Deep Learning Based Binary Classification for Alzheimer's Disease Detection using Brain MRI Images

Emtiaz Hussain‡, Mahmudul Hasan‡, Syed Zafrul Hassan‡, Tanzina Hassan Azmi‡,
Md Anisur Rahman*, Mohammad Zavid Parvez‡
*School of Computing & Mathematics, Charles Sturt University, Australia
‡Software Engineering and HCI Research Group, Department of Computer Science and Engineering,
BRAC University, Bangladesh
Email: ‡{ehrupok, hasanpiyal, hassanshozib, tanzina2090}@gmail.com,
*arahman@csu.edu.au, ‡zavid.parvez@bracu.ac.bd

***Abstract*—Alzheimer's disease is an irremediable, continuous brain disorder that gradually destroys memory and thinking skills and, eventually, the ability to carry out the simplest tasks. It has become one of the critical diseases throughout the world. Moreover, there is no remedy for Alzheimer's disease. Machine learning techniques, especially deep learning-based Convolutional Neural Network (CNN), are used to improve the process for the detection of Alzheimer's disease. In recent days, CNN has achieved major success in MRI image analysis and biomedical research. A lot of research has been carried out for the detection of Alzheimer's disease based on brain MRI images using CNN. However, one of the fundamental limitations is that proper comparison between a proposed CNN model and pre-trained CNN models (InceptionV3, Xception, MobilenetV2, VGG) was not established. Therefore, in this paper, we present a model based on 12-layer CNN for binary classification and detection of Alzheimer's disease using brain MRI data. The performance of the proposed model is compared with some existing CNN models in terms of accuracy, precision, recall, F1 score, and ROC curve on the Open Access Series of Imaging Studies (OASIS) dataset. The main contribution of the paper is a 12-layer CNN model with an accuracy of 97.75%, which is higher than any other existing CNN models published on this dataset. The paper also shows side by side comparison between our proposed model and pre-trained CNN models (InceptioV3, Xception, MobilenetV2, VGG). The experimental results show the superiority of the proposed model over the existing models.**

***Index Terms*—Alzheimer, Machine Learning, Deep Learning, CNN, MRI, OASIS-1, Confusion Matrix, Accuracy, ROC Curve.**

I. Introduction

According to the World Health Organization (WHO), Alzheimer's has become one of the major diseases throughout the world. There are approximately 50 million people all around the world who have Alzheimer's disease and different types of dementia. The total number of new cases of dementia each year worldwide is nearly 10 million, implying 1 new case every 3 seconds. The number of people with dementia is expected to increase to 82 million in 2030 and 152 million in 2050. Alzheimer's disease mainly causes memory loss [1]. Eventually, patients forget their own identity, like name, age, and addresses [2]. As it is a gradual process at some point, patients do not even remember their family members and relatives. For a successful diagnosis of Alzheimer's disease (AD), several examinations and assessments are required, like physical and neuro-biological exams, Mini-Mental State Examination (MMSE), and patient's detailed history. Doctors today use brain MRI images for the diagnosis of Alzheimer's disease.

4

Convolutional Neural Networks (CNN) have been successfully used in medical image analysis such as MRI [3], X-ray [4], CT scans [5], Ultrasonography [6], etc. CNN have also been very successful in natural language processing [7], computer vision [8], audio recognition [9] and speech recognition [10]. Furthermore, a neural network is a series of algorithms that recognize relationships in a set of data through a process that is very similar to human brain operation. This algorithm is very effective for pattern recognition and image processing. It takes images as input and builds a model that processes the images to extract the features from the images [11] and recognize a pattern [12]. By using the pattern, CNN identifies the similarities of a new input as accurately as possible. This algorithm is very popular because of its simple structure, adaptability, reduced training parameters, and this algorithm reduces the complexity of a network model.

As Alzhiemer's disease detection using CNN is a well-established research area, we have found some great research works in this field. In recent years some extraordinary research works have been done on the OASIS dataset using CNN or Artificial Neural Networks (ANN). In 2019, Khagi *et al.* used pre-trained CNN models namely Alexnet, GoogleNet, Resnet50 to detect healthy (i.e. Non-Demented) vs Alzheimer's patients (i.e. Demented) [13]. In their research, they used 28 Normal Controls and 28 Alzheimer's patients. In 2018, Wang *et al.* did classification of Alzheimer's disease based on an 8 layer CNN [14]. They used 98 AD patients and 98 Normal Controls for their research. The 8 layer-based CNN achieved an accuracy of 97.65%. Hon *et al.*

978-1-7281-5169-4/20/$31.00 ©2020 IEEE

detected Alzheimer's through transfer learning in 2017 [15]. Furthermore, in the same year, Islam *et al.* designed a deep neural network (i.e., CNN) to detect Alzheimer's disease [16]. Their designed deeper version of inception achieved an accuracy of 73.75%, whereas the traditional inception achieves an accuracy of 64.25%. Mahmood *et al.* did the detection and classification of Alzheimer's using principal component analysis (PCA) and ANN in 2013 [17]. Their research achieved nearly 90% accuracy. In another research, García-Sebastián used morphometry based features for Alzheimer's detection in 2009 [18]. In the same year Savio *et al.* used ANN for detection and classification of Alzheimer's disease [19].

All the research mentioned above works have been done in recent years and have achieved better results on the OASIS dataset [20]. However, there are some limitations in the existing research works mentioned as follows: (1) some of them have lower accuracy than others [16], (2) some models achieve higher accuracy but the performance of them have not been appropriately demonstrated against the pre-trained models [14], (3) another research focuses only on the performance of the pre-trained models for Alzheimer's detection, but they did not propose any new model [13]. Therefore, in this paper, we propose a model to overcome the above limitations. Our proposed CNN model is based on a 12-layer architecture, which consists of convolutional, max pooling, dense, and flatten layers and three activation functions, namely Sigmoid, ReLU, and Leaky ReLU. Our proposed model has been used for binary classification and detection of Alzheimer's disease. The performance of our proposed model is compared with some existing CNN models that demonstrate the superiority of our proposed model over the existing models. The main contribution of the paper is as follows:

1) A 12-layer CNN architecture, which has achieved an accuracy of 97.75%, which is higher than any other previous studies that have been done before on the OASIS dataset.
2) The performance of our proposed model is better than some pre-trained models, namely InceptionV3, Xception, MobilenetV2, and VGG19.

The rest of the paper is organized as follows. In section II, we have described our proposed model of work for the whole study. In section III, we have presented the experimental results and discussion. Finally, we conclude our study with future plans in section IV.

II. Our Proposed Method

In this section, we discuss our proposed method that consist of a new 12-layer CNN model and compare the performance of the model with the pre-trained models (InceptionV3, Xception, MobileNetv2, VGG19). Our proposed method is divided into seven basic steps as shown below:

Step 1: Dataset collection.
Step 2: Data pre-processing.
Step 3: Data labeling.
Step 4: Our proposed 12-layer CNN model.
Step 5: Demonstrate performance of our proposed model.
Step 6: Loading pre-trained CNN models.
Step 7: Demonstrate performance of the pre-trained CNN models and comparing results with our proposed model.

A block diagram on our proposed method is presented in Fig. 1 that shows our workflow of the proposed method from beginning to end. This block diagram also shows the seven basic steps of our proposed method.

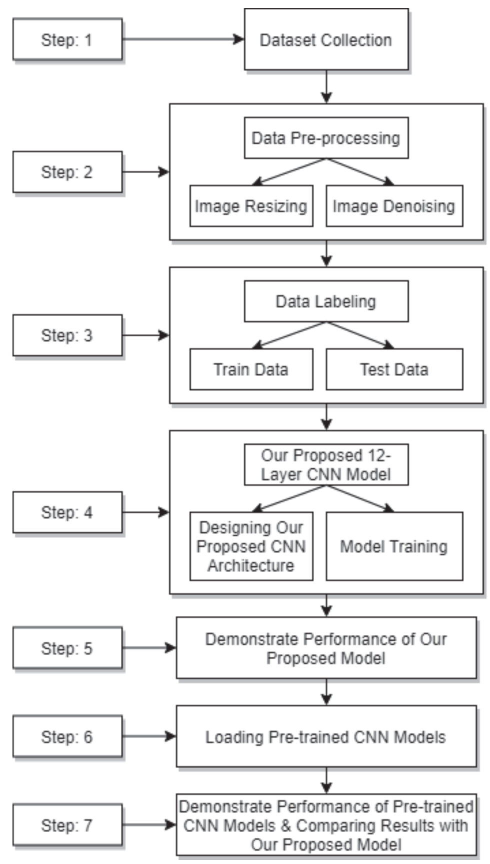

Fig. 1. A Block Diagram of Our Proposed Method.

A. Dataset collection

We have collected our data from the OASIS dataset [20]. OASIS stands for Open Access Series Of Imaging Studies. This dataset contains a cross-sectional collection of 416 subjects. These subjects are aged from 18 to 96 years. For every subject, 3 or 4 individual T1-weighted MRI scans are included that were acquired in single scan sessions. The subjects include both men and women, and all of them are right-handed. 100 out of the 416 subjects that are aged over 60 years have been diagnosed with Alzheimer's disease (AD), ranging from very mild to moderate level. In addition, for 20 Non-Demented subjects, a reliability data set is included that contains images of the following visit within 90 days of their initial session.

Fig 2 shows sample images of an Alzheimer's patient from the Oasis dataset.

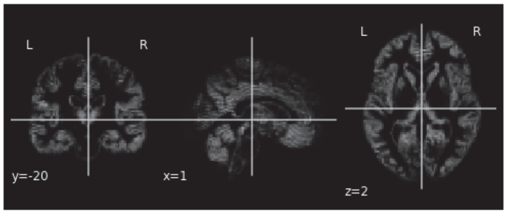

Fig. 2. OASIS-I MRI Image Data Example.

B. Data pre-processing

There are different sizes of images in the dataset. The different sizes of images can influence the architecture towards low accuracy.That's why we performed data pre-processing. There are two parts in our data pre-processing: a) Image resizing, and b) Image denoising. Image resizing reduces the time of neural network model training. We resized the images using OpenCV python. One of the fundamental challenges in image processing and computer vision is image denoising. What denoising does is to estimate the original image by suppressing noise from the image. We did image denoising on the brain MRI images from our OASIS dataset to get better performance of our model. We used python OpenCV3 to denoise images. Fig. 3 and Fig. 4 show images before and after denoising.

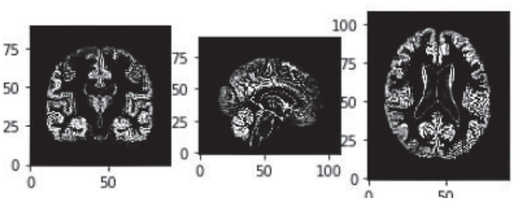

Fig. 3. Before Denoising.

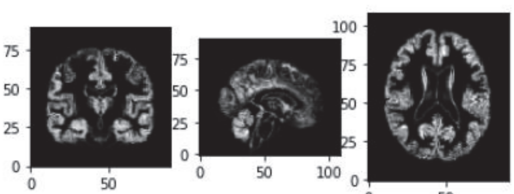

Fig. 4. After Denoising.

C. Data labeling

After pre-processing, we labeled our data for binary classification and fixed our sample size. As we are doing binary classification, therefore we have labeled the records of the dataset with Clinical Dementia Ratio(CDR) 0 or 1. Note that CDR 0 indicates healthy (i.e., Non-Demented), and CDR 1 indicates severe Alzheimer's (i.e., Demented). There are 28 patients with CDR 1. For classification purposes, we considered 28 Alzheimer's patients and 28 Non-Demented patients. For every patient, there are two images. Then we divided the dataset into an 8:2 ratio based on random selection. That means 80% of data are used for training, and 20% of data are used for testing purposes.

D. Our Proposed 12-layer CNN model

In this section, we discuss our proposed 12-layer CNN model for the detection and classification of Alzheimer's disease using brain MRI images. Our 12-layer CNN model has five steps:

1) **Convolutional layer selection:** In our proposed CNN model, we used Conv2D. We have used four conv2D layers in our model.
2) **Pooling layer selection:** In this model, we have used Maxpooling2D. For every Conv2D layer, we have used a Maxpool2D layer. Therefore, we have used four Max-Pool2D layers.
3) **Flatten Layer:** In our model, after using the pooling layer, we used a flatten layer to flatten the whole network.
4) **Dense Layer:** After the flatten layer, we have used two dense layers. The dense layers are also known as fully connected layers.
5) **Activation Function:** We have used Sigmoid function as shown in Eq. 1 with another dense layer and ReLU function, as shown in Eq. 2. We have also used a Leaky ReLU activation function as it has proved to give the best performance with Maxpooling2D. The three activation functions are shown as follows:

$$f(x) = \frac{1}{1 + exp(-x)} \tag{1}$$

$$ReLU = max(0, x) \tag{2}$$

$$f(x) = 1(x < 0)(x) + 1(x >= 0)(x) \tag{3}$$

In the above equations, x is a real-valued number, and is a small constant. The value of the sigmoid function ranges from 0 to 1. The sigmoid function is an S-shaped curve. It is comparatively easy, but it has a vanishing gradient problem. The output is not centered at zero; as a result, the gradient updates go too far in different directions, which makes the optimization harder. On the other hand, ReLU is an efficient and straightforward activation function as it avoids the vanishing gradient problem. Leaky ReLU has been used to solve the "dying ReLU" problem. Instead of the function being 0 where x <0, a Leaky ReLU will have a small negative slope.

Table I shows summary of our proposed 12-layer CNN model.

TABLE I
A SUMMARY OF OUR PROPOSED CNN MODEL.

Layer (type)	Output Shape	Param #
conv2d_64 (Conv2D)	(None, 89, 107, 32)	26240
max_pooling2d_64 (MaxPooling)	(None, 44, 53, 32)	0
conv2d_65 (Conv2D)	(None, 42, 51, 64)	18496
max_pooling2d_65 (MaxPooling)	(None, 21, 25, 64)	0
conv2d_66 (Conv2D)	(None, 19, 23, 128)	73856
max_pooling2d_66 (MaxPooling)	(None, 9, 11, 128)	0
conv2d_67 (Conv2D)	(None, 7, 9, 256)	295168
max_pooling2d_67 (MaxPooling)	(None, 3, 4, 256)	0
flatten_13 (Flatten)	(None, 3072)	0
dense_22 (Dense)	(None, 512)	1573376
leaky_re_lu_15 (LeakyReLU)	(None, 512)	0
dense_23 (Dense)	(None, 1)	513

E. Demonstrate performance of our proposed model

In order to analyze the performance of our proposed CNN model, we calculated precision, recall, F1 score, accuracy, and ROC curve. The equations of accuracy, f1-score, and precision are given as follows:

Accuracy equation [21]–[23]:

$$\begin{aligned} ACC &= \frac{TP+TN}{P+N} \\ &= \frac{TP+TN}{TP+TN+FP+FN} \end{aligned} \tag{4}$$

Where-

- ACC = Accuracy
- TP = True Positive
- TN = True Negative
- P = Condition Positive
- N = Condition Negative

F1-score equation [21]–[23]:

$$\begin{aligned} F1 &= 2 \times \frac{PPV \times TPR}{PPV + TPR} \\ &= \frac{2TP}{2TP+FP+FN} \end{aligned} \tag{5}$$

Where-

- PPV = Positive Predictive Value/ Precision
- TPR = True Positive Rate
- TP = True Positive
- FP = False Positive
- FN = False Negative

Precision equation [21]–[23]:

$$\begin{aligned} PPV &= \frac{TP}{TP+FP} \\ &= 1 - FDR \end{aligned} \tag{6}$$

Where-

- FDR = False Discovery Rate
- PPV = Positive Predictive Value/ Precision
- TP = True Positive
- FP = False Positive

Recall equation [21]–[23]:

$$\begin{aligned} PPV &= \frac{TP}{TP+FN} \\ &= 1 - FDR \end{aligned} \tag{7}$$

Where-

- FDR = False Discovery Rate
- PPV = Positive Predictive Value/ Precision
- TP = True Positive
- FN = False Negative

Table II shows the performance of our proposed CNN model, which will be discussed briefly in comparison with pre-trained CNN models, namely InceptionV3. Xception, MobilenetV2, and VGG19 in section III precision, recall and F1 score,

TABLE II
PERFORMANCE ANALYSIS OF THE MODEL

Precision	Demented	100%
	Non-Demented	93%
Recall	Demented	92%
	Non-Demented	100%
F1-Acore	Demented	97%
	Non-Demented	98%
Accuracy	97.75%	
ROC	99.21%	

In order to demonstrate the better performance of our proposed 12-layer CNN model, we have also plotted model accuracy in Fig 5 and model loss in Fig 6. These figures shows that our proposed 12-layer CNN model is neither under-fit nor over-fit. This explains why we achieved better results in terms of accuracy, precision, recall, f1- score, and area under ROC curve.

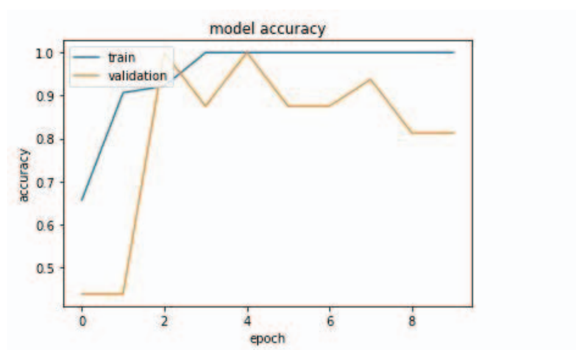

Fig. 5. Model Accuracy Graph.

F. Loading pre-trained CNN Models

In this section, we load the pre-trained CNN models, which are InceptionV3, Xception, MobileNetV2, and VGG19. It is noted that after loading each pre-trained model we added a flatten layer and two dense layers with each model. Activation

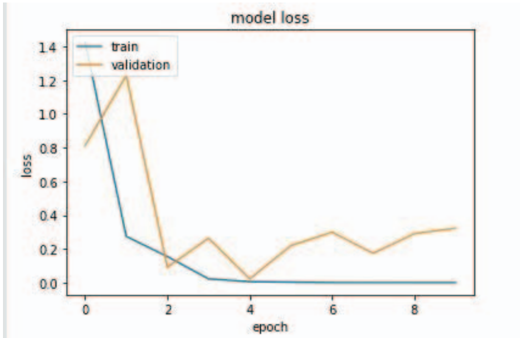

Fig. 6. Model Loss Graph.

functions sigmoid and relu have also been used with each dense layer.

III. Experimental Results and Discussion

We compare the performance of our proposed model with four pre-trained models, namely InceptionV3, Xception, MobileNetV2, and VGG19. In our experiment, we use the OASIS dataset that we collected from [20]. The experimental results are presented in Table III and Table IV. From Table III, we can see that our proposed 12-layer CNN model has achieved the highest accuracy of 97.75%, whereas the accuracy of InceptionV3, Xception, and MobileNetV2 are 90.62%, 84.37%, and 81.24%, respectively. Note that, VGG19 has performed very poorly on the dataset with only 50% accuracy. The performance of our proposed model in terms of f1-score and area under ROC curve is also better than the pre-trained models that we can see from Table III.

As we have mentioned earlier, the main purpose of this study is to establish a CNN model that achieves higher accuracy than any other existing models of CNN. We achieved our goal having an accuracy of 97.75% with our proposed 12-layer CNN model in binary classification and detection of Alzheimer's disease. Note that our proposed CNN model has achieved the highest accuracy on the OASIS dataset. To the best of our knowledge, the previously achieved highest accuracy of an 8-layer CNN model is 97.65% [14] on the OASIS dataset.

From Table IV, we can see that the performance of our proposed CNN model is better than the pre-trained models in terms of precision and recall. From the experimental results, we can mention that the performance of our proposed 12-layer CNN model is better than the pre-trained models in terms of all evaluation criteria that we used for comparison.

TABLE III
Performance Comparison in terms of F1-score, ROC, and Accuracy.

Model Name	f1-score		Area Under ROC Curve	Accuracy
	Demented	Non-Demented		
Proposed Model	97%	98%	99.21%	97.75%
InceptionV3	90%	91%	96.48%	90.62%
Xception	81%	86%	96.48%	84.37%
Mobile-NetV2	79%	83%	85.15%	81.24%
VGG19	0%	67%	50%	50%

TABLE IV
Performance Comparison in terms of Precision and Recall.

Model Name	Precision		ReCall	
	Demented	Non-Demented	Demented	Non-Demented
Proposed Model	100%	93%	92%	100%
InceptionV3	93%	88%	88%	94%
Xception	100%	76%	69%	100%
Mobile-NetV2	92%	75%	69%	94%
VGG19	0%	50%	0%	100%

IV. Conclusion

In this paper, we presented a 12-layer CNN model for binary classification and detection of Alzheimer's disease. We performed our study on the OASIS dataset. We used data pre-processing techniques, namely, Image resizing and Image denoising. Our proposed 12-layer CNN model is based on deep learning and machine learning algorithms. Our proposed model performs better than an existing 8-layer CNN model, and four pre-trained CNN models. Our future research plan is to perform multi-class classification on the OASIS dataset and early detection of Alzheimer's disease.

References

[1] K. Parsons-Suhl, M. E. Johnson, J. J. McCann, and S. Solberg, "Losing One's Memory in Early Alzheimer's Disease", Qualitative Health Research, vol. 18, no. 1, pp. 31–42, 2008.
[2] H. MacRae, "Managing Identity While Living With Alzheimer's Disease", Qualitative Health Research, vol. 20, no. 3, pp. 293–305, 2010.
[3] L. Zou, J. Zheng, C. Miao, M. J. Mckeown and Z. J. Wang, "3D CNN Based Automatic Diagnosis of Attention Deficit Hyperactivity Disorder Using Functional and Structural MRI," in IEEE Access, vol. 5, pp. 23626-23636, 2017.
[4] C. Liu et al., "TX-CNN: Detecting tuberculosis in chest X-ray images using convolutional neural network", 2017 IEEE International Conference on Image Processing (ICIP), Beijing, pp. 2314-2318, 2017.
[5] X. Zhao, L. Liu, S. Qi, Y. Teng, J. Li and W. Qian, "Agile convolutional neural network for pulmonary nodule classification using CT images", Int J Comput Assist Radiol Surg, Vol. 13, no. 1, pp. 585–595, 2018.
[6] Liu J. et al. "Integrate Domain Knowledge in Training CNN for Ultrasonography Breast Cancer Diagnosis," in Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 868-875, 2018.

[7] A. M. Alayba, V.Palade, M. England and R. Iqbal, ”A Combined CNN and LSTM Model for Arabic Sentiment Analysis.” in Proc International Cross-Domain Conference for Machine Learning and Knowledge Extraction, pp. 179-191, 2018.
[8] X. Penga and C. Schmid, ”Multi-region Two-Stream R-CNN for Action Detection.” in Proc. European Conference on Computer Vision, pp. 744-759, 2016.
[9] S. Zhang, S. Zhang, T. Huang, W. Gao and Q. Tian, ”Learning Affective Features With a Hybrid Deep Model for Audio–Visual Emotion Recognition,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 28, no. 10, pp. 3030-3043, 2018.
[10] O. Abdel-Hamid, A. Mohamed, H. Jiang and G. Penn, ”Applying Convolutional Neural Networks concepts to hybrid NN-HMM model for speech recognition,” in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Kyoto, pp. 4277-4280, 2012.
[11] Y. Chen, H. Jiang, C. Li, X. Jia and P. Ghamisi, ”Deep Feature Extraction and Classification of Hyperspectral Images Based on Convolutional Neural Networks,” IEEE Transactions on Geoscience and Remote Sensing, vol. 54, no. 10, pp. 6232-6251, 2016.
[12] L. Chen, S. Wang, W. Fan, J. Sun and S. Naoi, ”Beyond human recognition: A CNN-based framework for handwritten character recognition,” in Proc. 3rd IAPR Asian Conference on Pattern Recognition (ACPR), Kuala Lumpur, pp. 695-699, 2015.
[13] B. Khagi, B. Lee, J. Pyun and G. Kwon, ”CNN Models Performance Analysis on MRI images of OASIS dataset for distinction between Healthy (i.e. Non Demented) and Alzheimer’s patient (i.e. Demented),” 2019 International Conference on Electronics, Information and Communication (ICEIC), Auckland, New Zealand, pp. 1-4, 2019.
[14] S. Wang, P. Phillips, Y. Sui et al. ”Classification of Alzheimer’s Disease Based on Eight-Layer Convolutional Neural Network with Leaky Rectified Linear Unit and Max Pooling”, J Med Syst, vol. 42, no. 85, 2018.
[15] M. Hon and N. M. Khan, ”Towards Alzheimer’s disease classification through transfer learning,” in Proc. IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Kansas City, pp. 1166-1169, 2017.
[16] J. Islam, Y. Zhang, ”A Novel Deep Learning Based Multi-class Classification Method for Alzheimer’s Disease Detection Using Brain MRI Data,” in Proc. International Conference on Brain Informatics, pp. 213-222, 2017.
[17] R. Mahmood and B. Ghimire, ”Automatic detection and classification of Alzheimer’s Disease from MRI scans using principal component analysis and artificial neural networks,” in Proc. 20th International Conference on Systems, Signals and Image Processing (IWSSIP), Bucharest, pp. 133-137, 2013.
[18] M. García-Sebastián, A. Savio, M. Graña, J. Villanúa, ”On the Use of Morphometry Based Features for Alzheimer’s Disease Detection on MRI,” in Proc. 10th International Work-Conference on Artificial Neural Networks (IWANN), Salamanca, Spain, pp. 957-964, 2009.
[19] A. Savio, M. García-Sebastián, C. Hernández, M. Graña, J. Villanúa, ”Classification Results of Artificial Neural Networks for Alzheimer’s Disease Detection,” in Proc. Intelligent Data Engineering and Automated Learning (IDEAL), Burgos, Spain, 2009.
[20] D. S. Marcus, T. H. Wang, J. Parker, J. G. Csernansky, J. C. Morris, and R. L. Buckner, “Open access series of imaging studies (oasis): Cross-sectionalmri data in young, middle aged, nondemented, and demented older adults,” Journal of cognitive neuroscience, vol. 19, no. 9, pp. 1498–1507, 2007.
[21] C. Nicholson, Evaluation metrics for machine learning-accuracy, precision, recall, and f1 defined, 2019. [Online]. Available: https://pathmind.com/wiki/accuracy-precision-recall-f1.
[22] M. A. Rahman, *Automatic Selection of High Quality Initial Seeds for Generating High Quality Clusters without Requiring any User Inputs*. PhD dissertation, Charles Sturt University, Australia, 2014.
[23] M. A. Rahman, L.M. Ang and K.P. Seng, “Unique neighborhood set parameter independent density-based clustering with outlier detection,” IEEE Access, vol. 6, no. 1, pp. 44707-44717, 2018.