DOI: https://doi.org/10.33103/uot.ijccce.24.1.5

A Comparative Study Using Deep Learning Models And Transfer Learning for Detection And Classification of Alzheimer's Disease

Afrah Salman Dawood
Communication Engineering Department, University of Technology, Baghdad, Iraq
afrah.s.dawood@uotechnology.edu.iq

Abstract— *Recently, the burgeoning disciplines of Machine Learning (ML) and Deep Learning (DL) have experienced considerable integration across diverse scientific domains. Of significant note is their integration into the medical sector, specifically in the intricate methodologies of pathological categorization. Present-day innovations underscore the pivotal role of Deep Convolutional Neural Networks (DCNN) in mediating the tasks of image-based taxonomies and prognostications within this domain. In this research, a new DCNN with different modified intelligent architectures like CNN, modified VGG-16, VGG-19, ResNet50, and DenseNet121, besides the newly added classification layer, was implemented and tested for the detection and classification of Alzheimer's disease. The evaluation and performance metrics are accuracy, loss, f1-score, precision, and recall. Experiments were made on Kaggle-based dataset and test results show that the CNN-based model is the most accurate model, with the highest accuracy of 96% and the lowest loss of 9.92%. Finally, the average performance percentage of the overall proposed model is as follows: accuracy is 91%, loss is 19.75%, precision is 89.4%, F1-score is 88.83%, and recall is 90%.*

Index Terms— *Deep Learning, Transfer Learning, Alzheimer's Disease, DCNN, CNN.*

I. INTRODUCTION AND RELATED WORK

In previous decades, the medical staff relied on several symptoms alongside medical instruments to detect the medical condition. Nowadays the medical advancement with the aid of technical improvement, especially in deep learning which is commonly used for medical research, assists in detection of what the diagnosis of the diseases exactly is. For example, Alzheimer's disease (AD) is prevalent among the elderly and it could be differentiated from other diseases (i.e., vascular or non-AD dementia) clearly and accurately [1], [2]. Earlier detection of AD plays an important role in the therapeutic plan and helps in accurate treatment [3], [4].

Memory loss and cognitive decline are two common signs of Alzheimer's disease, a progressive neurological illness. Current explanations for its origin center on neurofibrillary tau tangles and amyloid beta plaques. The majority of Alzheimer's disease treatments recognized by the US Food and Drug Administration focus on cognitive performance [5]. The principal cause of dementia, Alzheimer's disease, is swiftly rising to the top of the list of the most expensive, deadly, and burdensome illnesses of the twenty-first century. The understanding of the underlying pathology, the identification of numerous protective and causative genes, the discovery of new blood-based and imaging biomarkers, and the first cautious signals of the beneficial effects of disease-modifying therapies and lifestyle interventions have all advanced significantly [6].

Received 25/July/2023; Accepted 29/August/2023

ISSN (Print) 1811-9212 *ISSN (Online) 2617-3352*

DOI: https://doi.org/10.33103/uot.ijccce.24.1.5

As we mentioned in the abstract, Deep learning (DL) [7] is widely used in the medical field. DL actually is a branch of Machine learning (ML) which itself is a subclass of Artificial Intelligence (AI) and this explains the correlation among them. Deep learning system is raised from (ANN); therefore, it’s got its learning ability from data. While on the other side, the ML system learns from its experience, due to its self-learning based on a specific algorithm. DL based on several architectures that assist to solve different problems in different fields like Convolutional Neural Networks (CNN), Recurrent Neural Network (RNN), etc. [8], [9].

CNN is an important classification model of deep learning architectures that are specifically used for feature extraction and image classification. CNN is made up of three layers, one input layer, many hidden layers and one output layer, or in other words we can classify them into convolutional layers, pooling layers and finally fully connected layers. In the Convolutional layer; first of all, the input images are convolved with filters (kernels) in order to extract the feature map as an output of this layer. Then the pooling layer would create a small feature map by shrinking the fed convolved output layer. Furthermore, the activation function in a fully connected layer is used to make a decision whether or not to fire a neuron according to a specific threshold value [10]–[12].

State of the art of CNN [13] architecture comes with many different models like VGG (Visual Geometry Group), ResNet (Residual Network) [14], DenseNet (Densely Connected Convolutional Networks) [15], [16], GoogleNet, inception and others [17]. VGG model is used for image classification and it consists of successive convolutional layers activated by activation function (usually ReLU) followed by pooling layers and finally merged to fully connected layers. VGG-16 [18] and VGG-19 [19] are similar to each other but differ in the number of convoluted layers (postfix number of VGG refers to the number of layers) [20]. Transfer learning is common in DL, it contributes to the transferring of the knowledge that gained in the training step of one model to another new model in order to solve completely new problems in the same field. This will assist in the transferring of a smaller dataset and improving the overall performance of the model [21].

Some related works and researches to the current study field were summarized in the following paragraphs: The death of brain cells causes memory loss and this leads to Alzheimer disease.

In (2019), silvia basaia et. al. Built a deep learning algorithm that predicts the diagnoses of AD based on the CNN. The comparisons in this proposal between AD and HC datasets show the higher accuracy, sensitivity and specificity (higher than 98%). These results obtained in the AD vs. HC classification using both ADNI dataset and the combined ADNI + Milan data set. The CNN in this research is able to distinguish between patients of c-MCI and HC with higher performance of (accuracy, sensitivity and specificity up to 86% [22].

Santos Bringas, et. al. developed a software in (Sep. 2020) to identify the level of AD the patients that suffer from through deep learning with mobility data. They based CNN in the proposal system to improve the accuracy of the identification of AD stages in comparison to common supervised learning models. Mobility data based on this research will facilitate the monitoring of the patients as well as facilitate the proper action to be taken in suitable time and this will help in providing optimal treatment. The based model used in this proposal achieved accuracy up to 90.91%, F1- score reached to 0.897 [23].

Received 25/July/2023; Accepted 29/August/2023

DOI: https://doi.org/10.33103/uot.ijccce.24.1.5

Recently R. C. Suganthe1 et. al. in (March 2021) proposed a research based on the incorporation of both the Resnet and inception formulation for magnetic resonance imaging (MRI) images. In this model the researchers get results of accuracy, Precision, Recall, AUC, and F1 Score metrics to test a classification model. This model achieved 79.12% Accuracy, 70.64% Precision, 28.22% Recall, 81.9%AUC, 39.91% F1 Score [24].

In another study by Xia Owen Chen that was released in (May. 2022), they proposed a combination of deep convolutional NNs with iterated random forest (RF) architecture. The mean training AUC of the three runs was 85.1% with 95% confidence intervals for the results of AD vs. MCI (CIs). For AD vs NC, the mean training AUC for the three runs was 90.6% in 95% CIs. For the three stratifications of AD, MCI, and NC, the F1 and MCC scores of both sets were 59.9% and 59.5%, respectively, although the relative accuracy, recall, and specificity for each category in the training test (TS) were all close to 89% [25].

Zhiwei Qin et. al. in (Aug. 2022) developed a model that was effective at extracting the important aspect of photos. The researchers suggested a 3D HA-ResUNet model, which incorporates a hybrid attention mechanism. The hybrid attention mechanism is paired with the backbone classification model's skip connection, which benefits from both channel and spatial attention. Accuracy, sensitivity, precision, F1 score, and G-mean are all improved by 4.88%, 10.52%, 0.94%, 6.17%, and 5.60%, respectively, with the inclusion of the hybrid attention module [26].

In this research, a five DCNN models (CNN, VGG-16, VGG-19, ResNet50 and DenseNet121 with transfer learning) were evaluated and tested for the classification and prediction of AD. The differences in the prediction accuracy of different models were compared and analyzed comprehensively. The main improvement was as follows: 1) we analyzed and compared five DCNN models for classification and prediction with transfer learning for an important medical disease. 2) We raised the models' accuracy % and investigated the causes of the various levels of influence. 3) We offered CNN model suggestions for various computational contexts. The rest of this paper is organized as follows: section II illustrates the proposed methods and materials; the specifications and preprocessing of the selected dataset is described besides the theoretical concepts of the proposed model and its architecture. This section also includes the evaluation metrics theory and hardware specifications used to run the model and extracts the results. Section III presents the experimental results and discussions and finally section IV discusses the conclusion.

II. PROPOSED METHODS AND MATERIALS

A. Dataset Specifications and Preprocessing

Image preprocessing is a technique that turns unprocessed picture data into a dataset that can be used. As the photos were gathered from multiple sources, the dataset is not in standard form. The dataset cannot be used for data analysis as a consequence. As a result, the picture collection has to be preprocessed to provide valid data in order to do feature extraction and prediction. This dataset is collected from Kaggle data science company [27] (a data science and machine learning online community operated by Google LLC.) and consists of preprocessed MRI (Magnetic Resonance Imaging) Images. It consists of 6400 MRI images resized into 128 x 128 pixels. There are four classes as follows (see *Fig. 1 (A)*):

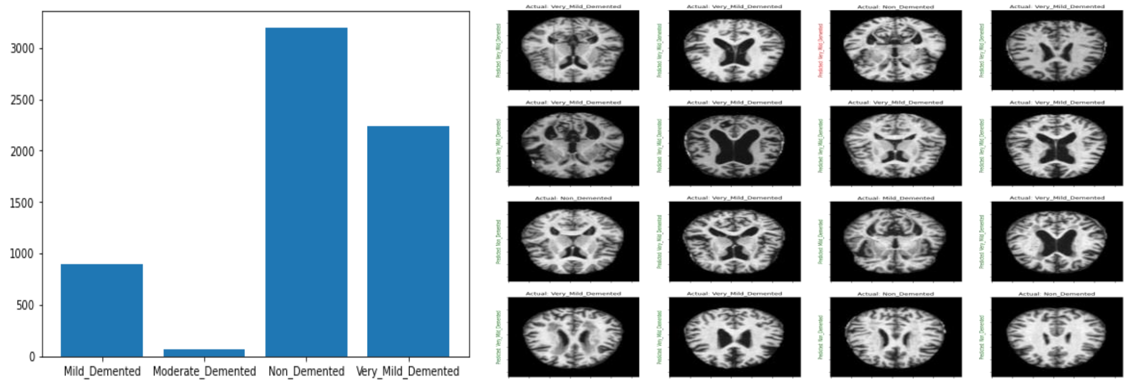
- Class - 1: Mild Demented (896 images)
- Class - 2: Moderate Demented (64 images)

Received 25/July/2023; Accepted 29/August/2023

DOI: https://doi.org/10.33103/uot.ijccce.24.1.5

- Class - 3: Non-Demented (3200 images)
- Class - 4: Very Mild Demented (2240 images)

A few samples of different classes of these images are shown in *Fig.1 (B)*. It is worth mentioning that the classification results discussed and explained in later sections are displayed in this figure to avoid figures' repetition.



A. CLASS DISTRIBUTION B. SAMPLE IMAGES OF DIFFERENT CLASSES WITH PREDICTIONS [27]

FIG. 1. AD DATASET SPECIFICATION.

B . Proposed DCNN Model

The input to our model is an RGB (Red Green Blue) image. In a wide view, the model is composed from three phases: First, the preprocessing phase where images are preprocessed through image rescaling strategy. Second, the convolutional layers that are mainly the DCNN algorithms which are: CNN, VGG-16, VGG-19, ResNet50 and DenseNet121 with eliminating the final layer in order to make use of the transfer learning strategy. The final phase is our proposed FC layer for the final image classification. First, transfer learning models and layers are explained briefly then the proposed layer that integrates the system will be explained after. The block diagram *Fig. 2* below shows the block diagram of the overall model.

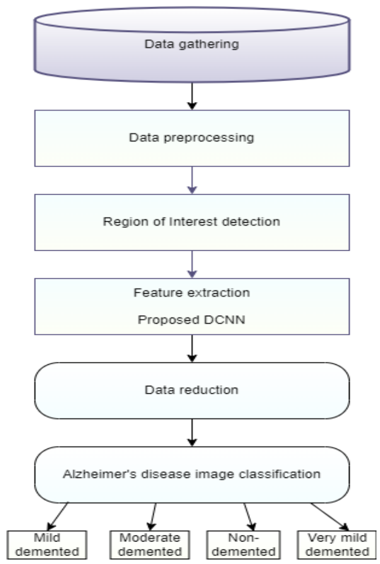


FIG. 2. PROPOSED MODEL'S BLOCK DIAGRAM.

Received 25/July/2023; Accepted 29/August/2023

DOI: https://doi.org/10.33103/uot.ijccce.24.1.5

1. CNN

The architecture of CNN [28], [29] in DL represents a class of feedforward NNs that learns directly from input data. It is a very effective approach for classifying image data. Mainly, the input image must be of size 28x28x1, then a series of convolution and pooling layers run to get the classified output image. The first layer is a convolution kernel with a valid padding (i.e., number of pixels added to an image while processing with CNN's kernel), the second layer is a max-pooling (a technique for generalizing features extracted by convolutional filters and helping in feature recognition). See reference [28] for more details on CNN architecture. The final layer is our suggested flatten Fully-Connected (FC) layer with a soft-max activation function. The flattening is used to convert 2D arrays into a single continuous linear vector. The dropout layer is a mask for eliminating some neurons towards the next layer and keeps others. Finally, the ReLU and soft-max activation functions are used through different layers see *Fig. 3.*

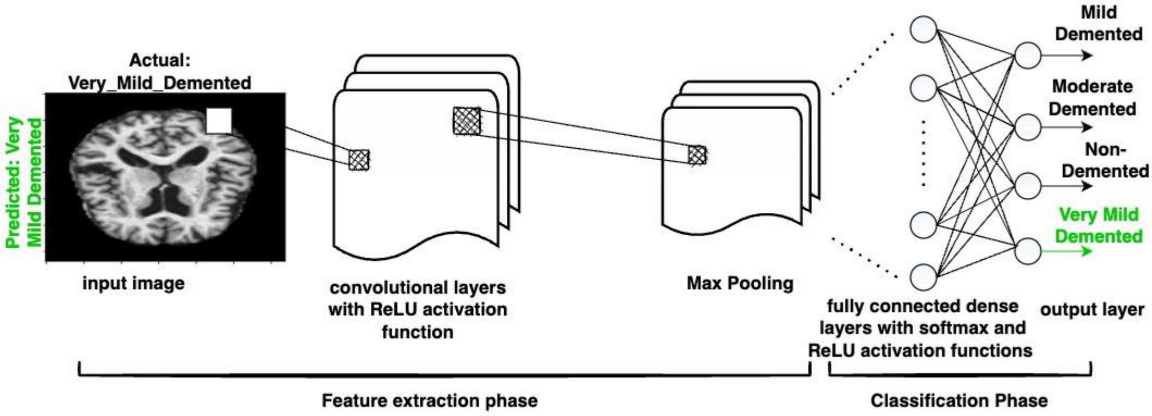


FIG. 3. CNN'S ARCHITECTURE [28].

2. VGG-16

VGG-16 [30], [31] is a CNN model used in object detection and image recognition but with increased depth by employing a stride 1 architecture with a very tiny (3x3) convolution filter and a stride 2 architecture with constant padding and maximum pool layer. for 16 weighted layers making it about 138 trainable parameters and this is the important feature about this type. It consists of 13 convolutional layers, 5 max pooling layers and 3 dense layers. The result is 21 layers but weight layers are on 16 where the convolution and max pool layers are arranged consistently. The input tensor size is 224x224 with 3 RGB channels. There are 64 filters in Conv-1 Layer, 128 filters in Conv-2, 256 filters in Conv-3, 512 filters in Conv-4, and 512 filters in Conv-5. A stack of convolutional layers is followed by 3 FC layers, the third of which conducts 1000-way ILSVRC classification and has 1000 channels. The first two FC layers have 4096 channels each. (one for each class), see references [32], [33]. Our recommended flatten Fully-Connected (FC) layer with a soft-max activation function is the top layer, see *Fig. 4.*

Received 25/July/2023; Accepted 29/August/2023

ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.24.1.5

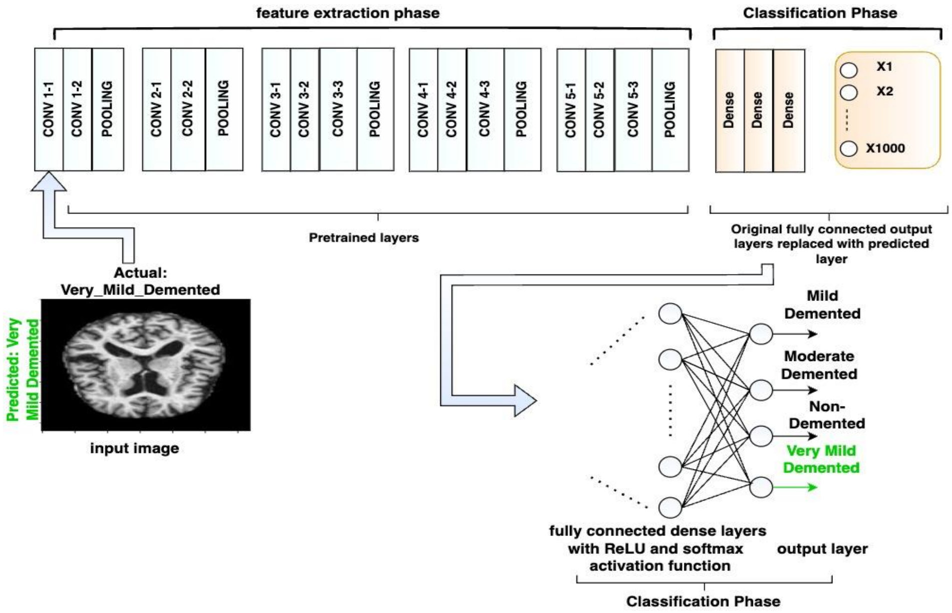


FIG. 4. VGG-16 ARCHITECTURE [30].

3. VGG-19

It is an extended version of VGG-16 by having 19 FC weighted with max pool and dropout layers as shown in *Fig. 4* below [32]. A fixed size of (224x224) RGB image was given as inputs with the same properties of VGG-16. The only preprocessing that is done is that they subtracted the mean RGB value from each pixel, computed over the whole training set. It uses kernels of size 3x3 with a stride size of 1 pixel to cover all images. spatial padding was used to preserve the spatial resolution of the image. Max pool was performed over 2x2 pixel windows with stride 2. Implemented 3 FC layers from which first two were of size 4096 and after that a layer with 1000 channels for 1000-way ILSVRC classification and the final layer is a soft-max function (refer to reference [34] for more details). *Fig. 5* illustrates the proposed model in terms of using this architecture with the suggested FC layer.

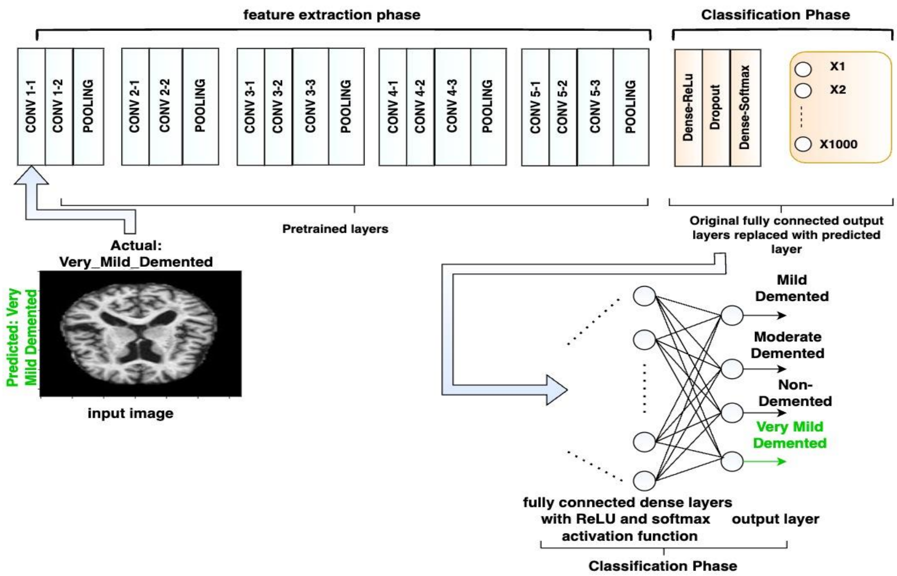


FIG. 5. VGG-19 ARCHITECTURE [32].

Received 25/July/2023; Accepted 29/August/2023

ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.24.1.5

4. ResNet50

A ResNet model version called ResNet50 includes 48 Convolution layers, 1 Max Pool layer, and 1 Average Pool layer. The total deep convolutional layers are 50 layers. The fundamental breakthrough with this model is that it allows to train extremely DNNs with 150+layers. There are two types of blocks in the architecture of this model which are identity blocks and convolutional blocks. There are also two kinds of mapping: identity mapping and residual mapping, refer to reference [35] for a detailed explanation about this model's architecture. *Fig. 6* illustrates the proposed model in terms of using this architecture with the suggested FC layer.

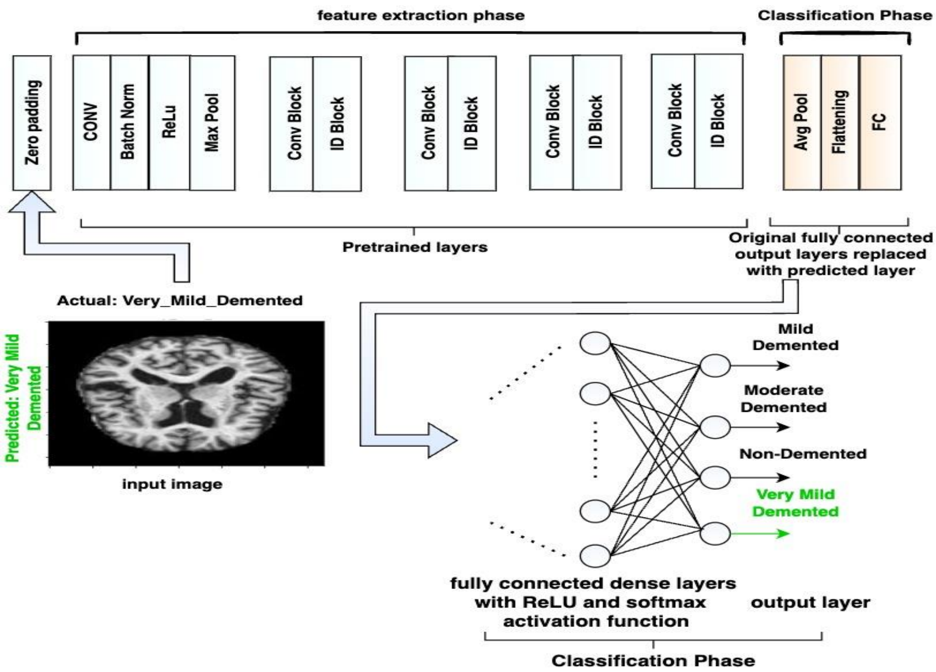


FIG. 6. RESNET50 ARCHITECTURE [35].

5. DenseNet121

This model simplifies the connectivity pattern among layers as compared to other architectures like ResNet and others by ensuring maximum information and gradient flow through connecting each layer directly to the others. Features map groupings cannot be achieved if their sizes are different no matter whether these groupings are an addition or a concatenation. Therefore, they are divided into Dense-Blocks, where feature maps' dimensions remain constant within each block, however, filters' numbers change among them. The in-between layers are called Transition Layers and are responsible for the down-sampling by applying a batch normalization, 1x1 convolution and 2x2 pooling layers. Refer to reference [36] for more details about the architecture of this model. *Fig. 7* illustrates the proposed model in terms of using this architecture with the suggested FC layer.

Received 25/July/2023; Accepted 29/August/2023

DOI: https://doi.org/10.33103/uot.ijccce.24.1.5

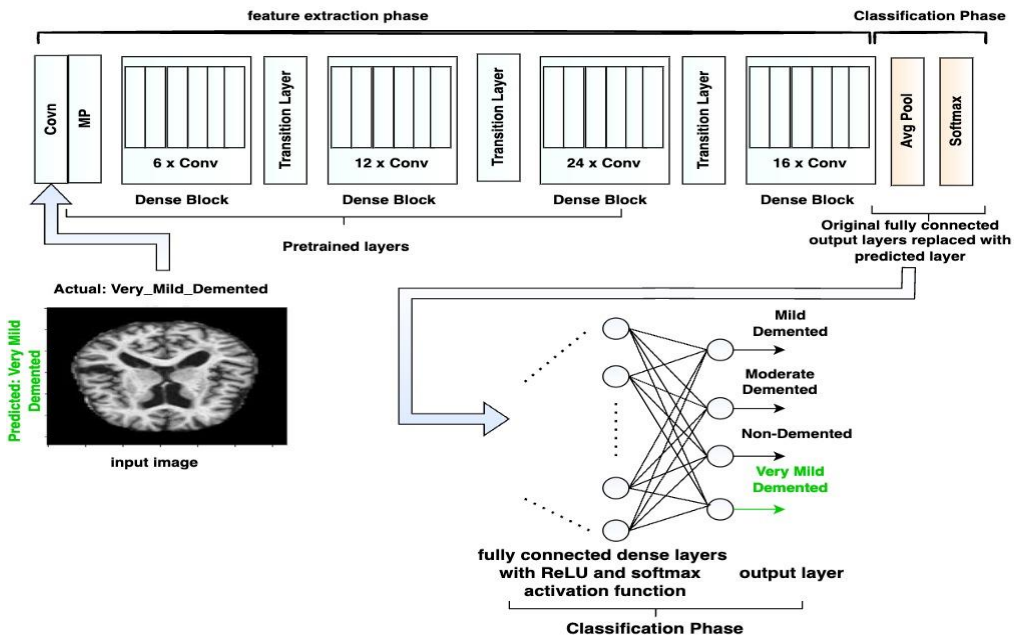


FIG. 7. DENSE121 ARCHITECTURE [36].

The final phase shown in *Fig. 2* through *Fig. 6* is our suggested flatten Fully-Connected (FC) phase with an Adam optimizer as shown in *Fig. 8*. According to the transfer learning, the top layer of the original architecture must be frozen and convolutional based layers must be set to untrainable. In this new phase there are four dense layers in which Each neuron in the preceding layer, three of which have ReLU activation functions, gets input from all other neurons in that layer. The last output layer has a softmax activation function. 4096 neurons make up the first layer, the second layer has 1072 neurons, the third layer has 512 neurons while the last layer has only four neurons. The dropout value was set to 0.25 to eliminate some neurons towards the next layer and keep others. The flattening is used to convert 2D arrays into a single continuous linear vector since a dense layer requires input in a one-dimensional shape.

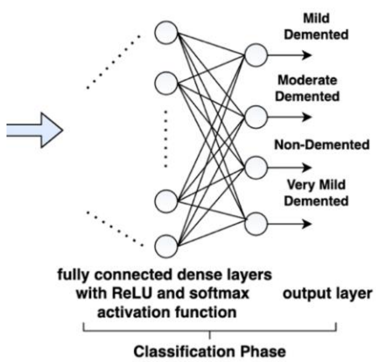


FIG. 8. SUGGESTED FC LAYER.

C. Evaluation Metrics

Numerous performance indicators were compiled throughout the examination of the algorithms that were deemed intelligent. The most popular measures from the relevant literature will be taken into account in the present estimation.

Received 25/July/2023; Accepted 29/August/2023
 ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.24.1.5

Recall (true positive rate) or responsiveness, which is connected to all positive members, pertains to the percentage of members who had a brain condition and were appropriately worried favorably. While working with unbalanced data, precision and recall are more reasonable to identify a model's faults. Precision refers to how many people who actually fall into this category have Parkinson's disease. Recall, on the other hand, shows the number of individuals who had Parkinson and are accurately predicted. Finally, f1-score represents the harmonic mean of the precision and recall and sums up the predictive performance of the model. Equations below represent the details of those parameters [37]:

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

$$F1score = \frac{2}{1/Recall+1/Precision} \quad (3)$$

where TN: true negative, TP: true positive, FP: false positive and FN: false negative. Model evaluation, on the other hand, is often based on the accuracy by describing the number of correct predictions overall predictions as shown in the following equation [38]:

$$\frac{\sum TP+\sum TN}{\sum TP+\sum TN+\sum FP+\sum FN} = \frac{No.\ of\ correct\ predictions}{No.of\ all\ predictions} = \frac{No.\ of\ correct\ predictions}{N} \quad (4)$$

Where N is the database size. In this research, the precision measures how many correct positive predictions (TP). Second, recall measures how many positive cases the classifier correctly predicted over all positive cases in the dataset. Finally, F1-score measures the average of precision and recall [39].

The proposed method was written in Python version 3.9 using Keras and other TensorFlow libraries with Jupyter notebook and scikit-learn tools run on Microsoft Visual Studio Code version 1.71.0 and Node.js: 16.14.2. In addition, all experiments and implementations were implemented on a Dell Precision laptop with Xeon processor, 64-GBps RAM and 1024-GBps Hard disk.

III. RESULTS AND DISCUSSION

In this section, DCNN models' performance were tested and evaluated using the Python 3.9 environment as mentioned in the previous section. Results were collected from the five previously explained architecture models which are CNN, VGG-16, VGG-19, ResNet50 and DenseNet121. The dataset was divided into two divisions, namely, 80% for training and 20% for validation. Concerning these models except CNN, the final classification phase was eliminated and the same fully connected output NN layers were used as part of the transfer learning philosophy in order for performance comparison to be justified. This phase consists of three dense layers of 4096, 1072 and 512 neurons with ReLU activation function, and a final dense layer of 4 neurons with softmax activation function optimizer and learning rate 0.002. *Fig. 9* on the left illustrates the classification report performance metrics (precision, recall and f1-score) of implemented DCNN models. Depending on these results CNN achieved the highest performance by recording 97.75% precision, 98% recall and 97.67% f1-score. DenseNet121, on the other hand, achieved the lowest performance by recording 77.75% precision, 83% recall and 77.50% f1-score.

Received 25/July/2023; Accepted 29/August/2023

ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

Iraqi Journal of Computers, Communications, Control & Systems Engineering (IJCCCE), Vol. 24, No. 1, March 2024

DOI: https://doi.org/10.33103/uot.ijccce.24.1.5

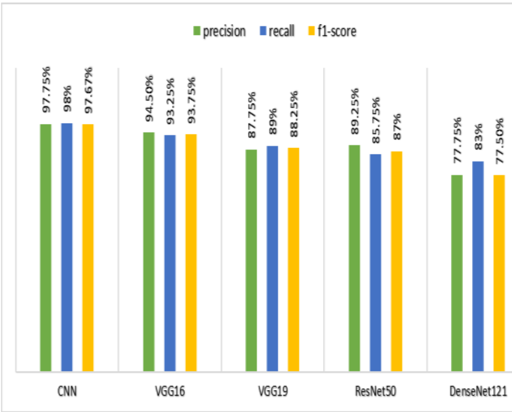


FIG. 9. CLASSIFICATION REPORT METRICS.

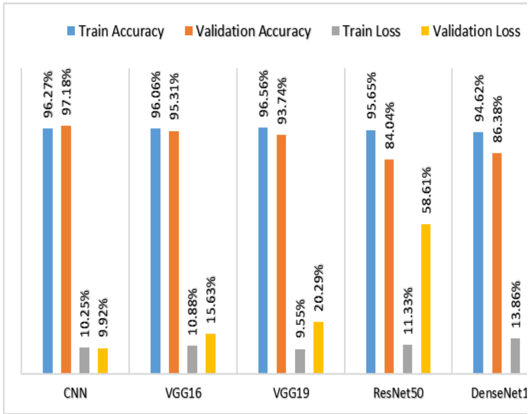


FIG. 10. ACCURACY AND LOSS PERCENTAGE.

Fig. 10 on the right, demonstrates the accuracy (test accuracy and validation accuracy) and loss (test loss and validation loss) on the proposed models. VGG-16 achieved the highest train accuracy of 96.56% while DenseNet121 had the lowest train accuracy of 94.62%. CNN recorded the highest test or validation accuracy of 97.18% while ResNet50 had the lowest validation accuracy of 84.04%. For train and validation loss metrics, VGG-19 had the lowest train loss of 9.55% while CNN had the lowest validation loss which is 9.92%. Finally, the highest train loss was 13.86% and highest validation loss was 37.16%, both achieved by DenseNet121.

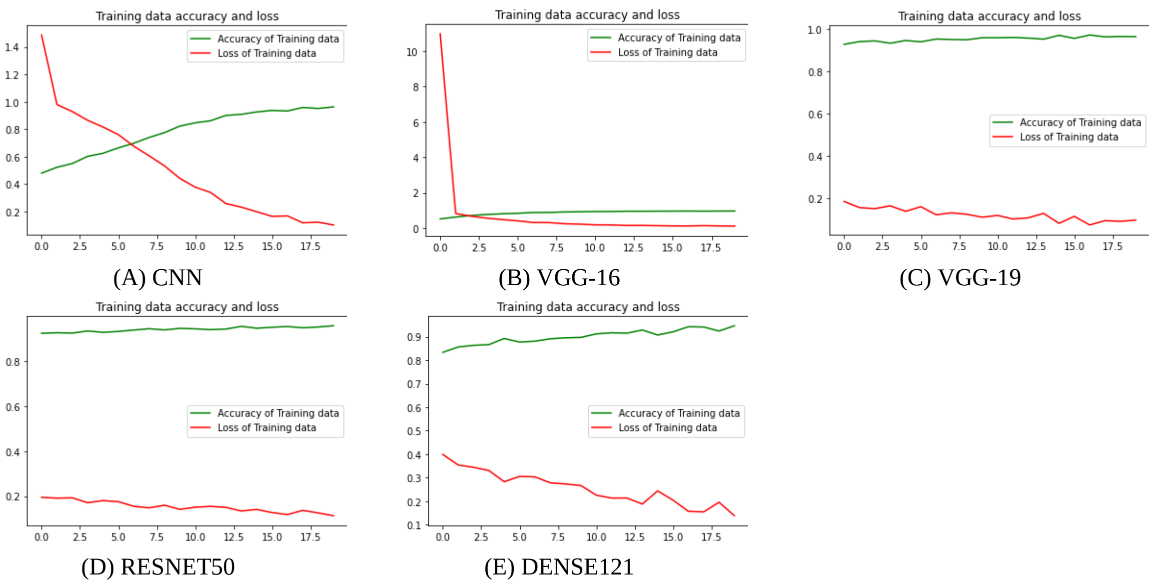


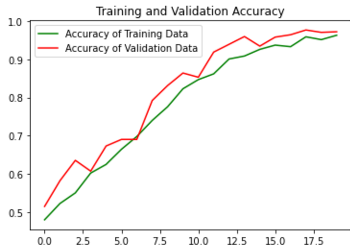
FIG. 11. TRAINING DATA ACCURACY AND LOSS.

Fig. 11 above, illustrates the performance curves in terms of training data accuracy and loss of the training data. From this figure, we can see the effect of different classification algorithms on reducing the loss of training data which is a very important factor to less than 0.1 with each epoch. Meanwhile, the accuracy curve moves in an opposite direction where it reaches ratios near 0.95 which represent important improvement as compared to other researches.

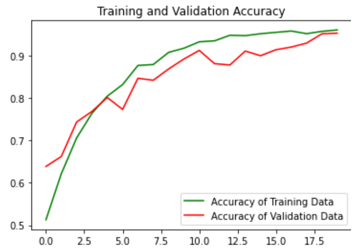
Received 25/July/2023; Accepted 29/August/2023

ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

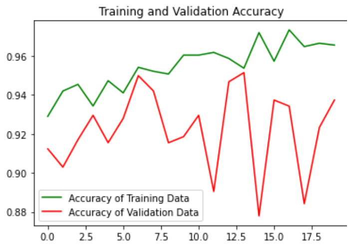
DOI: https://doi.org/10.33103/uot.ijccce.24.1.5



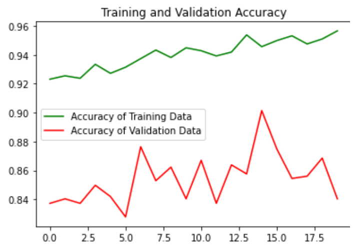
(A) CNN



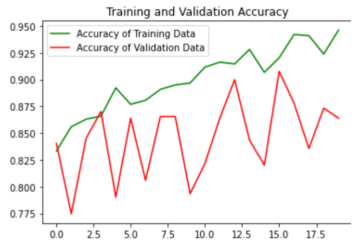
(B) VGG-16



(C) VGG-19



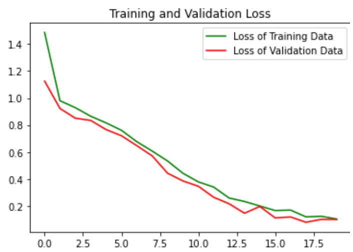
(D) RESNET50



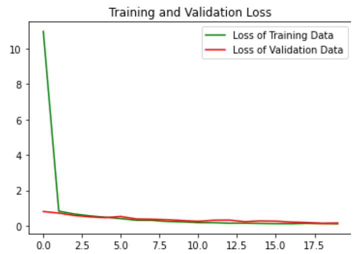
(E) DENSE121

FIG. 12. TRAINING AND VALIDATION ACCURACY.

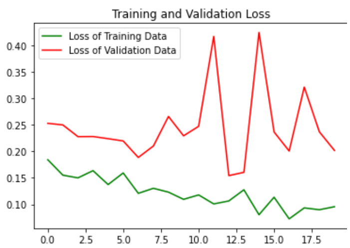
Fig. 12 above, interprets the difference between accuracy of training data and accuracy of validation or testing data. We can see that the stability of validation curves bounces with initial epochs and begins to stabilize for a higher number of epochs although it gets closer to the training accuracy curve. In this figure, we can see that the validation accuracy of CNN only is higher than its training accuracy because of the use of dropout philosophy, since the behavior is different between training and validating. The percentage of features when training is set to 25% because of using (Dropout (0.25)). In the validation phase, all features are used and scaled appropriately thus the model at validation time is more robust and results in higher validation accuracy.



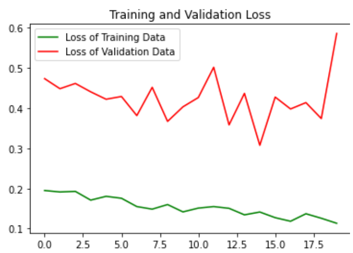
(A) CNN



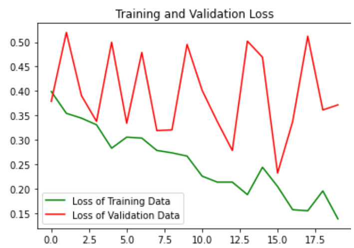
(B) VGG-16



(C) VGG-19



(D) RESNET50



(E) DENSE121

FIG. 13.TRAINING AND VALIDATION LOSS.

Fig. 13 above, shows the difference between training loss and validation loss. Results are different in each algorithm; we can see the high performance of CNN where the loss curve of the validation data has smaller values of the loss curve of the training data which

Received 25/July/2023; Accepted 29/August/2023
 ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.24.1.5

is the desired situation. VGG-16 has nearly the same curves while VGG-19 has a higher validation curve than training curve which restricts the performance of this algorithm although these loss values do not exceed 0.5. ResNet50 and Dense121, on the other hand, have similar characteristics of VGG-19 with validation loss values not exceeding 0.6 and training loss values not exceeding 0.4 in the first epochs and decreases to 0.1 in the last epochs.

As compared to other researches, our research provides an improved performance in terms of accuracy and loss metrics. Table I below presents a comparison with references [22], [23] where they achieved accuracy less than our research in 2% for the first one and 20.36% for the second reference. It is worth mentioning that the compared references didn't mention the loss metric of their models.

TABLE I. COMPARISON WITH OTHER RESEARCHES

Metric	**Proposed Model**	**[22]**	**[23]**
Accuracy	91%	70.64%	89%
Loss	19.75%	-	-
F1-score	88.83%	39.91%	59.9%

IV. CONCLUSIONS

Brain diseases like Alzheimer's represent an extreme difficulty in a human's life and surrounding relatives and it should be clearly diagnosed to avoid future reduplications or side effects. Nowadays, with the evaluation of artificial intelligence especially DL and different CNN-based architectures in image classifications and predictions, it is easier and more efficient to detect such kind of diseases and classify it in its different stages (i.e., mild demented, moderate demented, non-demented and very mild demented).

In this path, this work researches the viability of different DCNN (CNN, VGG-16, VGG-19, ResNet50 and DenseNet121) algorithms to appoint the most precise algorithm for predicting and classifying AD stages depending on sampled MRI images. The algorithms' performance evaluation has been calculated using accuracy, f1-score that totalizes precision and recall metrics. Those are the main metrics to compare and evaluate such types of algorithms besides uncovering the algorithm's authenticity and predictive capacity regarding AD.

During implementation, results were satisfying in terms of accuracy for such an image classification approach. Where, the average performance percentage of the overall proposed model is as follows: accuracy is 91%, loss is 19.75%, precision is 89.4%, F1-score is 88.83% and recall is 90%. Finally, it is worth mentioning that transfer learning extremely increases the accuracy of DCNN models.

CONFLICT OF INTEREST

Authors have no conflict of interest.

Received 25/July/2023; Accepted 29/August/2023

DOI: https://doi.org/10.33103/uot.ijccce.24.1.5

REFERENCES

[1] S. Gao and D. Lima, “A review of the application of deep learning in the detection of Alzheimer’s disease,” *International Journal of Cognitive Computing in Engineering*, vol. 3, pp. 1–8, Jun. 2022, doi: 10.1016/J.IJCCE.2021.12.002.

[2] K. Yang and E. A. Mohammed, “A Review of Artificial Intelligence Technologies for Early Prediction of Alzheimer’s Disease,” Dec. 2020, Accessed: Sep. 18, 2022. [Online]. Available: http://arxiv.org/abs/2101.01781

[3] S. Basheera and M. S. Sai Ram, “Convolution neural network–based Alzheimer’s disease classification using hybrid enhanced independent component analysis based segmented gray matter of T2 weighted magnetic resonance imaging with clinical valuation,” *Alzheimer’s & Dementia: Translational Research & Clinical Interventions*, vol. 5, no. 1, pp. 974–986, Jan. 2019, doi: 10.1016/j.trci.2019.10.001.

[4] H. Li, M. Habes, D. A. Wolk, and Y. Fan, “A deep learning model for early prediction of Alzheimer’s disease dementia based on hippocampal magnetic resonance imaging data,” *Alzheimer’s & Dementia*, vol. 15, no. 8, pp. 1059–1070, Aug. 2019, doi: 10.1016/j.jalz.2019.02.007.

[5] S. Asher and R. Priefer, “Alzheimer’s disease failed clinical trials,” *Life Sci*, vol. 306, p. 120861, Oct. 2022, doi: 10.1016/j.lfs.2022.120861.

[6] P. Scheltens *et al.*, “Alzheimer’s disease,” *Lancet*, vol. 397, no. 10284, p. 1577, Apr. 2021, doi: 10.1016/S0140-6736(20)32205-4.

[7] A. S. Dawood, “Machine Learning and Artificial Neural Network for Data Mining Classification and Prediction of Brain Diseases,” *International Journal of Reasoning-based Intelligent Systems*, vol. 1, no. 1, p. 1, 2023, doi: 10.1504/IJRIS.2023.10052940.

[8] N. Sharma, R. Sharma, and N. Jindal, “Machine Learning and Deep Learning Applications-A Vision,” *Global Transitions Proceedings*, vol. 2, no. 1, pp. 24–28, Jun. 2021, doi: 10.1016/j.gltp.2021.01.004.

[9] I. H. Sarker, “Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions,” vol. 2, p. 420, 2021, doi: 10.1007/s42979-021-00815-1.

[10] L. Alzubaidi *et al.*, “Review of deep learning: concepts, CNN architectures, challenges, applications, future directions,” *J Big Data*, vol. 8, p. 53, 2021, doi: 10.1186/s40537-021-00444-8.

[11] S. Albawi, T. A. Mohammed, and S. Al-Zawi, “Understanding of a convolutional neural network,” *Proceedings of 2017 International Conference on Engineering and Technology, ICET 2017*, vol. 2018-January, pp. 1–6, Mar. 2018, doi: 10.1109/ICENGTECHNOL.2017.8308186.

[12] S. Khan, H. Rahmani, S. A. A. Shah, and M. Bennamoun, “A Guide to Convolutional Neural Networks for Computer Vision,” *Synthesis Lectures on Computer Vision*, vol. 8, no. 1, pp. 1–207, Feb. 2018, doi: 10.2200/S00822ED1V01Y201712COV015.

[13] M. Hussain, J. J. Bird, and D. R. Faria, “A study on CNN transfer learning for image classification,” *Advances in Intelligent Systems and Computing*, vol. 840, pp. 191–202, 2019, doi: 10.1007/978-3-319-97982-3_16/COVER.

[14] S. Wang, X. Xia, L. Ye, and B. Yang, “Automatic detection and classification of steel surface defect using deep convolutional neural networks,” *Metals (Basel)*, vol. 11, no. 3, pp. 1–23, Mar. 2021, doi: 10.3390/MET11030388.

[15] P. Feng *et al.*, “Embranchment Cnn Based Local Climate Zone Classification Using Sar and Multispectral Remote Sensing Data,” *International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 6344–6347, Jul. 2019, doi: 10.1109/IGARSS.2019.8898703.

[16] Q. Ji, J. Huang, W. He, and Y. Sun, “Optimized deep convolutional neural networks for identification of macular diseases from optical coherence tomography images,” *Algorithms*, vol. 12, no. 3, 2019, doi: 10.3390/A12030051.

[17] D. Saravagi, S. Agrawal, M. Saravagi, J. M. Chatterjee, and M. Agarwal, “Diagnosis of Lumbar Spondylolisthesis Using Optimized Pretrained CNN Models,” *Comput Intell Neurosci*, vol. 2022, pp. 1–12, Apr. 2022, doi: 10.1155/2022/7459260.

[18] H. Perez and J. H. M. Tah, “Improving the accuracy of convolutional neural networks by identifying and removing outlier images in datasets using t-SNE,” *Mathematics*, vol. 8, no. 5, May 2020, doi: 10.3390/MATH8050662.

[19] J. Jaworek-Korjakowska, P. Kleczek, and M. Gorgon, “Melanoma thickness prediction based on convolutional neural network with VGG-19 model transfer learning,” *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, vol. 2019-June, pp. 2748–2756, Jun. 2019, doi: 10.1109/CVPRW.2019.00333.

[20] X. Du, Y. Sun, Y. Song, H. Sun, and L. Yang, “A Comparative Study of Different CNN Models and Transfer Learning Effect for Underwater Object Classification in Side-Scan Sonar Images,” *Remote Sens (Basel)*, vol. 15, no. 3, p. 593, Jan. 2023, doi: 10.3390/rs15030593.

Received 25/July/2023; Accepted 29/August/2023

DOI: https://doi.org/10.33103/uot.ijccce.24.1.5

[21] F. Zhuang *et al.*, "A Comprehensive Survey on Transfer Learning," *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, Jan. 2021, doi: 10.1109/JPROC.2020.3004555.

[22] S. Basaia *et al.*, "Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks," *Neuroimage Clin*, vol. 21, p. 101645, Jan. 2019, doi: 10.1016/J.NICL.2018.101645.

[23] S. Bringas, S. Salomón, R. Duque, C. Lage, and J. L. Montaña, "Alzheimer's Disease stage identification using deep learning models," *J Biomed Inform*, vol. 109, p. 103514, Sep. 2020, doi: 10.1016/J.JBI.2020.103514.

[24] R. C. Suganthe, M. Geetha, G. R. Sreekanth, K. Gowtham, S. Deepakkumar, and R. Elango, "Multiclass Classification of Alzheimer's Disease Using Hybrid Deep Convolutional Neural Network," *NVEO - NATURAL VOLATILES & ESSENTIAL OILS Journal | NVEO*, vol. 8, no. 5, pp. 145–153, Nov. 2021, Accessed: Sep. 19, 2022. [Online]. Available: https://www.nveo.org/index.php/journal/article/view/338

[25] X. Chen, M. Tang, A. Liu, and X. Wei, "Diagnostic accuracy study of automated stratification of Alzheimer's disease and mild cognitive impairment via deep learning based on MRI," *Ann Transl Med*, vol. 10, no. 14, pp. 765–765, Jul. 2022, doi: 10.21037/ATM-22-2961.

[26] Z. Qin, Z. Liu, Q. Guo, and P. Zhu, "3D convolutional neural networks with hybrid attention mechanism for early diagnosis of Alzheimer's disease," *Biomed Signal Process Control*, vol. 77, p. 103828, Aug. 2022, doi: 10.1016/J.BSPC.2022.103828.

[27] "Kaggle: Your Machine Learning and Data Science Community." https://www.kaggle.com/ (accessed Sep. 21, 2022).

[28] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects," *IEEE Trans Neural Netw Learn Syst*, pp. 1–21, 2021, doi: 10.1109/TNNLS.2021.3084827.

[29] C.-C. J. Kuo, "Understanding convolutional neural networks with a mathematical model," *J Vis Commun Image Represent*, vol. 41, pp. 406–413, Nov. 2016, doi: 10.1016/j.jvcir.2016.11.003.

[30] S. Ghosh, A. Chaki, and K. Santosh, "Improved U-Net architecture with VGG-16 for brain tumor segmentation," *Phys Eng Sci Med*, vol. 44, no. 3, pp. 703–712, Sep. 2021, doi: 10.1007/s13246-021-01019-w.

[31] T. Kaur and T. K. Gandhi, "Automated Brain Image Classification Based on VGG-16 and Transfer Learning," in *2019 International Conference on Information Technology (ICIT)*, IEEE, Dec. 2019, pp. 94–98. doi: 10.1109/ICIT48102.2019.00023.

[32] A. Bagaskara and M. Suryanegara, "Evaluation of VGG-16 and VGG-19 Deep Learning Architecture for Classifying Dementia People," in *2021 4th International Conference of Computer and Informatics Engineering (IC2IE)*, IEEE, Sep. 2021, pp. 1–4. doi: 10.1109/IC2IE53219.2021.9649132.

[33] Q. Guan *et al.*, "Deep convolutional neural network VGG-16 model for differential diagnosing of papillary thyroid carcinomas in cytological images: a pilot study," *J Cancer*, vol. 10, no. 20, pp. 4876–4882, 2019, doi: 10.7150/jca.28769.

[34] N. Habib, Md. M. Hasan, Md. M. Reza, and M. M. Rahman, "Ensemble of CheXNet and VGG-19 Feature Extractor with Random Forest Classifier for Pediatric Pneumonia Detection," *SN Comput Sci*, vol. 1, no. 6, p. 359, Nov. 2020, doi: 10.1007/s42979-020-00373-y.

[35] B. Li and D. Lima, "Facial expression recognition via ResNet-50," *International Journal of Cognitive Computing in Engineering*, vol. 2, pp. 57–64, Jun. 2021, doi: 10.1016/j.ijcce.2021.02.002.

[36] A. S. Vellaichamy, A. Swaminathan, C. Varun, and K. S, "MULTIPLE PLANT LEAF DISEASE CLASSIFICATION USING DENSENET-121 ARCHITECTURE," *INTERNATIONAL JOURNAL OF ELECTRICAL ENGINEERING AND TECHNOLOGY*, vol. 12, no. 5, May 2021, doi: 10.34218/IJEET.12.5.2021.005.

[37] V. Tong, H. A. Tran, S. Souihi, and A. Mellouk, "A novel QUIC traffic Classifier based on Convolutional Neural Networks," *2018 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, Sep. 2018, doi: 10.1109/GLOCOM.2018.8647128.

[38] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Inf Process Manag*, vol. 45, no. 4, pp. 427–437, Jul. 2009, doi: 10.1016/J.IPM.2009.03.002.

[39] T. Kanstren, "A Look at Precision, Recall, and F1-Score," *Towards Data Science*, Sep. 12, 2020. https://towardsdatascience.com/a-look-at-precision-recall-and-f1-score-36b5fd0dd3ec#:~:text=F1%2Dscore%20when%20Precision%3D0.8%20and%20Recall%20%3D%200.01%20to%201.0&text=Here%20precision%20is%20fixed%20at,varies%20from%200.0%20to%201.0. (accessed Jul. 25, 2022).

Received 25/July/2023; Accepted 29/August/2023