Multiclass-classification of Alzheimer's Disease using 3-D CNN and Hyper-Parameter Optimization of Machine Learning Algorithms

Chinmay Pathak[1], Abhay Jidge[2], Vishal Mourya[3], Omkaresh Kulkarni[4], Bharati Dixit[5]

[1, 2]Department of Computer Science and Technology, MIT College of Engineering, Pune, India

[3]Department of Information Technology, MIT College of Engineering, Pune, India

[4]Professor, Department of Computer Science, MIT World Peace University, Pune, India

[5]Professor, Department of Computer Science, MIT College of Engineering), Pune, India

Abstract: ***Dementia and its forms like Mild Cognitive Impairment and Alzheimer's Disease are posing great defiance to society claiming many lives. Alzheimer's Disease is the fifth leading cause of death in the world. The most frequently observed form of this dementia is Alzheimer's Disease. Diagnosis of this disease is a very tedious task for the doctors, which can lead to errors in the judgment. This paper proposes multi-class classification by using MRI Images of subjects into 3 different classes such as Normal, MCI (Mild Cognitive Impairment) and Alzheimer's Disease. The presented 3D CNN model can accurately classify Cognitive Normal patients and also MCI (Mild Cognitive Impairment) which may morph into Alzheimer's Disease to prevent the further disintegration and severity of the patient's condition. This paper also presents the comparative study of the performance of different Machine Learning models on the test diagnosis data of the subjects and suggests the most efficient approach to classify into 3 classes by data pre-processing and Hyperparameter optimization.***

Keywords: Alzheimer's Disease, 3-D CNN, Hyper-parameter optimization, Machine Learning

1. Introduction

Alzheimer's Disease is a disease that is very prevalent in humans. It is a progressive deterioration of the brain which causes problems like memory mutations and disorientation. The root of the deterioration is still unknown. The thing that is certain in Alzheimer's disease is that it cannot be diagnosed very easily and can become fatal if not detected early and treated accordingly [12].The reason for this is that at the early stage the patient can perform all the normal activities but as the disease progresses [2], it can impede many normal activities which takes many decades. Reports from NIA (National Institute of Aging) suggest that the number of people persuading the disease doubles every 5 years. Nearly 3.3 million women and 2 million men are victimized currently in the U.S [1] [9].

This paper describes the process of training a 3-dimensional convolutional neural network for classifying the disease using MRI images of the patient. The approach used here is peculiar as the model uses all the available slices (62) of the patient to classify the disease as either Normal or MCI or victimized by Alzheimer's Disease. The main intention of this research is to correctly identify convolutional layers to give accurate results. The paper compares the various Supervised Learning Algorithms to identify the algorithm which gives the most accurate results to predict the disease.

Machine Learning techniques are being comprehensively used in medical research tasks and also in neurodegenerative data. The recent amelioration in technology has made it possible to collect huge amounts of data for the research work in medical fields [11]. Hence, the extensive use of the different classification systems is on a constant rise. The direct application of the models may however, hamper the results as the data is mainly in the raw format and the results generated may not be acceptable. The use of data pre-processing techniques, feature-selection, selecting a balanced dataset can achieve promising results. The selection of the machine learning algorithm can be a major challenge in the classification task. The paper proposes a comparative study of the performance of the different models to cogently choose the correct model for the classification of the patients as Normal, affected with MCI (Mild Cognitive Impairment or Alzheimer's Disease. Hyper parameter optimization i.e. tuning the hyper parameters (the parameters that can be controlled through the learning process) is also performed and the improved accuracy of the model is depicted.

2. Related Work

Previously the diagnosis of Alzheimer's disease was done manually by the doctors which were based on the atrophy of the brain images. The different patterns of the structure could be identified to diagnose the disease. This technique was developed by Ph Scheltens et al. [2]. The method was subject to an accuracy of 86% of the images.

Neurological disorders among the children could be identified with the help of Machine Learning Models by using various symptoms like depression levels, apathy, epilepsy, seizures, etc. This method was developed by G. Reshma, Dr. P.V.S Lakshmi [3]. The accuracy of prediction the of the neurological disease was 87.5%.

Tejeswinee. Ka, Shomona Gracia Jacob, and Athilakshmi. Rc [4] developed a comparative study by using different

ML algorithms like SVM, Random Forest, Naive Bayes, Adaboost and KNN on Alzheimer's and Parkinson's disease. Feature selection was applied to the KEGG database and PROFEAT server before applying the models. It was observed that SVM gave the highest accuracy of 94%.

Konstantinos Kamnitsas, Christian Ledig [5] developed an efficient multi-scale 3D CNN for brain lesion segmentation. The model developed has an efficient 11-layers deep, multi-scale, 3D CNN architecture. Accuracy of the models for the prediction of Alzheimer's disease could be increased by using such segmentation approach.

Alzheimer's is a progressive disease and becomes worse as time passes. In order to early diagnosis of Alzheimer Diseases A. B. Rabeh [6] proposed an application in which they were using three sections: frontal to extract the Hippocampus (H), Sagittal to analyze the Corpus Callosum (CC) and axial to work with the various features of the Cortex(C). The model is trained using Support Vector Machine (SVM) and gives an accuracy of 90.66%.

Accurate classification of Alzheimer's Disease at each stage is important to prevent the progression of the disease. Feng Li

[7] proposed a deep learning model and method to optimize it which increases classification accuracy by 5.9% on average. In this model, they have used PET and MRI scans. Firstly, by applying the principal component analysis (PCA) they have extracted the most important features and then these features are passed to the neural network. Further to overcome the overfitting the have applied dropout techniques. Finally learned features were passed to Support Vector Machine (SVM) to classify the patient in AD/MCI class. Yang [8] proposed 3 approaches to generate visual explanations from 3D convolutional neural networks in Alzheimer's disease classification. In the first approach, they have conducted a sensitivity analysis on segmented 3D images. And in other two to visualize the network activations.

3. Proposed Computational Framework

a) Classification using 3-D CNN (Method 1)

1) Dataset Generation: For the collection of data to train the model ADNI (Alzheimer's Disease Neuroimaging Initiative) project was used. It provides the structural MRI scans of the subjects with their labels. To achieve the expected goal the i.e. multi-class classification the MRI images of 800 Cognitive Normal subjects, 800 subjects with Mild Cognitive Impairment (MCI) and 800 subjects affected with Alzheimer's disease were extracted. From the sagittal, coronal and axial views available T2 weighted 62 axial slices of each of these 2400 subjects were used. From the 800 subjects, 500 were used for training, 150 for validation and the remaining 150 were used for testing purposes.

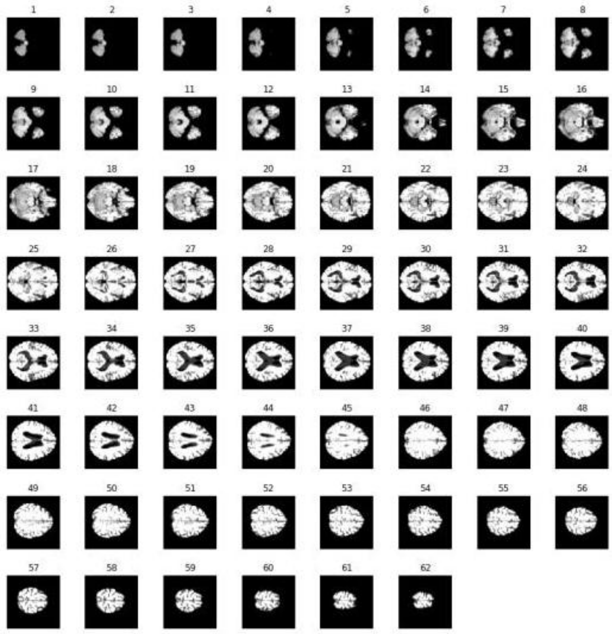

Figure 1: 62 Slices of brain used for training the model

The extracted images were in a raw format, so skull scripting and resizing were performed. All the 62 slices (Fig 1) of each subject was used for training purposes. Every slice of dimension 96x96 were stacked to form a NumPy array of dimension 1x62x96x96 which was used to train the 3-D CNN model. The computation was performed on 1x Tesla K80 GPU.The GPU has 12GB of memory and 2496 cores. The reason of using 800 subjects of each category was that the proposed 3-D CNN model was trained accurately. Earlier 500 subjects of each category were used on which the results obtained were not satisfactory. Further to reduce the computational complexity image rescaling was performed. Table 1 represents the summary of the dataset.

Table I: Summary of the Dataset

Diagnosis	Diseases		
	CN	MCI	AD
Male / Female	607 / 616	454 / 381	514 / 441
Age (mean STD)	$77{:}94_{4:99}$	$76{:}09_{7:66}$	$75{:}88_{7:80}$
Number of subjects	1223	835	955

Paper ID: SR20317205759 DOI: 10.21275/SR20317205759

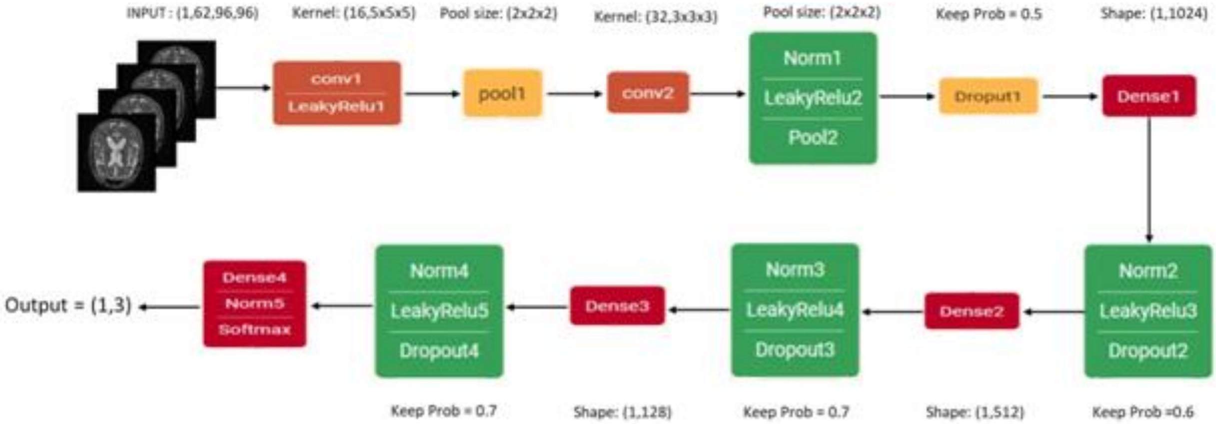

Figure 7: Architecture of 3-D CNN

2) Proposed Architecture: The model designed for the multi-class classification(Fig 2) which gave the best possible accuracy consisted of two convolutional layers and four dense layers followed by a softmax activation function to predict the probability of each of the desired three classes i.e. Cognitive Normal, MCI(Mild Cognitive Impairment) and Alzheimer's Disease. From the two convolutional layers, the first layer consists of 16 channels with a kernel size of 5x5x5 and the second layer consists of 32 channels with the kernel size of 3x3x3. In both these layers, the stride of 2x2x2 is used to create the feature map.

Each of the Convolution layers has a LeakyRelu activation function. The main challenge while improving the accuracy of the classifier was to reduce the noise and avoid the information loss. For this purpose, the Leaky Relu activation function was used, which gives the derivative of the negative part as a small fraction in the negative part and not entirely zero. Unlike Relu which gives the derivative as complete zero; which leads to information loss because of which the model's accuracy is hampered.

After the second convolutional layer there is one layer of dropout followed by a flatten layer. Lastly there are four Fully connected with a SoftMax activation function. Several experiments were performed to determine the optimized number of layers, filters for each layer, learning rate, pooling size, and dropout probability.

To train the neural network models, major people face the difficulty to determine the optimal number of epochs of training to avoid overfitting or underfitting of the model. Hence, to overcome this, Early stopping is used in order to find the efficient number of epochs. From figure [5], we can observe that after 45th epoch there is no significant change in the accuracy of the model. So, by using early stopping, the model automatically stopped the training at 45th epoch.

Another challenge people face is to determine the learning rate of the neural networks to potentially not miss the minima which otherwise hampers the accuracy [13]. The solution for the above problem is possible in keras by Reducing the learning rate on plateau. The advantage of this approach is, it can reduce the learning rate at the appropriate situation and provide the optimizer to get the best path to the minima. Before using Reduce LR On Plateau, the validation accuracy obtained was 84.60 % and after applying this approach was 88.98 %.

b) Classification using Machine Learning Models (Method 2)

1) Dataset generation: In our project implementation, we have assessed our model performance on Alzheimer's Dis-ease neuroimaging Dataset. This dataset consists of Cognitive Normal subjects, Mild Cognitive Impairment subjects and the subjects affected with Alzheimer's disease. A total of 5627 subjects were used for training the model. From these 5627 subjects 2200 were Normal, 2509 subjects of Mild Cognitive impairment and 918 patients affected with Alzheimer's Dis-ease. The total number of female subjects were 2372 and the total number of male subjects were 3255 with their mean age of 76.919 years. The total ten attributes namely Sex, Weight, APOE (apolipoprotein A1) score, APOE (apolipoprotein A2) score, Age, MSME Score, GD-Scale, Global CDR Score, FAQ Total Score, NPI-Q Total Score were used to predict the target variable. From the ten attributes one is categorical and remaining are Numerical in nature. The above-mentioned tests are used for diagnosis by doctors. Table 2 represents the attribute information used to train the model.

Table II: Attribute Summary

Attribute No.	Attribute Name	Data type (Values)
A1	Sex	Categorical
A2	Weight	Numerical (Continuous)
A3	APOE A1 Score	Numerical (Discrete)
A4	APOE A2 Score	Numerical (Discrete)
A5	Age	Numerical (Continuous)
A6	MMSE Score	Numerical (Discrete)
A7	GDScale Score	Numerical (Discrete)
A8	Global CDR Score	Numerical (Discrete)
A9	FAQ Total Score	Numerical (Discrete)
A10	NPI-Q Total Score	Numerical (Discrete)

2) Data Pre-processing: The data obtained was in a raw format hence data pre-processing was required to achieve the desired results. So, a series of steps were performed to obtain clean and balanced data to train the model.

3) Missing value Analysis and Outlier Treatment: The dataset had certain missing scores for the diagnosis tests. Initially the missing values were replaced by 'Nan' (Not a

number) and then many imputation techniques like mean and median were applied to handle the missing values. Initially mean was applied which gave poor results in terms of accuracy and then median technique was applied which gave the satisfactory results, hence we decided to go with median. The data also had certain outliers which were first analysed by using box plot (Fig 3) and then removed.

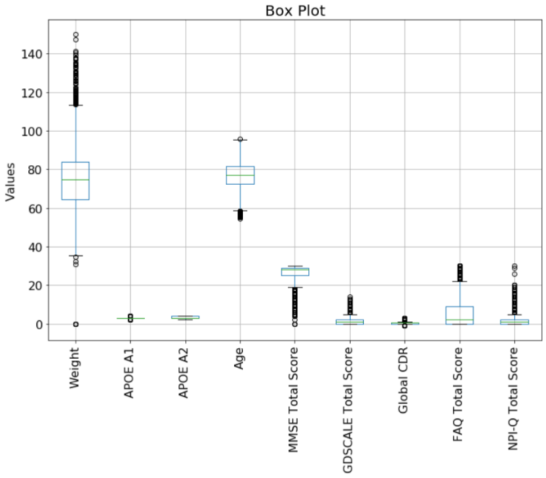

Figure 3: Box plot of the attributes

4) Correlation Analysis between Dependent and Independent Feature/Attributes: The last step which was performed was the analysis of correlation of the different attributes of the dataset. The analysis was performed using the correlation matrix. From the matrix (Fig 4) it was observed that Global CDR Score and FAQ Score are correlated with the MSME Scores in the data.

5) Learning Algorithms: We employed machine learning models like Support Vector Machine (SVM), Decision Tree, XGBoost and Random Forest for the multi class classification. Below is the brief overview of the models. SVM is a supervised machine learning algorithm which is mainly used for classification. In this algorithm the classification is mainly done by plotting the data points in n-dimensional space so each data point has a particular coordinate value. By using these coordinate values, it calculates optimal hyperplane which is separating the different classes. In order to calculate the optimal hyperplane, we take the projections of points on hyperplane and maximize the distance. For better classification results the parameters of the SVM model must be selected

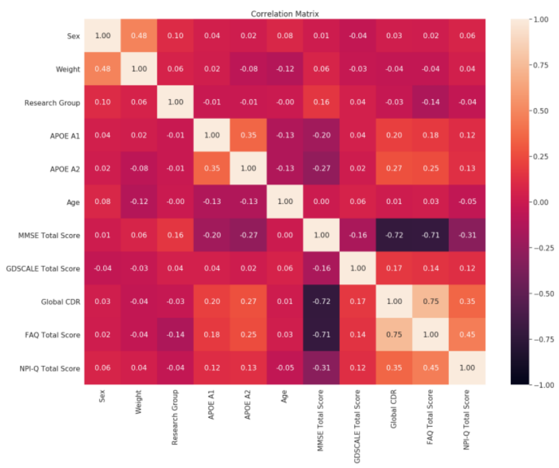

Figure 4: Correlation Matrix

properly. In order to convert the data points to the required format SVM uses various kernels including nonlinear kernels such as linear, polynomial, rbf and sigmoid. A proper Degree should be given for the polynomial kernel. The value of C and Gamma provides the soft margin for nonlinear SVM. The effect of misplaced data points can be overcome by proper choice of C and Gamma values.

DECISION TREE The decision tree is the supervised machine learning algorithm that can be used for classification or regression. In the training phase of the model the decision tree was created by learning simple decision rules. Initially, the entire training data is considered as the root node. Further the dataset splits recursively into sub-nodes based on attribute values and the whole decision tree is constructed. In the prediction phase of a record, it is assigned to the root node and follows the path by comparing the values with internal nodes till it reaches to the leaf node. In order to measure the impurity at each node, we have used Gini index and the other hyper parameters are tuned by using Grid Search CV. Appropriate selection of hyper parameters is important as the depth of the tree increases, the model overfits. For this model optimal hyper parameter values are max-features = auto, min-samples-leaf = 7, min-samples split = 15.

XGBOOST is an extreme Gradient Boosting. XGBoost is an algorithm that has recently gained immense adoration in machine learning. XGBoost is implemented using gradient boosted decision trees. This library is focused on efficient computational speed and model performance. It offers a number of advanced features like Gradient Boosting, Stochastic Gradient Boosting, Regularized Gradient Boosting. In order to maximize the accuracy of the output, parameter tuning is done using Grid Search CV. The parameters used for tuning are max-depth, n-estimators, min-child weight. For the XG Boost model, the optimal hyper parameters obtained are max-depth= 8, minchild weight = 5, nestimators = 100.

RANDOM FOREST overcomes the main drawback of the decision tree which is over fitting. The decision tree memorizes the training data by fitting it closely; to overcome this draw-back and make a more generalized model in random forest multiple decision trees were generated. Random forest is the ensemble method and multiple decision trees were created by random sampling training dataset and prediction is based on the majority voting For random forest models using the Grid Search CV, the highest accuracy obtained for the parameters max-depth=30, min-samples-leaf=1, min-samples-split=2, n-estimators=100.

4. Experimental Results

a) 3-D CNN Results

The performance of the proposed 3-D CNN architecture was carried out on the ADNI (Alzheimer's Disease Neuroimaging Initiative) dataset. The axial view of T2 weighted MRI images were used for training the 3-D CNN model and the accuracy was obtained on the same images. The performance of the proposed 3-D CNN architecture was evaluated for the multi-class classification of

AD/MCI/NC. The training accuracy of 96.39 was observed and the validation accuracy of 88.98% and testing accuracy was 83.15%

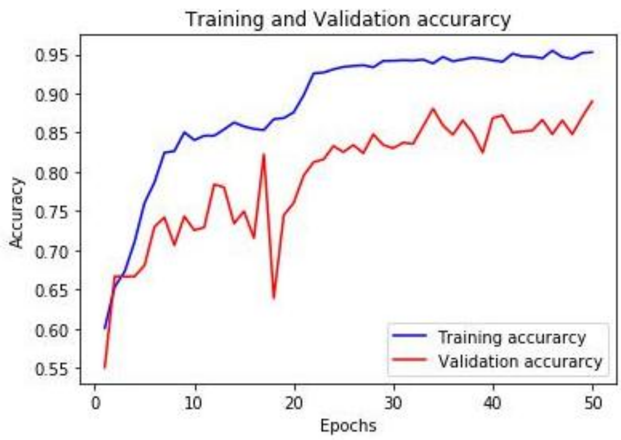

Figure 5: Training and Validation accuracy

From Fig 5 it can be inferred that the training accuracy increases uniformly up to 23 epochs and then achieves stability in the accuracy. In the validation part the accuracy is increasing, but with few spikes observed up to 20 epochs and the accuracy then accomplishes a stability with 85% and the standard deviation of 3%. The spikes are source to further improvement.

The Fig 6 conveys that validation loss and the training loss are continuously decreasing uniformly and the gap between the training and validation loss is very less which is proof that the model is a good fit.

b) Machine Learning Model Results

In this section, the performance of various machine learning models that were used for the training purpose is compared by

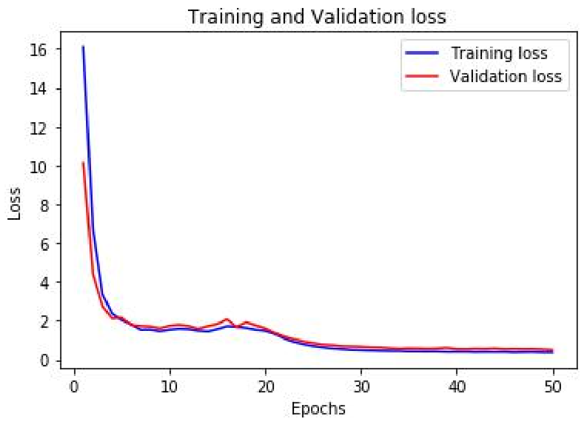

Figure 6: Training and validation loss

applying hyper parameter optimization. In this paper we have considered various combination of machine learning models and optimization techniques to get the optimal model for predicting the correct disease i.e. CN/MCI/AD. In order to analyse results the dataset is divided in training data and testing data in 75% and 25% respectively. The primary goal of the comparison is to evaluate various optimization techniques. Initially for the desired multi class classification among the available models we have applied Random Forest Classifier, Decision Tree Classifier, Support Vector Machine (SVM) and XGBoost. As Table 3 shows the accuracy of these models, without optimizing any hyper parameter (using the default values of hyper parameters) Random Forest, Decision Tree Classifier and SVM gives accuracy of 80.171%, 75.26%, 71.14% respectively. XGBoost performs marginally better than Random Forest Classifier. In order to increase the predictive capability of the models the hyper parameters must be optimized for each model. To choose the best hyper parameters for the model we have used the Grid search method. By applying the Grid Search method it can be seen from the table that there is an increase in the accuracy of each of the models, the highest one is of Random Forest Classifier with 82.302% accuracy. These results show that randomly chosen hyper parameters are more efficient than the manual search.

The precision, recall and F1 score are also calculated along with the accuracy. Before applying hyper parameter optimization, the highest precision was of XGBoost and highest recall was of Random Forest Classifier. Again, the same pattern as that of accuracy was that the precision and the F1 Score increased after Grid search was applied which again suggests that applying hyper parameter optimization is efficient method as compared to manually giving the parameters to the model.

In the confusion matrix (Fig 7), it can be seen that very few patients that are clinically normal (CN) are classified with Alzheimer's disease (AD) and vice-versa. The confusion matrix also shows that there is a high degree of unpredictability between patients that have mild cognitive impairment (MCI) and are classified with Alzheimer's disease (AD) and vice- versa.

Table 3: Comparison of accuracy of different models

Machine Learning Algorithms	Before Hyperparameter Turning				After Hyperparameter Turning			
	Accuracy	Precision	Recall	F1 score	Accuracy	Precision	Recall	F1 score
Random Forest Classifier	80.17	76.85	76.67	76.76	82.30	80.18	77.23	78.40
Decision Tree Classifier	75.26	71.78	72.84	72.25	75.83	71.76	70.21	70.83
SVM	71.14	68.71	60.34	60.47	80.24	79.07	73.95	75.65
Xgboost	81.66	79.50	76.30	77.53	81.94	79.51	77.00	78.01

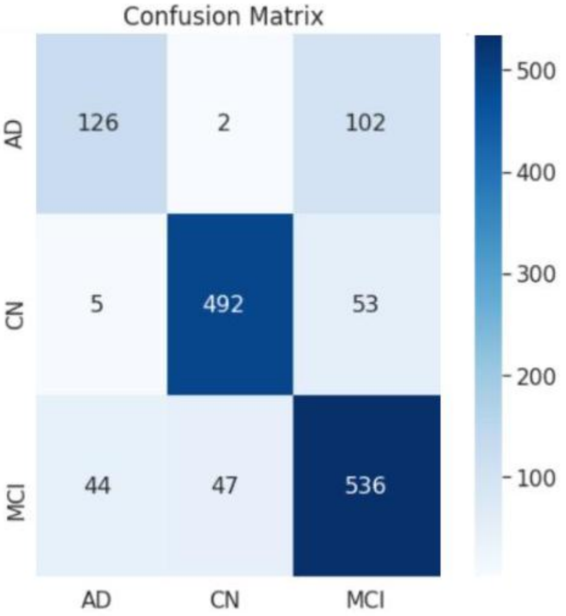

Figure 7: Confusion Matrix

5. Conclusion

Prediction of Alzheimer's disease at the onset is a difficult task to perform manually. The machine learning and deep learning techniques can help tremendously in this regard. In this paper, two ways of predicting the Cognitive Normal patients, patients affected with Mild Cognitive Impairment and Alzheimer's disease are shown. The first of which is through the MRI images of the patients. The deep learning model trained for the MRI images gives the accuracy of 88.98%.The other method in which a comparison is made between the machine learning models like Support Vector Machine, Decision tree and Random Forest on the various diagnostic tests performed by the doctors like MSME, Global CDR, FAQ etc shows that the machine learning models if trained by using hyper-parameter optimization techniques achieves a good accuracy. After applying hyper-parameter optimization random forest performed the best among the three models used with an accuracy of 88.30%.

References

[1] Alzheimer's Association et al., "2014 alzheimer's disease facts and figures, " Alzheimers Dement, vol. 10, no. 2, pp. e47–e92, 2014.
[2] S R, Bhagyashree Sheshadri, Holalu. (2014). An Approach in the Diagnosis of Alzheimer Disease - A Survey. International Journal of En-gineering Trends and Technology. 7. 41-43. 10.14445/22315381/IJETT-V7P229.
[3] G, Reshma P.V.S, Lakshmi. (2019). Prediction of Neurological Disorders among Children Using Machine Learning Techniques. HELIX. 9. 4775-4780. 10.29042/2019-4775-4780.
[4] Tejaswini K., Gracia Jacob, "Feature Selection Techniques for Prediction of Neuro-Degenerative Disorders: A Case-Study with Alzheimer's And Parkinson's Disease". 7th International Conference on Advances in Computing Communications, ICACC-2017, 22-24 August 2017, Cochin, India.
[5] K. Kamnitsas, C. Ledig, V. F. J. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert and B. Glocker, "Efficient multi -scale 3D CNN with fully connected CRF for accurate brain lesion segmentation, " Medical image analysis, vol. 36, pp. 61 -78, 2017.
[6] A.B. Rabeh, et al. Diagnosis of Alzheimer Diseases in Early Step Using SVM (Support Vector Machine). Computer Graphics, Imaging and Visualization (CGiV) 2016: 364-367.
[7] Feng Li et al., A robust deep model for improved classification of AD/MCI patients, IEEE J. Biomed. Health Inform., vol. 19, no. 5, pp. 16101616, 2015.
[8] Yang, Chengliang et al. "Visual Explanations From Deep 3D Convolu-tional Neural Networks for Alzheimer's Disease Classification." AMIA Annual Symposium proceedings. AMIA Symposium vol. 2018 1571-1580. 5 Dec. 2018
[9] Rombouts, SARB, Lazeron, RHC, Barkhof, F, Machielsen, WCM, Witter, MP, Sprenger, M, Valk, J Scheltens, P 1998, "Comparison of medial temporal lobe activation in young and elderly subjects and patients with Alzheimer's disease", Neuroimage, vol. 7, no. 4 PART II.
[10] Basavappa SR; Rao SL; Harish B; Shantala R Basavappa; Shobini L Rao Department of Electronics Communication, Manipal Institute of Technology, Manipal, India. Expert system for dementia/depression diagnosis. NIMHANS Journal. 1996 Apr; 14(2): 99-106.
[11] R. Cuingnet et al., "Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database, " NeuroImage, vol. 56, no. 2, pp. 766–781, 2011.
[12] Pasquier, F. Early diagnosis of dementia: neuropsychology. J Neurol 246, 6–15 (1999). https://doi.org/10.1007/s004150050299
[13] X. Glorot and Y. Bengio, "Understanding the difficulty of train- ing deep feedforward neural networks, " in AISTATS, 2010, pp.249–256.