International Journal of Reconfigurable and Embedded Systems (IJRES)
Vol. 13, No. 1, March 2024, pp. 179∼191
ISSN: 2089-4864, DOI: 10.11591/ijres.v13.i1.pp179-191

Deep convolutional neural network framework with multi-modal fusion for Alzheimer's detection

Manoj Kumar Sharma[1], M. Shamim Kaiser[2], Kanad Ray[3]
[1]Amity School of Engineering and Technology, Amity University, Rajasthan, India
[2]Institute of Information Technology, Jahangirnagar University, Dhaka, Bangladesh
[3]Amity School of Applied Sciences, Amity University, Rajasthan, India

Article Info

Article history:

Received Dec 21, 2022
Revised Apr 28, 2023
Accepted May 8, 2023

Keywords:

Alzheimer disease
Convolutional neural network
Magnetic resonance imaging
Nature-inspired
Particle swarm optimization algorithm
Positron emission tomography

ABSTRACT

The biomedical profession has gained importance due to the rapid and accurate diagnosis of clinical patients using computer-aided diagnosis (CAD) tools. The diagnosis and treatment of Alzheimer's disease (AD) using complementary multimodalities can improve the quality of life and mental state of patients. In this study, we integrated a lightweight custom convolutional neural network (CNN) model and nature-inspired optimization techniques to enhance the performance, robustness, and stability of progress detection in AD. A multi-modal fusion database approach was implemented, including positron emission tomography (PET) and magnetic resonance imaging (MRI) datasets, to create a fused database. We compared the performance of custom and pre-trained deep learning models with and without optimization and found that employing nature-inspired algorithms like the particle swarm optimization algorithm (PSO) algorithm significantly improved system performance. The proposed methodology, which includes a fused multimodality database and optimization strategy, improved performance metrics such as training, validation, test accuracy, precision, and recall. Furthermore, PSO was found to improve the performance of pre-trained models by 3-5% and custom models by up to 22%. Combining different medical imaging modalities improved the overall model performance by 2-5%. In conclusion, a customized lightweight CNN model and nature-inspired optimization techniques can significantly enhance progress detection, leading to better biomedical research and patient care.

Corresponding Author:

Manoj Kumar Sharma
Amity School of Engineering and Technology, Amity University
SP-1, Kant Kalwar, RIICO Industrial Area, NH-11C, Jaipur, Rajasthan, India
Email: vmanojsharma@gmail.com

1. INTRODUCTION

Alzheimer's disease (AD) is a neurological brain condition that permanently damages the brain cells that are responsible for thinking and remembering. In the United States, AD affects around 5.7 million people, making it the sixth biggest cause of mortality, according to facts and figures from 2018 [1]. The datasets of magnetic resonance imaging (MRI) as modality-1 and positron emission tomography (PET) as modality-2 are used to diagnose AD. These modalities are combined to produce a dataset that is significantly more varied and trustworthy [2]. The data-fused dataset contributes to the robustness of the deep learning (DL) models. Although there are several categorization algorithms in use today, DL has captured the attention of all academics due to its adaptability and ability to generate the best results [3].

The study also found that the DL network was able to detect early signs of Alzheimer's in brain images with high accuracy. These findings suggest that DL networks could be used to develop early detection and prediction tools for AD [4]. Accuracy increased with a hybrid architecture built on transfer learning. High-level features like edges, patterns, and other features are easily recognized by pre-trained models [5]. The deep neural networks (DNN) based models' performance is highly influenced by their hyper-parameters [6].

The performance of DNNs is heavily dependent on the hyper-parameters, which are values set before the training process, including the learning rate, batch size, and number of layers. The optimal hyper-parameters can result in a more efficient and robust model, but it is much harder to identify the optimal values for the hyper-parameters than it seems. With the advancements in the DNN's architectures, the need for effective optimization algorithms to search for optimal values also increased and became more important than ever. Recently, nature-inspired optimization algorithms are one such algorithm that has helped researchers immensely in this regard [7]. Therefore, researchers dealing with extensive and intricate datasets in DL models have come to rely on nature-inspired optimization algorithms as a crucial tool. The rest of the paper is structured as follows; section 2 includes a review of recent and highly relevant literature; section 3 explains the suggested approach; section 4 lists the availability of the data and materials used; section 5 describes the results and discussion; and sections 6 discuss the conclusion, limitations, and future scope of the current work respectively.

2. LITERATURE REVIEW

Islam and Zhang [8] created an extremely deep convolutional model and displayed the outcomes on the open access series of imaging studies (OASIS) database. Brain MRI dataset is used to detect and classify AD (critical neurological brain disorder) through the very deep convolutional network (DCN). The proposed model is based on the pre-trained CNN model named Inception and the parameters i.e., weights were optimized using a gradient-based optimization algorithm named root-mean-squared-propagation (RMSProp).

Ghoniem [9] proposed a DL approach to diagnosing liver cancer. These are two key contributions of this method. Firstly, segNet is used to separate the liver from the abdominal scans, U-net model is used for lesion extraction, and artificial bee colony (ABC) optimization named $SegNet+UNet+ABC$ is used for the proposed novel hybrid segmentation technique to extract liver lesions. Secondly, a hybrid technique proposed named $LeNet + 5 + ABC$ is used to extract features and classify the liver lesions. The final result shows that the $SegNet + UNet + ABC$ technique is better compared to other techniques regarding convergence time, dice index, correlation_coefficient, and jaccard index. The leNet-5/ABC model performs better regarding computational time, F-1 score, accuracy, and specificity. Ismael *et al.* [10] proposed an enhanced approach of residual networks to classify brain tumor types. The proposed model is evaluated on a benchmark dataset having 3,064 MRI images of three brain tumor types (meningiomas, gliomas, and pituitary). On the same dataset, the proposed model's accuracy of 98% was the highest. Joo *et al.* [11] developed a DL method for automatic detection and localization of intracranial aneurysms and evaluation of the performance. A three-dimensional framework (ResNet) related to the DL algorithm is determined by the trained set. The results gave positive predictive, sensitivity, and specificity of 91.5%, 85.7%, and 98.0% for the external testing set and 92.8%, 87.1%, and 92.0% for the internal testing set, respectively.

Kim *et al.* [12] developed a computer-assisted detection scheme with the help of a convolutional neural network (CNN)-based model on an image of 3D digital-subtraction angiography for smaller-size aneurysm ruptures. A retrospective dataset comprising 368 subjects was utilized as a training cohort for CNNs with the TensorFlow platform. Six-direction aneurysm image of each patient is attained and region-of-interest is extracted from each image. Jnawali *et al.* [13] presented DNN-based to predict brain hemorrhage, based on the CT imagery data. The presented architecture's first three-dimensional CNN is used to extract features and detect brain_hemorrhage using logistic function as the last layer of the network. Finally, proposed three different 3D CNN algorithms to improve the performance of machine learning (ML) algorithms. Shi *et al.* [14] proposed a specific DL based method that has a good understanding of image quality and is validated with various architectures. Several experiments are conducted in cohorts, externally, and internally, in which it achieves an improved lesion in terms of enhancement and sensitivity on the subject level. Chen *et al.* [15] presented an artificial intelligence technology to improve the performance of the magnetic-induction-tomography (MIT) inverse problem. Four DL methods, including stacked autoencoders (SAE), deep belief networks (DBN), denoising autoencoders (DAE), and restricted boltzmann machines (RBM) are used to solve the nonlinear recreation problem of MIT, and then the results of the back-projection method and DL methods are compared. Solorio-Ramírez *et al.* [16] presented a new pattern identification algorithm based on the implementation of minimalist-machine-learning (MML) and a higher relevant attribute selection technique called dMeans. Afterward, to conduct the identification through CT brain images the proposed algorithm performance is examined and compared with k-nearest neighbors (KNN), multilayer perceptron (MLP), Naïve Bayes (NB), AdaBoost, random forests (RF), and support vector machine (SVM) classifiers. Phan *et al.* [17] presented a new method based on the DL algorithm and hounsfield unit system. The proposed method not only describes the level and duration of hemorrhage but also classifies the brain hemorrhagic region on the MRI image. To select the most suitable method for classification three neural network systems are compared and evaluated. Due to its importance in medical diagnostics, computer vision, and the internet of things, multimodal medical imaging has become a hot research area in the scientific community in recent years [18]-[20]. In order to detect AD progression based on the late fusion of MRI, demographics, neuropsychological, and apolipoprotein E4 (APOe4) genetic data, Spasov *et al.* [21] suggested a multimodal single-task classification model based on a CNN. Kumar *et al.* [22] integration of anatomical and functional modalities for the early identification of malignant tissue is one of the significant clinical applications of medical imaging fusion.

3. PROPOSED METHOD

In this section, the proposed method framework for AD detection was provided by the author. The multi-modal datasets were downloaded from the website and stored on the hard drive. These stored Alzheimer's databases were manually separated into two modalities MRI and PET scan on the basis of patients with and without AD. Then these images are pre-processed with format changing, image registration, segmentation, and resizing done through MATLAB code. After pre-processing, the fusion process was implemented and the fused data were stored in a MATLAB drive. Using this augmented datastore of fused images, the DCNN custom and pre-trained networks are trained, validated, and evaluated. To achieve the best outcomes, the nature-inspired particle swarm optimization (PSO) and Bayesian algorithm are used with custom and pre-trained models for hyper-parameter tuning. Results were eventually gathered and evaluated. The multi-modal data fusion process and optimization workflow of the system is shown in Figure 1 and can be observed from top to bottom. In this paper, the author has used two databases Alzheimer's disease neuroimaging initiative (ADNI) and Kaggle. The pre-processing was done on MRI and PET.dicom images that were converted into the .jpg format using the MATLAB program. Images that have been converted to JPEG format can be analyzed and stored more effectively, which raises diagnostic accuracy [23]. Relying less on specialized DICOM image viewer tools to see medical images [24]. The workflow of the suggested technique is briefly outlined. The MRI and PET images were initially pre-processed and converted to JPEG format before being used in the multi-modal data fusion technique. After that, nature-inspired optimization techniques and conventional optimization techniques were utilized to optimize the hyper-parameters. After that, the custom and pre-trained models were trained with and without optimized hyper-parameters on the fused datasets. Finally, these trained models were tested and the outcomes of each model can be compared and evaluated to determine the most effective approach. This workflow has the potential to improve the accuracy and reliability of ML models in medical imaging applications, allowing for more precise diagnoses and treatment planning.

MRI scans provide a detailed description of the brain, including gray and white matter, and PET scans measure levels of certain metabolites in the brain. Combining these two data sources provides a powerful tool for accurately diagnosing and predicting AD [25]. Then these two multi-modal images were undergone through the fusion process and the fused database was created. The use of multiple modalities in data collection helps to mitigate the impact of any inherent biases that may exist in a single modality. By merging different sources of data, a more holistic perspective of the subject matter can be attained, resulting in a more thorough comprehension of it. Figure 2 is a pictorial representation of the steps followed from data collection to categorizing the fused datasets in train and test folders for both ADNI and Kaggle fused datasets. An interactive and simple fusion process is implemented in MATLAB, as demonstrated in Figure 3. This graphical user interface (GUI) in MATLAB, which was made using the MATLAB app designer named data fusion, is used to achieve the fusion process. These fused datasets were utilized to train the pre-trained deep convolutional neural networks (DCNN) like custom CNN, AlexNet, MobileNetV2, and GoogLeNet using a DL toolbox in MATLAB. Additionally, the use of a GUI for data fusion can help reduce the time and effort required for data preprocessing, enabling more efficient experimentation and analysis of multi-modal data.

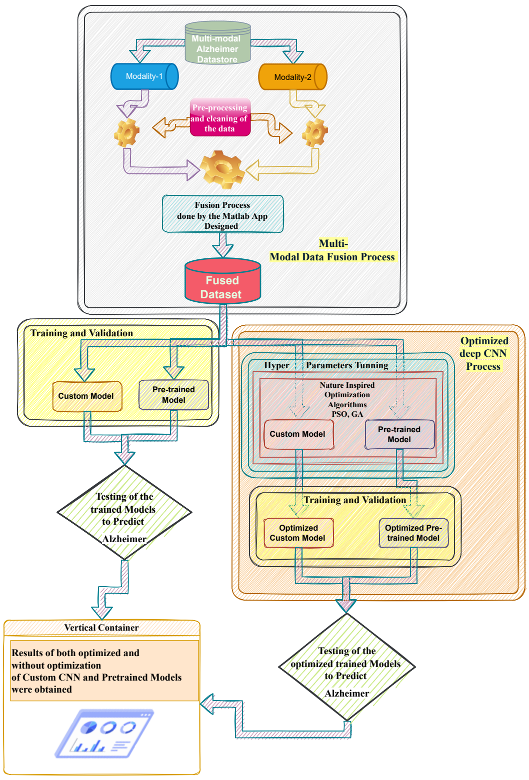


Figure 1. The multi-modal data fusion process and nature-inspired hyper-parameters optimization workflow of our proposed framework for diagnosing AD

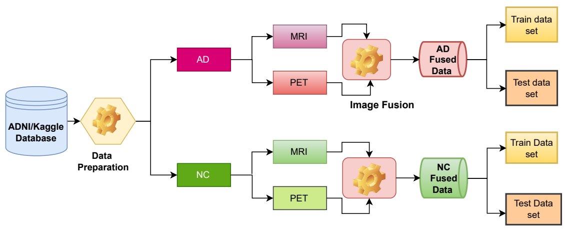


Figure 2. Steps to achieve multi-modal fusion with ADNI and Kaggle databases

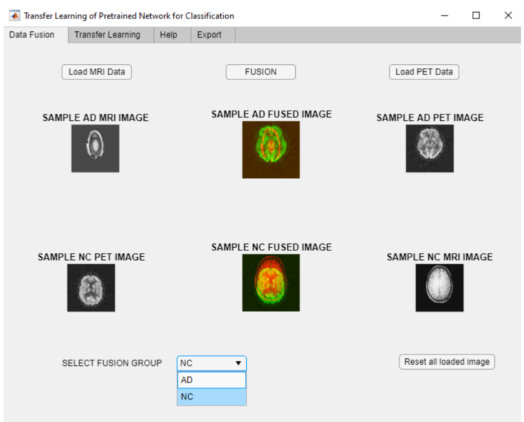


Figure 3. A fusion MATLAB app interface is shown to achieve multi-modal fusion with ADNI and Kaggle databases

The procedures for training and testing the optimized pre-trained models are shown in Figure 4. Before training the pre-trained CNN models, they are optimized using a nature-inspired algorithm. Then a test dataset was used to test whether the trained model was performing well or not. If not, then further iterations were required to optimize the hyper-parameters of the selected DCNN network.

The concept of the transfer learning approach is illustrated in Figure 5. In transfer learning, pre-trained weights are transferred to predict a new, similar task with some changes in the last layers. This was accomplished with MATLAB to speed up the training and testing process using transfer learning.

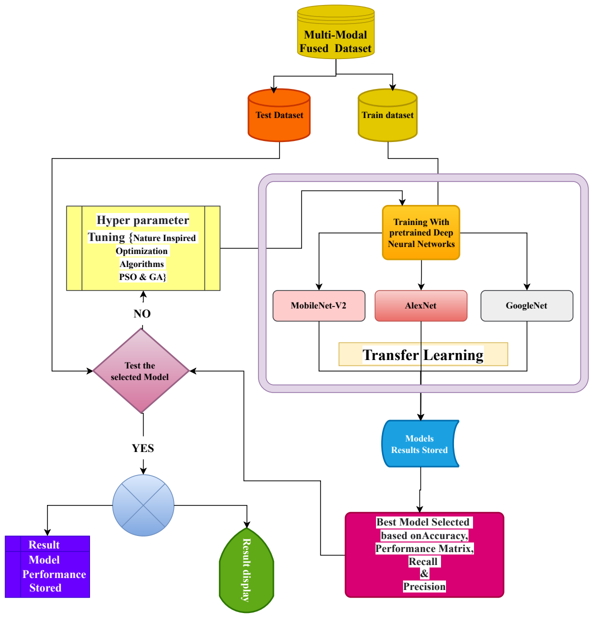


Figure 4. The transfer learning approach used on to the DCNN with hyper-parameter optimization techniques with multi-modal fused datasets

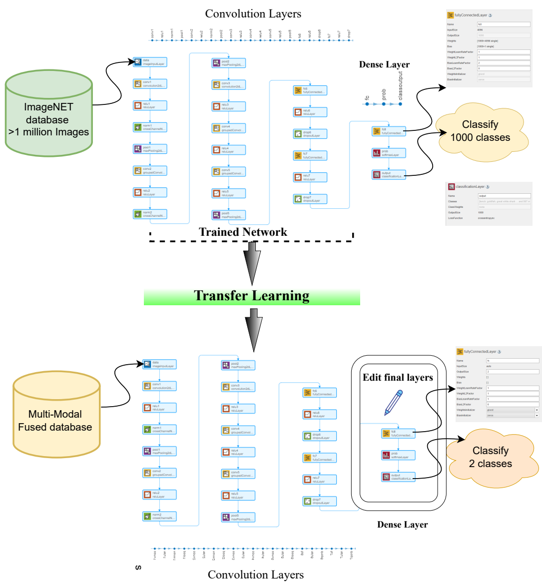


Figure 5. The transfer learning approach used to increase the performance of selected DCNN

3.1. Pre-trained models architectures

3.1.1. AlexNet

AlexNet was the winner of ILSRVRC'2012 challenge [26]. It has an 8-layer deep architecture, which consists of five convolutional layers and three max pooling layers after the first, second, and fifth layers respectively, and ReLu is used as an activation function. The max pooling layers are overlapped with strides 2 and a filter size of 3×3 to reduce the error. These layers are followed by two dense layers with softmax to perform the predictions. The AlexNet architecture has been used for image classification, scene recognition, and object detection [27]-[29].

3.1.2. GoogLeNet

ILSRVRC'2014 was won by Google architecture, which had fewer errors than the runner-up VGG, and the previous winner AlexNet. The architecture of GoogLeNet consists of 22 layers [30]. The architecture is a combination of 1×1 convolutional layers, an inception module, global average pooling layers, and auxiliary classifiers. The concept of 1×1 was used to minimize the parameters, i.e., weights and biases, to lower the computational cost with a much deeper network. The inception module consists of different sizes of CNN layers, i.e., 1×1, 3×3, and 5×5, and a max pooling layer of size 3×3, working in parallel to extract deep features from the objects of different sizes on a larger scale. The auxiliary classifiers are used by the inception architecture to calculate the loss at different stages during the training and add them to the final loss with weights valued at 0.3 to generate the overall loss. The auxiliary classifiers assist in overcoming the gradient vanishing problem and in regularization. Google has been widely used in object detection and face recognition [31], [32].

3.1.3. MobileNetV2

MobileNetV2 is also known as the "lightweight" model, which has a comparatively much lower complexity cost that makes it suitable for mobile devices. The architecture consists of depth-wise convolution and point-wise convolution. In the depth-wise convolution, a single convolutional filter is applied to each input signal to perform lightweight filtering, whereas, in the point-wise convolution, 1×1 convolution-based is performed to extract deep features by computing linear combinations between the input channels. Table 1 summarizes and the pre-trained CNN architectures are compared. That was already utilized for object detection and recognition across vast numbers of classes [33].

Table 1. Comparison of the pre-trained CNN architectures

Pre-trained CNN models	Depth	Size (MB)	Parameters (Million)	Input size	layers
AlexNet	8	227	61	227×227×3	25
GoogLeNet	22	27	7	224×224×3	144
ResNet-18	18	44	11.7	224×224×3	71
MobileNetV2	53	13	3.5	224×224×3	154

3.2. Custom convolutional neural network model

A traditional CNN model consists of convolutional layers followed by pooling layers to extract the deep features. The multi-dimensional features are then flattened into 1-dimensional features, followed by fully connected layers to perform classification. A block diagram for a typical CNN is shown in Figure 6. In this paper, the customized CNN model consists of three convolutional layers, with a max-pooling layer coming after each. The initial CNN layer contains 32 kernels of 5×5 size with an l2 regularizer; the subsequent layers contain 8 filters of the same size. The sigmoid is used as an activation function in each layer.

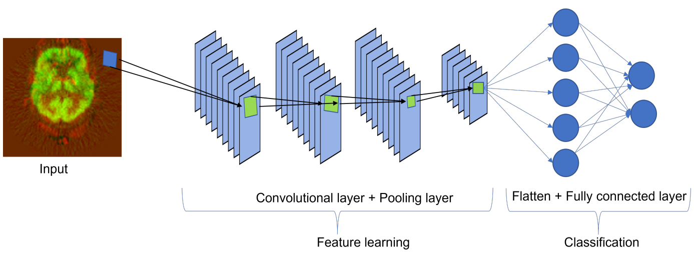


Figure 6. CNN block diagram

3.3. Optimization algorithms

Optimization algorithms are an essential component of DL model development. They help to search for the optimal values for the hyper-parameters of the DNNs, such as l1 regularization, l2 regularization, learning rate, and the number of filters. The choice of the optimization algorithm can significantly impact the model's performance, and researchers have developed various algorithms to optimize the hyper-parameters. In this study, we will focus on two popular optimization algorithms: PSO and the Bayesian optimization algorithm.

3.3.1. Particle swarm optimization

PSO is inspired by the movement of a flock of birds or a group of fish, where all of the individuals can benefit from the discovery of one of the fish or birds. PSO doesn't require a gradient, unlike other statistical optimization algorithms, which means the differentials are also not needed, which makes it simple and computationally cheap. In the PSO algorithm, a position vector of a $i^t h$ particle at iteration *t* i.e., $X^i(t)$=$(x^i(t), y^i(t))$, which has the coordinates and the velocity of each particle i.e., $V^i(t)$=$(v^i_x(t), v^i_y(t))$ are used to locate and update the position of particle after each iteration as shown in (1) and (2), until the optimal value, is not achieved or the global minimum of some function $f(x, y)$ is not found. The pseudocode for the algorithm is shown in Figure 7.

$$X^i(t+1)=X^i(t) + V^i(t+1) \quad (1)$$

$$V^i(t+1)=wV^i(t) + c_1r_1(pbest^i - X^i(t)) + c_2r_2(gbest^i - X^i(t)) \quad (2)$$

In 2, the $pbest^i$ and $gbest^i$ the ideal nearby location discovered by a $i^t h$ particle and global best position by all the particles in the swarm.

```
for each particle do
    Initialize particle
end for
Do
for each particle do
    Calculate fitness value
    if the fitness value is better than the best fitness value (pBest) in history then
        set current value as the new pBest
    end if
    Choose the particle with the best fitness value of all the particles as the gBest
    for each particle do
        Calculate particle velocity according to equation    (2)
        Update particle position according to equation    (1)
    end for
    While maximum iterations or minimum error criteria is not attained
```

Figure 7. PSO pseudocode

3.3.2. Bayesian optimization

Bayesian optimization works on Bayes theorem as in (3) to direct search for the optimal solutions. This algorithm uses the acquisition function, i.e., expected improvement, to select a sample from the space and the objective function, i.e., Gaussian process regression, to compute the cost or root mean squared error (RMSE). After the cost calculation, the data is updated, and the process is repeated until the global maximum is not reached.

$$P(A/B)=\frac{P(B/A) * P(A)}{P(B)} \quad (3)$$

After the simplification, by removing the normalizing factor i.e., *P(B)*, to make it a proportional quantity, and also the object is to optimize the quantity, not to calculate the individual probability, the (3) becomes (4).

$$P(A/B)=P(B/A) * P(A) \quad (4)$$

3.3.3. Hyper-parameters optimization

Hyper-parameters are essential factors that influence the performance of DL models. To achieve optimal values for these hyper-parameters, optimization algorithms are applied. This study employs two optimization algorithms, PSO, and Bayesian optimization algorithm, for the purpose. Two types of models, custom models, and pre-trained models were tested and trained. For a CNN, the initial values, and acceptable range of values for the hyper-parameters depend on the specific network architecture, data, and task.

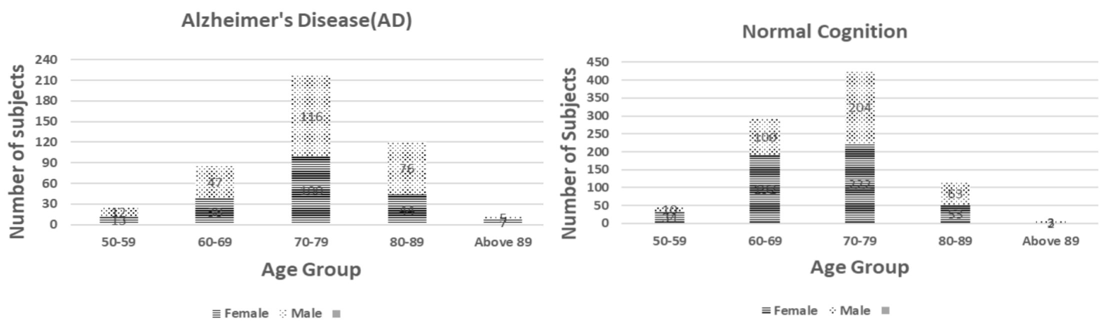
The study conducted fine-tuning experiments on both custom models and pre-trained models, using various hyper-parameters. For the custom models, the hyper-parameters included the number of convolutional layers, the learning rate, the number of kernels or filters, and the L2 regularization parameter. The initial range for these hyper-parameters was as follows: i) the number of convolutional layers ranged from 1 to 8, ii) the learning rate ranged from $1e^{-}2$ to 1, iii) the number of kernels or filters ranged from 1 to 32, and iv) the L2 regularization parameter ranged from $1e^{-}10$ to $1e^{-}2$.

On the other hand, for the pre-trained models (GoogLeNet, MobileNetV2, and AlexNet), the hyper-parameters used for fine-tuning were the number of filters and convolutional layers, the learning rate, and the L2 regularization parameter. Specifically, the number of filters and convolutional layers were determined based on Table 1, while: i) the learning rate ranged from $1e^{-}2$ to 1, and ii) the L2 regularization parameter ranged from $1e^{-}10$ to $1e^{-}2$. After applying PSO optimization to the obtained optimal hyper-parameter values with 0.922 initial learning rate, 1.0779 convolutional layers, and 0.0035 L2 regularisation, the model produced the best results. In terms of computational time, the performance of PSO and Bayesian optimization was compared on several test functions. While results varied across the experiments, it was generally observed that PSO converged to the global optimum in an average of 10-20 iterations, while Bayesian optimization required around 50-100 iterations.

4. ACCESS TO DATA AND MATERIALS

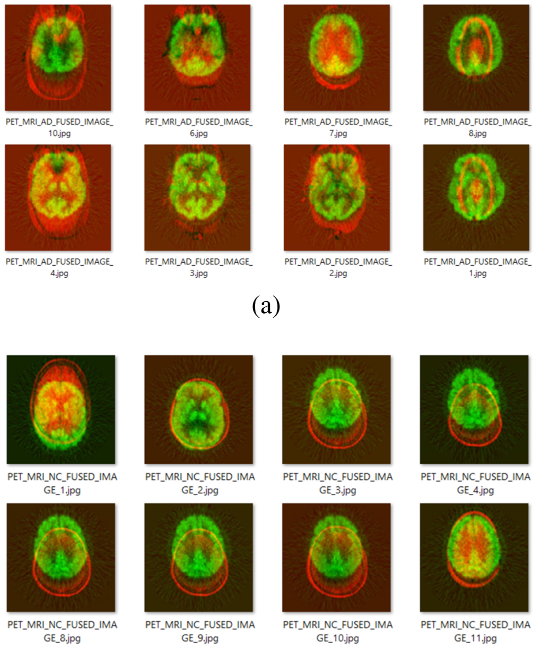
Our study developed an optimized DCNN with a multi-modal fusion approach for detecting AD, using two datasets: the ADNI dataset [34] and the Alzheimer's dataset on Kaggle [35]. Figure 8 illustrates the age distribution of participants based on gender and group. Specifically, Figure 8(a) shows the age distribution of male and female participants with AD, while Figure 8(b) shows the age distribution of male and female participants in the normal control (NC) group.

The Kaggle database is made up of training and testing folders with around 5,000+ photos each, which are divided into four classes according to the severity of Alzheimer's: i) MildDemented, ii) VeryMildDemented, iii) NonDemented, and iv) ModerateDemeneted. Except for NonDemented, which was maintained in a different folder for NC categories, all the images were kept in one folder for AD categories. The PET scans of ADNI AD and NC were then combined with these two distinct Kaggle datasets. These fused databases were stored for later processing in the test and train folders. The ADNI dataset was split and processed into distinct databases, and multi-modal fusion pre-processing was performed on the ADNI dataset to create fused databases for AD and NC categories. Figure 9 shows montages of samples taken from these databases, with Figure 9(a) displaying samples from the AD category and Figure 9(b) displaying samples from the NC category.



(a) (b)

Figure 8. Distribution of participants based on gender and age (a) with AD and (b) in the NC group



(a)

(b)

Figure 9. Montages of MRI and PET fused images for both AD and NC categories sample of (a) PET_MRI_AD_FUSED_IMG image and (b) PET_MRI_NC_FUSED_IMG image

5. RESULTS AND DISCUSSION

In this section, the obtained results from numerous and divergent experimentation's are discussed. As the pre-trained models have large architectures and some layers need to be frozen down in order to have fast and effective training. And, also the pre-trained weights to be loaded as the retraining of the model having complicated architecture on larger datasets like ImageNet comprising millions of images with 1,000 different classes, requires a lot of computation. Thus it is not suitable for smaller datasets to use larger architectures, that's why pre-trained weights are loaded, and then transfer learning is performed to make the architecture suitable for a custom dataset. So, initially three famous pre-trained architectures i.e., AlexNet, GoogLeNet, and MobileNetV2 are trained and tested on both ADNI and Kaggle datasets and results are reported. Secondly, a custom model, with comparatively fewer complications, is also trained on a custom dataset and the obtained results are documented. Thirdly, optimization algorithms i.e. Bayesian, PSO, and GA are applied to custom models to optimize their hyper-parameters. In general, it is observed that the optimization algorithm results in improvement from 2 to 7%. Secondly, in the case of the custom model, which is at least 4 to 6 times lighter than pre-trained models, over 20% of improvement was observed i.e., test accuracy of 67% was improved to 91.02%, which was higher than AlexNet and MobileNetV2 by over 3 to 5%, as illustrated in Table 2. Table 3 is giving performance metrics results on ADNI fused dataset.

Similarly, considering the datasets, it is observed that the fused dataset of MRI and PET results in an improvement of 2 to 5% as shown in Table 4. According to Shanmugam *et al*. [36], GoogLeNet, AlexNet, and ResNet-18 have achieved 96.39%, 94.08%, and 97.51% accuracy in detecting AD using Uni-Modal (MRI) images. The multi-modal fusion-based approach using GoogLeNet and AlexNet improves the results by 0.92% and 5.6% respectively.

Stochastic gradient descent with momentum (SGDM) is more suitable for this problem than adaptive moment estimation (Adam), as shown in Table 5. The ADNI fused dataset resulted in an average increase of 3% accuracy in all four pre-trained models. Using the PSO optimization algorithm with the ADNI and Kaggle fused datasets improved results by over 23% and 16%, respectively Table 6. The GoogLeNet and AlexNet multi-modal fusion approach improved results by 0.92% and 5.6%, respectively.

The performance comparison of custom and pre-trained DL models on the ADNI fusion dataset is shown in Figure 10. Figures 10(a) to 10(d) respectively depict the performance of the custom model, GoogLeNet model, MobileNetV2 model, and AlexNet model. The use of the PSO optimization algorithm improves the performance of all four models, as observed in the figures. Specifically, the PSO algorithm improves the performance of the custom model in Figure 10(a), GoogLeNet model in Figure 10(b), MobileNetV2 model in Figure 10(c), and AlexNet model in Figure 10(d).

Table 2. Obtained results on ADNI fused dataset

Parameter	Model name	Training accuracy (%)	Validation accuracy (%)	Test accuracy (%)	Optimizer	Optimization algorithm
Before optimization	GoogLeNet	97.19	95.61	96.09	SGDM	
		96.33	95.6	92.96	Adam	
	AlexNet	**92**	**90.12**	**90.23**	Adam	
		99.87	97.56	97.26	SGDM	-
	MobileNetV2	62.64	60	62.89	Adam	
	Custom model	77.12	59	67	Adam	
		70	67	67.45	SGDM	
After optimization	GoogLeNet	97.77	97.98	96.88	Adam	Bayesian
	AlexNet	**100**	**100**	**68.7912**	Adam	PSO
	MobileNetV2	65.62	62	63.12		PSO
	Custom model	93.28	89.76	91.02	Adam	PSO
		72	65.37	65.51	Adam	Bayesian

Table 3. Performance metrics results on ADNI fused dataset

Parameter	Model name	Precision	Recall	Optimizer	Optimization algorithm
Before optimization	GoogLeNet	0.9775	0.9158	SGDM	
		1	0.8317	Adam	
	AlexNet	**0.9325**	**0.9111**	Adam	
		0.9213	1	SGDM	-
	MobileNetV2	0.0112	0.125	Adam	
	Custom model	0.55	0.62	Adam	
		0.64	0.66	SGDM	
After optimization	GoogLeNet	-	-	-	-
	AlexNet	-	-	Adam	PSO
	MobileNetV2	1	0.3435		PSO
	Custom model	0.89	0.92	Adam	PSO
		0.71	0.73	Adam	Bayesian

Table 4. Comparative results of using uni-modal and multi-modal datasets in diagnosing Alzheimer

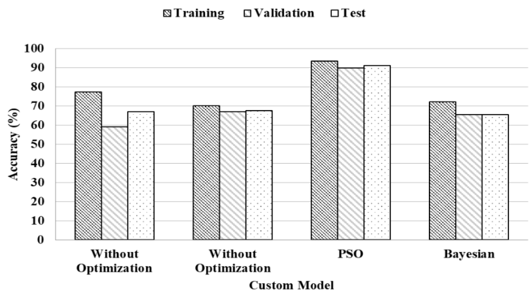
DCNN models	Uni-modal (MRI) average accuracy (%)	Multi-modal fused (MRI+PET) average accuracy (%)
GoogLeNet	96.3	97.19
AlexNet	94.39	99.98
ResNet-18	97.51	75.4
MobileNetV2	-	61.84
Custom model	-	68.15

Table 5. Results on the basis of optimizer on multi-modal ADNI fused dataset

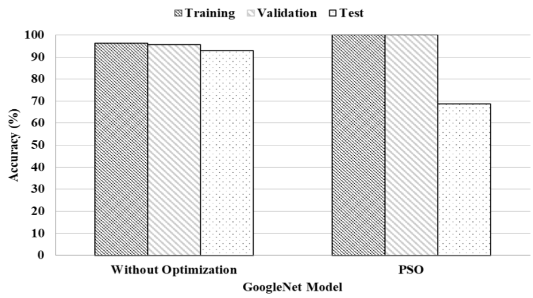
DCNN models	SGDM average accuracy (%)	Adam average accuracy (%)
GoogLeNet	96.296	94.96
AlexNet	98.23	90.78
MobileNetV2	62.34	61.84
Custom model	68.15	67.706

Table 6. PSO optimization on the custom model compared with ADNI and Kaggle fused datasets

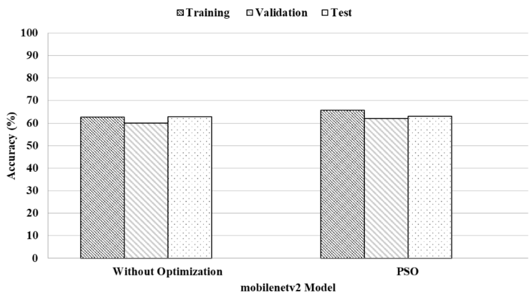
Model name	ADNI fused Average accuracy (%)	Kaggle fused Average accuracy (%)	Optimization algorithm
Custom model	68.15	67.706	Without PSO
	91.35	83.77	With PSO



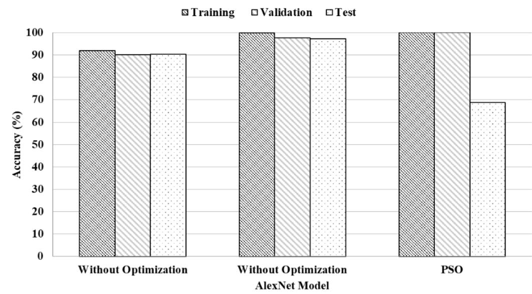
(a)



(b)



(c)



(d)

Figure 10. A comparison of the performance of custom and pre-trained deep learning models on the ADNI fusion dataset with and without the use of the optimization algorithm (a) the custom model, (b) the GoogLeNet model, (c) the mobilenetv2 model, and (d) the AlexNet model

6. CONCLUSIONS

In this article, optimized DL models based on an automatic computer-aided AD detection approach are proposed. The different pre-trained models, including AlexNet, GooLeNet, MobileNetV2, and a custom model, are assessed using the ADNI and Kaggle datasets. Two optimization algorithms, Bayesian and PSO are used to optimize the hyper-parameters of the models and the results before and after the optimization are reported. The performance is evaluated in terms of training accuracy, testing accuracy, validation accuracy, precision, and recall. It is found that the nature-inspired optimization algorithm i.e., PSO provides better results on some of the pre-trained models. But when the PSO is applied to the very light custom model can outperform in comparison to larger pre-trained architectures. This shows that for mobile application development, lighter customized models should be utilized. The PSO and Bayesian are found to have improved the results by 15% on average i.e., 2 to 5% in the case of pre-trained models and up to 22% for a custom model. Similarly, the fused dataset of PET and MRI also contributed to the improvement of overall performance by up to 5%.

ACKNOWLEDGEMENTS

We also acknowledge the support of the ADNI data and other open-source databases for approving the access request. We are assisted in our research effort by software and relevant MATLAB webinars. We are grateful to MIT, a private research university in Cambridge, Massachusetts, for offering some of the research linked to our work that is open-access.

REFERENCES

[1] Alzheimer's Association, "2019 Alzheimer's disease facts and figures," *Alzheimer's & Dementia*, vol. 15, no. 3, pp. 321–387, Mar. 2019, doi: 10.1016/j.jalz.2019.01.010.

[2] C. Catana, A. R. Guimaraes, and B. R. Rosen, "PET and MR imaging: the odd couple or a match made in heaven?," *Journal of Nuclear Medicine*, vol. 54, no. 5, pp. 815–824, May 2013, doi: 10.2967/jnumed.112.112771.

[3] D. Prakash, N. Madusanka, S. Bhattacharjee, H.-G. Park, C.-H. Kim, and H.-K. Choi, "A comparative study of Alzheimer's disease classification using multiple transfer learning models," *Journal of Multimedia Information System*, vol. 6, no. 4, pp. 209–216, Dec. 2019, doi: 10.33851/JMIS.2019.6.4.209

[4] A. Ebrahimi, S. Luo, and R. Chiong, "Introducing transfer learning to 3D ResNet-18 for Alzheimer's disease detection on MRI images," in *2020 35th International Conference on Image and Vision Computing New Zealand (IVCNZ)*, Nov. 2020, pp. 1–6, doi: 10.1109/IVCNZ51579.2020.9290616.

[5] S. R. Chandaran, G. Muthusamy, L. R. Sevalaiappan, and N. Senthilkumaran, "Deep learning-based transfer learning model in diagnosis of diseases with brain magnetic resonance imaging," *Acta Polytechnica Hungarica*, vol. 19, no. 5, pp. 127–147, 2022, doi: 10.12700/APH.19.5.2022.5.7.

[6] H. Bertrand, R. Ardon, M. Perrot, and I. Bloch, "Hyperparameter optimization of deep neural networks: combining hyperband with Bayesian model selection state of the art," in *Conference sur l'Apprentissage Automatique*, 2017, pp. 1–5.

[7] X.-S. Yang, N*ature-inspired optimization algorithms*, 2nd ed. Academic Press, 2020.

[8] J. Islam and Y. Zhang, "A novel deep learning based multi-class classification method for Alzheimer's disease detection using brain MRI data," in *Lecture Notes in Computer Science()*, Cham: Springer, 2017, pp. 213–222, doi: 10.1007/978-3-319-70772-3_20.

[9] R. M. Ghoniem, "A novel bio-inspired deep learning approach for liver cancer diagnosis," *Information*, vol. 11, no. 2, p. 1-36, Jan. 2020, doi: 10.3390/info11020080.

[10] S. A. A. Ismael, A. Mohammed, and H. Hefny, "An enhanced deep learning approach for brain cancer MRI images classification using residual networks," *Artificial Intelligence in Medicine*, vol. 102, Jan. 2020, doi: 10.1016/j.artmed.2019.101779.

[11] B. Joo *et al*., "A deep learning algorithm may automate intracranial aneurysm detection on MR angiography with high diagnostic performance," *European Radiology*, vol. 30, no. 11, pp. 5785–5793, Nov. 2020, doi: 10.1007/s00330-020-06966-8.

[12] H. C. Kim *et al*., "Machine learning application for rupture risk assessment in small-sized intracranial aneurysm," *Journal of Clinical Medicine*, vol. 8, no. 5, pp. 1-10, May 2019, doi: 10.3390/jcm8050683.

[13] K. Jnawali, M. R. Arbabshirani, N. Rao, and A. A. Patel, "Deep 3D convolution neural network for CT brain hemorrhage classification," in *Medical Imaging 2018: Computer-Aided Diagnosis*, Feb. 2018, doi: 10.1117/12.2293725.

[14] Z. Shi *et al*., "A clinically applicable deep-learning model for detecting intracranial aneurysm in computed tomography angiography images," *Nature Communications*, vol. 11, no. 1, pp. 1-11, Nov. 2020, doi: 10.1038/s41467-020-19527-w.

[15] R. Chen, J. Huang, Y. Song, B. Li, J. Wang, and H. Wang, "Deep learning algorithms for brain disease detection with magnetic induction tomography," *Medical Physics*, vol. 48, no. 2, pp. 745–759, Feb. 2021, doi: 10.1002/mp.14558.

[16] J.-L. Solorio-Ramírez, M. Saldana-Perez, M. D. Lytras, M.-A. Moreno-Ibarra, and C. Yáñez-Márquez, "Brain hemorrhage classification in CT scan images using minimalist machine learning," *Diagnostics*, vol. 11, no. 8, Aug. 2021, doi: 10.3390/diagnostics11081449.

[17] A.-C. Phan, H.-P. Cao, T.-N. Trieu, and T.-C. Phan, "Detection and classification of brain hemorrhage using hounsfield unit and deep learning techniques," in Future Data and Security Engineering. Big Data, Security and Privacy, Smart City and Industry 4.0 Applications. FDSE 2020, Singapore: Springer, 2020, pp. 281–293, doi: 10.1007/978-981-33-4370-2_20.

[18] S. Bebortta *et al*., "Evidence of power-law behavior in cognitive IoT applications," *Neural Computing and Applications*, vol. 32, no. 20, pp. 16043–16055, Oct. 2020, doi: 10.1007/s00521-020-04705-0.

[19] W. Tan, P. Tiwari, H. M. Pandey, C. Moreira, and A. K. Jaiswal, "Multimodal medical image fusion algorithm in the era of big data," *Neural Computing and Applications*, Jul. 2020, doi: 10.1007/s00521-020-05173-2.

[20] F. Amato, V. Moscato, A. Picariello, F. Piccialli, and G. Sperlí, "Centrality in heterogeneous social networks for lurkers detection: An approach based on hypergraphs," *Concurrency and Computation: Practice and Experience*, vol. 30, no. 3, Feb. 2018, doi: 10.1002/cpe.4188.

[21] S. Spasov, L. Passamonti, A. Duggento, P. Liò, and N. Toschi, "A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to Alzheimer's disease," *NeuroImage*, vol. 189, pp. 276–287, Apr. 2019, doi: 10.1016/j.neuroimage.2019.01.031.

[22] N. Kumar, N. Hoffmann, M. Oelschlägel, E. Koch, M. Kirsch, and S. Gumhold, "Structural similarity based anatomical and functional brain imaging fusion," in *Multimodal Brain Image Analysis and Mathematical Foundations of Computational Anatomy. MBIA MFCA 2019*, Cham: Springer, 2019, pp. 121–129, doi: 10.1007/978-3-030-33226-6_14.

[23] D. R. Varma, "Managing DICOM images: tips and tricks for the radiologist," *Indian Journal of Radiology and Imaging*, vol. 22, no. 01, pp. 4–13, Jan. 2012, doi: 10.4103/0971-3026.95396.

[24] S. Desai and B. S. Usha, "Medical image transcoder for telemedicine based on wireless communication devices," in *2011 3rd International Conference on Electronics Computer Technology*, Apr. 2011, pp. 389–393, doi: 10.1109/ICECTECH.2011.5941629.

[25] C. Marcus, E. Mena, and R. M. Subramaniam, "Brain PET in the diagnosis of Alzheimer's disease," *Clinical Nuclear Medicine*, vol. 39, no. 10, pp. e413–e426, Oct. 2014, doi: 10.1097/RLU.0000000000000547.

[26] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.

[27] M. Shaha and M. Pawar, "Transfer learning for image classification," in *2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, Mar. 2018, pp. 656–660, doi: 10.1109/ICECA.2018.8474802.

[28] L. Xiao, Q. Yan, and S. Deng, "Scene classification with improved AlexNet model," in *2017 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE)*, Nov. 2017, pp. 1–6, doi: 10.1109/ISKE.2017.8258820.

[29] L. Zhu, Z. Li, C. Li, J. Wu, and J. Yue, "High performance vegetable classification from images based on AlexNet deep learning model," *International Journal of Agricultural and Biological Engineering,* vol. 11, no. 4, pp. 190–196, 2018, doi: 10.25165/j.ijabe.20181104.2690.

[30] C. Szegedy *et al*., "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition*

(CVPR), Jun. 2015, pp. 1–9, doi: 10.1109/CVPR.2015.7298594.
[31] R. Anand, T. Shanthi, M. S. Nithish, and S. Lakshman, "Face recognition and classification using GoogLeNet architecture," in *Soft Computing for Problem Solving. Advances in Intelligent Systems and Computing*, Singapore: Springer, 2020, pp. 261–269.
[32] G. Sapijaszko and W. B. Mikhael, "An overview of recent convolutional neural network algorithms for image recognition," in *2018 IEEE 61st International Midwest Symposium on Circuits and Systems (MWSCAS)*, Aug. 2018, pp. 743–746, doi: 10.1109/MWSCAS.2018.8623911.
[33] "Pretrained deep neural networks," https://in.mathworks.com/help/deeplearning/ug/pretrained-convolutional-neural-networks.htmlbvnyev7-1 (accessed May 09, 2023).
[34] "Alzheimer's disease meuroimaging initiative." https://ida.loni.usc.edu/home/projectPage.jsp?project=ADNI (accessed May 09, 2023).
[35] "Alzheimer's dataset (4 class of images)." https://www.kaggle.com/datasets/tourist55/alzheimers-dataset-4-class-of-images (accessed May 09, 2023).
[36] J. V. Shanmugam, B. Duraisamy, B. C. Simon, and P. Bhaskaran, "Alzheimer's disease classification using pre-trained deep networks," *Biomedical Signal Processing and Control*, vol. 71, p. 103217, Jan. 2022, doi: 10.1016/j.bspc.2021.103217.

BIOGRAPHIES OF AUTHORS

Manoj Kumar Sharma is working as an assistant professor at the Vidya College of Engineering, Meerut, Uttar Pradesh, India. After completion of the masters in VLSI engineering at the Shobhit University, Meerut. He worked in industries for more than 2 years. In the year 2013, he joined the Department of Electronics and Communication Engineering, Vidya College of Engineering, Meerut. He has supervised various post-graduate theses. He is also mentoring a startup embedded company named Inst!ller in Meerut with his student. In the year 2015 again, he started a journey as a researcher in the field of neuroscience and try to build an AI-based disease diagnosis medical system to provide better living conditions for people. He can be contacted at email: vmanojsharma@gmail.com.

Dr. M. Shamim Kaiser is currently serving as a professor at the Institute of Information Technology, Jahangirnagar University, Bangladesh. He obtained his bachelor's and master's degrees in applied physics electronics and communication engineering from the University of Dhaka, Bangladesh in 2002 and 2004 respectively, and completed his Ph.D. in telecommunication engineering from the Asian Institute of Technology (AIT) Pathumthani, Thailand, in 2010. He has also worked as a postdoc fellow at Anglia Ruskin University, UK in 2017-2018 and as a special research student at Tohoku University, Japan in 2008. His interests include various areas such as data analytics, machine learning, wireless networks and signal processing, cognitive radio networks, big IoT data, healthcare, neuroinformatics, and cyber security. He has published over 150 papers in different journals and conferences. He is also actively involved in various professional organizations, including being an academic editor of Plos One Journal and an associate editor of the IEEE Access and Cognitive Computation Journal. He is a life member of the Bangladesh Electronic Society, the Bangladesh Physical Society, and NOAMI, and a senior member of IEEE, USA, and IEICE, Japan. Additionally, he volunteers for the IEEE Bangladesh Section and founded the IEEE Bangladesh Section Computer Society Chapter. He can be contacted at email: mskaiser@juniv.edu.

Dr. Kanad Ray is a professor and head of physics at the Amity School of Applied Sciences Physics Amity University Rajasthan (AUR), Jaipur, India. He has obtained M.Sc. and Ph.D. degrees in physics from Calcutta University and Jadavpur University, West Bengal, India. In an academic career spanning over 22 years, he has published and presented research papers in several national and international journals and conferences in India and abroad. He has authored a book on the Electromagnetic Field Theory. He current research areas of interest include cognition, communication, electromagnetic field theory, antenna and wave propagation, microwave, computational biology, and applied physics. Presently he is guiding 8 Ph.D. scholars in various interdisciplinary fields. He has served as editor of the Springer Book Series. Presently he is an associated editor of the Journal of Integrative Neuroscience published by IOS Press, Netherlands. He has traveled to more than a dozen countries on different academic missions. He has established an MOU between his University and the University of Montreal, Canada for various joint research activities. He has also established collaboration with National Institute for Materials Science(NIMS), Japan for joint research activities and visits NIMS as a visiting scientist. He organizes international conference series such as SoCTA and ICOEVCI as general chair. He is currently an IEEE Executive Committee member of the Rajasthan Subsection. He can be contacted at email: kray@jpr.amity.edu.