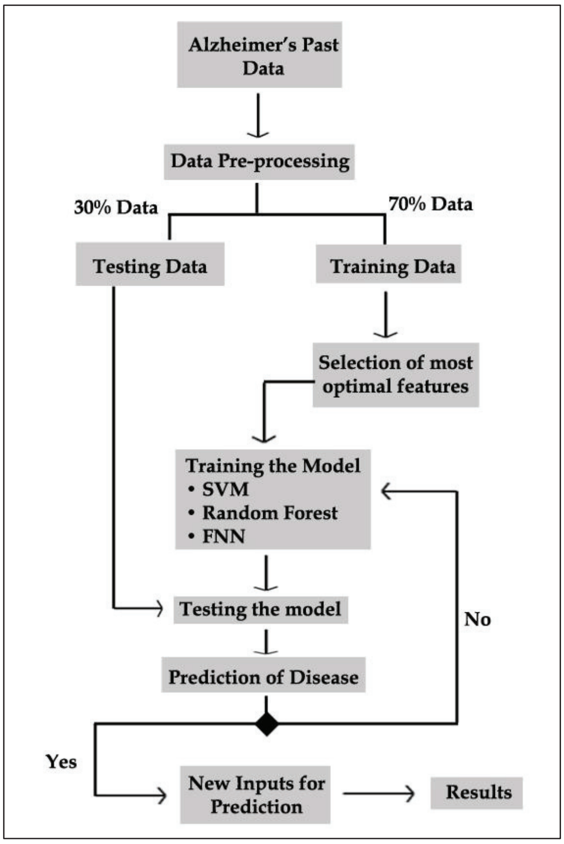

Fig. 1. Workflow of the model

B. Architecture Of The Model

A basic architecture for the software model we build acts as the foundation upon which we build our prediction model. This makes the development of the model easier and more

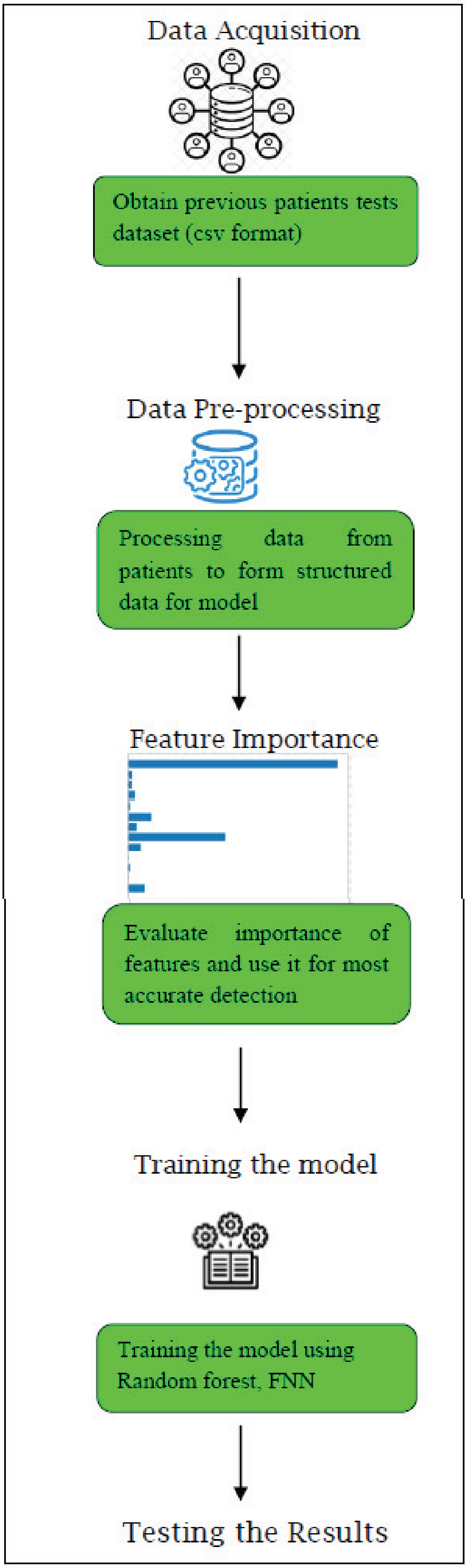

Fig. 2. Architecture of the model

Authorized licensed use limited to: University of West Florida. Downloaded on December 12,2024 at 22:21:04 UTC from IEEE Xplore. Restrictions apply.

efficient, as it gives us an in-depth understanding of the processes in the model. It also helps in backtracking the processes in the model, as it gives us a clear goal for evaluation in case of errors in the model's development.

The system architecture is built based on the following steps:

- Data Acquisition.
- Data Pre-processing.
- Feature Extraction.
- Training the model using machine learning algorithms
- Testing the Results

1) Data Acquisition

This is the step in which all the data from the test records of past Alzheimer's disease patients is obtained. Various details in the data, like the education status of the patient (EDUC), socio-economic status (SES), gender, age, estimated total Total volume (eTIV), mini mental state examination (MMSE), etc., are gathered together to form a complete dataset.

Our final dataset for this problem has 8 columns of data, i.e., 8 types of features of patients tested for Alzheimer's disease.

2) Data Pre-processing

In this stage, the data obtained is pre-processed to remove any null or unwanted values that are detrimental to the accuracy of the solution model. We also convert the data labels Demented, Non-Demented, and Converted from strings to integer values 1, 0, and -1. This process turns the unstructured form of data into a structured format. This makes it easier for the machine to understand and learn the attributes of the patients in the dataset.

3) Feature Importance

Every feature of a patient, such as age, gender, status of income, MMSE, eTIV, etc., influences the possibility of having Alzheimer's disease. Hence, among them, we must find the features or labels that influence the possibility of the disease and their impact in producing the most optimal model for the satisfaction and accuracy of the problem objective.

4) Training the model using machine learning algorithms

The data is divided into training data (70% of the data) and testing data (30% of the data) in this step. The training data is trained using algorithms such as SVM, Random Forest, and FNN. The model is then ready to be used for the prediction of Alzheimer's disease.

5) Testing the Results

Now, we can test the data using the test results. If the model has high accuracy, we can go on to test it using other new data and then send it to the healthcare department.

As a result, our goal of developing a machine model for Alzheimer's disease detection in its early stages has been met.

IV. DATASET & ALGORITHMS USED FOR TRAINING THE MODEL

A. Dataset & Tool Used

The dataset we currently use is obtained from the Kaggle Internet website, which is a platform for getting datasets for various types of projects. The dataset has eight labels for various tests of the disease, which are used for the complete training and testing data for the research of our project.

TABLE I contains information about the features in the dataset.

TABLE I. FEATURES DESCRIPTION

Feature	Abbreviation	Range
Age	Age	1-100
EDUC	Years of Education	1-25
SES	Socio-economic Status	1-5
MMSE	Mini Mental State Examination	1-30
CDR	Clinical Dementia Rating	0-3
eTIV	Estimated Total Intracanial Volume	1100-2000
nWBV	Normalized Whole Brain Volume	0-1
ASF	Atlas Scaling Factor	0-2

The software tool used for the training of the model can either be Jupyter Notebook, Google Colaboratory, or any other Python interpreter and compiler.

B. Algorithms Used

1) Support Vector Machine(SVM)

SVM is one of the more popular machine learning training algorithms. The reason it *stands out is due to the concept of supporting vectors and the SVM* classifier, which help in choosing an N-dimensional space for classifying the data. The algorithm uses these components to perform operations needed by the user.

Kernels are used in this algorithm to convert a low-level dimensional space to a higher-level dimensional space. This helps make classification and other processes easier. In our research, we used two kernels for the training of data.

2) Random Forest(RF)

Random Forests is a supervised machine learning algorithm based on the decision tree concept. In this method, the algorithm divides the data into different decision trees, thereby simplifying the complex data. The model is then trained using this data. The final optimal solution is the one that satisfies the majority of the conditions in the decision tree.

3) Feed-forward Neural Network(FNN)

The concept of neural networks is based on the neural system of the human body. The algorithm method uses the method of layers to perform the learning process for the machine. In this algorithm, the layers contain neurons consisting of data from the patients, which are forwarded to the next layer's nodes, or neurons, for training. The output layer gives the final prediction model. The advantage of this algorithm is that it has a much higher capacity for independent work compared to the other algorithms.

Authorized licensed use limited to: University of West Florida. Downloaded on December 12,2024 at 22:21:04 UTC from IEEE Xplore. Restrictions apply.

V. RESULTS

Analysis of the dataset is carried out using various methods to find the impact of the labels or attributes in the dataset. This gives us a measure of the weights of each attribute for the training of the model for predicting the state of patient with Alzheimer's disease. The various analytics carried out are as follows:

A. In Dataset taken-

Let the various states of the disease be taken as the follows

In figure-3

Demented(1) , Non-Demented(0), Converted (-1)

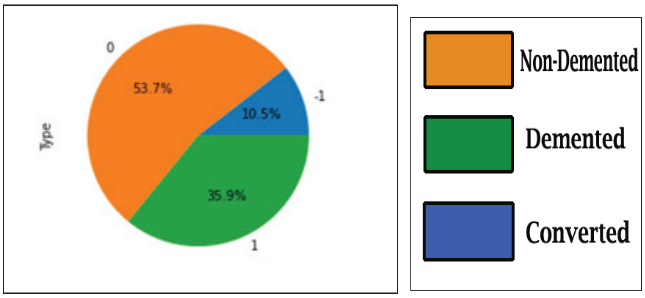

Fig. 3. Density of disease states in the population in dataset

B. Gender vs Population Density in each state -

In Figure-4, using the dataset, the gender of patient and number of the patients in each state of the disease is plotted.

Blue(0) -> Male, Orange(1) -> Female

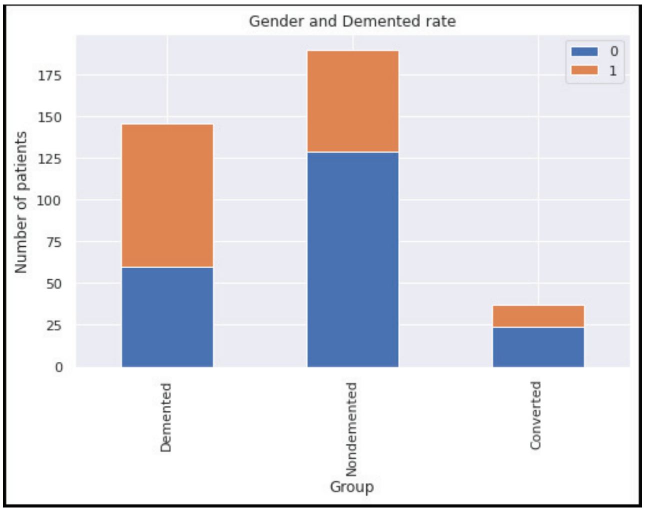

Fig. 4. Gender vs Density of disease

C. MMSE vs Population Density in each state-

MMSE is a medical test conducted on the patients to determine their mental state. Figure-5 plot the Population density at each stage in the MMSE score obtained by the MMSE test.

In the plot,
-1 -> Converted, 0 -> Non-Demented,
1 -> Demented

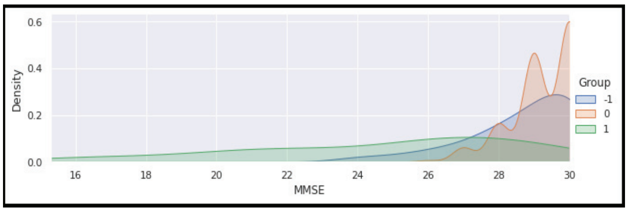

Fig. 5. Population Density vs MMSE

D. ASF vs Population Density in each state -

Atlas Scaling factor is also a medical test for testing dementia. Figure-6 below plots the value of ASF to the population density of each state of disease in the dataset.

In the plot,

-1 -> Converted, 0 -> Non-Demented,

1 -> Demented

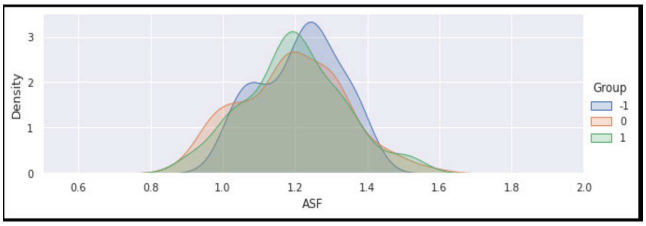

Fig. 6. ASF vs Population Density

E. eTIV vs Population Density in each state –

eTIV is a measure of brain volume obtained through medical tests. This helps in understanding the changes in volume of the brain in various states of disease. Figure-7 below plots the comparison.

In the plot,

-1 -> Converted, 0 -> Non-Demented,

1 -> Demented

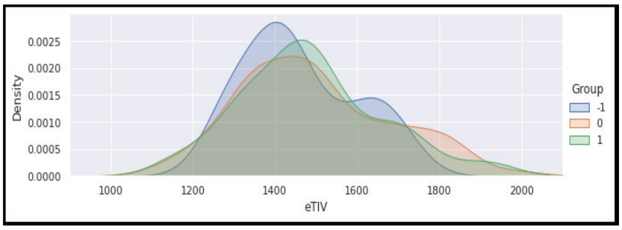

Fig. 7. eTIV vs Population Density

F. nWBV vs Population Density in each state –

nWBV is also a medical test for the brain volume of a patient. Figure-8 plots the comparison of nWBV to the state of disease in the patient.

In the plot,

-1 -> Converted, 0 -> Non-Demented,
1 -> Demented

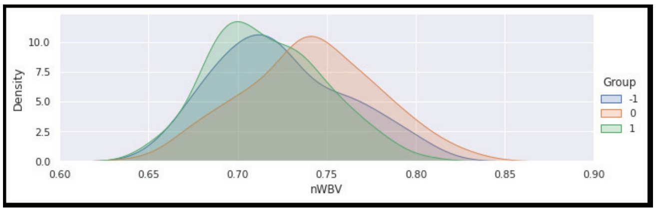

Fig. 8. nWBV vs Population Density

Authorized licensed use limited to: University of West Florida. Downloaded on December 12,2024 at 22:21:04 UTC from IEEE Xplore. Restrictions apply.

G. Age vs Population Density in each state –

Figure-9 plots the comparison of patient's age and their state of disease in the population.

In the plot,

-1 -> Converted, 0 -> Non-Demented,

1 -> Demented

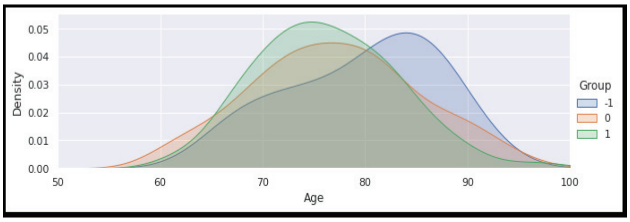

Fig. 9. Age vs Population density

H. EDUC vs Population Density in each state –

Figure-10 plots the comparison of patient's education status to their state of disease in the population.

In the plot,

-1 -> Converted, 0 -> Non-Demented,

1 -> Demented

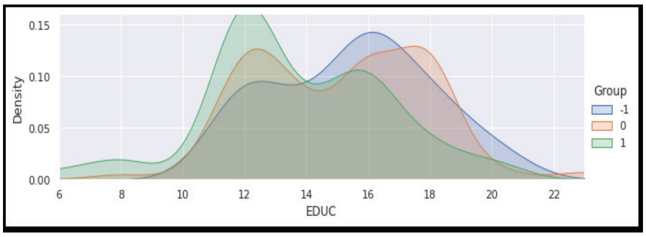

Fig. 10. EDUC vs Population Density

All the above analytics give us in-depth information about various attributes, such as at what scale of an attribute the possibility of a particular state of a disease is high or low, etc.

Now, for the training of model we understand and tally the importance of features in the dataset for prediction of the disease state. Figure 11 below shows the importance of each feature in the prediction of the state of the disease.

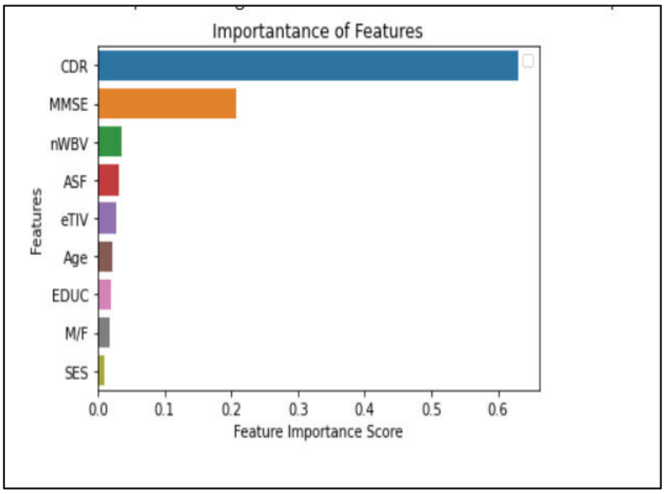

Fig. 11. Feature Importance Score

A correlation analysis is used to derive the impact of each individual feature on the output of the model. Using this score as parameters for the model, we train the model for prediction of the Alzheimer's disease state of a patient.

Now, after training the model using SVM, RF, and FNN techniques, the performance measures for the models using the different algorithms are given in Figure 12.

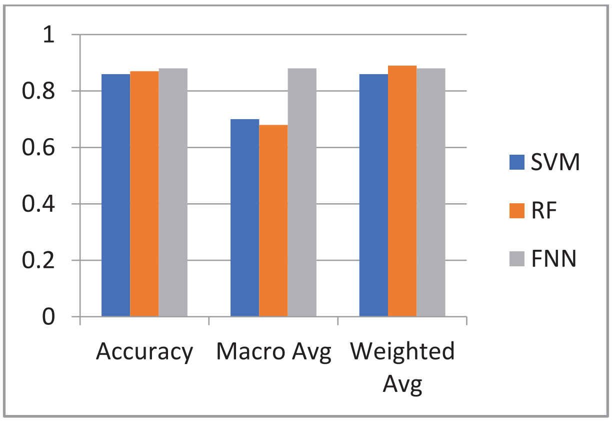

Fig. 12. Comparison of SVM, Random Forest and FNN trained models

VI. CONCLUSION

From the above observations, we get more accurate predictions from the FNN model (88%), compared to the SVM (86%), and the Random Forest (87%). Furthermore, the patient's CDR is frequently the most important factor in determining a patient's disease possibility or stage. The various graphs comparing the disease population density to the features have also been displayed. This gives us a better understanding of the intensity of a feature based on the disease population density.

TABLE II. ACCURACY RESULTS

Model Method	**Accuracy (%)**
SVM	86
Random Forest	87
FNN	88

The required trained machine model for detecting Alzheimer's disease has been accomplished. This research helped a lot in understanding various Python libraries and the ways to use them. The study of various machine learning algorithms increased our mastery in the machine learning field of study. Further insights on how to improve the current work have been realized.

The future work is to include MRI brain scan images of the patients in the dataset to get a more accurate assessment of the disease. The detection of various stages of Alzheimer's disease is also to be included in the work. Finally, a web application is to be developed so that patients can assess the results themselves online by accessing the web application and giving the test records as input. This is the plan for the near future. Additional enhancements will be made in accordance with the work.

Authorized licensed use limited to: University of West Florida. Downloaded on December 12,2024 at 22:21:04 UTC from IEEE Xplore. Restrictions apply.

REFERENCES

[1] Kishore, P., Ch Usha Kumari, M. N. V. S. S. Kumar, and T. Pavani. "Detection and analysis of Alzheimer's disease using various machine learning algorithms." *Materials today: proceedings* 45 (2021): 1502-1508.

[2] Orouskhani, Maysam, Sahar Rostamian, Firoozeh Shomal Zadeh, Mehrzad Shafiei, and Yasin Orouskhani. "Alzheimer's Disease Detection from Structural MRI Using Conditional Deep Triplet Network." *Neuroscience Informatics* (2022): 100066.

[3] Zhao, Xinxing, Candice Ke En Ang, U. Rajendra Acharya, and Kang Hao Cheong. "Application of Artificial Intelligence techniques for the detection of Alzheimer's disease using structural MRI images." *Biocybernetics and Biomedical Engineering* 41, no. 2 (2021): 456-473.

[4] Sharma, Abhilash, Sukhkirandeep Kaur, Naz Memon, A. Jainul Fathima, Samrat Ray, and Mohammed Wasim Bhatt. "Alzheimer's patients detection using support vector machine (SVM) with quantitative analysis." *Neuroscience Informatics* 1, no. 3 (2021): 100012.

[5] Kumar, L. Sathish, S. Hariharasitaraman, Kanagaraj Narayanasamy, K. Thinakaran, J. Mahalakshmi, and V. Pandimurugan. "AlexNet approach for early stage Alzheimer's disease detection from MRI brain images." *Materials Today: Proceedings* 51 (2022): 58-65.

[6] Balne, Sridevi, and Anupriya Elumalai. "Machine learning and deep learning algorithms used to diagnosis of alzheimer's." *Materials Today: Proceedings* 47 (2021): 5151-5156.

Authorized licensed use limited to: University of West Florida. Downloaded on December 12,2024 at 22:21:04 UTC from IEEE Xplore. Restrictions apply.