International Journal of Computational and Experimental Science and ENgineering
(IJCESEN)
Vol. 11-No.1 (2025) pp. 97-108
http://www.ijcesen.com

ISSN: 2149-9144

Research Article

An efficient hybrid Deep Learning-Machine Learning method for diagnosing neurodegenerative disorders

Johnsymol Joy[1]*, Mercy Paul Selvan[2]

[1]Research Scholar, Reg No. 2021191104, Dept. of Computer Science and Engineering,Sathyabama Institute of Science and Technology, Jeppiaar Nagar, Chennai – 600 119, India
* **Corresponding Author Email:** johnsymoljoy07@gmail.com-**ORCID:** 0009-0003-4402-5760

[2]Associate Professor, Dept. of Computer Science and Engineering,Sathyabama Institute of Science and Technology, Jeppiaar Nagar, Chennai – 600119, India
Email:mercypaulselvan.cse@sathyabama.ac.in-**ORCID:**0000-0001-8950-849X

Article Info:

DOI: 10.22399/ijcesen.701
Received : 28 September 2024
Accepted : 02 December 2024

Keywords :

Machine Learning,
Deep Learning,
Multi-class classification,
CNN,
ResNet,
Transfer Learning.

Abstract:

A neurodegenerative illness known as Alzheimer's causes the loss of brain cells and the progressive atrophy of brain tissue. It badly affects a person's normal life. However, if we are able to detect it early and treat it, most patients will be able to recover to some degree and lead a normal life with some dependence. Continuous clinical assessment is needed for diagnosing this type of disorder. Medical diagnosis today extensively relies on deep learning approaches. However, medical image data analysis has a lot of constraints. One of the major constraints faced during medical image analysis is data scarcity and data imbalance. In light of these concerns, the current study sets out to create a hybrid deep learning model that can effectively categorise various Alzheimer's disease variants using magnetic resonance imaging (MRI) data.
For solving data imbalance, first, we blur and sharpen all the images, and finally, we pass all these images along with the original images through a predefined CNN (Convolutional Neural Network) model that was trained using mnist weight for extracting features, then pass these features to an extra-tree classifier for feature reduction, and finally input these reduced features to a customised deep learning model. This work used different pre-trained models for extracting features for our customised DNN (Deep Neural Network) and compared those models with the cutting-edge model chosen as the base model. The results state that our proposed model, which was pre-trained using ResNet with a dropout concept, got the highest values of training accuracy (98.20) and validation accuracy (92.61). This model also effectively addresses the problem of overfitting.

1. Introduction

Alzheimer's disease, a neurodegenerative illness, leads to the death of brain cells and the shrinking of brain tissue. Dementia is an unavoidable consequence that occurs when there is a gradual decline in cognitive, behavioural, and social abilities, leading to a reduced capacity for independent functioning. Alzheimer's disease has affected more than six million Americans. By 2050, this figure is expected to exceed 13 million. Alzheimer's disease or related dementia accounts for one out of every three elderly deaths, surpassing the combined fatalities caused by prostate and breast cancer [1]. Alzheimer's disease can be detected by analyzing different biomarkers [2]. Medical professionals and researchers can observe various indicators on different types of brain scans that aid in diagnosing Alzheimer's disease or similar dementia. These scans are also utilized to detect other potential issues like tumors or strokes, which can assist in the diagnostic process. Three different methods of brain imaging are used to look for dementia: Magnetic Resonance Imaging (MRI), Computed Tomography (CT), and Positron Emission Tomography (PET) scans.
The majority of individuals with Alzheimer's disease eventually encounter consistent symptoms, such as memory loss, disorientation, challenges in performing previously familiar tasks, and difficulties in making judgments. Despite the fact that its consequences are comparable, the condition is divided into two basic forms, Alzheimer's disease

with early and late-onset [3]. The illness is typically first identified in people in their 40s or 50s. Early-onset Alzheimer's is exceedingly rare, affecting only up to 5% of all cases. Down syndrome sufferers are more likely to get it. Late-onset Alzheimer's affects people who are 65 and older and is the most commonly dominant type. The type of gene that may cause it hasn't been pinpointed by researchers yet.

Currently, there is no established remedy for AD (Alzheimer's disease). Despite this, there are treatments for AD that can lessen symptoms or halt their development. So Catching AD in its prodromal stage is critical. Today, medical diagnostics commonly employ deep learning techniques for detecting this type of disorders. Computer-aided systems are used for precise and early diagnosis of AD to lessen the high care costs associated with AD patients [4]. The prime motto of this research is to make a quick and reliable method for diagnosing Alzheimer's disease before any symptoms manifest. MCI (Mild Cognitive Impairment) is one of the phases of AD that manifests as an illness in the prodromal stage. AD has other stages as well. MCI is a stage of cognitive decline in adults who remain capable of doing the majority of their everyday chores on their own but are losing their memory or other cognitive abilities. The main stages are Cognitive Normal (CN)- this include individuals with safe category, Mild Cognitive Impairment (MCI)-includes individuals in early stage , Late Mild Cognitive Impairment (LMCI)-includes severely affected people, and Alzheimer's Disease-this includes completely affected people [5]. One of the neuroimaging techniques that help researchers to categorize this type of disease is MRI, which is commonly employed for scanning brain tissue. When it comes to illnesses that need large volumes of input data to diagnose and categorise, deep learning is clearly the method to use. When it comes to diagnosing AD, deep learning algorithms outperform more traditional methods since they can learn picture characteristics on their own.

Machine learning is the practice of training a system to understand and use various datasets and algorithms to make predictions or categorise new data. For machine learning, there are three main types of algorithms: supervised, unsupervised, and reinforcement learning.

Classification comes under supervised learning. Classifications are binary classification and multi-class classification. In contrast to multi-class classification, which divides data into more than two classes, binary classification divides data into two classes [6]. Deep learning techniques implement a neural network with three or more layers for solving different critical problems and are categorized as a subset of Machine Learning. Deep neural networks handle data in complicated manners by using advanced mathematical modelling. Neural networks aim to emulate the functioning of the human brain. This paper proposes a novel hybrid DL-ML (Deep Learning –Machine Learning) models for classifying Alzheimer's disease. A reduced feature set extracted using different feature extraction and feature reduction techniques applied on MRI images, are supplied to a customized Deep Neural Network for classifying AD stages

For the literature review, several papers that use different DL techniques were revised. A. Mehmood et al. [7] used layer-wise transfer learning and brain tissue segmentation from images to diagnose AD early. VGG (Visual Geometry Group) is utilized for the layer-wise transfer learning process. Each subject had their tissue segmented in order to remove the grey matter (GM) tissue. The suggested method was tested on pre-processed data to determine its validity. It effectively differentiate EMCI and LMCI patients with an accuracy value 83.72%.

Raw structural MRI scans were first segmented into different anatomical structures, or Regions of Interest (ROIs), using FreeSurfer, as suggested by Da Ma et al. [8]. Afterwards, super-pixel patches of varying sizes were created by further subdividing each ROI. We recovered volume and cortical thickness as multi-scale, multi-type features at each level of the patch. In order to perform differential diagnosis and distinguish NC patients from those with AD and FTD (Frontotemporal Dementia), a Generative Adversarial Network with a variety of scale and type features was developed. The accuracy of this system was 88.28%.

The most widely used anatomical MRI of the brain, T1WI, was used by Dan Pan et al. [9] to test their unique deep learning approach, which integrated CNN with Ensemble Learning. The two goals of their study were to classify AD as MCIc vs. HC (Healthy Controls) or MCIc (MCI conversion) vs. MCInc (MCI non-conversion) and identify the complicated change patterns connected to AD. This technique employs a data-driven, CNN-based adaptive representation learning method to find distinguishable features from the MR images. The two-stage ensemble approach used by the suggested method enhanced generalisation and resilience. The model attained an average accuracy of 84%.

For the purpose of analysing single-type MRI data, Yan Wang et al. [10] developed a CNN-based multimodal MRI analytical approach. Before the CNN could be trained, the human brain network connectivity matrix was extracted from the multimodal MRI data. After that, a special CNN framework was used to analyse the network matrix and categorise people into three groups: those with Alzheimer's disease (AD), those with amnestic mild

cognitive impairment (aMCI), and healthy controls. They improved classification accuracy by combining CNN convolution kernel with multimodal MRI data. A precision of 92.06% was achieved by this model.
The authors of the study [11] use a feature extractor for their classification job that is pre-trained on the ImageNet dataset: the VGG-16 CNN architecture. They include the mathematical model PFSECTL into it. Here, the Alzheimer's Disease Neuroimaging Initiative (ADNI) database serves as the basis for the investigations. According to the results, the suggested strategy achieves a three-way classification accuracy of 95.73% on the validation set.
S. El-Sappagh et al. [12] presented a CNN-BilSTM ensemble multimodal multitask deep learning model. The main goal is to learn AD multiclass classification and four cognitive score regression together. For the purpose of capturing time series variations and extracting longitudinal features from each modality, the BiLSTM subnetworks are utilised. To the contrary, local characteristics from specific time series are extracted using the CNN subnetworks. It received scores for accuracy (92.62 ±2.41), precision (94.02± 3.26), recall (98.42± 1.38), and f1-score (92.56 ±2.38).

2. Material and Methods

2.1 Outline

The general structural architecture of the proposed work is illustrated in Figure 1. For handling data imbalance, we first blur and sharpen the whole image and pass the original image plus the modified images to different pre-trained CNN models like ResNet, DenseNet, and XceptionNet for feature extraction. Then the extracted features are passed to an Extra Tree classifier, and finally the reduced feature set to our customized DNN model. A customized Deep Neural Networks (DNN) is initially created. This model acts as the classifier

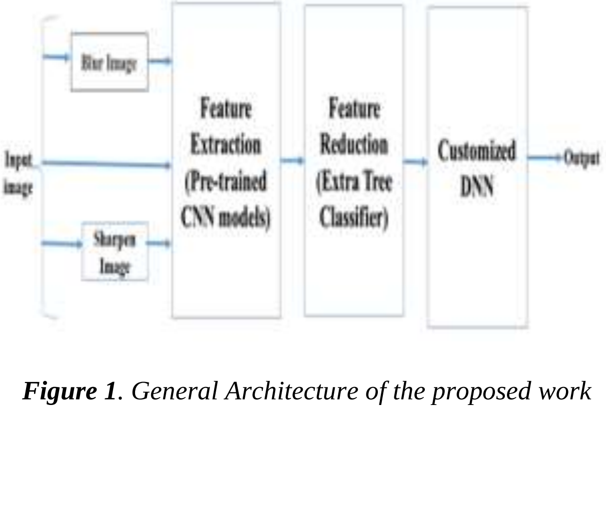


Figure 1. *General Architecture of the proposed work*

module for our proposed models. The structure of Customized DNN model is illustrated in Figure 2. This DNN model is created by adding layers of neural networks in sequential mode. The structure consists of many thick layers, one flattening layer, and one average pooling layer. We first check the performance of this DNN model and then further developed the proposed models by adding data augmentation, feature extraction, and feature reduction modules.

Layer (type)	Output Shape	Param #
input_7 (InputLayer)	[(None, 27419)]	0
dense (Dense)	(None, 10000)	274200000
dense_1 (Dense)	(None, 5000)	50005000
dense_2 (Dense)	(None, 2000)	10002000
dense_3 (Dense)	(None, 500)	1000500
dense_4 (Dense)	(None, 300)	150300
dense_5 (Dense)	(None, 200)	60200
dense_6 (Dense)	(None, 120)	24120
dense_7 (Dense)	(None, 30)	3630
dense_8 (Dense)	(None, 30)	930
dense_9 (Dense)	(None, 4)	124

Figure 2. *Structure of Customized DNN*

2.2 Data Augmentation

The term "data augmentation" describes a method of artificially increasing the amount of data by adding new data points to an existing dataset. To increase the amount of the dataset, we may either make small adjustments to the data or apply deep learning or machine learning models to generate additional data points in the raw data's latent space. It is used to increase model accuracy, decrease operational costs associated with labelling and cleaning the original dataset, prevent models from overfitting, address issues with the initial training set, and more [13]. This research uses a multiclass classification image Alzheimer's dataset from Kaggle [14] to analyse the effectiveness of the aforementioned proposed work. Classes are uneven, which is the problem with this dataset. The suggested approach employs data augmentation methods to rectify the data imbalance. To address the aforementioned issue, blurring and sharpening techniques are applied to the original pictures in this dataset. This work uses convolution masks 3X3 and 5X5 with a strength of less than 50 to sharpen the images. Additionally, different methods such as Gaussian blur with different kernel sizes, stack blur with a different radius, radial blur with a different size, and finally motion blur with a different degree are also used.

2.3 Feature Extraction

The term "feature extraction" describes a method for extracting numerical characteristics from unstructured data without modifying the source data. When opposed to using machine learning on the raw data alone, feature extraction produces far more precise results. The term "Transfer Learning" [15] refers to the process of applying a previously learned model to a new situation. By applying what it has learned from one job to another, a computer may improve its generalizability. This process is called transfer learning. To illustrate, let's say you want to apply the network's learnt weights vector from "Task A" to a completely new "Task B." The proposed model uses different predefined CNN models like ResNet, DenseNet, and Xception pre-trained with Mnist weight as feature extractors. Although transfer learning has other advantages as well, its key advantages are its reduced training time, improved neural network performance by using less data.

ResNet (Residual Network):
Deep residual networks, the well-known ResNet-50 model, are convolutional neural networks (CNNs) with a minimum of 50 layers. In an Artificial Neural Network (ANN), also known as a ResNet, the residual blocks are piled on top of one another to form a network [16]. Each layer of ResNet consists of several blocks. This is due to the fact that ResNets often add more operations to each block as they go deeper, while the overall number of layers remains constant at 4. A convolution, a batch normalisation, and a ReLU activation of an input are all referred to as operations

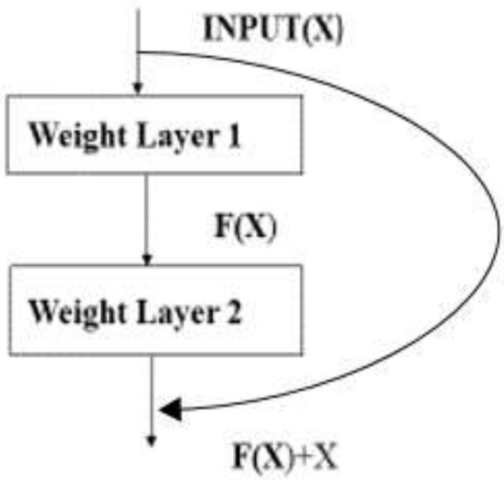


Figure 3. *Building block of a ResNet.*

in this context, with the exception of the final operation in the block, which lacks the ReLU. The ResNet [17] design introduces a simple notion of adding an intermediate input to the output of many convolutional blocks. Figure 3 shows a typical residual block, and Figure 4 shows the ResNet-50 design. An interconnected layer is the last component of this design, which also includes input pre-processing, Cfg [0] blocks (Configuration [0]), Cfg [1] blocks, Cfg [2] blocks, and Cfg [3] blocks. In ResNet, "CFG" typically refers to the network configuration or architecture. All neurons do not need to activate at every epoch for the ResNet design. This significantly decreases training time and improves accuracy. It employs a highly innovative tactic that greatly increases the efficacy of model training by focusing on learning new characteristics rather than trying to learn previously learned features repeatedly. Each residual block has a skip connection that joins its input and output, allowing data to pass from one layer to the next without any intermediate layers. Because of this, we can finally put an end to the vanishing gradient issue and start building very deep neural networks that are top-notch at many computer vision tasks like picture classification and object recognition.

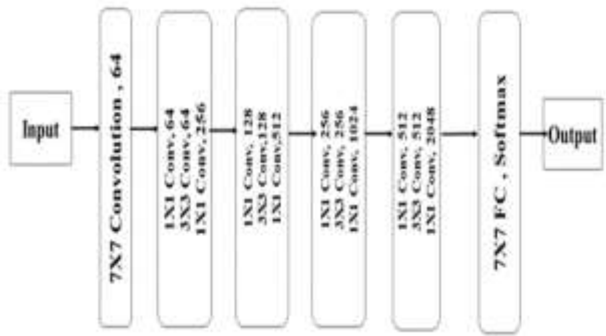


Figure 4. *Architecture diagram.-ResNet-50*

The architecture of the first suggested model, which is a Hybrid DL-ML (Deep Learning-Machine Learning) model with ResNet as a feature extractor and an extra-tree classifier as a feature reducer, is illustrated in Figure 5. Initially, the original images are blurred and sharpened and then along with the original images those converted images are passed to the feature extraction module- which uses pre-trained ResNet. Then pass these extracted features to an extra-tree classifier for feature reduction. And finally the reduced feature set is sent to a customized DNN module.

DenseNet (Dense convolutional Networks):
A specific type of convolutional neural network known as a densely connected convolutional

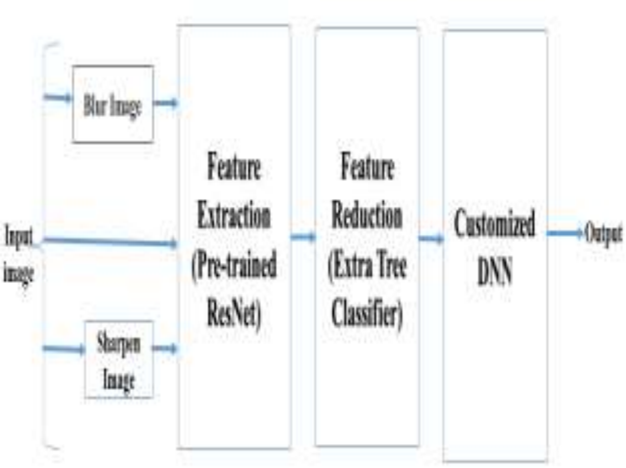


Figure 5. *Architecture of the proposed Model 1-Hybrid DL-ML model with ResNet as feature extractor and extra-tree classifier as feature reducer.*

network, or DenseNet [18], uses dense connections between layers by directly connecting each one using dense blocks. Compared to conventional deep CNNs, DenseNet has one key advantage: the data won't be lost or washed out before it reaches the network's end. In conventional convolutional networks, each layer is connected to the one below it by n connections. Since every layer in DenseNet feeds forward connections to every other layer, the total number of connections is n(n+1)/2. Each layer takes as inputs the feature maps of all the layers below it, and every layer above that uses its own feature maps to train its model.

The architecture of a typical DenseNet model is presented in Figure 6. We have the dense blocks at the highest level. Batch normalisation, a ReLU layer, and a convolution layer make up each layer individually. A collection of pooling and convolution layers links each of these Dense Blocks together [18]. Each layer can also be altered by selecting the number of feature maps and the filter size that best suits our needs. Compared to other architectures, this model uses fewer parameters and has better information flow, which means it has implicit "Deep supervision" since the filters are smaller and the input is preserved, allowing gradients and information to easily flow through the network. Figure 7 shows the architecture of the second

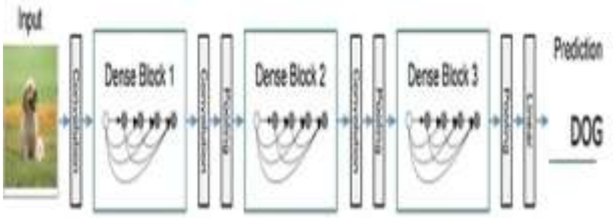


Figure 6. *Architecture of a DenseNet model*

proposed model, a hybrid DL-ML model that uses Extra-tree classifiers as feature reducers and DenseNet as a feature extractor. In this model initially all the images are blurred and sharpened, and then these modified images, along with our original images, are passed to a module that extracts features using a pre-trained DenseNet. This extracted feature set is then passed to an extra-tree classifier for feature reduction, and the reduced feature set is finally sent to our customised DNN module.

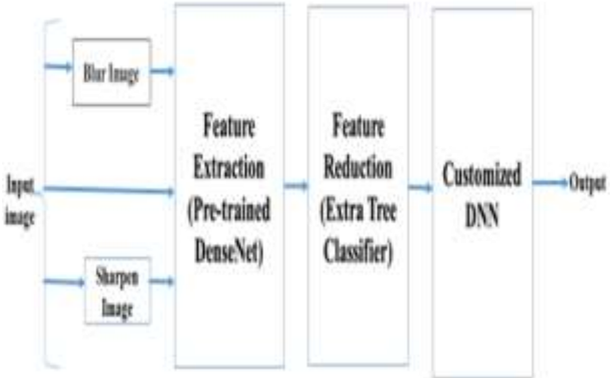


Figure 7. *Architecture of the proposed Model 2-Hybrid DL-ML model with DenseNet as feature extractor and extra-tree classifier as feature reducer.*

Xception:

Extreme Inception is the abbreviation for "Xception" [19]. The feature extraction basis of the Xception architecture is a network of 36 convolutional layers. Attributed to [20] [21], Figure 8 shows the Xception schematic. Starting with the input flow, the data is then sent through eight repetitions of the middle flow, and lastly, it is passed via the exit flow. This design uses batch normalisation after every Convolution and Separable Convolution layer. Typically, the entrance flow is responsible for collecting basic features, the middle flow for learning more complex ones, and the exit flow for applying the acquired features to a classification or regression problem. When it comes to picture classification and object identification, among other computer vision applications, the Xception network's architectural architecture allows it to provide state-of-the-art performance. The architecture of Model 3, which makes use of a pre-trained Xception model, is depicted in Figure 9. The process begins with sharpening and blurring the photos, followed by sending them and the originals to a module that use a pre-trained Xception to extract the features. The extracted feature set is then delivered to an extra-tree classifier for feature reduction before being sent to a customised DNN module.

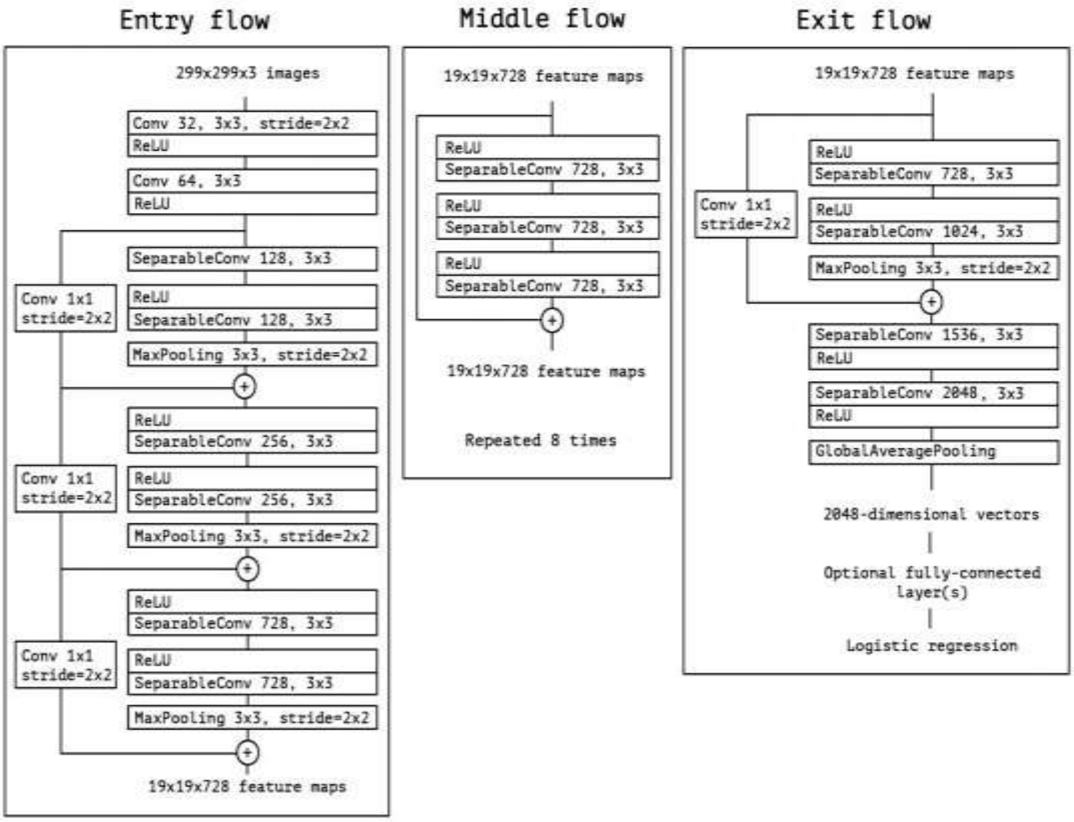


Figure 8. *Xception architecture*

***Table 1**. Performance measures for different proposed models.*

Models		**Accuracy values**		**Precision**	**Recall**	**F1 score**
		Training	**Validation**			
Model 1(Hybrid DL-ML model with ResNet as feature extractor)	Without dropout	96.30	91.76	92	92	92
	With dropout	98.20	92.61	93	93	93
Model 2 (Hybrid DL-ML model with DenseNet as feature extractor)	Without dropout	97.76	86.51	88	87	87
	With dropout	97.74	89.77	89	89	90
Model 3(Hybrid DL-ML model with Xception as feature extractor)	Without dropout	95.98	77.96	78	78	78
	With dropout	92	75.68	77	76	76

2.4 Feature Reduction module

Along with the feature extraction module, the proposed work also uses feature reduction module. The goal of feature reduction, also known as dimensionality reduction, is to decrease the number of features used in a resource-intensive calculation while preserving all relevant information. Because there are fewer variables in a simpler operation, a computer can do it more quickly and with fewer features. The objective of feature reduction is to make the system more efficient by decreasing the amount of features it needs to process. The computer can perform more because computation takes less time and requires less storage space. Feature reduction in deep learning improves the accuracy of the model. This work uses a feature selection technique called the extra-tree classifier.

Extra-tree classifier

In this work, we used an augmented dataset. The dataset used is a multi-class dataset, and each class has around 8000 images, so in total, the system has to process 32000 images. Our final module, Customised DNN, is not able to handle this wide set of features extracted from different pre-trained CNN networks. So we have to reduce our extracted feature set by keeping all relevant features. For that, we use a feature selection or reduction method called an extra-tree classifier. So in this proposed study, choices about the significance of the features are made using tree-based supervised models. With the training dataset used to randomly seed several tree models, the Extremely Random Tree Classifier [22], also known as the extra tree classifier, combines the features via majority voting. It differs from conventional decision tree approaches in that it uses random split points to partition the nodes and fits the decision trees to the whole dataset without bootstrapping replicas. The number of levels of decision trees is directly proportional to the entropy, or degree of unpredictability, in the sub-nodes. The component tree models use all of the dataset's variables to partition the nodes, with the best split producing the most similar offspring. This lowers the model's variance and makes overfitting less likely.

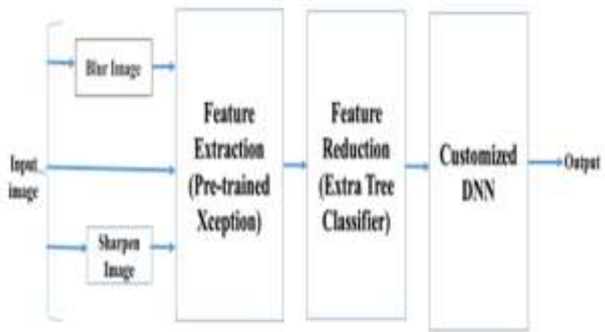


Figure 9. *Architecture of the proposed Model 3-Hybrid DL-ML model with Xception as feature extractor and extra-tree classifier as feature reducer.*

3. Results and Discussions

3.1 Datasets and Methodology used

To evaluate how well the aforementioned proposed works performed, this paper uses a multiclass classification image Alzheimer's dataset from Kaggle [14]. The data is manually gathered from multiple websites, and each label is confirmed. Images from MRIs make up this dataset. Images are categorized into four groups: mild, moderate, non-demented, and very mildly demented in both the training and testing sets. The original dataset comprises 896 photos for Mild Demented, 64 for Moderate Demented, 3200 for Non Demented and 2240 for Very Mild Demented. This dataset has an imbalance, which is a problem. We start by adjusting the original images' blur and sharpness levels to fix the data imbalance. There are 8700 pictures for mild dementia, 7200 for moderate dementia, 9600 for non-dementia, and 8900 for very mild dementia in the updated dataset.

Our models run on a system with the following specifications: Windows 10 operating system, NVIDIA Quadro GV100 graphics card, 512GB SSD hard disc for operating system, and 6TB SATA hard disc for data storage. Intel Xeon Gold 6134 CPU, 24.75M Cache, 3.20 GHz, 3.19 GHz (2 processors). The workstation has been redesigned by the NVIDIA® Quadro® GV100 to fulfil the needs of AI-enhanced design and visualisation operations. The recommended model was implemented using the Keras library, which has TensorFlow as a backend.

3.2 Performance Metrics

Visual representations of the proposed models' performance are provided via Loss and Accuracy graphs, in addition to a number of performance assessment metrics like as Accuracy, Precision, Recall, and F1-score [23]. In this case, the x-axis represents the total number of epochs, while the y-axis displays the accuracy and loss, respectively. As the number of epochs rises, the validation and training accuracy curves both show an increasing trend. The main reason for the discrepancy between training and validation accuracy is the unbiased data, so to address this, we also include a notion called dropout. The term "dropout" in deep learning refers to the practice of randomly ignoring some nodes in a layer during training. It is a regularisation strategy that makes sure no units are dependent on one another in order to prevent overfitting.

Accuracy

Accuracy is defined as the proportion of cases out of all examples that are correctly classified. In general, it measures how well the model predicts the future. Here is how it works:

$$Accuracy\ value = (TP + TN) / (TP + TN + FP + FN) \quad (1)$$

Where TP stands for the total number of positively identified events, TN for the total number of negatively identified events, FP for the total number of incorrectly positively identified events, and FN for the total number of incorrectly negatively identified events.

Precision

This word describes the proportion of favorably predicted events that really happen. It determines the percentage of projected good outcomes that were really positive.

$$Precision = TP / (TP + FP). \quad (2)$$

Recall

Among all optimistically expected occurrences, it is the fraction that really takes place. It finds out how many favorable outcomes are really positive relative to all the others. This is the equation:

$$Recall = TP / (TP + FN). \quad (3)$$

F1-Score

By using false positives and false negatives, the F1 score achieves a balance between recall and accuracy by representing the harmonic mean of the two. One way to represent the F1 score is:

$$F1 - Score = 2 * (Precision * Recall) / (Precision + Recall). \quad (4)$$

Since this paper uses a multi-class classification dataset and even though we have used various augmentation techniques to address data imbalance, the reduced feature set from the extra-tree classifier still produce results that are not entirely biased. For examining the performance of the aforementioned proposed models, macro-averaged values of performance assessment criteria like accuracy, precision, recall, f1-score, etc. are preferable. The metrics is calculated for each class separately before being averaged in macro-averaging [24]. With this method, each class receives the same amount of weight regardless of how frequently it appears in the dataset. Models developed to solve problems of multi-class classification are evaluated using macro average statistics. Macro averaging is employed when a class is unbalanced (different numbers of instances are associated to different class labels). The arithmetic average of all the precision values for various classes is known as the macro average precision. The calculation for the data's macro average would be as follows [25]:

Macro average precision= (Precision 1^{st} + Precision 2^{nd} +…+Precision $n^{th)}$ /n, (5)

Where n denotes the number of classes.

The arithmetic average of all recall scores across all classes is known as the macro average recall and is calculated as:

Macro average Recall = (Recall 1^{st} +Recall 2^{nd} +…+Recall class $n^{th)}$ /n, (6)

Where n denotes the number of classes.

3.3 Test results and analysis

The recommended models were tested using the Al Shehri [26] study, which makes use of the ResNet-50 and DenseNet-169 architectures. Dementia, very mild dementia, mild dementia, and moderate dementia are the four categories that make up the multiclass Alzheimer's disease dataset that he used. The DenseNet-169 architecture outperformed the alternatives during testing and training. While DenseNet-169 achieved 97.7 and 83.82 accuracy during training and testing, ResNet-50 achieved 88.70 and 81.92 respectively.

The final module of the proposed work is a classification module that is a customised DNN that includes an average pooling layer, a flattening layer, and several dense layers. Customized DNN alone has an accuracy of 82% for our original dataset, while its accuracy for the augmented dataset was 89%. We then implement our proposed models by adding feature extraction and feature reduction modules and thereby sending the reduced dataset to this customised DNN (final classification module). All the proposed model works on the augmented dataset. Table 1 shows performance measures for different proposed models with and without dropouts. The first proposed model which is a hybrid DL-ML model with ResNet as feature extractor and extra-tree classifier as feature reducer got training accuracy 96.30, validation accuracy 91.76, and recall, precision and F1 score 92. The accuracy and loss graphs for the model without dropout and with dropout are depicted in Figures 10 and 11, respectively. The variation in validation and training accuracy is insignificant, even though we can use dropout regularisation techniques to reduce this

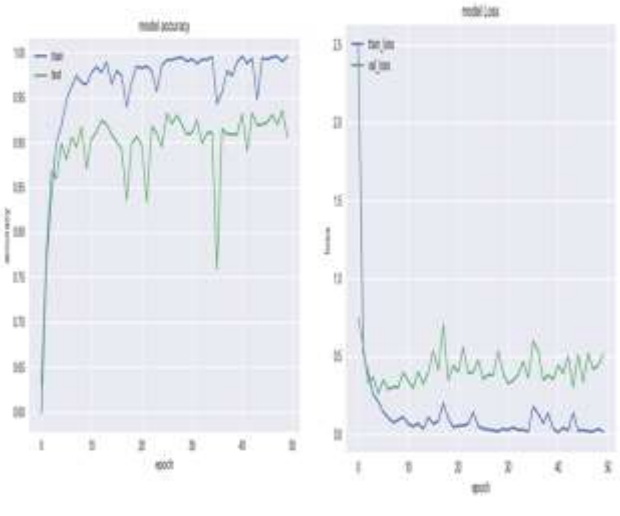


Figure 10*. Graph depicting the performance of Model 1(without dropout) in terms of accuracy and loss*

Discrepancy. This paper adds a 20 and 30 percent dropout ratio in the hidden layer, which gives a training accuracy value of 98.20, a validation accuracy value of 92.61, a recall value of 93, a precision value of 93, and an F1 score of 93. From this result, we can observe that the dropout strategy successfully increased the validation accuracy and decreased the discrepancy between validation and training accuracy. Training accuracy was 97.76, validation accuracy was 86.51, precision was 88, recall was 87, and the F1-score was 87 for the second suggested model, a Hybrid DL-ML model using DenseNet as the feature extractor and extra-tree classifier as the feature reducer. This model achieved a validation accuracy of 89.77, a precision value of 89, a recall value of 89, and an F1-score of 89 with dropout. Its training accuracy was 97.74. Model 2 with and without dropout are shown in Figures 12 and 13, respectively, along with their respective accuracy and loss graphs.

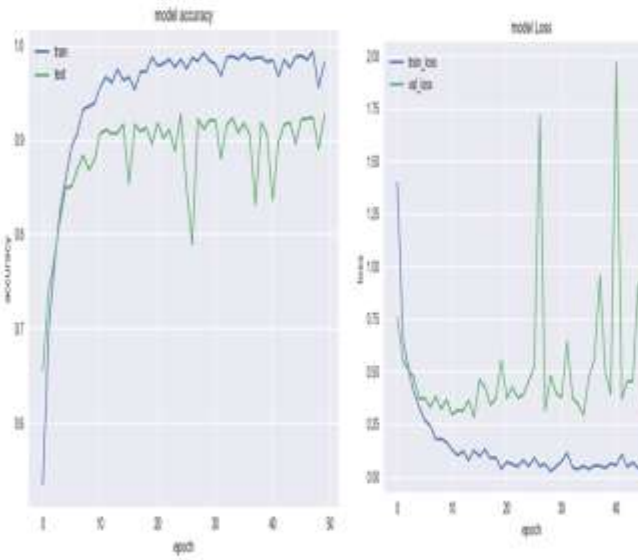


Figure 11. *Graph depicting the performance of Model 1 (with dropout) in terms of accuracy and loss*

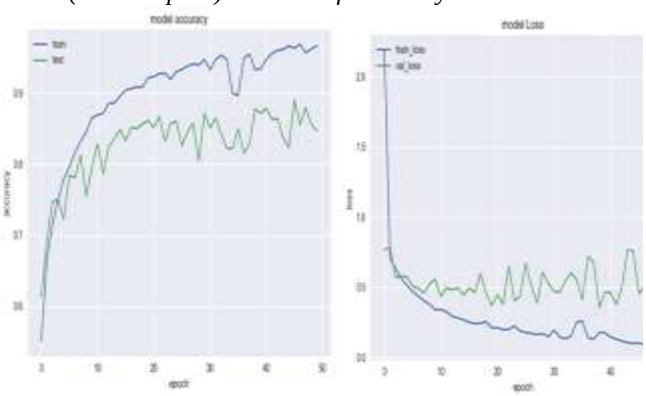


Figure 12. *Graph depicting the performance of Model 2 (without dropout) in terms of accuracy and loss*

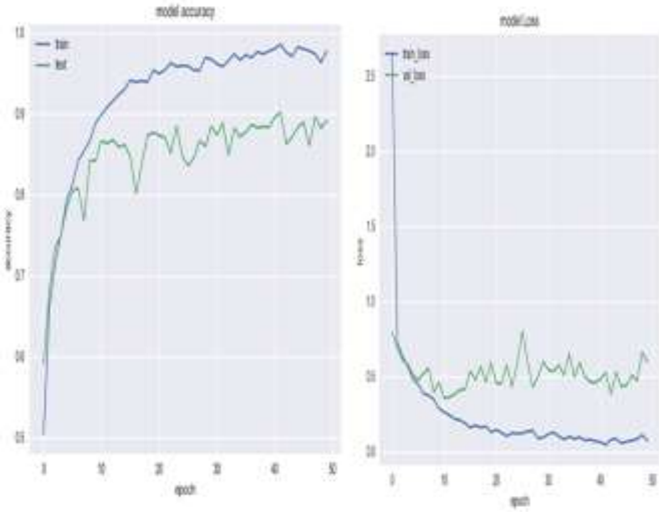


Figure 13. *Graph depicting the performance of Model 2 (with dropout) in terms of accuracy and loss*

The third suggested model- Hybrid DL-ML model with Xception as feature extractor and extra-tree classifier as feature reducer got a training accuracy value of 95.98, validation accuracy of 77.96, precision value of 78, recall value of 78, and F1-score of 78. The model with dropout got a training accuracy value of 92, validation accuracy of 75.68, precision value of 77, recall value of 76, and F1-score of 76. Figures 14 and 15 depicts the accuracy and loss graphs for Model 3 without dropout and with dropout. The accuracy metric was employed to evaluate the performance of the proposed models. They were compared with the advanced model introduced by Al Shheri [26], and the comparative results are visualized in Figure 16. From the above graph, we can conclude that our proposed model, which uses ResNet as a feature extractor shows better performance for this multi-class classification AD dataset. The discrepancy

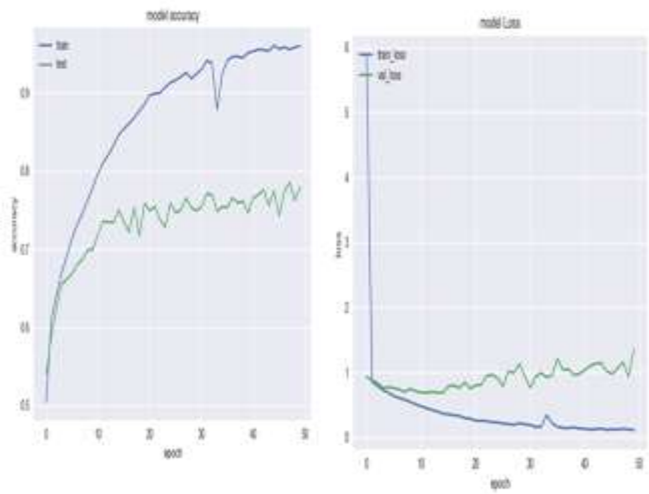


Figure 14. *Graph depicting the performance of Model 3 (without dropout) in terms of accuracy and loss*

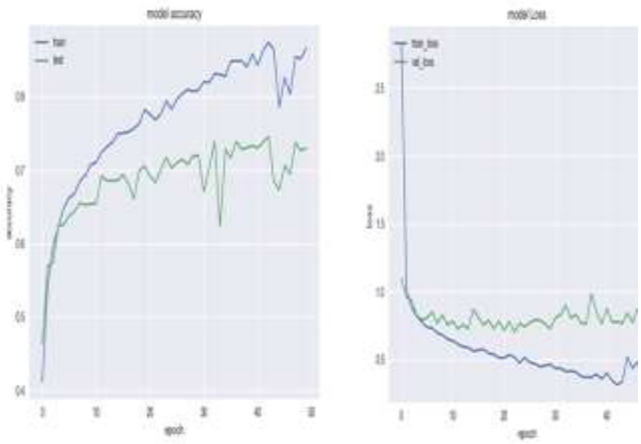
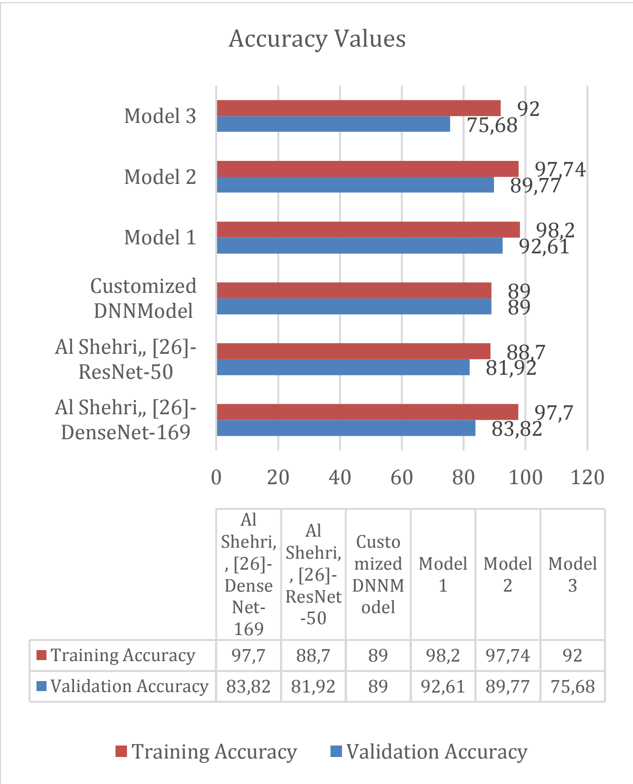


Figure 15. *Graph depicting the performance of Model 13 (with dropout) in terms of accuracy and loss*



	Al Shehri,, [26]-DenseNet-169	Al Shehri,, [26]-ResNet-50	Customized DNNModel	Model 1	Model 2	Model 3
Training Accuracy	97,7	88,7	89	98,2	97,74	92
Validation Accuracy	83,82	81,92	89	92,61	89,77	75,68

Figure 16. *Comparison Graph-Accuracy values*

between validation and training accuracy is less compared to the base paper by Al Shheri [26], which implies that our proposed model, Model 1-Hybrid DL-ML model with ResNet as feature extractor, effectively addresses the problem of overfitting also.

4. Conclusions

Neurodegenerative disorders are a diverse group of diseases that gradually lead to the degeneration of the central nervous system or peripheral nervous system, affecting their structure and functions progressively. This category of disorders includes diseases like Parkinson's and Alzheimer's. If we detect it early and treat it correctly, people can lead a 90% normal life. Continuous clinical evaluations are necessary for diagnosing this category of illness. Fortunately, we have a variety of neuroimaging biomarkers, including MRI, to quickly diagnose these conditions. To aid in the early identification of these diseases or to predict their stage, a number of deep learning algorithms are now available. This paper suggests different useful hybrid DL-ML models that include modules for dealing with data imbalance and data scarcity, a pre-trained feature extractor for extracting features, an extra tree classifier for feature reduction, and lastly, a customized DNN model for dealing with multiclass classification. Highest validation accuracy of 92.61% is achieved by the proposed model, which uses a pre-trained ResNet feature extractor. The similar interesting topic were discussed in literature [27-42].

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper
- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.
- **Author contributions:** The authors declare that they have equal right on this paper.
- **Funding information:** The authors declare that there is no funding to be acknowledged.
- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

[1] 'Alzheimer's Facts and Figures Report | Alzheimer's Association'. https://www.alz.org/alzheimers-dementia/facts-figures (accessed Apr. 21, 2023).

[2] J. Cummings. (2019). The Role of Biomarkers in Alzheimer's Disease Drug Development. *Adv. Exp. Med. Biol.* 1118; 29. DOI:10.1007/978-3-030-05542-4_2.

[3] 'Types of Alzheimer's: Early-Onset, Late-Onset, and Familial'.https://www.webmd.com/alzheimers/guide/alzheimers-types (accessed Apr. 21, 2023).

[4] L. Nanni et al. (2020). Comparison of Transfer

Learning and Conventional Machine Learning Applied to Structural Brain MRI for the Early Diagnosis and Prognosis of Alzheimer's Disease. *Front. Neurol.* 11; 1345. DOI:10.3389/FNEUR.2020.576194/BIBTEX.

[5] F. Ramzan et al. (2020). A Deep Learning Approach for Automated Diagnosis and Multi-Class Classification of Alzheimer's Disease Stages Using Resting-State fMRI and Residual Neural Networks. *J. Med. Syst.* 44(2). DOI:10.1007/s10916-019-1475-2.

[6] J. Joy and M. P. Selvan. (2022). A comprehensive study on the performance of different Multi-class Classification Algorithms and Hyperparameter Tuning Techniques using Optuna. *Proc. Int. Conf. Comput. Commun. Secur. Intell. Syst. IC3SIS 2022*. DOI:10.1109/IC3SIS54991.2022.9885695.

[7] A. Mehmood et al. (2021). A Transfer Learning Approach for Early Diagnosis of Alzheimer's Disease on MRI Images. *Neuroscience*. 460; 43–52. DOI:10.1016/J.NEUROSCIENCE.2021.01.002.

[8] D. Ma, D. Lu, K. Popuri, L. Wang, and M. F. Beg. (2020). Differential Diagnosis of Frontotemporal Dementia, Alzheimer's Disease, and Normal Aging Using a Multi-Scale Multi-Type Feature Generative Adversarial Deep Neural Network on Structural Magnetic Resonance Images. *Front. Neurosci.* 14. DOI:10.3389/FNINS.2020.00853.

[9] D. Pan, A. Zeng, L. Jia, Y. Huang, T. Frizzell, and X. Song. (2020). Early Detection of Alzheimer's Disease Using Magnetic Resonance Imaging: A Novel Approach Combining Convolutional Neural Networks and Ensemble Learning. *Front. Neurosci.* 14; 259. DOI:10.3389/FNINS.2020.00259/BIBTEX.

[10]Y. Wang et al. (2018, Oct.). A Novel Multimodal MRI Analysis for Alzheimer's Disease Based on Convolutional Neural Network. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Int. Conf.* 2018; 754–757. DOI:10.1109/EMBC.2018.8512372.

[11]R. Jain, N. Jain, A. Aggarwal, and D. J. Hemanth. (2019). Convolutional neural network based Alzheimer's disease classification from magnetic resonance brain images. *Cogn. Syst. Res.* 57; 147–159. DOI:10.1016/J.COGSYS.2018.12.015.

[12]S. El-Sappagh, T. Abuhmed, S. M. Riazul Islam, and K. S. Kwak. (2020). Multimodal multitask deep learning model for Alzheimer's disease progression detection based on time series data. *Neurocomputing*. 412; 197–215. DOI:10.1016/J.NEUCOM.2020.05.087.

[13]'A Complete Guide to Data Augmentation | DataCamp'.https://www.datacamp.com/tutorial/complete-guide-data-augmentation (accessed Apr. 22, 2023).

[14]'Alzheimer's Dataset (4 class of Images) | Kaggle'.https://www.kaggle.com/datasets/tourist55/alzheimers-dataset-4-class-of-images (accessed Jan. 26, 2023).

[15]R. Jha, V. Bhattacharjee, and A. Mustafi. (2022). Transfer Learning with Feature Extraction Modules for Improved Classifier Performance on Medical Image Data. *Sci. Program.* 2022. DOI:10.1155/2022/4983174.

[16]'ResNet: The Basics and 3 ResNet Extensions'. https://datagen.tech/guides/computer-vision/resnet/ (accessed Jan. 14, 2023).

[17] 'The Annotated ResNet-50. Explaining how ResNet-50 works and why… | by Suvaditya Mukherjee | Towards Data Science'. https://towardsdatascience.com/the-annotated-resnet-50-a6c536034758 (accessed Jan. 26, 2023).

[18]T. Zhou, X. Ye, H. Lu, X. Zheng, S. Qiu, and Y. Liu. (2022). Dense Convolutional Network and Its Application in Medical Image Analysis. *Biomed Res. Int.* 2022. DOI:10.1155/2022/2384830.

[19]'XCeption Model and Depthwise Separable Convolutions- https://maelfabien.github.io/deeplearning/xception/#what-does-it-look-like (accessed Jan. 26, 2023).

[20]'ImageNet: VGGNet, ResNet, Inception, and Xception with Keras - PyImageSearch'. https://pyimagesearch.com/2017/03/20/imagenet-vggnet-resnet-inception-xception-keras/ (accessed Jan. 26, 2023).

[21]F. Chollet, 'Xception: Deep Learning with Depthwise Separable Convolutions'.

[22]'sklearn.ensemble.ExtraTreesClassifier — scikit-learn 1.2.2 documentation'. https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.ExtraTreesClassifier.html (accessed Apr. 22, 2023).

[23]'Evaluation Metrics for Machine Learning - Accuracy, Precision, Recall, and F1 Defined | Pathmind'.http: wiki.pathmind.com/accuracy-precision-recall (accessed Apr. 22, 2023).

[24]'Micro-average and Macro-average Scoring Metrics - Data Analytics'. https:/vitalflux.com/micro-average-macro-average-scoring-metrics-multi-class-classification-python/ (accessed Apr. 22, 2023).

[25]'Understanding Micro, Macro, Weighted Averages - Scikit-Learn metrics in multi-class classification with example'. http://iamirmasoud.com/2022/06/19/understanding-micro-macro-and-weighted-averages-for-scikit-learn-metrics-in-multi-class-classification-with-example/ (accessed Apr. 19, 2023).

[26]W. Al Shehri. (2022). Alzheimer's disease diagnosis and classification using deep learning techniques. *PeerJ Comput. Sci.* 8; e1177. DOI:10.7717/PEERJ-CS.1177/SUPP-2.

[27]Nuthakki, praveena, & Pavankumar T. (2024). Comparative Assessment of Machine Learning Algorithms for Effective Diabetes Prediction and Care. *International Journal of Computational and Experimental Science and Engineering*, 10(4)1337-1343. https://doi.org/10.22399/ijcesen.606

[28]MOHAMMED, H. A., Adem, Şevki, & SAHAB, K. S. (2024). Estimation the Biochemical parameters Changes in Blood of Corona Virus Patients in Iraq in Order to Support the Timely Decision Needs . *International Journal of Applied Sciences and Radiation Research*, 1(1). Retrieved from https://ijasrar.com/index.php/ijasrar/article/view/4

[29]M, V., V, J., K, A., Kalakoti, G., & Nithila, E. (2024). Explainable AI for Transparent MRI Segmentation: Deep Learning and Visual Attribution in Clinical Decision Support. *International Journal of*

Computational and Experimental Science and Engineering, 10(4);575-584. https://doi.org/10.22399/ijcesen.479

[30]Özlen, M. S., Cuma, A. B., Yazıcı, S. D., Yeğin, N., Demir, Özge, Aksoy, H., … Günay, O. (2024). Determination of Radiation Dose Level Exposed to Thyroid in C-Arm Scopy. *International Journal of Applied Sciences and Radiation Research,* 1(1). Retrieved from https://ijasrar.com/index.php/ijasrar/article/view/13

[31]ÖZNACAR, T., & ERGENE, N. (2024). A Machine Learning Approach to Early Detection and Malignancy Prediction in Breast Cancer. *International Journal of Computational and Experimental Science and Engineering,* 10(4);911-917. https://doi.org/10.22399/ijcesen.516

[32]J. Anandraj. (2024). Transforming Education with Industry 6.0: A Human-Centric Approach . *International Journal of Computational and Experimental Science and Engineering*, 10(4);1851-1862. https://doi.org/10.22399/ijcesen.732

[33]ERTEKİN, R., RODOPLU, H., & GÜRSEL, S. (2024). The Use of Artificial Intelligence in Aviation: A Bibliometric Analysis. *International Journal of Computational and Experimental Science and Engineering,* 10(4);1863-1872. https://doi.org/10.22399/ijcesen.747

[34]Türkmen, G., Sezen, A., & Şengül, G. (2024). Comparative Analysis of Programming Languages Utilized in Artificial Intelligence Applications: Features, Performance, and Suitability. *International Journal of Computational and Experimental Science and Engineering,* 10(3);461-469. https://doi.org/10.22399/ijcesen.342

[35]Venkatraman Umbalacheri Ramasamy. (2024). Overview of Anomaly Detection Techniques across Different Domains: A Systematic Review. *International Journal of Computational and Experimental Science and Engineering*, 10(4);898-910. https://doi.org/10.22399/ijcesen.522

[36]Jafar Ismail, R., Samar Jaafar Ismael, Dr. Sara Raouf Muhamad Amin, Wassan Adnan Hashim, & Israa Tahseen Ali. (2024). Survey of Multiple Destination Route Discovery Protocols. *International Journal of Computational and Experimental Science and Engineering,* 10(3);420-426. https://doi.org/10.22399/ijcesen.385

[37]guven, mesut. (2024). Dynamic Malware Analysis Using a Sandbox Environment, Network Traffic Logs, and Artificial Intelligence. *International Journal of Computational and Experimental Science and Engineering,* 10(3);480-490. https://doi.org/10.22399/ijcesen.460

[38]Serap ÇATLI DİNÇ, AKMANSU, M., BORA, H., ÜÇGÜL, A., ÇETİN, B. E., ERPOLAT, P., … ŞENTÜRK, E. (2024). Evaluation of a Clinical Acceptability of Deep Learning-Based Autocontouring: An Example of The Use of Artificial Intelligence in Prostate Radiotherapy. *International Journal of Computational and Experimental Science and Engineering,* 10(4);1181-1186. https://doi.org/10.22399/ijcesen.386

[39]Günoğlu, K., & AKKURT, İskender. (2023). Gamma-ray attenuation properties carbide compounds (WC, Mo2C, TiC, SiC, B4C) using Phy-X/PSD software. *International Journal of Applied Sciences and Radiation Research*, 1(1), 1–8. Retrieved from https://ijasrar.com/index.php/ijasrar/article/view/6

[40]S. Esakkiammal, & K. Kasturi. (2024). Advancing Educational Outcomes with Artificial Intelligence: Challenges, Opportunities, And Future Directions. *International Journal of Computational and Experimental Science and Engineering,* 10(4);1749-1756. https://doi.org/10.22399/ijcesen.799

[41]J. Prakash, R. Swathiramya, G. Balambigai, R. Menaha, & J.S. Abhirami. (2024). AI-Driven Real-Time Feedback System for Enhanced Student Support: Leveraging Sentiment Analysis and Machine Learning Algorithms. *International Journal of Computational and Experimental Science and Engineering,* 10(4);1567-1574. https://doi.org/10.22399/ijcesen.780

[42]MOHAMMED, H. A., Adem, Şevki, & SAHAB, K. S. (2023). Optimal Examination Ways to follow up patients effected by COVID-19: case study in Jalawla General Hospital in Iraq. *International Journal of Applied Sciences and Radiation Research,* 1(1). Retrieved from https://ijasrar.com/index.php/ijasrar/article/view/5