A Binary Classifier Using Fully Connected Neural Network for Alzheimer's Disease Classification

Rukesh Prajapati[1], Goo-Rak Kwon[1*]

Abstract

Early-stage diagnosis of Alzheimer's Disease (AD) from Cognitively Normal (CN) patients is crucial because treatment at an early stage of AD can prevent further progress in the AD's severity in the future. Recently, computer-aided diagnosis using magnetic resonance image (MRI) has shown better performance in the classification of AD. However, these methods use a traditional machine learning algorithm that requires supervision and uses a combination of many complicated processes. In recent research, the performance of deep neural networks has outperformed the traditional machine learning algorithms. The ability to learn from the data and extract features on its own makes the neural networks less prone to errors. In this paper, a dense neural network is designed for binary classification of Alzheimer's disease. To create a classifier with better results, we studied result of different activation functions in the prediction. We obtained results from 5-folds validations with combinations of different activation functions and compared with each other, and the one with the best validation score is used to classify the test data. In this experiment, features used to train the model are obtained from the ADNI database after processing them using FreeSurfer software. For 5-folds validation, two groups: AD and CN are classified. The proposed DNN obtained better accuracy than the traditional machine learning algorithms and the compared previous studies for AD vs. CN, AD vs. Mild Cognitive Impairment (MCI), and MCI vs. CN classifications, respectively. This neural network is robust and better.

Key Words: Activation Functions, Alzheimer's Disease, Dense Neural Network, FreeSurfer, MRI.

I. INTRODUCTION

In 1906, Alzheimer's disease (AD) was first defined by Alois Alzheimer [1]. AD is a brain disease that is a major health problem in elderly humans. Death is inevitable in AD, and it is the sixth disease that causes most deaths around the world [3-4]. The estimated number of people with AD is about 30 million [3]. Destruction of neurons causes changes in the brain. Only after few years, the symptoms of AD are seen in patients [2]. The symptoms include changes in personality, poor judgment, abnormalities in mood, and sleep [2]. Identifying the transition in the progressive disorder of AD is challenging. AD patient's diagnosis includes a collection of his/her history, neurological and laboratory examinations, and neuroimaging [1].

First, deposits of amyloid β-peptide (Aβ) appear to be neuropathological features of AD [5]. Amyloid precursor protein (APP) generates Aβ peptide. The activity of β-secretase and γ-secretase generates the Aβ peptide Aβ42, which is a longer species, initiates the deposition of amyloid [1]. Another feature of AD appears due to the accumulation of neurofibrillary tangles [5]. Tau protein is responsible for this phenomenon. Tangle formation is proportional to the severity of the disease. More tau tangles lead to severe conditions of AD [1]. Mild Cognitive Impairment (MCI) can be referred to as a phase of AD, in which the symptoms can be observed. Most commonly, the symptoms include a decrease in the ability to learn, unable to perform daily actions which were performed in past, require assistance for the daily activities [6].

For the analysis of the brain, the most used technique is Magnetic resonance imaging (MRI). MRI produces images in a non-invasive imaging technique. With the repetition of varied excitation, MRI produces an image of contrast in tissue for structures of interest [7]. There is a successive pulse sequence time interval, called Repetition Time (TR) and the time interval between delivery and reception of RF pulse and echo signal respectively is Time to Echo (TE). T1-weighted images generate common structural analytical annotation of tissues [8]. In T1-weighted images, both TR and TE times are short [8].

Information about the brain can be represented as an atlas. Desikan-Killiany-Tourville (DKT) atlas uses information based on curvature to define the regions of the brain. For defining the cortical regions in the DKT atlas, it is assumed that the curvature information on the inflated surface would

help to gain high accuracy [9]. By using the mean distance of "mislabeling", DKT atlas detects the geographical mismatch between regions [9]. By the study of brain atlas, the significant changes in the brain due to AD had been noticed [10]. For the analysis of the brain, atlas choice is very important. Researches have shown that the DKT atlas is able to identify significant differences between AD and CN group. In [10], 11 out of 62 structures were found to have a larger percentage difference between AD and CN.

The most commonly used workflow of studies includes feature extraction followed by classification. Many machine learning methods are available for classification. In [11], Support Vector Machine (SVM) classified four groups related to AD. Some studies [12, 13, 32, 25] classified groups using multiple classifiers. In these studies, the basic idea is to find the best classifier or to combine them for better results. Like in [13], classification using Softmax classifier, K-Nearest Neighbor (KNN), SVM, and Naive Bayes (NB) is compared for the best classifier. In [12], the combination of SVM, KNN, and Random Forest (RF) got higher accuracy than using them individually. SVM is found to be widely used in most cases. Modified SVM like Twin-SVM is also used in the classification [14].

Deep Neural Network (DNN) is the most common method which allows machines to perform multiple functions such as classification, analysis, and prediction [15]. Generally, a dense neural network also referred to as Fully Connected (FC) layer, consists of a large portion of the parameters of DNN [16]. In [17], the classification of AD is studied using an FC neural network. In most of the traditional AD analysis methods, many modalities are used at once to improve the performance of the system. Processing multiple modalities require an enormous amount of time followed by several feature selection methods. Only then some levels of satisfactory results were obtained. But these better results are not consistent among different datasets or different traditional classification models. In order to address these problems related with requirement of multiple numbers of modalities, classifiers and, feature extraction methods we propose a binary classifier based on a dense neural network. For this study, we used MRI images of only one modality (T1-weighted). The purpose of testing and comparing activation functions is to choose suitable activation functions that will help the model to learn even from the negative values.

In this study, we develop an FC neural network for the improvement in classification of AD with binary classification task. First, we processed the 3D images obtained from ADNI dataset using the FreeSurfer software. From Free-Surfer, we get atlas as features of the brain. The feature extraction process is followed by feature filtering using Principal Component Analysis (PCA) which allows feature selection, which is then followed by the classification. The

proposed FC neural network comprises two hidden layers. Within those hidden layers, three different activation functions are tested for validation. A combination of activation functions from Parametric ReLU (PReLU), Leaky ReLU, and Exponential Linear Unit (ELU) is selected which has the highest validation accuracy. We performed 5-fold cross validation. After identifying, the model with the best validation accuracy, we used the ssame model to classify the test data for different group classification. This FC neural network performs as a binary classifier.

The contributions of this paper are summarized as follows:

- We propose the combination of two out of three activation functions in the dense neural network with the best validation accuracy.
- We propose use of combination of PCA with the dense neural network for the dimension reduction and feature selection to reduce manual task of filtering features.
- We compare result of our model with the traditional machine learning methods on the same data and compared the result with the previous studies done with the same kind of data and processes.

Our contribution is to introduce a novel method of binary classification for AD detection with higher accuracy than other traditional methods. Furthermore, this system can be utilized in the early diagnosis of various stages of AD patients. Our aim is to develop a system that requires fewer resources but performs better than previous methods. The effectiveness of our proposed model is shown using accuracy, sensitivity, specificity, and bar plot for the comparison with traditional machine learning models.

The paper is arranged as follow: Section II consists of information about materials and methods along with the proposed classifier model; Section III conducts experiments using different activations functions and their results; Section IV shows the comparison of the proposed model with previous models and discussion; finally, the conclusion with the summarization of the paper is in Section V.

II. MATERIAL AND METHOD

2.1. Data

In our study, we accessed data available on Alzheimer's Disease Neuroimaging Initiative (ADNI). It was initiated with the primary objective to investigate whether imaging modalities can measure the progression or early detection of AD.

The dataset consists of 3 groups: AD, MCI, and Cognitively Normal (CN). Total of 178 subjects: 58 AD subjects (21 female, 37 male; age±SD = 75.3±7.9 years; education level = 15.1±3.4), 60 MCI (34 female, 26 male; age±SD =

Table 1. Baseline clinical and socio-demographic characteristics of the studied population (from the ADNI database).

Group	AD	MCI	CN
Nos. of subject	58	60	60
Female/male	21/37	34/26	27/33
Age	75.3±7.9	74.5±3.8	76.4±4.5
Education	15.1±3.4	15.2±2.5	15.5±2.8
CDR	0.7±0.2	0.5	0

74.5±3.8 years; education level = 15.2±2.5), and 60 CN subjects (27 female, 33 male; age±SD = 76.4±4.5 years; education level = 15.5±2.8) as shown in Table 1.

We used equal number of subjects in all groups, to have performance with unbiased estimations. The dataset is split into two parts, 80:20 ratio for training and for testing parts, respectively. The training data is further divided for the training and validation process. The model with the best validation accuracy score is then trained and finally, used to predict test data.

2.2. Features Extraction

After data collection, we extracted features from those images in the next step. Fig. 1 shows the proposed method with the remaining processes. For this study, we ran Free-Surfer using the full recon-all pipeline to compute the DKT atlas which consists of cortical volumetric features. DKT atlas comprises 31 regions from each hemisphere. All regions are listed in Table 2. We used features from both hemispheres, which gave us 62 regions of each subject.

This atlas also provides 9 different anatomical calculations of each region as listed in Table 3. For this study, we used all 62 regions from both hemispheres and all features of each region. After, features extraction process we got 31 x 2 x 9 features of each subject.

Table 2. Regions extracted from each hemisphere using FreeSurfer.

Regions
1. Caudal anterior-cingulate cortex
2. Caudal middle frontal gyrus
3. Cuneus cortex
4. Entorhinal cortex
5. Fusiform gyrus
6. Inferior parietal cortex
7. Inferior temporal gyrus
8. Isthmus-cingulate cortex
9. Lateral occipital cortex
10. Lateral orbital frontal cortex
11. Lingual gyrus
12. Medial orbital frontal cortex
13. Middle temporal gyrus
14. Parahippocampal gyrus
15. Paracentral lobule
16. Pars opercularis
17. Pars orbitalis
18. Pars triangularis
19. Pericalcarine cortex
20. Postcentral gyrus
21. Posterior-cingulate cortex
22. Precentral gyrus
23. Precuneus cortex
24. Rostral anterior cingulate cortex
25. Rostral middle frontal gyrus
26. Superior frontal gyrus
27. Superior parietal cortex
28. Superior temporal gyrus
29. Supramarginal gyrus
30. Temporal pole
31. Transverse temporal cortex

2.3. Features Selection

After feature extraction, we normalized the data, which results in zero mean and unit variance of all features. This process helps in improving data integrity and also reduces data redundancy [12]. For matrix, X the normalized matrix is given by
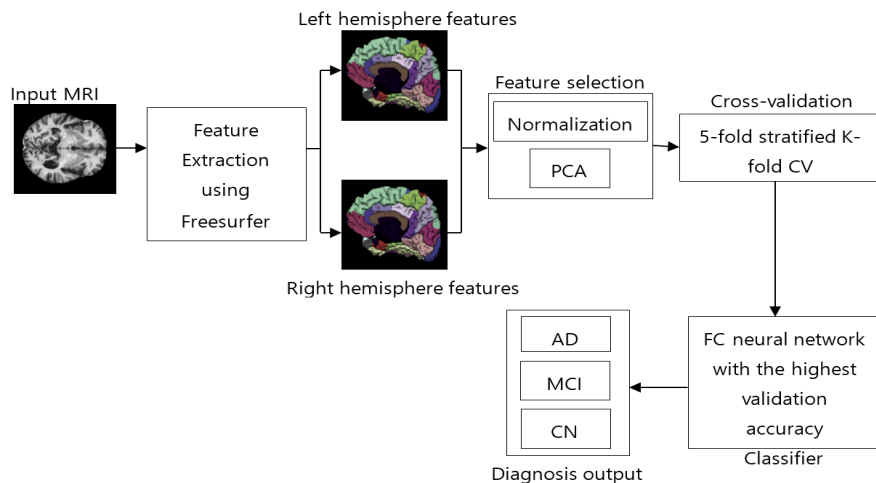

Fig. 1. Block diagram of the proposed method.

Table 3. Anatomical fields calculation of each region included in DKT atlas.

Calculation field
1. Number of vertices
2. Surface area
3. Gray matter volume
4. Average thickness
5. Thickness standard deviation
6. Integrated rectified mean curvature
7. Integrated rectified Gaussian curvature
8. Folding index
9. Intrinsic curvature index

$$X_{norm} = \frac{x_{(i,j)} - mean(X_j)}{std(X_j)},$$

where X_j represents a jth column of X. The columns of matrix X are features and rows are subjects.

Principal Component Analysis (PCA) is a non-parametric, dimensionality-reduction method. PCA helps to extract relevant information from large datasets by reducing the complexity of the dataset and simplifying the structure [18]. PCA creates new features called Principal Components (PCs). In PCA, initial features are combined to create new features. These new features are uncorrelated. the first components are formed by compressing initial variables which comprise most of the information [12]. For this study, we maintained 99% of the variance and determined the number of PCs required. Fig. 2 shows that 104 PCs are required to preserve 99% of the variance.

2.4. Proposed FC Neural Network

The proposed FC neural network comprises 2 hidden layers. After feature selection, we obtained 104 compressed features. Thus, the input layers have 104 nodes. For the study, we used 50 nodes in the first hidden layer and 25 nodes in the second hidden layer as shown in Fig. 3.

The proposed FC network is built using the Keras library. A combination of three different activation functions was
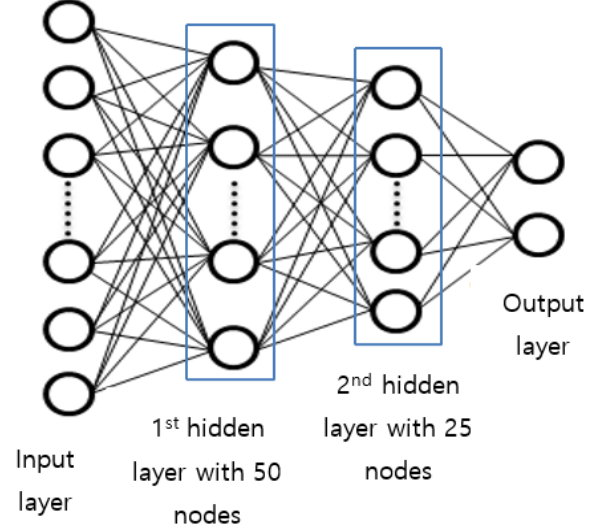

Fig. 3. Proposed FC neural network.

tested. As shown in Fig. 4, these activation functions Leaky ReLU, ELU, and PReLU do not eliminate values for negative values, which overcomes the dying ReLU problem.

For negative values, Leaky ReLU has a slope.

$$f(x) = \begin{cases} x, & if\ x > 0, \\ 0.1x, & otherwise. \end{cases}$$

Instead of the fixed slope, PReLU treats the slope as a parameter.

$$f(x) = \begin{cases} x, & if\ x > 0, \\ ax, & otherwise, \end{cases}$$

where a is a learnable parameter with a negative slope.

In ELU, a straight line for negative value is replaced by a logarithmic curve. For positive inputs, ELU is similar to ReLU.

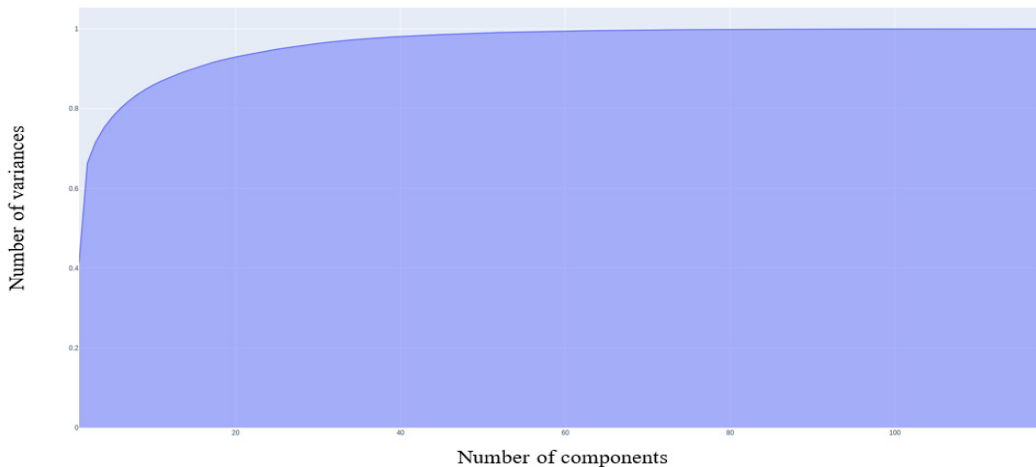$$f(x) = \begin{cases} x, & if\ x > 0, \\ a(e^x - 1), & otherwise. \end{cases}$$

Fig. 2. Number of principal components vs number of variance for AD vs CN group comparison.
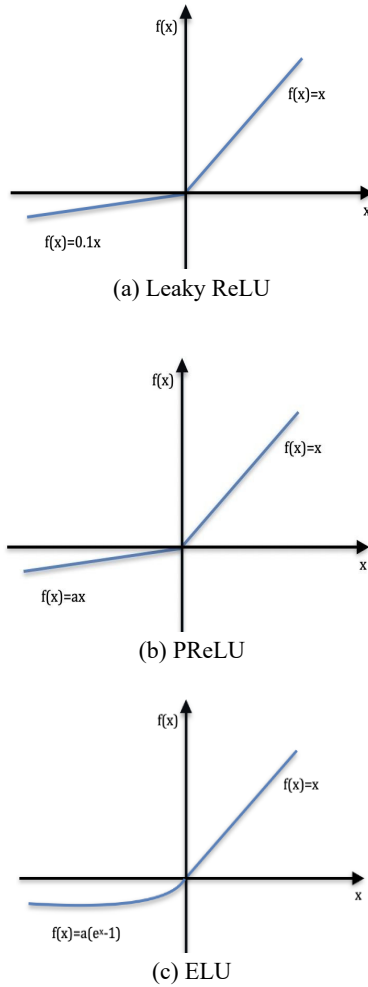
(a) Leaky ReLU

(b) PReLU

(c) ELU

Fig. 4. Activation functions.

III. EXPERIMENT AND RESULTS

3.1. Selection of Activation Functions

For the classification of the test data, there are two parts. The first one is the selection of activation functions based on validation accuracy obtained with these functions. Only training data is used for the validation accuracy. And the second part involves the classification of the test data.

In the validation process, the training data is further divided into two parts: training data and validation data. For this study, we performed three different experiments with ELU, PReLU, and Leaky ReLU in the first hidden layer respectively. Then in the second hidden layer, the same activation functions were used one by one and obtained validation accuracy for a different number of epochs. We obtained validation accuracy for 100, 200, 300, and 400 epochs. Table 4, Table 5, and Table 6 shows the validation accuracy of different FC networks for different epochs.

We performed 5-folds cross-validation. Validation is performed using training data only. Testing data is not used during this process. In this way, we test the model later with

Table 4. Validation accuracy with ELU as activation function in the first hidden layer.

1st hidden layer	2nd hidden layer	Epoch number	Accuracy rate
ELU	ELU	100	76.67%
		200	72.40%
		300	70.41%
		400	68.25%
ELU	PReLU	100	70.35%
		200	72.51%
		300	75.67%
		400	73.51%
ELU	Leaky ReLU	100	65.03%
		200	71.23%
		300	72.40%
		400	**79.94%**

Table 5. Validation accuracy with Leaky ReLU as activation function in the first hidden layer.

1st hidden layer	2nd hidden layer	Epoch number	Accuracy rate
Leaky ReLU	ELU	100	72.57%
		200	72.46%
		300	67.13%
		400	70.18%
Leaky ReLU	PReLU	100	64.91%
		200	77.78%
		300	79.90%
		400	79.88%
Leaky ReLU	Leaky ReLU	100	67.02%
		200	72.46%
		300	74.50%
		400	71.46%

Table 6. Validation accuracy with PReLU as activation function in the first hidden layer.

1st hidden layer	2nd hidden layer	Epoch number	Accuracy rate
PReLU	ELU	100	68.25%
		200	72.40%
		300	69.30%
		400	67.08%
PReLU	PReLU	100	63.98%
		200	70.18%
		300	75.67%
		400	74.44%
PReLU	Leaky ReLU	100	72.46%
		200	72.34%
		300	74.62%
		400	72.34%

the data it has never seen before. Given validation accuracy for each epoch is the average of accuracies obtained from 5-folds cross-validation.

The validation accuracy comparison is performed on the AD vs. CN training data. Comparing the results from the above tables, we can see that the validation accuracy of a model with ELU in the first hidden layer and Leaky ReLU in the second hidden layer has the highest accuracy score of 79.94%. This model is then selected for classification in the next step.

3.2. Classification

From previous process, a model with the highest score is then used to classify the test data. In our study, we found that the ELU and Leaky ReLU in the first and second hidden layers respectively has better performance than others.

The accuracy (ACC) score of the binary classifier on test data is obtained and then evaluated using a confusion matrix. As shown in Fig. 5(a), four elements of the confusion matrix: true positive (TP), true negative (TN), false positive (FP), and false negative (FN) are used to measure additional performance metrics: sensitivity (SEN), and specificity (SPEC).

For the classification of AD vs MCI, the classifier obtained 83.33% accuracy with 75.00% sensitivity, and 91.67% specificity. Similarly, for AD vs CN, the scored 87.50% accuracy with 83.33% sensitivity and 91.70% specificity. And for the final classification group; MCI vs CN, the accuracy score was 79.17% with a sensitivity of 100% and specificity of 58.30%.

3.3. Comparison with Other Methods

For the comparison of the proposed model, we used traditional machine learning methods from scikit-learn [19] to classify the same testing data. We used six classification algorithms: Linear Discriminant Analysis (LDA), KNN, NB, Classification and Regression Trees (CART), Logistic Regression (LR), and SVM. We compared the accuracy score, sensitivity, and specificity of these models with the new proposed model.

In AD vs MCI, our proposed network has 83.33% accuracy, which is higher than other listed classifiers. Scores of each model for AD vs MCI is listed in Table 7 and Fig. 6 (a) graphically represented bar graphs of data from the same table. In the case of sensitivity and specificity, the scores are not the highest one but the other classifier which has a higher score in sensitivity or specificity has a lower score in other performance. Like in this case, KNN has the highest specificity but lower sensitivity and accuracy. Our proposed model was able to maintain an almost similar score in all performance matrices along with the highest accuracy.

Similarly, in the classification of AD vs CN, the proposed

(a) Confusion matrix

(b) AD vs MCI

(c) AD vs CN

(d) MCI vs CN

Fig. 5. Confusion matrix for the classification result obtained from proposed binary classifier.

model still scored the highest accuracy of 87.50%. From the data in Table 8 and Fig. 6(b), we can see that the problem with another classifier still remained in this test data as others classifiers could not score high accuracy or maintain the same performance in other parameters too.

In the final group, the accuracy is lower than the other two groups but still, it has the highest accuracy than other methods in the group as shown in Table 9 and Fig. 6(c). For MCI vs CN, the model scored 79.17% accuracy. The reason

Table 7. Classification of result of AD vs. MCI.

Classifier	ACC	SEN	SPEC
LR	79.20	75.00	83.33
LDA	50.00	58.30	41.67
KNN	75.00	50.00	100.0
SVM	75.00	58.30	91.70
NB	41.70	41.70	41.70
CART	58.30	67.00	50.00
Proposed model	**83.33**	**75.00**	**91.67**

Table 8. Classification of result of AD vs. CN.

Classifier	ACC	SEN	SPEC
LR	79.00	83.33	75.00
LDA	50.00	58.30	41.67
KNN	70.80	50.00	91.70
SVM	50.00	0.0	100.0
NB	67.00	91.70	41.70
CART	62.50	67.00	58.30
Proposed model	**87.50**	**83.33**	**91.70**

Table 9. Classification of result of MCI vs. CN.

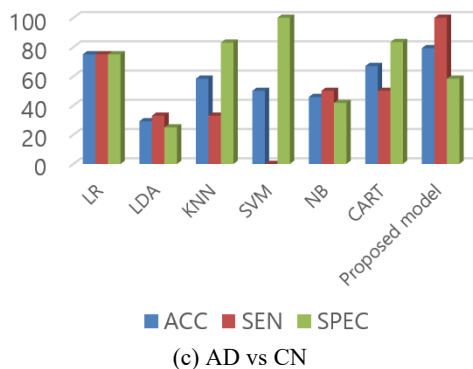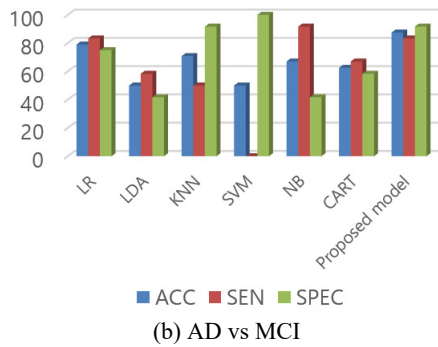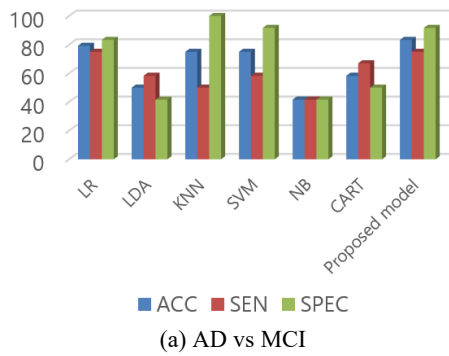| Classifier	ACC	SEN	SPEC
LR	75.00	75.00	75.00
LDA	29.20	33.00	25.00
KNN	58.30	33.00	83.00
SVM	50.00	0.0	100.0
NB	45.80	50.00	41.70
CART	67.00	50.00	83.33
Proposed model	**79.17**	**100.0**	**58.30**

(a) AD vs MCI

(b) AD vs MCI

(c) AD vs CN

Fig. 6. Classification result of each group with performance measure of accuracy, sensitivity, and specificity.

for the poor performance in this group might be because of the reason that there is no vast change in the brain between MCI and CN.

The workstation used in this experiment has Intel® Core™ i5-9600K 3.70GHz CPU, 32 GB RAM, and NVIDIA RTX 2070 GPU.NVIDIA RTX 2070 GPU.

IV. DISCUSSION

In this paper, we proposed a novel technique to classify data for the classification of three different groups of AD. In addition, earlier studies aimed to classify using traditional machine learning classifiers and shallow neural networks with the same type of activation functions in layers. In this paper, we studied the combination of different activation functions in the neural network layers. Three activation functions with non-zero values for negative values are considered. The classification validation accuracy of dense neural networks with a combination of PReLU, ELU, and Leaky ReLU are compared.

Many studies have classified the AD groups using different classifiers. However, direct comparison with state-of-the-art methods is difficult as the studies use different mo-

Table 10. comparison of classification performance for the proposed method with published methods for AD vs. CN.

Authors	Classifiers	ACC	SEN	SPEC
Zhang et al. [20]	SVM	83.1%	80.5%	85.1%
Lin et al. [22]	MLP	82.86%	77.72%	92.31%
Zhang et al. [23]	KSVM	86.71%	85.71%	86.99%
Chyzhyk et al. [24]	DC	74.25%	96%	52.5%
Proposed method	FC-neural network	**87.50%**	**83.33%**	**91.70%**

dalities and datasets. Comparison with the multiclass classification studies is not suitable for our binary classification model. Classification results in Table 10 compare the proposed model with other studies. The study by Zhang et al. [20] extracted landmark-based features and obtained 83.1% accuracy (80.5% sensitivity and 85.1% specificity) in the classification of AD vs. CN using SVM as a classifier.

The same study obtained classification accuracy of 73.6 %(75.3% sensitivity and 69.7% specificity) for the MCI vs. CN. In another study by Lin et al. [23] used inter-class variance (ICV) for key slices selection and eigenbrain was generated, which was followed by Welch's t-test (WTT) to obtain most important eigenbrain (MIE) and used kernel-SVMs (KSVM) as a classifier. They obtained a classification accuracy of 86.71% (85.71% sensitivity and 86.99% specificity). A study by Chyzhyk et al. [24], reduced dimension by Lattice Independent Component Analysis (LICA) and used dendritic computing (DC) for binary classification of AD groups. In that study, an accuracy of 74.25% (96% sensitivity and 52.5% specificity) was achieved.

A method similar to our proposed method was used in a study by Lin et al. [22] for the classification of AD groups. A multi-layer perceptron (MLP) of 5 layers with 3 hidden layers classified AD group in that study. However, the features used for the classification are not the same as in our study. Also, the dataset is different and in our study, we maintained a number of subjects to be equal in every group. The study [22] used radial blood pressure waveform (BPW) and finger photoplethysmography signals to train the MLP network and scored classification accuracy of 82.86% (77.72% sensitivity and 92.31% specificity).

Finally, the proposed FC-neural network achieved an accuracy of 87.50% accuracy, a sensitivity of 83.33%, and a specificity of 91.70% for AD vs. CN group. Comparing accuracy scores, our proposed method outperforms other methods listed in Table 10. We also performed a comparison with the state-of-the-art methods using the same data. The results are shown in Table 7, Table 8, and Table 9. Given that the same data are provided to the classifiers, our proposed FC-neural network achieved a higher accuracy score in all three AD groups (AD vs. CN, AD vs. MCI, and MCI vs. CN). Hence, the obtained results from our model are better and comparable to other models.

In Fig. 6, we can see that the sensitivity and specificity of the proposed method is not the best one. We can see that the KNN and SVM has higher specificity. Similarly, LR and NB has higher sensitivity than proposed method in Fig. 6 (a) and (b) respectively. But as we know that sensitivity is obtained from the true positive and specificity is obtained from the true negative predictions. The models which have higher specificity have lower sensitivity, which means that those classifiers were unable to fit the data properly or we can say that it was able to learn from the true negative only

which can be interpreted as overfitting for those kinds of data only. In same way, for the classifiers with higher sensitivity, they have lower specificity than the proposed model. This indicates that the models are overfitted with either true positive or true negative only. A model or classifier should be able to classify data properly in different test or group classification. From the Fig. 6, it is clear that the specificity and sensitivity of the proposed model is consistent and doesn't vary with huge difference for different group classification. This indicates that the proposed model can classify true positive as well as true negative better than other classifiers pointing out that it is the robust in nature than the others.

Recently, there has been significant improvement in the classification using deep learning models. Comparing the results from the machine learning models and the deep learning models, it is clear that the deep learning models are superior to the machine learning algorithms in case of feature extraction and classifications [26-31]. Although, the deep learning process has many advantages over the machine learning, they require huge data to train the model. In deep learning, more the data more better result is obtained. Along with requirement of huge number of data, the deep learning models also require more computational time and better computational setup with graphics processing units (GPU). Our proposed method has advantage of training and testing with comparatively lower number of subjects as well as lower computational time and computational setup.

V. CONCLUSION

In this paper, a binary classifier using dense neural network is proposed. This method alleviates the problem of necessity of multiple modalities and processes. We designed a fully connected dense neural network with two hidden layers to perform binary classification of AD. After comparing the validation of the model with different activation functions in the hidden layers, the model was finalized. The proposed model is compared with six different traditional machine learning methods. Maintaining specificity and sensitivity, the model scored the highest accuracy in all three groups: AD vs CN, AD vs MCI, and MCI vs CN. For AD vs CN, AD vs MCI, and MCI vs CN the accuracy scores are 87.50%, 83.33%, and 79.17% respectively. Finally, we used only T1-weighted images to extract features and after feature reduction, we classified the data with the proposed model. Comparatively, our model does not require a multiple numbers of modalities as input, and multiple models and processes. From the obtained results, we can say that the proposed method with FC model performance is better than other classification methods used for comparison.

Requirement of lower computational time and lower

computational setup is the advantage of our proposed model. However, deep learning models have shown more robust and better result in the classification of AD. In our future work, we will implement and classify AD subjects using deep learning models and compare it with the machine learning models. In addition, we aim to use different activation function in deep learning models and get a robust and better classifier.

ACKNOWLEDGMENT

REFERENCES

[1] A. S. Schachter and K. L. Davis, "Alzheimer's disease", *Dialogues in Clinical Neuroscience*, vol. 2, no. 2, pp. 91-100, 2000.

[2] Z. Longhe, "2020 Alzheimer's disease facts and figures", *Alzheimer's & Dementia*, vol. 16, no. 3, pp. 391-460, 2020.

[3] R. Haque and A. Levey, "Alzheimer's disease: A clinical perspective and future nonhuman primate research opportunities", in *Proceedings of the National Academy of Sciences*. vol. 116, no. 52, pp. 26224-26229, 2019.

[4] I. J. Deary and L. J. Whalley, "Recent research on the causes of Alzheimer's disease", *BMJ (Clinical research ed.)*, vol. 297, no. 6652, pp. 807-810, Oct. 1988.

[5] M. Silva, C. Loures, L. Alves, L. Cruz de Souza, K. Borges, and M. Carvalho, "Alzheimer's disease: Risk factors and potentially protective measures", *Journal of Biomedical Science,* vol. 26, no. 1, pp. 33, May 2019.

[6] M. S. Albert, S. T. DeKosky, D. Dickson, B. Dubois, H. H. Feldman, and N. C. Fox, 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", *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, vol. 7, no. 3, pp. 270-279, 2011.

[7] S. Bauer, R. Wiest, L. P. Nolte, and M. Reyes, "A survey of MRI-based medical image analysis for brain tumor studies", *Physics in Medicine and Biology*, vol. 58, no. 13, pp. 97-129, 2013.

[8] Q. Lyu and G. Wang, "Quantitative MRI: Absolute T1, T2 and proton density parameters from deep learning", *arXiv: Medical Physics*, 2018.

[9] R. S. Desikan, F. Ségonne, B. Fischl, B. T. Quinn, B. C. Dickerson, and D. Blacker, et al, "An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest", *NeuroImage*, vol. 31, no. 3, pp. 968-980, Mar. 2006.

[10] S. N. Yaakub, R. A. Heckemann, and S. S. Keller, "On brain atlas choice and automatic segmentation methods: A comparison of MAPER & FreeSurfer using three atlas databases", *Scientific Reports*, vol. 10, no. 1, pp. 1-15.

[11] Y. Gupta, R. K. Lama, G. R. Kwon, and Alzheimer's Disease Neuroimaging Initiative, "Prediction and classification of Alzheimer's disease based on combined features from apolipoprotein-E genotype, cerebrospinal fluid, MR, and FDG-PET imaging biomarkers", *Frontiers in Computational Neuroscience,* vol. 13, no. 72, 2019.

[12] Y. Gupta, K. H. Lee, K. Y. Choi, J. J. Lee, B. C. Kim, and G. R. Kwon, et al., "Early diagnosis of Alzheimer's disease using combined features from voxel-based morphometry and cortical, subcortical, and hippocampus regions of MRI T1 brain images", *PloS One*, vol. 14, no. 10, 2019.

[13] Y. Gupta, K. H. Lee, K. Y. Choi, J. J. Lee, B. C. Kim, and G. R. Kwon, "Alzheimer's disease diagnosis based on cortical and subcortical features", *Journal of Healthcare Engineering*, vol. 2019, pp. 2040-2295, Mar. 2019.

[14] S. Alam, G. R. Kwon, J. I. Kim, and C. S. Park, "Twin SVM-based classification of Alzheimer's disease using complex dual-tree wavelet principal coefficients and LDA", *Journal of Healthcare Engineering*, vol. 2017.

[15] A. Eldem, H. Eldem, and D. Üstün, "A model of deep neural network for iris classification with different activation functions," in *2018 International Conference on Artificial Intelligence and Data Processing (IDAP)*, pp. 1-4, 2018.

[16] S. H. Basha, S. Dubey, V. Pulabaigari, and S. Mukherjee, "Impact of fully connected layers on performance of convolutional neural networks for image classification", *Neurocomputing*, vol. 378, pp. 112-119, Feb. 2020.

[17] D. Jha, J. I. Kim, and G. R. Kwon, "Diagnosis of Alzheimer's disease using dual-tree complex wavelet transform, PCA, and feed-forward neural network", *Journal of Healthcare Engineering*, vol. 2017, 2017.

[18] J. Shlens, "A tutorial on principal component analysis", *ArXiv, abs/1404.1100*, 2014.

[19] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, and M. Blondel, et al., "Scikit-learn: Ma-

chine learning in python*", Journal of Machine Learning Research*, vol. 12, no. 85, pp. 2825-2830, 2011.

[20] J. Zhang, Y. Gao, B. C. Munsell, and D. Shen, "Detecting anatomical landmarks for fast Alzheimer's disease diagnosis", *IEEE Transactions on Medical Imaging*, vol. 35, no. 12, pp. 2524-2533, 2016.

[21] N. Kumar, Dr. Manhas, and V. Sharma, "Comparative study to measure the performance of commonly used machine learning algorithms in diagnosis of Alzheimer's disease", *Journal of Multimedia Information System*, vol. 6, no. 2, pp. 75-80, 2019.

[22] S. K. Lin, H. Hsiu, H. S. Chen, and C. J. Yang, "Classification of patients with Alzheimer's disease using the arterial pulse spectrum and a multilayer-perceptron analysis", *Scientific Reports*, vol. 11, no. 1, 2021.

[23] Y. Zhang, Z. Dong, P. Phillips, S. Wang, G. Ji, and J. Yang, et al., "Detection of subjects and brain regions related to Alzheimer's disease using 3D MRI scans based on eigenbrain and machine learning", *Frontiers in Computational Neuroscience*, vol. 9, 2015.

[24] D. Chyzhyk, M. Graña, A. Savio, and J. Maiora, "Hybrid dendritic computing with kernel-LICA applied to Alzheimer's disease detection in MRI", *Neurocomputing*, vol. 75, no. 1, pp. 72-77, 2012.

[25] F. R. Faisal, U. Khatri, and G. R. Kwon, "Diagnosis of Alzheimer's disease using combined feature selection method," *Journal of the Korean Society for Multimedia*, vol. 24, no. 5, pp. 667-675, 2021.

[26] S. J. Park, B. G. Kim, and N. Chilamkurti, "A robust facial expression recognition algorithm based on multi-rate feature fusion scheme", *Sensors*, vol. 21, no. 21, 2021.

[27] S. Mukherjee, P. Kumar, R. Saini, P. P. Roy, D. P. Dogra, and B. G. Kim, "Plant disease identification using deep neural networks," *Journal of Multimedia Information System*, vol. 4, no. 4, pp. 233-238, 2017.

[28] D. Prakash, N. Madusanka, S. Bhattacharjee, H.-G. Park, C. H. Kim, and H. K. Choi, "A comparative study of Alzheimer's disease classification using multiple transfer learning models," *Journal of Multimedia Information System*, vol. 6, no. 4, pp. 209-216, 2019.

[29] D. Jeong, B. G. Kim, and S. Y. Dong, "Deep Joint Spatiotemporal Network (DJSTN) for efficient facial expression recognition", *Sensors*, vol. 20, no. 7, 2020.

[30] J. Manhas, R. K. Gupta, and P. P. Roy, "A review on automated cancer detection in medical images using machine learning and deep learning based computational techniques: Challenges and opportunities", *Archievs of Computational Methods in Engineering*, pp.1-41, 2021.

[31] J. H. Kim, B. G. Kim, P. P. Roy, and D. -M. Jeong, "Efficient facial expression recognition algorithm based on hierarchical deep neural network structure," in *IEEE Access*, vol. 7, pp. 41273-41285, 2019.

[32] N. kumar, J. manhas, and V. sharma, "Comparative study to measure the performance of commonly used machine learning algorithms in diagnosis of Alzheimer's disease," *Journal of Multimedia Information System*, vol. 6, no. 2, pp. 75-80, 2019.

AUTHORS

Rukesh Prajapati received his BE degree in the Depart ment of Electronics and Communication Engineering from Khwopa Engineering College, Purbanchal University, Nepal, in 2017. In 2020, he joined the Department of Information and Communication Engineering for pursuing his Master's degree at Chosun University.

His research interests include image processing, image segmentation, deep learning, computer vision.

Goo-Rak Kwon received a Ph.D. from the Department of Mechatronic Engineering, Korea University, in 2007. He served as Chief Executive Officer and the Director of Dalitech Co. Ltd. from 2004 to 2007. He joined the Department of Electronic Engineering, Korea University, from 2007 to 2008, where he was a Postdoctoral Researcher supporting the BK21 Information Technique Business. He has been a Professor with Chosun University, since 2017. He has also been an Associate Dean with the Industry-academic Cooperation Foundation, since 2018. He has contributed 55 and 81 articles to journals and conference proceedings, respectively. He also holds 27 patents on medical image analysis and the security of multimedia contents for digital rights management. He was a member of the IEICE and IS&T international institutes. In domestic institutes, he was a member of the signal processing society in the IEIE, KMMS, KIPS, and KICS. His research interests include medical image analysis, A/V signal processing, video communication, and applications.