IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 14 September 2022, accepted 15 October 2022, date of publication 21 October 2022, date of current version 28 October 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3216393

RESEARCH ARTICLE

Data Complexity Based Evaluation of the Model Dependence of Brain MRI Images for Classification of Brain Tumor and Alzheimer's Disease

ANIMA KUJUR[1], ZAHID RAZA[1], ARFAT AHMAD KHAN[2], AND CHITAPONG WECHTAISONG[3]

[1]School of Computer and Systems Sciences, Jawaharlal Nehru University, New Delhi 110067, India
[2]Department of Computer Science, College of Computing, Khon Kaen University, Khon Kaen 40002, Thailand
[3]School of Telecommunication Engineering, Suranaree University of Technology, Nakhon Ratchasima 30000, Thailand

Corresponding author: Arfat Ahmad Khan (arfatkhan@kku.ac.th)

This work was supported by the College of Computing, Khon Kaen University, Khon Kaen, Thailand.

ABSTRACT The convolutional neural networks (CNN) have shown promising results for various classification problems over the past years. However, selecting various CNN architectures is still challenging as each architecture performs differently with the same dataset. This research aims to evaluate the dependence of brain MRI on various predictive models of CNN based on the complexity of the data for Brain Tumor and Alzheimer's Disease. Our proposed approach has three parts. First part is the pre-processing of the data which mainly focuses on class balancing and the estimation of data complexity. The second part uses stratified k-fold cross-validation for more reliable results. The last part corresponds to the implementation of four CNN models applying described methods. This paper compares the classification performance of rigorous experimentation on four CNN variants namely S-CNN (CNN trained from scratch), ResNet50, InceptionV3, and Xception over two brain MRI image datasets evaluated with and without the use of Principal Component Analysis (PCA). The work benchmarks CNN models by comparing the average scores of Accuracy, Precision, Recall, F1 score, and AUC score from the stratified five-fold cross-validation.

INDEX TERMS Magnetic imaging resonance (MRI), brain tumor, alzheimer's disease, S-CNN, ResNet50, inceptionV3, Xception, stratified k-fold, principal component analysis (PCA).

I. INTRODUCTION

Brain Tumor is a life-threatening disease that is a result of uncontrolled and abnormal growth of brain cells leading to Brain Tumors disrupting the operation of the brain. Brain Tumors have been reported as the 10th leading cause of deaths worldwide [1]. The main causes of Brain Tumors are still unknown and the successful treatment, detection, and identification of Brain Tumors at an early phase becomes vital. Alzheimer's disease is a neurological brain degenerative disorder that affects the mental health of people, particularly older people more than 65 years of age [2]. It is a common form of dementia that affects patients to lose their memory. It is an irreversible disease and the patients need full-time assistance. There is no cure for dementia, but its progression can be delayed and therefore, the early diagnosis of Alzheimer's becomes very important to ensure proper treatment and patient care.

The use of deep learning algorithms has tremendously increased in all fields of medical research as it facilitates learning from the given data without any human supervision and accordingly assists in image classification. Over the past few decades, CNN variants have become popular and are being used in various classification problems using the

The associate editor coordinating the review of this manuscript and approving it for publication was Jinhua Sheng.

transfer learning technique where a network trained over natural images is reused for medical images [3]. Due to various CNN architectures, selecting the most suitable one among them becomes a challenging task as trying out all CNN variants is time-consuming and demands high computational resources. The size and type of training data also have an important role in the deep learning methods. A model trained with small data can lead to poor generalization performance and wrong computer-aided disease diagnosis. Another important challenge is class imbalance where the representation of one class is more than the other [4]. Training models over imbalanced dataset results in an unstable model with biased performance towards the overrepresented class.

Over the past few years, many state-of-the-art CNN architectures have been developed and have shown promising results on natural image classification problems [5]. Furthermore, the state-of-the-art CNN architectures are trained over millions of natural training images [6]. Training these models on medical images is challenging as medical image data is scarce [7]. However, a technique called transfer learning is used to transfer the knowledge from models trained on natural image classification problems to medical image classification and this technique has gained significant results [7], [8]. Due to the many popular CNN variants, it is always desired to choose the one that is suitable for the data under investigation. Literature review in the domain suggests that many CNN variants used for Brain Tumor and Alzheimer's disease classification have been selected arbitrarily. Moreover, none of the reported studies on Brain Tumors and Alzheimer's disease have applied the k-fold cross-validation.

This work considers two brain MRI datasets belonging to two disease systems with similar goals of binary classification as either *positive* or *negative*. The first dataset corresponds to Brain Tumor detection and the second one for Alzheimer's disease detection. For each dataset, the work implemented one CNN which is trained from scratch, and the remaining three CNN variants were implemented using transfer learning techniques. In the first step, the complexity of data was determined using unsupervised k-means clustering [9] which helped estimate the number of samples that can be correctly labeled. Medical image datasets are imbalanced in which one target class has more observations than the other class where unhealthy image samples are more than the healthy image samples. The classification models trained with imbalanced data perform poorly on the minority class because the classifier aims to optimize the model's overall accuracy without considering the class ratio distribution. Therefore, the oversampling method, Synthetic Minority Oversampling Technique (SMOTE) [10] is used to balance the class labels where samples from minor classes are randomly selected to create the synthetic data. Further, the work implements the stratified k-fold cross-validation method [11] on balanced data to get a stable generalization performance for a more reliable computer-aided disease diagnosis.

The proposed approach in this work has three parts where the first is the pre-processing of the data which mainly focuses on class balancing and the estimation of data complexity. The second is stratified k-fold cross-validation and the last is an implementation of four CNN models applying the mentioned methods. The objective of the work is to evaluate the dependence of brain MRI on various predictive models of CNN based on the complexity of the data under investigation. This is achieved by comparing the classification performance of rigorous experimentation on four CNN variants namely S-CNN (trained from scratch) [12], ResNet50 [13], InceptionV3 [14], and Xception [15] and all four CNN models were compared with proposed PCA-CNN model. The aim is to benchmark CNN models by comparing the average scores of Accuracy, Precision, Recall, F1 score, and AUC score from the stratified five-fold cross-validation with the final goal of early detection of Brain Tumor and Alzheimer's as a proactive approach.

The remainder of the paper is as follows: Section II discusses the related work presented in the literature and contributions of this work. Section III presents the proposed methodology for the Brain MRI classification model. Section IV discusses the experimental results. Section V describes the Principal Component Analysis method, proposed PCA-CNN framework and its implementation. Section VI presents the simulation study of PCA-CNN and discussion of the results. The FLOPs computation is presented in the section VII. Finally, Section VIII presents the conclusions and future remarks.

II. RELATED WORK

Brain Tumor is a deadly disease that originates in the brain where the cells in a mass of tissue get multiplied. Tumors can be formed at any location of the brain and depending upon the location different cells can cause the tumor. Various studies have been done on the Brain Tumor dataset. The transfer learning technique is one of the most used techniques that have been adopted to develop robust and efficient methods for the classification of Brain Tumors. Reference [1] used brain MRI data from Kaggle to detect tumors that contain 253 brain MRI images out of which 155 image samples are tumorous and 98 images are non-tumorous. The authors have implemented three image augmentation techniques, namely flipping, rotation and translation, and produced 3700 new augmented images. The augmented dataset is further implemented using various state-of-the-art CNN models using transfer learning technique and reported that VGG-16 has the highest accuracy with 96%. Similar works have been presented in [16] where apart from VGG-16, AlexNet, ResNet, InceptionV3, and DenseNet have been used. Arbane et al. [17] proposed three CNN models, ResNet-50, Xception, and MobileNet-V2 for the detection of Brain Tumors on 253 brain MRI images and reported that MobileNet-V2 outperformed with an F1 score of 98.42%.

Alzheimer's disease is a neurological disorder and a common form of dementia mostly found in elderly people which initially destroys memory and can eventually cause the death of the person. Therefore, the early detection of disease can

help the afflicted to get the right treatment. Many studies have been reported on Alzheimer's disease detection. Again, various pre-trained CNN networks have been adopted using the transfer learning technique for the classification of Alzheimer's disease. Zaabi et al. [18] Proposed CNN architecture with 5 convolutional layers where each convolutional layer has 32 filters of size (3,3) and implemented on the Open Access Series of Imaging Studies dataset [19] for the detection of Alzheimer's disease on 166 brain MRI image samples with 86 unhealthy images and 80 healthy images. Further, the AlexNet model was implemented using transfer learning, and both models were tested on 84 images and reported that AlexNet outperformed the proposed CNN (88.10%) with 92.86%. It is to be noted that reported literature on Alzheimer's disease detection mostly illustrates multi-class classification problems. On the other hand, the proposed work focuses on binary classification. The major limitation of the existing work is that none of the works exploits the stratified k-fold cross-validation. Another limitation gathered from the literature survey indicates that none of the reported studies have applied data complexity-based implementation.

The contributions of the work have been summarized below.

i. Various CNN architectures have been evaluated with implementation over the Brain Tumor and Alzheimer's disease datasets.
ii. The stratified 5-fold cross-validation has been implemented for the CNNs to ensure robust and reliable results.
iii. A data complexity-based evaluation of the CNN architectures.
iv. Model complexity reduction by leveraging input feature dimensionality reduction using the Principal Component Analysis (PCA) method has been performed.
v. Floating Point Operations (FLOPs) have been estimated to evaluate the complexity of the models along with standard performance measures.

III. BRAIN MRI CLASSIFICATION MODEL

This work uses the images obtained from a publicly available data source, Kaggle [https://www.kaggle.com]. The proposed work makes use of the S-CNN, ResNet50, InceptionV3, and Xception predictive models that have been applied to the pre-processed Brain MRI datasets. The flow diagram of the proposed Brain MRI classification model comprises five phases viz. Data collection, Date pre-processing, and Data split followed by Model training and Model evaluation as shown in Fig. 1. The performance of these models has been evaluated over the standard performance metrics viz. Accuracy, Precision, Recall, F1 Score, and AUC-ROC Score.

A. DATA COLLECTION

The dataset implemented in this work is MRI images for Brain Tumor and Alzheimer's disease. The MRI technique uses a magnetic field and radio waves generated from the

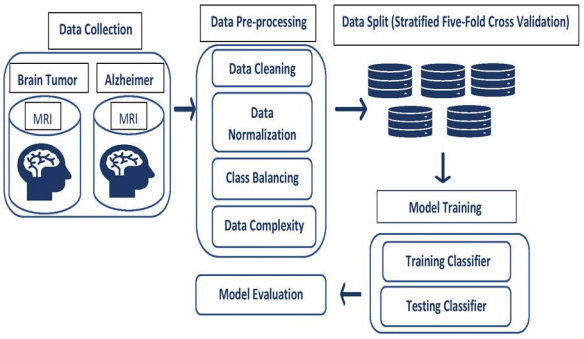

FIGURE 1. Flow diagram for brain MRI image classification.

computer to create very detailed high-resolution images of the brain. The MRI imaging technique is a non-invasive method that is used to detect the abnormalities of the organs and tissues of the body and is also used for monitoring the treatment. MRI technology produces three-dimensional images. Brain MRI images are 3D images. There are techniques for 3D to 2D image patch conversion [20]. The preprocessed 2D image dataset is collected from a publicly available data source Kaggle. The Brain Tumor dataset has 3533 images, which contains 599 non-tumorous images, and 2934 tumorous images. Some of these images presenting a healthy brain without a tumor have been shown in Fig. 2 and an unhealthy brain with a tumor have been shown in Fig. 3.

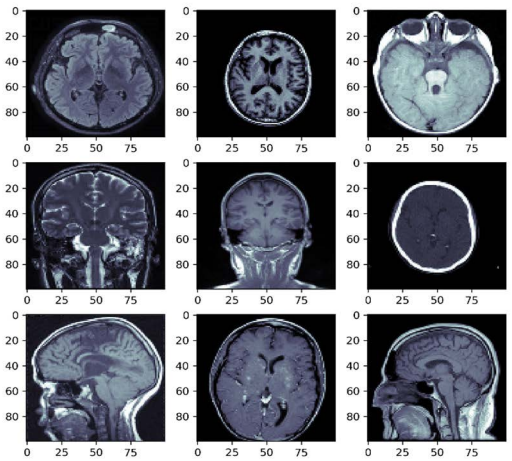

FIGURE 2. Healthy brain MRI images without tumor.

Advanced neuroimaging technologies such as MRI are also used to diagnose Alzheimer's disease, which is the most common type of dementia. Brain MRI images for Alzheimer's disease detection are 3D images. There are techniques for 3D to 2D image patch conversion [20]. The preprocessed 2D image dataset of Alzheimer's disease has been collected from a publicly available data source Kaggle [https://www.kaggle.com]. Alzheimer's disease dataset has 6400 images, which contains 3200 images with dementia and 3200 images without dementia. Some of these images

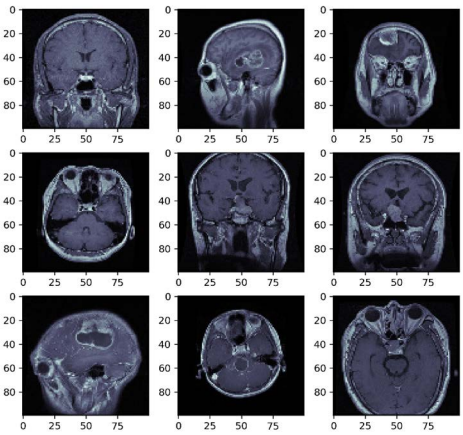

FIGURE 3. Unhealthy brain MRI images with tumor.

depicting a healthy brain without Alzheimer's disease have been shown in Fig. 4 and an unhealthy brain with Alzheimer's disease has been shown in Fig. 5.

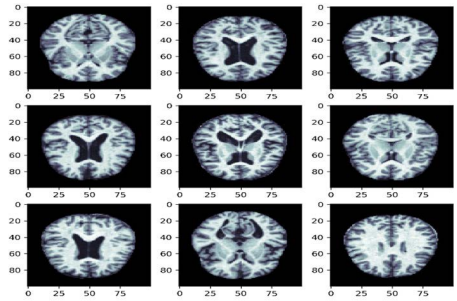

FIGURE 4. Healthy brain MRI images of alzheimer's disease without dementia.

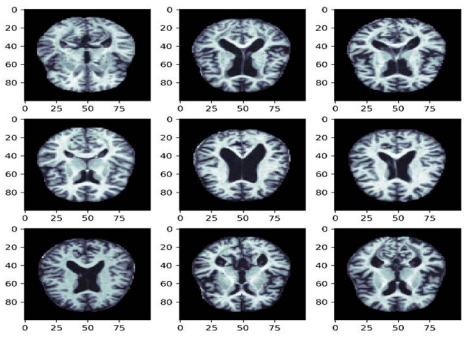

FIGURE 5. Unhealthy brain MRI images of alzheimer's disease with dementia.

B. DATA PRE-PROCESSING

The datasets discussed above have images of different dimensions. Each image size differs from the other. As a first step toward data pre-processing is data cleaning, images were resampled and resized into standard one dimension such that all images have the same height and width. The images which were not labeled were discarded. To normalize the image pixel values which range from 0 to 255, each pixel is divided by 255 to get the pixel range from 0 to 1. This work uses (224,224,3) as input image shape for ResNet50, InceptionV3, Xception, and (128,128,3) for S-CNN. The Brain Tumor dataset suffers from the class imbalance problem in which the tumor class has more observations than the ones with no tumor.

The classification models trained with imbalanced data perform poorly on the minority class because the classifier aims to optimize the model's overall accuracy without considering the class ratio distribution. The imbalanced data can cause poor classification performance. To resolve this, among various oversampling techniques, we chose the SMOTE technique to balance the Brain Tumor dataset [10]. SMOTE uses the k-nearest neighbors (KNN) algorithm to generate synthetic images. From the minority class, some samples are randomly selected to compute k-nearest neighbor images and then a line segment is generated between the selected images and the computed KNN images. Then, a selection of the synthetic images falling within the line segments is done which are then added to the original dataset. The resultant dataset is further used for implementation. SMOTE has the advantage that it creates synthetic data points which are slightly different from the original data points and does not create mere duplicates.

The images have overlapping features or pixel intensities between classes and this leads to a difficult learning process. To determine a priori data complexity before training, this work simply clustered the images into two groups using k-means clustering [21] over pixel intensities. The labels are assigned to the clustered data and estimated the number of samples that can be correctly labeled. This work assumes that higher clustering label prediction from a simple Euclidean distance metric indicates lower complexity of the dataset and vice versa. The cluster with a high prediction label has lower complexity and the cluster with a low prediction label has high complexity. In the case of Brain Tumor data, 66 percent of the image samples are correctly labeled. Therefore, the complexity of the Brain Tumor data is 66%. For Alzheimer's disease detection, only 50 percent of the image samples are correctly labeled. Therefore, the complexity of the Alzheimer's data is 50%.

C. DATA SPLITTING (STRATIFIED K-FOLD CROSS VALIDATION)

The work implements the stratified k-fold cross-validation technique [11] to train and test the CNN models. The stratified k-fold preserves the class ratio which means it includes an equal number of positive and negative samples in each fold. In this work, using the validation technique, each dataset was divided into five disjoint folds or subsets as shown in Fig. 6. A total of 1173 images in each fold were considered for Brain Tumor detection, and 1280 images in each fold for Alzheimer's disease detection. All CNN models were trained five times iteratively over four folds and tested over

the remaining fold. In this process, each fold is at most given one chance to be tested. In each iteration, 4692 images were used for training, and 1173 images were used for validation of the Brain Tumor dataset. Similarly, for Alzheimer's disease, 5120 images were used for training and 1280 images were used for validation.

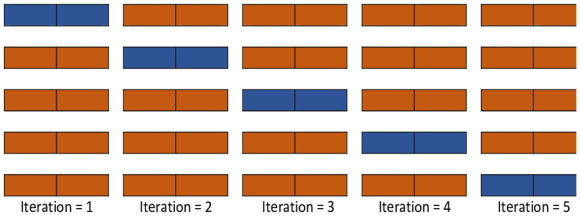

FIGURE 6. Stratified five-fold cross-validation.

D. CONVOLUTIONAL NEURAL NETWORK MODELS FOR CLASSIFICATION

The classification performance has been evaluated for the MRI dataset on four CNN variants, namely S-CNN (trained from scratch) [12], ResNet50 [13], InceptionV3 [14], and Xception [15].

1) S-CNN

The Convolutional neural network is a deep learning algorithm most utilized for image classification and computer vision tasks [22]. It provides a more scalable approach to object recognition and image classification tasks. CNN can capture the spatial and temporal dependencies in an image and it performs a better fitting due to the reduction in the number of parameters involved and the reusability of weights. A simple CNN consists of one Convolutional layer, one pooling layer, and followed by one classification layer. The convolutional layer is used for feature extraction. The pooling layer is further used for reducing the feature dimensions and the last classification layer is used for the classification of objects. S-CNN used in this work stands for CNN trained from scratch.

Fig. 7 depicts the S-CNN architecture implemented in this work. It takes the input image of shape (128,128,3), then the convolution operation is applied with 256 kernels of shape (3,3). The convolutional layer reduces the number of image features into a form that is easier to process by preserving the features critical for a good prediction. The output of the convolutional layer is called a feature map. Further, max-pooling of size (2,2) is applied to the 256 feature maps of shape (126,126) which reduces the features by pooling out the essential features. The pooled 256 feature maps of size (63,63) are flattened into the single dimensional array of shape 1,016,064 and passed to the fully connected layer with 128 nodes for the final classification.

2) ResNet50

It is a variant of Residual Neural Network (ResNet) that utilizes skip connections. Skip connections are shortcuts to

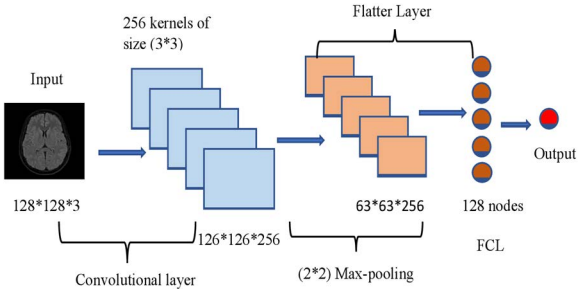

FIGURE 7. S-CNN architecture.

jump over some layers. ResNet50 is CNN with 50 layers deep where it has 48 convolution layers along with max-pooling layer and one average pooling layer and was originally trained on the ImageNet database [13]. The feature extraction layer is very deep as it learns features at multiple levels.

3) InceptionV3

Inception is a module used in convolutional neural networks which allows more efficient computations. The simplest inception module performs a convolution on three different filter sizes (1×1, 3×3, 5×5) followed by max-pooling and then outputs are concatenated to be sent to the next layer. InceptionV3 is the third edition of the Inception family originally introduced for the ImageNet recognition challenge. InceptionV3 is 48 layers deep [14].

4) Xception

It is a convolutional neural network that is 71 layers deep with a modified depth-wise separable convolution. It was proposed by Chollet Francis and is the extension of Inception model architecture. Xception slightly outperformed InceptionV3 on the ImageNet dataset and it has the same number of model parameters as Inception [15]. The Xception has replaced the inception module with a modified depth-wise separable convolution on the Inception model.

This work implemented ResNet50, InceptionV3, and Xception using a transfer learning technique. In the transfer learning, only the base layers are taken with the pre-trained weights. The top layer, which is a problem-specific layer of the model, is replaced by the new classification layer. The top layer includes one global average pooling layer, one fully connected layer with 256 nodes, and the last one is the output layer with one node. During the training of the model, only the top layers is trained. The feature extraction layers are kept frozen as it uses the pre-trained weights to extract the features from the input image.

E. MODEL TRAINING AND EVALUATION

The S-CNN has one convolutional layer with 256 filters, a relu activation function, one max pooling layer of size (2,2), and one dense layer with 128 nodes, followed by a sigmoid activation function on a binary classification layer with 1 node. For the transfer learning of ResNet50, InceptionV3,

and Xception model, feature extraction layers were kept frozen. The feature extraction layers also known as base layers are not being trained during model training because the pre-trained model weights are being used. The addition of the top layer called the problem-specific layer has been done in the work which includes one max global average pooling layer, one dense layer with 256 nodes, and the output layer with 1 node. All three pre-trained CNN variants were trained with Adam optimizer with a learning rate of 0.0001, 50 epochs with a batch size of 32.

F. PERFORMANCE METRICS

Trained CNN models were evaluated based on the elements of the confusion matrix, which is a (2,2) matrix of true and predicted labels as presented in Fig. 8. The confusion matrix is the most widely used metric to find the correctness of the model. It is a square matrix having A_{ij} as the elements, where i denote the true label and j denotes the predicted label. The elements are true positive (A_{11}), false positive (A_{10}), false negative (A_{01}), and true negative (A_{00}). The trained CNN models were compared based on five performance metrics, namely accuracy, precision, recall, F1 score, and AUC – ROC score [23].

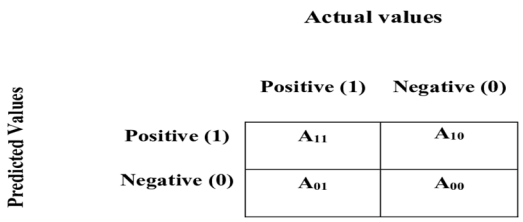

FIGURE 8. Confusion matrix.

1) ACCURACY

It is defined as the ratio of the number of correct predictions to the total number of predictions made by the classifier evaluated by the following equation:

$$\text{Accuracy} = \frac{A_{11} + A_{00}}{A_{11} + A_{01} + A_{00} + A_{10}} \tag{1}$$

2) PRECISION

It is defined as the ratio of the number of true positive predictions to the total number of predicted positives and it is evaluated by the following equation:

$$\text{Precision} = \frac{A_{11}}{A_{11} + A_{10}} \tag{2}$$

3) RECALL

It is defined as the ratio of true positive predictions to the actual positive samples. It is calculated by the following equation:

$$\text{Recall} = \frac{A_{11}}{A_{11} + A_{01}} \tag{3}$$

4) F1 SCORE

It is the harmonic mean of precision and recall. F1 score is used to create the balance between the false positive and false negative classes, which is evaluated as the following equation:

$$\text{F1 Score} = 2 \times \frac{Precision + Recall}{Precision + Recall} \tag{4}$$

5) AUC-ROC SCORE

The receiver operating characteristics (ROC) is a technique to depict visualization and select the classifier for all the classification thresholds. It is plotted between the true positive rates and false-positive rates. The area under the ROC curve (AUC-ROC) is the measure of the performance of the binary classifier.

IV. EXPERIMENTAL RESULTS

This section presents the evaluation of the results first for the Brain Tumor dataset and then for Alzheimer's disease dataset. The simulation study was performed using the hardware/software specifications as mentioned in Table 1. Table 2 presents the parameters used by CNN variants used in the work.

TABLE 1. Hardware/ software specifications.

Hardware/Software Specifications	Version
GPU	K80, T4, and P100
RAM	52GB
Python	3.7.13
Tensorflow	2.8.0
Keras	2.8.0
Scikit Learn	1.0.2
Pandas	1.3.5
Numpy	1.21.6
Matplotlib	3.2.2
Seaborn	0.11.2
Imblearn	0.8.1
Cv2	4.1.2

To estimate the a priori data complexity of Brain Tumor and Alzheimer's dataset, the images were clustered into two groups using a k-means clustering algorithm. The k- means clustering algorithm is an unsupervised machine learning algorithm that uses Euclidean distance metric to cluster the images using their pixel intensities. The observed cluster labels were then compared with the true labels. Thus, k-means clustering helped to determine how many images are correctly labelled. Percent of images that are correctly

TABLE 2. Parameters used in CNN classifiers.

S-CNN Parameters					
Parameter	Value	Parameter	Value	Parameter	Value
filters	256	Filter size	(3,3)	Input shape	(128,128,3)
Nodes	129	loss	Binary cross-entropy	Pooling size	(2,2)
Activation function	Relu, sigmoid	Beta_2	0.999	loss_weights	None
Learning Rate	0.00001	Epsilon	1e-07	Shuffle	True
Beta_1	0.9	Verbose	1	Steps_per_epoch	None
epochs	50	Batch size	32	Callbacks	None

ResNet50 Parameters					
Parameter	Value	Parameter	Value	Parameter	Value
Weights	imagenet	Include top	False	loss_weights	None
Nodes	257	Loss	Binary cross-entropy	Weighted_metrics	None
Activation function	Relu, sigmoid	Beta_2	0.999	Input shape	(224,224,3)
Learning Rate	0.0001	Epsilon	1e-07	Shuffle	True
Beta_1	0.9	Verbose	1	Steps_per_epoch	None
epochs	50	Batch size	32	Callbacks	None

InceptionV3 Parameters					
Parameter	Value	Parameter	Value	Parameter	Value
Nodes	257	Include top	False	loss_weights	None
Learning Rate	0.0001	Loss	Binary cross-entropy	Weighted_metrics	None
Activation function	Relu, sigmoid	Beta_2	0.999	Input shape	(224,224,3)
Beta_1	0.9	Epsilon	1e-07	Shuffle	True
epochs	50	Batch size	32	Callbacks	None

TABLE 2. *(Continued.)* Parameters used in CNN classifiers.

Xception Parameters					
Parameter	Value	Parameter	Value	Parameter	Value
Nodes	257	Include top	False	loss_weights	None
Beta_1	0.9	Loss	Binary cross-entropy	Weighted_metrics	None
Activation function	Relu, sigmoid	Beta_2	0.999	Input shape	(224,224,3)
Learning Rate	0.0001	Epsilon	1e-07	Shuffle	True
epochs	50	Batch size	32	Callbacks	None

labelled for Brain Tumor data was observed to be 66% and for Alzheimer's data to be 50% of the images were correctly labelled. Thus, it is observed that the data complexity of the Brain Tumor data and Alzheimer's data is 66% and 50% respectively.

A. BRAIN TUMOR CLASSIFICATION RESULTS

This section contains measured metric scores for Brain Tumor dataset for all the four CNN classifiers implemented in the work. Table 3 to table 7 present the performance of four classifiers for various performance metrics discussed in Section III(F). For the same, the work considered the computation to observe the minimal values (Min), maximal values (Max), mean values (Mean), and standard deviation (SD) for getting a comprehensive view of the results.

Table 3 shows the accuracy of the four classifiers under study for the Brain Tumor dataset. In the case of mean accuracy, it is observed that Xception reports the highest value of 98.51% with a minimum standard deviation. InceptionV3 also reports a value that is closest to Xception among the remaining three classifiers. ResNet50 performs the worst in this case with the least accuracy and a high standard deviation. Thus, it can be inferred that if Accuracy is the key parameter for judging the performance of the classifiers for Brain Tumors, it can be done best with Xception or InceptionV3 as the second choice.

Table 4 shows the precision scores of the four classifiers for the Brain Tumor dataset. In the case of mean precision, Xception performed comparatively better reporting the highest precision of 98.79% than peers with a significantly lesser standard deviation. Inception reports the precision value closer to Xception but has a high standard deviation while ResNet50 reports the most inferior results. Thus, if precision is the key parameter, Xception and InceptionV3 prove to be better classifiers for the dataset.

TABLE 3. Accuracy of classifiers for brain tumor detection.

Classifiers	Min	Max	Mean	SD
S-CNN	0.9693	0.9787	0.9735	0.0034
ResNet50	0.8901	0.9105	0.8994	0.0070
InceptionV3	0.9786	0.9863	0.9820	0.0029
Xception	0.9812	0.9897	0.9851	0.0029

TABLE 4. Precision scores of classifiers for brain tumor detection.

Classifiers	Min	Max	Mean	SD
S-CNN	0.9647	0.9827	0.9712	0.0071
ResNet50	0.8417	0.9242	0.8838	0.0340
InceptionV3	0.9779	0.9847	0.9832	0.0034
Xception	0.9862	0.9897	0.9879	0.0015

Table 5 shows the Recall scores of classifiers for the Brain Tumor dataset. Again, it is observed that Xception offers the best mean Recall scores followed by InceptionV3. Thus, if Recall is the key parameter, both Xception and InceptionV3 classifiers could be used for effective Brain Tumor classification.

TABLE 5. Recall scores of classifiers for brain tumor detection.

Classifiers	Min	Max	Mean	SD
S-CNN	0.9693	0.9812	0.9760	0.0050
ResNet50	0.8841	0.9608	0.9229	0.0315
InceptionV3	0.9761	0.9880	0.9806	0.0034
Xception	0.9778	0.9897	0.9842	0.0042

Table 6 shows the F1 scores of the classifiers for the Brain Tumor dataset. Here, it can be seen that the mean F1 score reported by Xception and InceptionV3 are again the best in the category with S-CNN reporting a value close to them. ResNet50 again performed poorly. Thus, if the F1 score is the key parameter, both Xception and InceptionV3 serve as the potential classifiers.

In order to ascertain the performance of Xception and InceptionV3 as compared to the peers, the area under the ROC curve (AUC-ROC) was computed. Table 7 presents the AUC-ROC scores of the classifiers for the Brain Tumor dataset. The AUC-ROC score of the Xception is observed to be the highest amongst all the classifiers followed by InceptionV3. ResNet50 again reports the least value of the AUC-ROC score. The results ascertain the performance of Xception and InceptionV3 for Brain Tumor detection again.

TABLE 6. F1 Scores of classifiers for brain tumor detection.

Classifiers	Min	Max	Mean	SD
S-CNN	0.9693	0.9787	0.9736	0.0033
ResNet50	0.8901	0.9090	0.9003	0.0060
InceptionV3	0.9786	0.9863	0.9820	0.0029
Xception	0.9811	0.9897	0.9850	0.0029

TABLE 7. AUC- ROC scores of classifiers for brain tumor detection.

Classifiers	Min	Max	Mean	SD
S-CNN	0.9693	0.9787	0.9735	0.0034
ResNet50	0.8901	0.9105	0.8994	0.0070
InceptionV3	0.9786	0.9863	0.9820	0.0029
Xception	0.9812	0.9897	0.9851	0.0029

Table 8 shows the summary of all the measured metrics for four CNN variants used in the work. It has been observed that the Xception model consistently reported the best performance for all the parameters with the best standard deviations. The other classifier which came close to the performance of Xception was InceptionV3. ResNet50 reported the worst performance throughout with S-CNN reporting a mediocre performance. The InceptionV3 and Xception models have the same standard deviation for all measured metrics except for recall. Thus, it is established that the best classification and detection of Brain Tumor can be done with Xception and InceptionV3 as the preferable classifiers.

TABLE 8. Summary table of classifiers for brain tumor detection.

Measures	Max (Min)	Max (Max)	Max (Mean)	Min (SD)
Accuracy	Xception	Xception	Xception	InceptionV3, Xception
Precision	Xception	Xception	Xception	Xception
Recall	Xception	Xception	Xception	Inception
F1 score	Xception	Xception	Xception	InceptionV3, Xception
AUC-ROC	Xception	Xception	Xception	InceptionV3, Xception

B. ALZHEIMER'S DISEASE CLASSIFICATION RESULTS

This section contains all measured metric scores for Alzheimer's disease dataset by CNN classifiers considered in the work viz. S-CNN, ResNet50, InceptionV3 and Xception. Table 9 - Table 13 presents the performance of these four

classifiers for various performance metrics as discussed in Section III(F). Again, as was the case in Section IV(A), the work considered the computation to observe the minimal values (Min), maximal values (Max), mean values (Mean), and standard deviation (SD) for getting a comprehensive view of the results.

Table 9 shows the accuracy of the classifiers for Alzheimer's disease dataset. It is observed that ResNet50 reports the least accuracy as compared to peers. In the case of mean accuracy, it is observed that InceptionV3 reports the highest value of 84.26%. Xception also reports a value that is closest to InceptionV3 among the remaining three classifiers. ResNet50 performs better than the S-CNN in this case with the minimum standard deviation but reports a much lower mean accuracy. Thus, it can be inferred that if Accuracy is the key parameter for judging the performance of the classifiers for Alzheimer's disease, it can be done best with InceptionV3 or with Xception as the second choice.

TABLE 9. Accuracy of classifiers for alzheimer's disease detection.

Classifiers	Min	Max	Mean	SD
S-CNN	0.6546	0.7187	0.6912	0.0238
ResNet50	0.6671	0.6941	0.6810	0.0093
InceptionV3	0.8273	0.8656	0.8426	0.0131
Xception	0.8000	0.8296	0.8130	0.0104

Table 10 shows the precision scores of four classifiers for Alzheimer's disease dataset. In the case of mean precision, Xception performed comparatively better reporting the highest precision of 83.43%. Inception reports the precision value closer to Xception and has a lesser standard deviation than the Xception. ResNet50 reports the most inferior results, but with a standard deviation closer to InceptionV3. The precision value of S-CNN is better than ResNet50 with a minimum standard deviation among the remaining three classifiers. Thus, if precision is the key parameter, Xception proves to be a better classifier for the dataset with InceptionV3 as a second choice.

TABLE 10. Precision scores of classifiers for alzheimer's disease detection.

Classifiers	Min	Max	Mean	SD
S-CNN	0.6671	0.6941	0.6810	0.0093
ResNet50	0.6437	0.6936	0.6666	0.0167
InceptionV3	0.7873	0.8342	0.8128	0.0199
Xception	0.7967	0.8582	0.8343	0.0230

Table 11 shows the recall values of the classifiers for Alzheimer's disease dataset. The mean recall value of InceptionV3 is reported highest among the peers with the least standard deviation. The performance of S-CNN is the least among the group. Thus, if Recall is the key parameter, InceptionV3 classifier offers the best choice for an effective Alzheimer's disease classification.

TABLE 11. Recall values of classifiers for alzheimer's disease detection.

Classifiers	Min	Max	Mean	SD
S-CNN	0.5859	0.8000	0.7187	0.0787
ResNet50	0.6687	0.8640	0.7799	0.0762
InceptionV3	0.8640	0.9125	0.8912	0.0199
Xception	0.7187	0.8468	0.7840	0.0492

Table 12 shows the F1 scores of the classifiers for Alzheimer's disease dataset. The mean F1 score of S-CNN is again the least among the peers with the mean F1 score reported by InceptionV3 being the best in the category with Xception reporting a value close to it. Thus, the if F1 score is the key parameter, InceptionV3 serves as the potential classifier.

TABLE 12. F1 score of classifiers for alzheimer's disease detection.

Classifiers	Min	Max	Mean	SD
S-CNN	0.6291	0.7398	0.6973	0.0403
ResNet50	0.6809	0.7378	0.7164	0.0279
InceptionV3	0.8385	0.8716	0.8499	0.0116
Xception	0.7823	0.8325	0.8068	0.0173

Table 13 shows the AUC-ROC scores of the classifiers for Alzheimer's disease detection. The mean AUC-ROC score of InceptionV3 is observed to be the highest amongst all the classifiers followed by the results from Xception. S-CNN reports the most inferior performance among the remaining three classifiers. The results ascertain the performance of InceptionV3.

Table 14 presents the overall summary of all the measured metrics for Alzheimer's disease detection for the four classifiers viz. S-CNN, ResNet50, InceptionV3 and Xception. It has been observed that InceptionV3 offers the best overall performance for the performance metrics over the remaining three models considering Max (Min), Max (Max), Max (Mean), and Min (SD) suggesting to be the most suitable choice for the classification of Alzheimer's disease. The Xception model can be considered a reasonable next best choice. ResNet50 and S-CNN offer the most inferior performance for all the parameters and can be avoided.

TABLE 13. AUC–ROC score of classifiers for alzheimer's disease detection.

Classifiers	Min	Max	Mean	SD
S-CNN	0.6546	0.7187	0.6912	0.0238
ResNet50	0.6734	0.7203	0.6937	0.0153
InceptionV3	0.8273	0.8656	0.8426	0.0131
Xception	0.8000	0.8296	0.8130	0.0104

TABLE 14. Summary table of classifiers for alzheimer's disease detection.

Measures	Max (Min)	Max (Max)	Max (Mean)	Min (SD)
Accuracy	InceptionV3	InceptionV3	InceptionV3	ResNet50
Precision	Xception	Xception	Xception	S-CNN
Recall	InceptionV3	InceptionV3	InceptionV3	InceptionV3
F1 score	InceptionV3	InceptionV3	InceptionV3	InceptionV3
AUC-ROC	InceptionV3	InceptionV3	InceptionV3	InceptionV3

V. PRINCIPAL COMPONENT ANALYSIS (PCA)

PCA [24] is a dimensionality reduction method used to reduce the dimension of a large dataset which has been applied to the Brain Tumor and Alzheimer's disease dataset.

The PCA is used to reduce the dimension of the input pixels and constructs the principal components with only the most contributing image pixels while removing the redundant ones. Thus, PCA significantly reduces the input feature dimensions.

PCA transforms a larger set of features into a smaller set of features that contains maximum information from the larger dataset thus enabling the exploratory data analysis of the dataset. Smaller datasets are easier to explore, train, and visualize. The first step to apply PCA is to standardize the range of input features as if the difference is large between the range of feature variables, features with a large range dominate the smaller range variables. The second step is to compute the covariance matrix which tells whether the features are correlated or not. Highly correlated features contain redundant information. From the covariance matrix, the third step of PCA is to compute eigenvectors and eigenvalues to identify principal components. Putting computed eigenvectors into descending order allows for finding principal components according to their order of significance. The principal components are the new features constructed from the linear combinations of the initial features in such a way that they squeeze the maximum possible information from the initial features into the first component, then the maximum remaining information into the second component, and so on until the given number of principal components. Moreover, the first component explains the largest possible variance. Constructed principal components are uncorrelated. PCA helps to reduce the dimensionality without losing much information by removing components with low information [25]. The principal components do not carry real meaning and are less interpretable as they are the linear combinations of initial variables. The fourth step is to compute the feature vectors by discarding components of low eigenvalues that are of lesser significance and with the remaining feature vectors, a matrix of vectors is formed which is called a feature vector.

The standard CNN takes the entire image pixels as input which causes high model complexity. On the other hand, PCA-CNN takes the principal components as the input leading to the reduced model complexity. Fig. 10 depicts the architecture of the proposed PCA-CNN architecture. It has two major phases. In the first phase, the dimensionality reduction using PCA is performed. It takes the entire image as an input, and the number of principal components is then constructed. Further, the number of principal components is converted into the 2D matrix. In the second phase, the CNN model is trained for classification. CNN takes a 2D matrix as an input, and convolution operation is applied with 32 filters with size (3,3). Then, (1,1) pooling size is taken as it is unnecessary to reduce the features further because it has already been done using PCA in the beginning. The pooled features are flattened into a one-dimensional array and then passed to a fully connected layer with 16 nodes for the final classification.

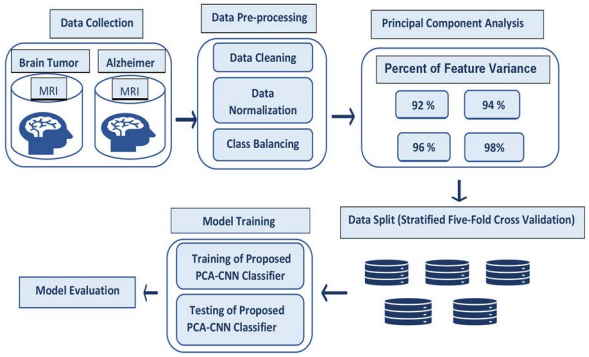

FIGURE 9. Proposed PCA–CNN framework.

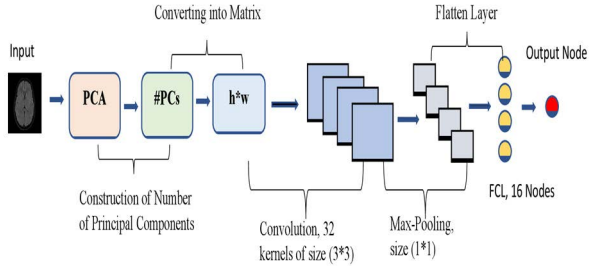

FIGURE 10. PCA-CNN architecture.

A. PCA IMPLEMENTATION

The framework for the proposed PCA-CNN has been presented in Fig. 9. The processing comprises six stages. In the

first stage, the data for both Brain Tumor and Alzheimer's disease is collected, which is preprocessed in the form of cleaning, normalization, and class balancing in the second stage. The third stage corresponds to the PCA analysis of the preprocessed data. The stage four corresponds to the validation of the data using Stratified Five-fold cross-validation which is presented to the PCA-CNN classifier for training and testing. Finally, the evaluation of the results takes place in stage six.

The initial dimension of the image is (100,100,1) viz. the number of features for each image is 10000. Experiments were performed to select the most uncorrelated number of features that will contain the maximum data variance. Only the number of principal components which had data variance of more than 90% was taken into consideration. Four different sets of principal components 400, 529, 784, and 1369 were implemented that contain data variance of 92%, 94%, 96%, and 98% respectively for Brain Tumor data. Similarly, for Alzheimer's data, the four different sets of principal components implemented were 441, 576, 784, and 1296 with data variance of 92%, 94%, 96%, and 98% respectively. Table 15 presents the reduced number of features, corresponding input shape, and percent of data variance for both datasets.

TABLE 15. Different sets of principal components.

	Brain Tumor	Alzheimer
Percent of Data Variance	Input shape / Number of Principal Components	Input shape / Number of Principal Components
92%	20x20, 400	21x21, 441
94%	23x23, 529	24x24, 576
96%	28x28, 784	28x28, 784
98%	37x37, 1369	36x36, 1296

Table 16 presents the hardware and software specifications for the PCA implementation setup. The parameters for PCA- CNN classifier for Both MRI datasets have been presented in Table 17.

VI. THE SIMULATION STUDY

This section contains all measured metric scores for both Brain Tumor and Alzheimer's disease classification using the PCA-CNN framework for the CNN classifier implemented in the work.

A. BRAIN TUMOR DATASET

Table 18 to table 22 presents the performance of PCA-CNN classifiers for various performance metrics as defined in section III (f). For the same, the work considered the computation to observe the mean values (Mean) and standard deviation (SD) for getting a comprehensive view of the results.

TABLE 16. Hardware and software specifications.

Software/Hardware Specifications	Version
CPU	Intel® Xeon® CPU @ 2.20GHz
RAM	12.68 GB
Python	3.7.13
Tensorflow	2.8.0
Keras	2.8.0
Scikit Learn	1.0.2
Pandas	1.3.5
Numpy	1.21.6
Matplotlib	3.2.2
Seaborn	0.11.2
Imblearn	0.8.1
Cv2	4.1.2

The accuracy measurement using the PCA-CNN classifier for the brain tumor dataset has been presented in Table 18. It is observed that a classifier trained with 400 principal components that contain 92% of data variance has a minimum accuracy of 98.03%. When the number of principal components is increased to study the effect of data variance of 94% and 96%, the accuracy of the classifier is again closer to each other. The highest accuracy was achieved with 1369 principal components that contain 98% of data variance.

Table 19 shows the precision scores of the PCA-CNN classifier for the Brain Tumor dataset. It can be seen that a classifier trained with 400 and 1369 principal components that contain 92% and 98% of data variance respectively have almost similar precision scores. When the number of principal components is 529 and 784 which have data variance of 94% and 96%, the precision scores of the classifier become lower, but still close to each other as 98.65% and 98.83% respectively. The highest precision score is achieved with 1369 principal components that contain 98% of data variance.

Table 20 presents the recall values of the PCA-CNN classifier for the Brain Tumor dataset. It is observed that a classifier trained with 400 principal components that contain 92% data variance offers the minimum recall value. When the number of principal components is 529 and 784 which have data variance of 94% and 96%, the recall values of the classifier are 97.67% and 98.66% respectively. The highest recall value is achieved again with 1369 principal components that contain 98% of data variance.

Table 21 presents the F1 score of the PCA-CNN classifier for the Brain Tumor dataset. It is observed that a classifier trained with 1369 principal components that contain 98% of

TABLE 17. Parameters for PCA-CNN classifier for both MRI datasets.

Brain Tumor Dataset		Alzheimer's Dataset	
Parameter	Value	Parameter	Value
Input shape	(20,20,1), (23,23,1), (28,28,1), (37,37,1)	Input shape	(21,21,1), (24,24,1), (28,28,1), (36,36,1)
Filters	32	Filters	32
Nodes	17	Nodes	17
Activation function	Relu, sigmoid	Activation function	Relu, sigmoid
Learning Rate	0.0001	Learning Rate	0.0001
Beta_1	0.9	Beta_1	0.9
epochs	30	epochs	30
Filter size	(3,3)	Filter size	(3,3)
Loss	Binary cross-entropy	Loss	Binary cross-entropy
Beta_2	0.999	Beta_2	0.999
Epsilon	1e-07	Epsilon	1e-07
Batch size	32	Batch size	32
Pooling size	(1,1)	Pooling size	(1,1)
Shuffle	True	Shuffle	True
Steps_per_epoch	None	Steps_per_epoch	None

TABLE 18. Accuracy.

Percent of Data Variance	Input shape / Number of Principal Components	Mean	SD
92%	20x20, 400	0.9803	0.0043
94%	23x23, 529	0.9817	0.0018
96%	28x28, 784	0.9875	0.0013
98%	37x37, 1369	0.9927	0.0030

the data variance has the highest F1 score. For the remaining three cases the F1 score is similar to each other.

Table 22 shows the ROC-AUC score of the PCA-CNN classifier for the Brain Tumor dataset. It can be seen that a classifier trained with 1369 principal components that contain 98% of the data variance has the highest ROC-AUC score. For the remaining three cases the ROC-AUC score is similar to each other.

TABLE 19. Precision.

Percent of Data Variance	Input shape / Number of Principal Components	Mean	SD
92%	20x20, 400	0.9912	0.0032
94%	23x23, 529	0.9865	0.0029
96%	28x28, 784	0.9883	0.0038
98%	37x37, 1369	0.9928	0.0048

TABLE 20. Recall.

Percent of Data Variance	Input shape / Number of Principal Components	Mean	SD
92%	20x20, 400	0.9692	0.0086
94%	23x23, 529	0.9767	0.0031
96%	28x28, 784	0.9866	0.0043
98%	37x37, 1369	0.9927	0.0022

TABLE 21. F1 score.

Percent of Data Variance	Input shape / Number of Principal Components	Mean	SD
92%	20x20, 400	0.9801	0.0044
94%	23x23, 529	0.9816	0.0019
96%	28x28, 784	0.9875	0.0013
98%	37x37, 1369	0.9927	0.0030

TABLE 22. ROC-AUC.

Percent of Data Variance	Input shape / Number of Principal Components	Mean	SD
92%	20x20, 400	0.9803	0.0043
94%	23x23, 529	0.9817	0.0018
96%	28x28, 784	0.9875	0.0013
98%	37x37, 1369	0.9927	0.0030

Table 23 shows the summary of all the measured metrics of the PCA-CNN classifier for the Brain Tumor dataset. It has

been observed that 1369 principal components with 98% data variance consistently reported the best performance for all the parameters. The minimum standard deviation varies for different metrics.

TABLE 23. Summary.

Measures	Max (Mean)	Min (SD)
Accuracy	98% Variance	96% Variance
Precision	98% Variance	94% Variance
Recall	98% Variance	98% Variance
F1 Score	98% Variance	96% Variance
AUC-ROC	98% Variance	96% Variance

B. ALZHEIMER'S DISEASE DATASET

This section presents all measured metric scores for Alzheimer's disease detection for all the proposed PCA-CNN classifiers implemented in the work. Table 24 to table 28 presents the performance of PCA-CNN classifiers for various performance metrics as defined in Section III(F). As mentioned earlier, the work considered the computation to observe the mean values (Mean) and standard deviation (SD) for getting a comprehensive view of the results.

Table 24 presents the accuracy of the PCA-CNN classifier for Alzheimer's disease dataset. It can be seen that a classifier trained with 441 principal components contains 92% of data variance and with 576 principal components containing 94% of data variance have similar accuracy. When the number of principal components is increased to have a data variance of 96%, the accuracy of the classifier improves to 97.54%. The highest accuracy is achieved with 1296 principal components that contain 98% of data variance.

TABLE 24. Accuracy.

Percent of Data Variance	Input shape / Principal Number of Components	Mean	SD
92%	21x21, 441	0.9682	0.0006
94%	24x24, 576	0.9694	0.0060
96%	28x28, 784	0.9754	0.0057
98%	36x36, 1296	0.9802	0.0027

Table 25 presents the precision scores of the PCA-CNN classifier for Alzheimer's disease dataset. It can be observed that a classifier trained with 441,784 and 1296 principal components that contain 92%, 96%, and 98% of data variance respectively have similar precision scores that do not differ significantly. When the number of principal components is 576 which has a data variance of 94%, the precision score of the classifier is the least. The highest precision score is achieved with 1296 principal components that contain 98% of data variance.

TABLE 25. Precision.

Percent of Data Variance	Input shape / Principal Number of Components	Mean	SD
92%	21x21, 441	0.9704	0.0039
94%	24x24, 576	0.9638	0.0146
96%	28x28, 784	0.9761	0.0035
98%	36x36, 1296	0.9778	0.0029

Table 26 presents the recall values of the PCA-CNN classifier for Alzheimer's disease dataset. The classifier trained with 441 principal components that contain 92% data variance has the minimum recall value. When the number of principal components is 576 and 784 which have data variance of 94% and 96%, the recall values of the classifier are closer to each other. The highest recall value is achieved with 1296 principal components that contain 98% of data variance.

TABLE 26. Recall.

Percent of Data Variance	Input shape / Principal Number of Components	Mean	SD
92%	21x21, 441	0.9658	0.0047
94%	24x24, 576	0.9759	0.0055
96%	28x28, 784	0.9746	0.0114
98%	36x36, 1296	0.9827	0.0035

Table 27 presents the F1 score of the PCA-CNN classifier for Alzheimer's dataset. The classifier trained with 1296 principal components that contain 98% of the data variance has the highest F1 score. Out of the remaining three cases, two with data variance of 92% and 94% the F1 score is similar to each other and the F1 score of one with a data variance of 96% is 97.53%.

Table 28 shows the ROC-AUC score of the PCA-CNN classifier for the Alzheimer's dataset. It can be seen that a classifier trained with 1296 principal components that contain 98% of the data variance has the highest ROC-AUC score. For the remaining three cases the ROC-AUC score is below 98%.

Table 29 shows the summary of all the measured metrics of the PCA-CNN classifier for the Alzheimer's dataset. It has been observed that 1296 principal components with 98% data

TABLE 27. F1 score.

Percent of Data Variance	Input shape / Principal Number of Components	Mean	SD
92%	21x21, 441	0.9682	0.0006
94%	24x24, 576	0.9697	0.0056
96%	28x28, 784	0.9753	0.0058
98%	36x36, 1296	0.9803	0.0027

TABLE 28. ROC-AUC.

Percent of Data Variance	Input shape / Principal Number of Components	Mean	SD
92%	21x21, 441	0.9682	0.0006
94%	24x24, 576	0.9694	0.0060
96%	28x28, 784	0.9754	0.0057
98%	36x36, 1296	0.9802	0.0027

variance consistently reported the best performance for all the parameters. Only the minimum standard deviation varies.

TABLE 29. Summary of performance for alzheimer's disease dataset.

Measures	Max (Mean)	Min (SD)
Accuracy	98% Variance	92% Variance
Precision	98% Variance	98% Variance
Recall	98% Variance	98% Variance
F1 Score	98% Variance	92% Variance
AUC-ROC	98% Variance	92% Variance

Table 30 shows the performance comparison between different classifiers for the Brain Tumor dataset. There is a significant difference between the number of input features used by the classifiers. The state-of-the-art pre-trained CNN networks have a large number of features. The number of features used in S-CNN is far lesser than in pre-trained CNN networks. The proposed PCA-CNN reported significantly improved results with a smaller number of features for all measured metrics. The Xception classifier has the best performance among S-CNN, ResNet50, and InceptionV3 for the Brain tumor dataset. However, PCA-CNN surpasses the performance of Xception on all counts and offers the best classification results.

Table 31 shows the performance comparison between different classifiers for Alzheimer's disease detection. There

TABLE 30. Comparison table for classifiers for brain tumor dataset.

Classifier	#Features	Input Dimension	Accuracy	F1 Score	AUC-ROC
S-CNN	49152	128x128 x3	0.97	0.9736	0.9735
ResNet50	150528	224x224 x3	0.8994	0.9003	0.8994
Inception V3	150528	224x224 x3	0.9820	0.9820	0.9820
Xception	150528	224x224 x3	0.9851	0.9850	0.9851
PCA-CNN	1369	37x37	0.9927	0.9927	0.9927

is a large difference between the number of input features of classifiers. The state-of-the-art pre-trained CNN networks have a large number of features. The number of features used in S-CNN is far lesser than in pre-trained CNN networks, but still larger than PCA-CNN. The proposed PCA-CNN classification again reported superior results with a significantly smaller number of features that consistently performed best for all measured metrics in comparison with the results obtained from peers. The performance of PCA-CNN is reported as around 13% better than the InceptionV3 which otherwise performed best among S-CNN, ResNet50, and Xception classifiers. Thus, PCA-CNN proves to be an efficient classification mechanism for Alzheimer's disease detection.

TABLE 31. Comparison table for classifiers for alzheimer's disease dataset.

Classifier	#Features	Input Dimension	Accuracy	F1 Score	AUC-ROC
S-CNN	49152	128x128x 3	0.6912	0.6973	0.6912
ResNet50	150528	224x224x 3	0.6810	0.7164	0.6937
InceptionV 3	150528	224x224x 3	0.8426	0.8499	0.8426
Xception	150528	224x224x 3	0.8130	0.8068	0.8130
PCA-CNN	1296	36x36	0.9802	0.9803	0.9802

C. DISCUSSION

The previously reported studies on similar Brain MRI datasets are presented in table 32. It has been observed that the evaluation metrics given in the previous works are not consistent as some models have been evaluated only using

the F1 score and other models using accuracy or classification rate. However, the proposed work has used accuracy, precision, recall, F1 score, and AUC-ROC score to evaluate the performance of the various CNN architectures. To classify Brain Tumors, [17] implemented ResNet50 and Xception on 253 Brain MRIs. The performance of the model is evaluated using only the F1 score. On the same dataset of 253 MRIs, [26] implemented ResNet50 and InceptionV3 while evaluating the model using accuracy, precision, recall, and F1 score. Reference [27] have implemented ResNet50 on 3064 Brain CE-MRIs using only accuracy. For the classification of Alzheimer's disease [28] implemented, ResNet50 and InceptionV3 models on the ADNI dataset. The models were evaluated only for accuracy. In [18], CNN has been implemented and evaluated while considering only the classification rate on the OASIS dataset. However, the proposed work implemented 3533 Brain MRIs for Brain Tumor classification and 6400 Brain MRIs for Alzheimer's disease classification over various CNN architectures. All five performance measures viz accuracy, precision, recall, F1 score, and AUC-ROC score were noted to be consistently better as compared to the peers. The superior performance scores of the PCA-CNN indicate the efficiency of the classification.

VII. FLOPS COMPUTATION FOR CNN CLASSIFIERS

The FLOPs are used to measure the computing complexity of the CNN models. The number of floating-point operations performed by the models, represent the computational requirements to perform a given task. Therefore, it is desired for any computational model to have superior performance while using a lesser number of operations. FLOPs are calculated using the Keras library called keras-flops (https://pypi.org/project/keras-flops).

The total number of parameters is calculated by using keras library called model summary which calculates the total number of trainable and non- trainable parameters. The detailed configuration about the models have been discussed in Section III(D).

The pre-trained models ResNet50, InceptionV3 and Xception are deep CNN architectures in which the layers are structured in a specific way. These three models are trained using transfer learning techniques. S-CNN is the basic CNN with three layers trained from scratch. The proposed PCA-CNN also has three layers trained from scratch, where the input features are the principal components. The number of parameters for the proposed PCA-CNN is significantly less than the remaining CNN models. Furthermore, there is a massive difference in the complexities of the model in terms of the total number of floating-point operations required.

Table 33 presents a comparison between various CNN architectures used in work in terms of FLOPs. The input shape for S-CNN is (128,128,3) and (224,224,3) for ResNet50, InceptionV3, and Xception. The input shape for PCA-CNN is (37,37,1) for Brain Tumor and (36,36,1) for Alzheimer's disease. The number of trainable parameters for S-CNN is larger than the pre-trained networks because in transfer learning, entire feature extraction layer is kept frozen as they do not learn new weights. The number of FLOPs for pre-trained networks are higher than the S-CNN even though

TABLE 32. Comparison of the performance of the proposed method with previous work.

Reference	Classification	Architecture	Performance
[17]	Brain Tumor	ResNet50 Xception	83.80% F1 score 97.82% F1 score
[18]	Alzheimer's Disease	CNN	88.10% Classification Rate
[26]	Brain Tumor	ResNet50	Accuracy 89%, precision 87%, recall 93%, and 90 % F1 score
		InceptionV3	Accuracy 75%, precision 77%, recall 71%, and 74 % F1 Score
[27]	Brain Tumor	ResNet50	97.65% Accuracy
[28]	Alzheimer's Disease	ResNet50 InceptionV3	70% Accuracy 85% Accuracy
Proposed Method	**Brain Tumor**	**PCA-CNN**	**Accuracy 99.27%, precision 99.27%, recall 99.27%, 99.27% F1 score and 99.27 AUC-ROC score**
	Alzheimer's Disease	**PCA-CNN**	**Accuracy 98.02%, precision 97.78%, recall 98.27%, 98.03 % F1 score, and 98.02% AUC-ROC score**

TABLE 33. Floating-point operations of CNN classifiers.

Architecture	No. of layers	No. of trainable parameters	Total no. of FLOPs	Total no. of GFLOPs
S-CNN	3	130,063,617	15,606,755,360	15.61
ResNet50	50	24,059,393	248,066,952,864	248.07
Inception V3	48	22,327,585	182,218,318,752	182.22
Xception	71	21,331,753	292,319,606,560	292.32
PCA-CNN (Brain Tumor)	**3**	**627,553**	**65,230,368**	**0.065230368**
PCA-CNN (Alzheimer)	**3**	**592,225**	**61,556,256**	**0.061556256**

they have the lower the number of trainable parameters as compared to S-CNN. This is because ResNet50, InceptionV3, and Xception are very deep and complex networks.

Furthermore, the number of trainable parameters for S-CNN, ResNet50, InceptionV3, and Xception is significantly lesser than the PCA-CNN. The Xception has the highest GFLOPs, followed by ResNet50. The reported GFLOPs for InceptionV3 are between ResNet50 and Xception. GFLOPs for S-CNN are comparatively far lesser than the three pre-trained CNNs. However, PCA-CNN reports a significantly lower number of GFLOPs than the remaining four CNNs owing to the fewer layers and the parameters.

VIII. CONCLUSION AND FUTURE REMARK

This work evaluates the performance of CNN models for the Brain Tumor dataset and Alzheimer's disease dataset with and without using PCA. Without considering the PCA-based analysis, it is observed that the performance of the Xception and InceptionV3 models do not significantly differ from each other for both the Brain Tumor dataset and Alzheimer's disease dataset. In particular, the Xception model reported higher scores for all five performance measures for the Brain Tumor dataset with very close results followed by InceptionV3. In the case of the Alzheimer's disease dataset, however, the InceptionV3 model reported the best performance scores on account of all five performance measures followed by Xception as the second-best candidate on all performance counts.

For the Brain Tumor data and the Alzheimer's disease dataset, the number of images getting correctly labelled were observed to be 66% and 50% respectively. It means that the Alzheimer's disease dataset has higher data complexity as compared to the Brain Tumor dataset by 16%. The highest performance score of accuracy for the Brain Tumor dataset is reported to be 99.27%. The results show that the highest accuracy for Alzheimer's dataset is 98.02%. The performances on both Brain MRI images are significantly different with the better performance achieved for the dataset with lesser complexity. Thus, we can say that the complexity of the data under investigation has an important role in the performance of the CNN models. The less complex data is yielding higher performance scores than highly complex data. This work can be extended to perform experiments for more than two MRI images of different diseases to get more generalized results.

REFERENCES

[1] H. Alsaif, R. Guesmi, B. M. Alshammari, T. Hamrouni, T. Guesmi, A. Alzamil, and L. Belguesmi, "A novel data augmentation-based brain tumor detection using convolutional neural network," *Appl. Sci.*, vol. 12, no. 8, p. 3773, Apr. 2022, doi: 10.3390/app12083773.

[2] T. Jo, K. Nho, and A. J. Saykin, "Deep learning in Alzheimer's disease: Diagnostic classification and prognostic prediction using neuroimaging data," *Frontiers Aging Neurosci.*, vol. 11, p. 220, Aug. 2019, doi: 10.3389/fnagi.2019.00220.

[3] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A comprehensive survey on transfer learning," *Proc. IEEE*, vol. 109, no. 1, pp. 43–76, Jan. 2021.

[4] V. H. Barella, L. P. F. Garcia, M. P. de Souto, A. C. Lorena, and A. de Carvalho, "Data complexity measures for imbalanced classification tasks," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2018, pp. 1–8, doi: 10.1109/IJCNN.2018.8489661.

[5] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. 25th Int. Conf. Neural Inf. Process. Syst. (NIPS)*, vol. 1, Dec. 2012, pp. 1097–1105.

[6] K. Simonyan and A. Zisserman. (2015). *Very Deep Convolutional Networks for Large-Scale Image Recognition*. [Online]. Available: http://www.robots.ox.ac.uk/

[7] S. Ahmed et al., "Brain tumor MRI medical images classification with data augmentation by transfer learning of VGG16," *J. Eng. Sci. Technol.*, pp. 21–32, Dec. 2021.

[8] K. Oh, Y.-C. Chung, K. W. Kim, W.-S. Kim, and I.-S. Oh, "Classification and visualization of Alzheimer's disease using volumetric convolutional neural network and transfer learning," *Sci. Rep.*, vol. 9, no. 1, Dec. 2019, doi: 10.1038/s41598-019-54548-6.

[9] O. J. Oyelade, O. O. Oladipupo, and I. C. Obagbuwa. (2010). *Application of K-Means Clustering Algorithm for Prediction of Students' Academic Performance*. [Online]. Available: http://sites.google.com/site/ijcsis/

[10] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," 2002, *arXiv:1106.1813*.

[11] X. Zeng and T. R. Martinez, "Distribution-balanced stratified cross-validation for accuracy estimation," *J. Experim. Theor. Artif. Intell.*, vol. 12, no. 1, pp. 1–12, 2000, doi: 10.1080/095281300146272.

[12] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, L. Wang, G. Wang, J. Cai, and T. Chen, "Recent advances in convolutional neural networks," 2015, *arXiv:1512.07108*.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778. [Online]. Available: http://image-net.org/challenges/LSVRC/2015/

[14] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 2818–2826, doi: 10.1109/CVPR.2016.308.

[15] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1251–1258.

[16] P. Kora, S. Mohammed, M. J. S. Teja, C. U. Kumari, K. Swaraja, and K. Meenakshi, "Brain tumor detection with transfer learning," in *Proc. 5th Int. Conf. I-SMAC (IoT Social, Mobile, Analytics, Cloud)*, 2021, pp. 443–446, 2021, doi: 10.1109/I-SMAC52330.2021.9640678.

[17] M. Arbane, R. Benlamri, Y. Brik, and M. Djerioui, "Transfer learning for automatic brain tumor classification using MRI images," in *Proc. 2nd Int. Workshop Hum.-Centric Smart Environ. Health Well-Being (IHSH)*, Feb. 2021, pp. 210–214, doi: 10.1109/IHSH51661.2021.9378739.

[18] M. Zaabi, N. Smaoui, H. Derbel, and W. Hariri, "Alzheimer's disease detection using convolutional neural networks and transfer learning-based methods," in *Proc. 17th Int. Multi-Conf. Syst., Signals, Devices (SSD)*, Jul. 2020, pp. 939–943, doi: 10.1109/SSD49366.2020.9364155.

[19] 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-sectional MRI data in young, middle aged, nondemented, and demented older adults," *J. Cognit. Neurosci.*, vol. 19, no. 9, pp. 1498–1507, Sep. 2007, doi: 10.1162/JOCN.2007.19.9.1498.

[20] M. Hamghalam, B. Lei, and T. Wang, "Convolutional 3D to 2D patch conversion for pixel-wise glioma segmentation in MRI scans," in *Proc. Int. MICCAI Brainlesion Workshop*, Oct. 2020, pp. 3–12, doi: 10.1007/978-3-030-46640-4_1.

[21] S. Na, L. Xumin, and G. Yong, "Research on K-means clustering algorithm: An improved K-means clustering algorithm," in *Proc. 3rd Int. Symp. Intell. Inf. Technol. Secur. Informat.*, Apr. 2010, pp. 63–67, doi: 10.1109/IITSI.2010.74.

[22] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9, doi: 10.1109/CVPR.2015.7298594.

[23] K. Hajian-Tilaki, "Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation," *Caspian J. Internal Med.*, vol. 4, no. 2, pp. 627–635, 2013.

[24] C. Syms, "Principal components analysis," in *Encyclopedia of Ecology*, vol. 3, 2nd ed., Jan. 2008, pp. 2940–2949, doi: 10.1016/B978-008045405-4.00538-3.

[25] S. B. Gaikwad and M. S. Joshi, "Brain tumor classification using principal component analysis and probabilistic neural network," *Int. J. Comput. Appl.*, vol. 120, no. 3, pp. 5–9, Jun. 2015, doi: 10.5120/21205-3885.
[26] H. A. Khan, W. Jue, M. Mushtaq, and M. U. Mushtaq, "Brain tumor classification in MRI image using convolutional neural network," *Math. Biosci. Eng.*, vol. 17, no. 5, pp. 6203–6216, Sep. 2020, doi: 10.3934/MBE.2020328
[27] R. Chelghoum et al., "Transfer learning using convolutional neural network architectures for brain tumor classification from MRI images," in *Proc. IFIP Int. Conf. Artif. Intell. Appl. Innov.*, vol. 583, 2020, pp. 189–200, doi: 10.1007/978-3-030-49161-1_17.
[28] M. T. Abed, U. Fatema, S. A. Nabil, M. A. Alam, and M. T. Reza, "Alzheimer's disease prediction using convolutional neural network models leveraging pre-existing architecture and transfer learning," in *Proc. Joint 9th Int. Conf. Informat., Electron. Vis. (ICIEV) 4th Int. Conf. Imag., Vis. Pattern Recognit. (icIVPR)*, Aug. 2020, pp. 1–6, doi: 10.1109/ICIEVicIVPR48672.2020.9306649.

ANIMA KUJUR received the B.Sc. degree in PCM (physics, chemistry, and mathematics), in 2013, the M.Sc. degree in applied mathematics, in 2016, the master's degree in computer application, in 2019. She is currently pursuing the Ph.D. degree in computer science and technology with the Jawaharlal Nehru University, New Delhi, India.

ZAHID RAZA received the master's degree in electronics and the master's and Ph.D. degrees in computer science. He was the Gold Medalist in master's degree in electronics.

He is currently working as an Associate Professor with the School of Computer and Systems Sciences, Jawaharlal Nehru University, India. Prior to joining Jawaharlal Nehru University, he served as a Lecturer at Banasthali Vidyapith University, Rajasthan, India. His research interests include parallel and distributed systems and machine learning. He has proposed various resource provisioning models for job scheduling for computational grid, cloud, the IoT, and parallel systems.

ARFAT AHMAD KHAN received the B.Eng. degree in electrical engineering from the University of Lahore, Pakistan, in 2013, the M.Eng. degree in electrical engineering from the Government College University Lahore, Pakistan, in 2015, and the Ph.D. degree in telecommunication and computer engineering from the Suranaree University of Technology, Thailand, in 2018.

From 2014 to 2016, he was a RF Engineer with Etisalat, UAE. From 2018 to 2022, he worked as a Lecturer and a Senior Researcher at the Suranaree University of Technology. He is currently working as a Senior Lecturer and a Researcher with Khon Kaen University, Thailand. His research interests include optimization and stochastic processes, channel and mathematical modeling, wireless sensor networks, ZigBee, green communications, massive MIMO, OFDM, wireless technologies, signal processing, and advance wireless communications.

CHITAPONG WECHTAISONG received the B.Eng. and M.Eng. degrees in telecommunication engineering from the Suranaree University of Technology, Thailand, in 2008 and 2014, and the Ph.D. degree in information and communication engineering from the Shibaura Institute of Technology, Japan, in 2016.

He is currently an Assistant Professor with the School of Telecommunication Engineering, Institute of Engineering, Suranaree University of Technology, Thailand. His research interests include wireless network design and optimization, network traffic localization, and global engineering education.