Multimedia Tools and Applications
https://doi.org/10.1007/s11042-020-10114-1

Diagnosis of Parkinson's disease using deep CNN with transfer learning and data augmentation

Sukhpal Kaur[1] · Himanshu Aggarwal[1] · Rinkle Rani[2]

Received: 9 May 2020 / Revised: 25 September 2020 / Accepted: 19 October 2020

Abstract
Parkinson's disease (PD) is one of the main types of neurological disorders affected by progressive brain degeneration. Early detection and prior care may help patients to improve their quality of life, although this neurodegenerative disease has no known cure. Magnetic Resonance (MR) Imaging is capable of detecting the structural changes in the brain due to dopamine deficiency in Parkinson's disease subjects. Deep learning algorithms provide cutting-edge results for various machine learning and computer vision tasks. We have proposed an approach to classify MR images of healthy and Parkinson's disease patients using deep convolution neural network. However, these algorithms require a large training dataset to perform well on a particular task. To this effect, we have applied a deep convolution neural network classifier that incorporates transfer learning and data augmentation techniques to improve the classification. To increase the size of training data, GAN-based data augmentation is used. A total of 504 images are collected, and 360 images are used to augment data. The increased data set of this model is as many as 4200 images, and the produced images are of good quality by using this data set for the detection of peak signal-to-noise ratio (PSNR) having an innovative value in the norm of real images. The pre-trained Alex-Net architecture helps in refining the diagnosis process. The MR images are trained and tested to provide accuracy measures through the transfer learned Alex-Net Model. The results are addressed to demonstrate that the fine-tuning of the final layers corresponds to an average classification accuracy of 89.23%. The experimental findings show that the proposed method offers an improved diagnosis of Parkinson's disease compared to state-of-the-art research.

Keywords Parkinson's disease · Generative Adversarial Network · Alex-Net · Transfer learning · Overfitting

✉ Sukhpal Kaur
sukhpal91@gmail.com

Extended author information available on the last page of the article

1 Introduction

Parkinson's disease is an emerging neurodegenerative disease that is induced by early damage caused to the dopaminergic neurons in a sizeable neural region [43]. Dopaminergic dysfunction of the neurons causes different Symptoms of motor and non-motor. The motor signs include sweating, pain, slow movement, and walking problems. While the disease's cardinal characteristics are motor, anxiety, hysteria, worsening fatigue, genitourinary problems, and sleep disturbances are the main non-motor symptoms [13]. All the motor and non-motor symptoms are associated with the aging factors and affect the quality of life [8]. The proper preventive steps that accurate, slow, or stop disease progression are keys during earlier phases of the PD. Although attempts are made to develop economic and easily available PD biomarkers, there have been no proven blood or laboratory tests to diagnose PD and its progression to date [40]. Physicians use a medical history and cognitive assessment, such as the staging scale of the Unified Parkinson Disease Rating Scale (UPDRS) and Hoehn and Yahr (1967) mini-mental state test (MMSE) for initial diagnosis of PD [62]. Nevertheless, these clinical evaluation scales are based on professional clinician's expertise and yield subjective judgment. It takes a significant amount of time and diligent cooperation to gather all the relevant information. Neuroimaging can track the pathophysical changes that assess the dysfunction in the dopaminergic pathways for the diagnosis of the PD [55]. It is a visual diagnostic tool for measuring and quantifying the loss of neuronal cells in the different lobes of the brain. A variety of neuro-imaging studies are carried out to quantify early PD in vivo [1]. Various modes of neuroimaging are used for PD diagnoses, for example, magnetic resonance imaging (MRI), Single-photon emission computed tomography (SPECT), positron emission tomography (PET) [44]. Nevertheless, such neuroimaging methodologies can only identify the disorder once 80% of the neurons degenerate, and the illness presents unusual symptoms [35]. Methods of brain imaging for magnetic resonance (MR) have recently shown promising results in early-stage PD diagnosis, and it is expected to be more sensitive than normal clinical measures [44]. MRI may track structural changes in the brain [61] and detect iron build-up in the significant nigra [20]. Quantifying Structural modifications should enable the assessment of the development of the disease. But, due to the Inherent intricacy of brain development and slight variability in severity, the neurodegeneration in the brain is difficult to recognize visually. Presently, most of the clinicians interpret these images where the possibility of human error cannot be ruled out. One such study[44] reports only an 80.6% pooled accuracy of the medical diagnostics of Parkinson's disease. We have come up with a new and practical approach that helps in the timely detection of Parkinson's disease where it automatically employing image processing and Deep Neural Network techniques.Therefore, a reliable, cost-effective, and non-invasive computer-aided diagnostic (CAD) method is required for the diagnosis of PD and the monitoring of progression of the disease using MRI, which could be useful for physicians [53]. Machine learning approaches have recently become a standard early-stage diagnostic tool for locating the disease inside the brain (i.e., finding disease markers) in recent years. Deep learning algorithms yield state-of-the-art results for various tasks involving computer vision and machine learning [38]. The CNN architectures have generated impressive results in a wide range of medical imaging applications, which include Alzheimer's disease diagnosis [22] and drug detection using MR images [56]. However, CNN's have some shortcomings: CNN training needs a large amount of labeled data [39]. CNN's, particularly contain several million parameters that involve potent GPUs to train to accelerate the process of training [19].

Therefore, the training is time-consuming as well as difficult. The usage of deep CNN Models to prepare small scale datasets leads to overfitting [31] even when using preventive methods such as dropout [17].Any framework for machine learning needs to understand the issues of overfitting and underfitting. The key goal is to create a model that the training data are neither under fitted nor overfit. In general, the purpose of each predictive model (in supervised learning) is to produce an effective function out of the training instances. If the training data is not suitable, the model is under fitted. The model is therefore supposed to show poor performance on both training and test results. On the other hand, if the model matches the training data very closely but does not perform well on the test data, this means that the model is over-fitted [15]. This may be due to discovering unnecessary or noisy train data features, or due to inadequate network configuration with excessive parameters. There are several hyper-parameters in Deep CNN such as the number of epochs, hidden layers, hidden nodes, activation function, drop-out, learning rate, batch size, etc. which affect the model's performance [64]. Therefore, we need to look for the best subset of features and optimally balanced neural networks. In the tuning of hyper-parameters, experiments are performed with a certain number of hidden layers, epochs, various activation functions, and learning rates [65]. We search for optimum network layout by using an exhaustive grid search or random search techniques to prevent overfitting and underfitting of the network [65]. For the highest precision, the optimum state is reached conveniently after fine-tuning the model. The dropout callbacks and lower learning rates on the plateau functions are used to minimize the risk of overfitting and to increase classification process efficiency [4].

Transfer learning provides a solution for the problem of training deep CNNs on a limited collection of data [24]. The simplified transfer learning approach [2] for deep CNNs is to take advantage of a pre-trained CNN model and adjust the classification layer to fine-tune the target dataset weight parameters. This method can speed up training and enhance efficiency despite having some shortcomings. Data augmentation is another approach to enhance the impact of training [34].

While training the network the validation loss is monitored. When consistency comes in validation loss, the training ceases to avoid overfitting [10]. The early stoppage is a technique for evaluating an arbitrary number of training epochs and stopping training until model performance with a validation dataset stops improving. For e.g., stop the algorithm when the accuracy exceeds a specified threshold, or when the loss value hits a minimum or reaches the maximum number of iterations. The Accuracy and loss value becomes stable once the network reaches the convergence stage. Thus, the model architecture, hyper-parameter tuning, and data Augmentation are important in minimizing model overfitting and helps in building a more stable convolution neural network model [25]. In this work, we proposed a system for creating the best configuration of hyperparameters for a transfer-learned model in Keras (https://www.keras.io/). To select the set of hyperparameters and improve the accuracy of the model, a systematic grid search optimization method was developed. Also, transfer learning and data augmentation techniques are used for minimizing the overfitting and to improve the generalization of the model. This paper provides two methods for these issues, where one of the perspectives proposes a model based on transfer learning depending on a pre-trained Alex-Net architecture. The second perspective concentrates on a novel strategy for the data augmentation dependent on the Generative Adversarial Network's. Despite the two enhancements, the PD classification model gets an improved accuracy whereby, to a certain extent, overfitting may be done away with.

1.1 Related works

Gil et al. [18] proposed an approach for classifying PD based on the Convolution neural network. Results obtained from 195 images of 31 patients were used to determine the potential of the approach proposed. The authors of this paper got an accuracy rating of 80.2%. Focke et al. [12] developed a classification system utilizing MR images to diagnose Parkinson's disease. The author of this paper [12] has used various regions of the brain, such as White Matter and Gray Matter, utilizing the Support vector machine classifier for an individual's diagnosis of PD. For preprocessing and extraction of the feature, voxel-based morphometry (VBM) was used. Nevertheless, the precision of the classification obtained was just 39.53% and 41.86%, respectively. Babu et al. [7] had proposed a Computer-Aided Diagnostic system for the diagnosis of Parkinson's disease. Their approach involves two main steps: extraction of the features and classification. For the first section, the VBM is used to create feature data over GM. Recursive feature elimination (RFE) was used to select the most discriminatory features. Classification is performed in the last stage using projection-based learning and radial-based Meta cognitive function, resulting in 81.21% accuracy. The drawback of this work is that VBM is univariate, and RFE is costly in computational terms. Salvatore et al. [48] assessed the feasibility of PCA with the SVM classifier. The PCA is applied to extract features from MR images after normalization. At that point, SVM was utilized as the classifier, bringing about 85.8% precision. Rana et al. [45] identified attributes around the three primary tissues of the brain that consist of WM, GM, and CSF. They then used a t-test for selecting the feature and SVM for classification in the next move. That resulted in GM and WM accuracy of 86.67%, and CSF accuracy of 83.33%. Spectral function selection based on the graph theory approach was used in their other work Rana et al. [46] to pick a collection of discriminating features from the entire brain volume. Usage of SVM as a classifier with a cross-validation leave-one-out scheme, a decision model was built, giving 86.67%accuracy. Adeli et al. [5] suggested an early-stage PD diagnostic method using a Joint FunctionSample Selection (JFSS) process to pick the best features to construct a model. They undertook the development of a robust classification model to construct the CAD for PD diagnosis. Synthetic and publicly accessible MR images of PD datasets are used for evaluation showing a high precision ranking. Another study carried out by Abos et al. [3] proposed an AI approach utilizing extracted features from the individual image. A Support Vector Machine (SVM) classifier with a quadratic kernel of 10-fold cross-validation is used to identify images, but overall accuracy achieved is only 80%. Amoroso et al. [6] developed a therapeutic help scheme using an Artificial neural network. Complex network measurements obtained from brain lesions were used as PD measurements for this analysis. Another research school focuses on Region of Interest Strategies (ROI), where specific brain regions such as gray matter, hippocampal volume, and cortical thickness are extracted from a priori understanding of their effect on brain functioning and cognition. Shinde et al. [52] explored the various machine learning algorithms and indicated that the Convolution Neural Network (CNN) is a superior classification algorithm for classifying Parkinson's disease MR Images with regards to the lower error rate on PPMI datasets. Cigdem et al. [9] proposed Principal Component Analysis (PCA) for extraction of the Features and a finely tuned CNN architecture for Parkinson's disease classification. Sivaranjini et al. [54] used CNN le-Net with transfer learning for the classification of MR images of people with PD and healthy individuals. The summary of existing Parkinson's disease identifications methods using MR images is shown in Table 1.

Table 1 Review of the existing strategies for recognizing Parkinson's disease

Author	Number of subjects	Selected Region of Brain	Data description	Method employed	Accuracy (%)
Gil et al. [18]	PD = 28, NC = 28	The selected region of interest	MR images	CNN	80.2
Focke et al. [12]	PD = 20, NC = 20	The selected region of interest	MR images, DT images	SVM	41.86
Babu et al. [7]	PD = 28, NC = 28	The selected region of interest	SPECT images	VBM, RFE	81.21
Salvatore et al. [48]	PD = 25, NC = 25	The selected region of interest	MR images	PCA, SVM	83.2
Rana et al. [45]	PD = 28, NC = 28	WM, GM, and CSF	MR images	SVM	86.67
Adeli et al. [5]	PD = 274, NC = 170	The selected region of interest	MR images	Joint feature sample selection	81.9
Abos et al. [3]	PD = 28, NC = 40	MRI based Atlas of Substantia nigra	MR images, DT images	SVM	80
Amoroso et al. [6]	PD = 274, NC = 170	The selected region of interest	MR images	ANN	83
Shinde et al. [52]	PD = 248, NC = 204	The selected region of interest	MR images	CNN + SVM	80
Cigdem et al. [9]	PD = 40, NC = 40	Whole-brain image analysis	MR images	PCA + CNN	84.25
Sivaranjini et al. [54]	PD = 82, NC = 100	Whole-brain image analysis	MR images	Transfer learned Le-Net	86.9
Proposed Approach	PD = 67 NC = 85	Whole-brain image analysis	MR images	Transfer learned Alex-Net+ data augmentation	89.23

For the detection of Parkinson's disease by utilizing MR-Images, state-of-the-artwork did by Cigdem et al. [9], and Sivaranjini et al. [54] serve the basis for our precision model. The best results achieved by Cigdem et al. [9] utilizing a PCA and CNN approach is 84.25%, and by Sivaranjini et al. [54] consolidating the transfer learned CNN classifier has a precision level of 86.9% where both of them utilize human-engineered extraction to train the Convolution Neural Networks (CNN) have proved their worth recently. The main focal point of this paper to construct and validate a deep transfer learned Alex-Net with a data augmentation approach that can forecast PD diagnosis depending on a given cross-sectional brain skeletal MRI scan.

1.2 Contribution/Novelty

Though all these studies on the PPMI dataset have achieved comparable accuracy. One of the most important caveats of previous works is that they have examined PD trends centered on the regions of interest chosen [18, 12, 7, 48, 45, 46, 5, 3, 6, 52]. Nonetheless, the limitation is that most of these techniques may be impaired by knowledge loss. Also, these strategies are potentially skewed by segmentation errors, which may affect the efficiency of classification [3, 6].In light of this, one of the main goals and novelty of our study is to overcome this limit by using whole-brain image processing as an approach to reduce the Spatial structure without having to rely on hand-made features while preserving the essential details. The present study aimed to build and validate a deep learning algorithm (specifically convolutional neural networks [CNN]) that can predict the individual diagnosis of Parkinson's disease patients based on a single cross-sectional brain structural MRI scan. For CNN's as a Machine Learning model, the supervised learning protocol aims to reduce the training error by conducting experiments with a specified dataset. The real goal, however, is to perform well on the unseen dataset. To assess this speculation capacity, a validation set is utilized during the training. Some strategies are usually implemented to minimize the gap between training and validation errors, such as transfer learning [34, 10, 25], data augmentation [2], and fine-tuning [65]. Accordingly, this is the novelty of this research to incorporate the idea of transfer learning and data augmentation technique in increasing the identification rate for early detection of Parkinson's disease due to the small number of MR images of patients having Parkinson's disease.

The major contributions of this research towards improving prediction efficiency include: -.

- The application of transfer learning and data augmentation methods in developing the early-stage diagnosis of Parkinson's' disease as preferred over the limited availability of PD patient's MR images.
- The preprocessing steps are employed to improve the classification efficiency to a further extent. These pre-processing steps help to minimize noise and retain valuable information; therefore, it leads to improved performance in high-level learning applications.
- To increase the size of the dataset and to enhance our model efficiency, we used the technique of data augmentation based on GAN.
- The proposed approach works very well, even when a limited number of training images are used.
- The phenomenon of overfitting is being investigated with less training data and their effect on classification results.
- Significant performance improvement is observed when the data-augmentation method is used with a transfer learned classifier.
- Compared with the existing state of the art approaches, the proposed approach has achieved better classification accuracy.

2 Proposed method

The proposed approach for classifying Parkinson's disease based on the Deep CNN architecture is discussed in this section. Figure 1 demonstrates the overall workflow of the proposed approach. The MR images of healthy people and patients with Parkinson's disease collected from the PPMI database (https://www.ppmi-info.org) are used as an input. The preprocessing

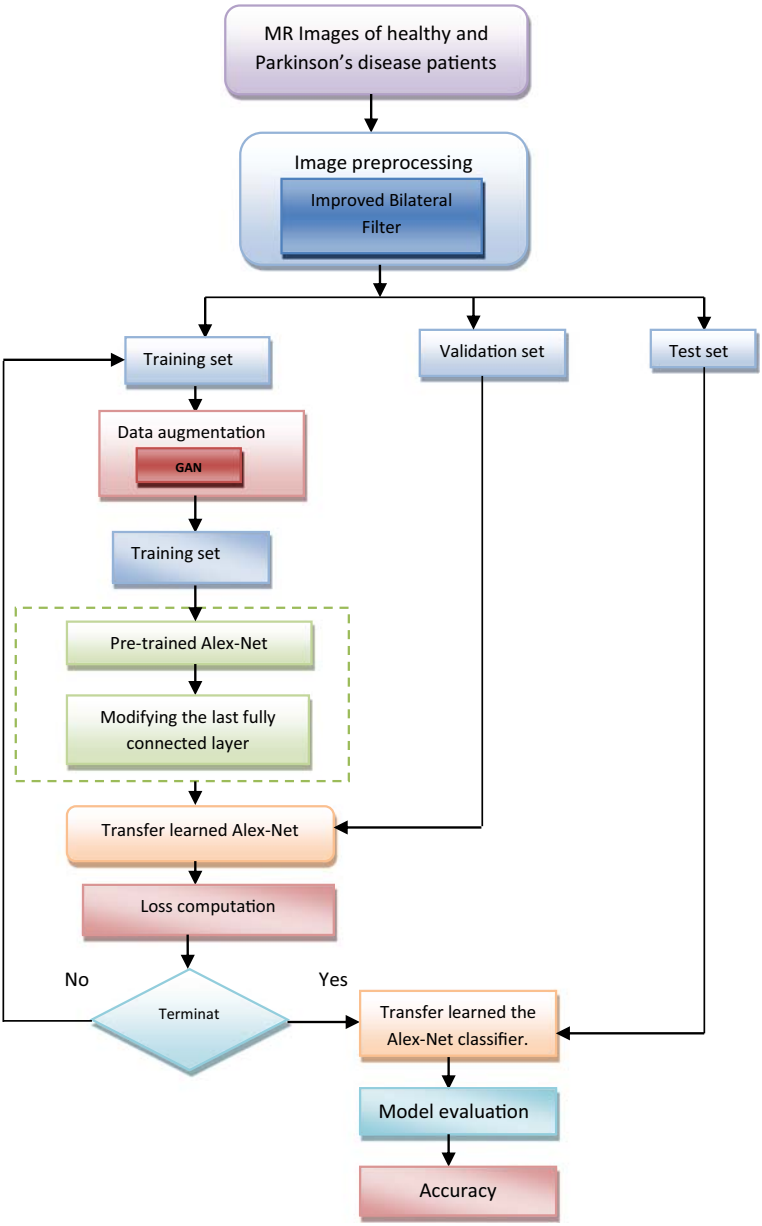

Fig. 1 Overall workflow of the proposed approach

of images is carried out in the first step to accommodate the variety of contrasting images and to reduce some noise. Data augmentation procedures are performed in the second stage to address the issue of the dataset's limited size and to improve the performance of the proposed approach. In the third step, transfer learning is applied to the pre-trained Alex-Net, and the last few layers are replaced to accommodate new categories of images for our application. Finally, the performance of the model proposed is evaluated on test MR images of HC and PD patients.

2.1 Images preprocessing

The most critical step in achieving the desired features and reasonable levels of classification is the preprocessing of images. MR images obtained from the PPMI database are chosen in the proposed approach. The MR images contain variability in comparison and brightness; basically, there is also some induced noise [41]. To minimize such an effect that adversely affects the Deep CNN process, image normalization should be performed by taking the difference between the MR image's intensity values [16]. Considering I1 as the minimum intensity value and I2 as the highest intensity value, the distinction between them gives the resulting intensity as I = I2-I1. Before resizing the images, the dimensions are 195×365, and the Alex-Net architecture of the input layer specifies that the image size should be $227 \times 227 \times 3$. The images are resized to 227×227 to preserve the uniformity of structured images, which is then used for training. The training is performed using the entire images without the application of ROI that would affect the ranking performance. The resized image is then analyzed with the improved bilateral filter to hold the edges, provided both the temporal and the spectrum [42]. The Pseudocode of an improved bilateral filter is given in algorithm 1.

Algorithm 1: Improved bilateral filter

1. Input: -The MR images of healthy Individuals and PD patients
2. Output: - Improved Bilateral filtered MR images of healthy and PD patients
3. $G[I_m](X_k,X_l,D)=G_{\sigma a}*I_m((X_k,X_l,D)$ // $G_{\sigma a}$ corresponds to gaussian filter
   //compute Gaussian filtering on entire image components
4. Bi $[I_m]=(\check{D}I_m/D)$ //$((\check{D}I_m/D)$ denotes pixel value at (X_k,X_l,D)
5. Return Bi $[I_m]$

2.2 GAN based data augmentation

An important technique in CNN to minimize the problem of overfitting is the augmentation process [51]. The augmentation [58] of images plays a critical role in the medical image analysis due to the lack of a large number of labeled images. To create new images, we used the GAN based data augmentation technique [60]. Generative Adversarial Networks (GANs) is a generative system comprised of two opposing networks: a generator network and a discriminator network. The generator generates synthetic data to deceive the discriminator, while the discriminator separates real data from synthetic [57].

2.2.1 Generator

An architecture based on six layers has been selected as a generator network which is represented in Fig. 2. The data fed to the generator network is a sampled vector out 100 randomly distributed invariant values in [0; 1] array [29]. The well-connected

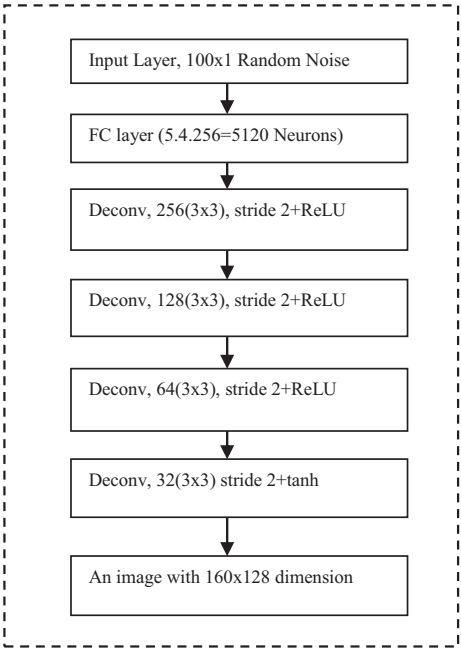

Fig. 2 Generator network

layer of 5120 units is directly fed by input data and is thus altered to reach the size of $16 \times 16 \times 256$. It passes from a four-layered network of deconvolutional layers and referred to as two dimensional transposed convolutions of 256 filters as well as kernel length. A 3×3 configured kernel and similar padding have opted for both the layer forms [36]. This assists in the multiplication of input spatial measurements. Besides the last layer, a ReLU activation function is applied to all the layers. The last layer is comprised of a hyperbolic tangent activation function so that the need for the output being bound makes it able in emitting an object. A comparison is made amongst the Tanh and Sigmoid function being based on 0 that show improvement at the time of training [21]. Lastly, the MR images are developed ranging a width of 160×128 following five deconvolution layers.

2.2.2 Discriminator

A five-layer-based architecture applied in this study i.e. the Discriminator network is represented in Fig. 3. The Discriminative input layer is comprised of a set of real image samples as well as a vector image component. The input fed to the Discriminator is a 160×128 image of a single channel. The input image is thus transferred through an alternate layering of convolutions five times with a stride of 2. The latter is utilized for subsampling as there being no pooling layers in the architecture. There is a random generation of weight and bias parameters [21, 33]. The last two layers are the FC layers. Besides the ending layer being without any activation function, the leaky ReLU activation function activates all the other layers.

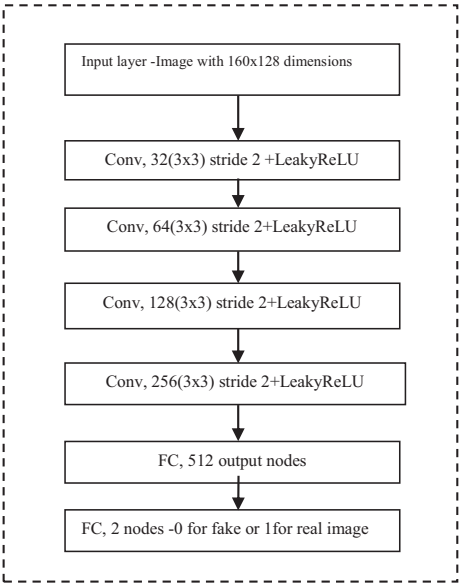

Fig. 3 Discriminator network

2.2.3 GAN Training

Algorithm 2 describes the training relevant iterative process of GAN. The stochastic gradient is used as an optimization method for the generative and an adversarial network. The weight and bias parameters are generated randomly. The GAN network carries n iterations, and at the end of each iteration loss matrix is calculated, and the entire weight and bias are updated until the generator and discriminator network becomes stable. We use the antagonistic effect [14] of GAN to produce objects that are close to the actual information in the iterations. The backpropagation method supports both the Generative and the Discriminator models to change the settings.

Algorithm 2: Generative adversarial Network

1. Input: -Real MR images, Random noise (Z)
2. Output: -synthetic MR images
3. Generation of synthetic images
   $I_m=G_j\ (Z_j)$
4. Discrimination between real and MR images
   $D_j(X_R,\ I_m)$
5. Updation of the Generator network using stochastic gradient descent

   $\nabla\Theta_d \frac{1}{k}\sum_{j=1}^{k}(\log(D(X_{tr}))\ +\log\ (1-D\ (G(Z_j))$
6. Updation of the Discriminator network using stochastic gradient descent

   $\nabla\Theta_d \frac{1}{k}\sum_{j=1}^{k}(\log(1-D(G\ (Z_j))$
7. The neural network performs n training iterations and returns the results
(a) The loss function is determined at the end of each iteration, and the weight matrix is modified accordingly until the training ends.
(b) The training stops whenever the generator and discriminator network appears to be stable.

The stochastic gradient descent [23] algorithm calculates the cumulative loss value for every variable to achieve the optimal overall solution. Even if the loss value obtained in each iteration is not ideal globally during the iteration process, the result obtained is typically similar to the optimal solution for the entire program. In this paper, the GAN learning rate is set at 0.001, the number of iterations is 200, and the size of the batch is 8. Both Generator and Discriminator networks undergo concurrent or adversarial learning until equilibrium is achieved. The deviation in the loss function can be seen in the results via the trend chart as shown in Fig. 4a and b.

The Generator Network's change in loss function is progressively smaller, while the loss function of the adversarial model is gradually increasing, which means the degree of modeling and bias against their current data has varying change levels and is advantageous for the entire model. The images generated out of the generative model was very unstable in the initial iteration process, as seen in Fig. 10 in the variation diagram of the loss function of the generative model. It started to decrease after 150 iterations and eventually became steady after 190 iterations, the total equilibrium remained stable. It can be shown that the value shift of the loss function of the adversarial network begins to increase from about 0.05 as it comes to iterating periods in Fig. 4b. It rises quickly at the start. As seen by the above two figures, however, the training of the GAN network is highly unpredictable and the generator and the discriminator can either be diminished or rapidly increased throughout the entire phase.

The images are faulty and distorted in the 100th iteration as shown in Fig. 5a, b. Contours and textures appeared after the 150th iterations Fig. 5c. We can see in Fig. 5d that after 190 iterations, the personalized characteristics are improved, to a certain degree close to the real MR images. After 190 iterations, the approach proposed can be balanced. We can see that the generated images not only mirrored the similarities but also demonstrated ingenuity on the latest trend base as well. However, if an image is enlarged, the final image produced is identical in style and tone to the image created, but it carries defects also. The granulation of an image is more extreme and noise occurs. We use table form for showing the generated effect in the synthesized images. PSNR is an unbiased measurement tool for evaluating image quality or optimum signal-to-noise ratio, and its classification is almost similar to the human eye detection process [50]. The equation of the PSNR method is shown below. The higher value of PSNR corresponds to less distortion in the image.

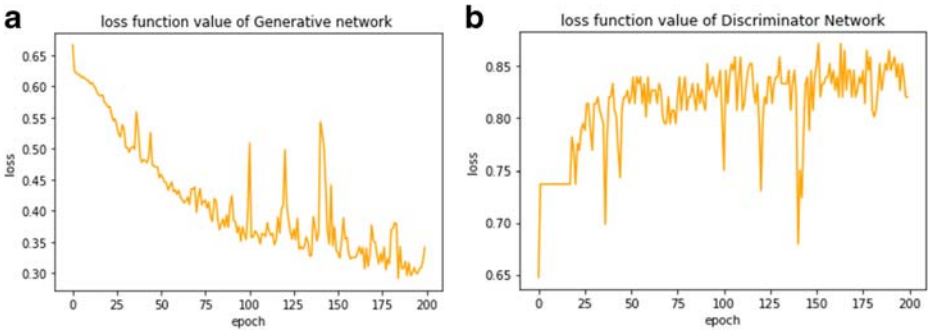

Fig. 4 loss value of (**a**) Generator network (**b**) Discriminator network

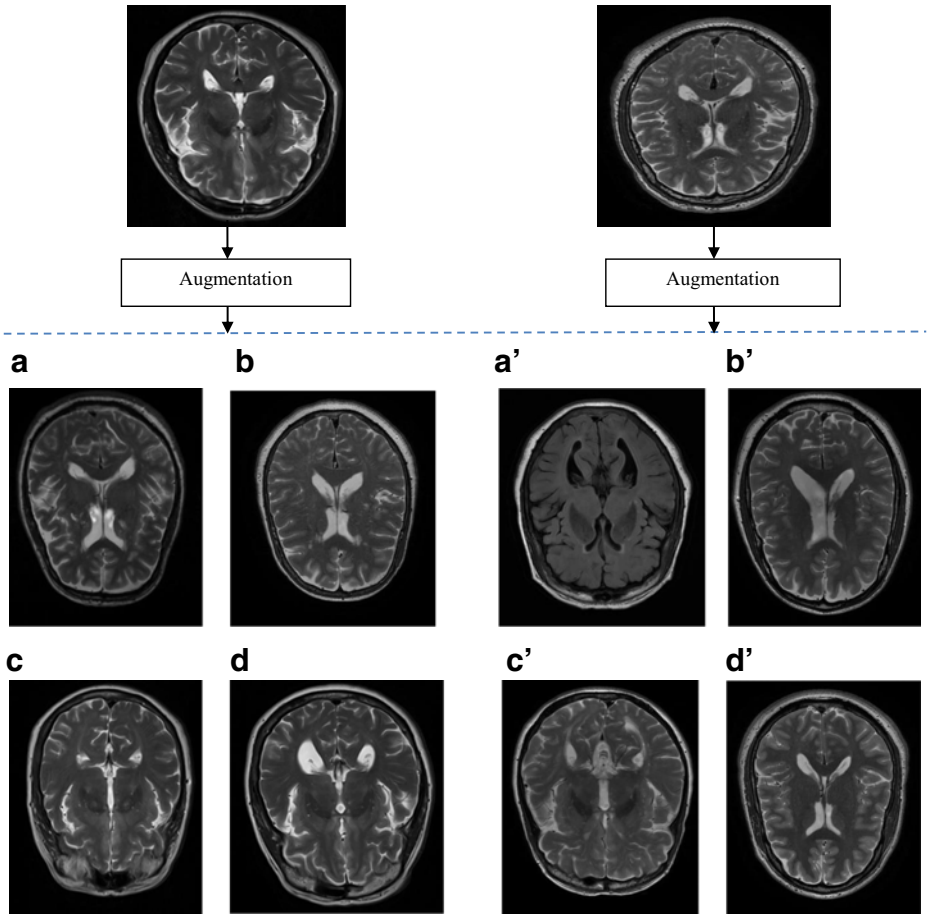

Fig. 5 Images generated during Augmentation. The images at the top are the original images of PD and HC, and the at the bottom is the synthetic images after (**a**, a') 50th iteration (**b**, b') 100th iteration (**c**, c') 150th iteration (**d**, d') 190th iteration

$$\text{Peak signal} - \text{to} - \text{noise ratio} = \frac{10\log(2^{m}-1)^{2}}{\text{MSE}} \tag{1}$$

Usually, the resulting images may vary in some way from the original image after compression of the file. As can be seen from the PSNR recognition result, the higher the number of variations, the greater the influence of object processing, which is compatible with the effect of human eye perception.

Besides, the value of the PSNR must exceed 30dB, so we can conclude that the image is above the norm. As shown in Table 2, the goal can only be reached after more than 150 iterations.

As we analyze Table 2, the increased number of iterations refines out the quality of images thus generated by the experiment. When the number of iterations reaches 190, there comes stability in the images along with the high value of PSNR. It can be observed that when the feature extraction's convolution kernel is at 3×3 and the deep

Table 2 PSNR value at different iterations

Iterations	PSNR value(dB)
1	0
50	18.52
100	23.14
150	27.71
190	33.15
200	33.23

CNN network's iteration number is at 190, the best possible results are achievable. So, the essential elements tend to influence the original model to a great extent with a sufficient influential ability to generate images.

2.3 Transfer learning to pre-trained Alex-Net

In deep neural networks, it is imperative to select the proper network architecture to implement CNN [63]. For the classification of MR images of HC and PD patients, the pre-trained Alex-Net model with transfer learning [49, 26, 59] is considered. Also, CNN would need a large number of MR images for the generation and upgrading of weights. Therefore, transfer to CNN will produce desirable output and increase the rate of convergence by moving weights from the pre-trained model [47]. The mechanism of transfer learning using Alex-Net models is briefly described in Fig. 6.

Alex-Net is a deep, eleven-layered, network, as shown in Fig. 7. It comprises a total of 650,000 nodes, 60 million parameters, and 630 million linkages [30]. The input layer collects the MR images from the user and passes the images to the layer below, as defined by the pre-trained model. The features of CNN are more generic in the first layers and specifically relate to the Parkinson data set in the later layers. Such layers perform the requisite operations to classify the input images into HC and PD. The fully connected layers are fine-tuned with the relevant hyperparameters, and thus high-level image characteristics are learned in MR images of healthy and PD patients. Figure 3 shows the image transformation for each of the following layers of the pre-trained Alex-Net. A first convolution layer comprising 96 kernels of size 11 to 11 with four strides and no padding is provided to the input image in dimension 227 to 227 to 3. The stride is defined as the number of pixels in the image matrix moved by the filter. The four strides of the first convolution layer are four pixels shifted through the filter. The resulting images taken from the first convolution layer is 55×55 by combining the image with 96 kernels.

The max-pooling is the process of discretization and it is used to reduce the dimension of an input image. The image is divided into different areas of bundling, and the overall pixel value is known as the performance of this layer for each field. The max-pooling is accomplished through the selection of maximum maps in the convolution network, which minimizes computational complexity. The sizes of kernel 3×3 of Max pool1 are two strides and zero paddings. This layer has $27 \times 27 \times 96$ performance measurements. In the second convolution layer composed of 256 kernels of size 5×5, the output from max- pool one is injected. The image is translated to 256 kernels. The output dimensions of the second convolution layer are $27 \times 27 \times 256$ that is given by kernels $3 \times$

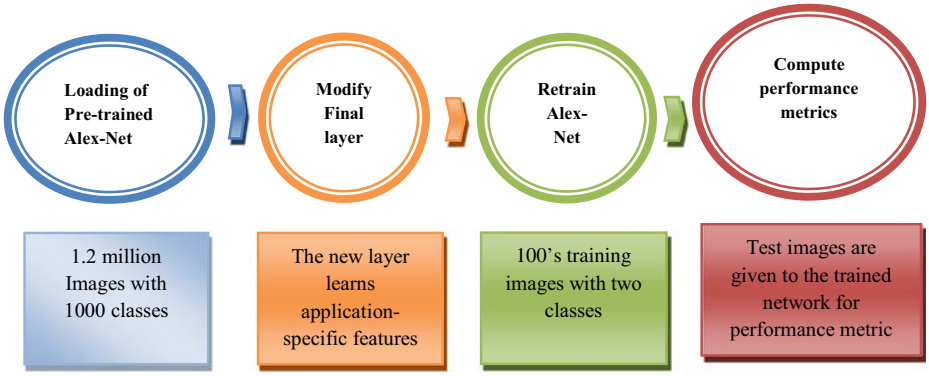

Fig. 6 A brief description of the transfer-learning process of the Alex-Net framework

3 with two strides to the second max-pooling layer. The size of the image is reduced to 13 × 13 × 256 as a function of the max-pooling-layer. The third layer of 384 kernels with one stride is inserted onto the convolution layer. After conversion, the size of the image is 13 × 13 × 384.In the next convolution layer, the cycle in the third convolution layer repeats itself. The output of the fourth convolution layer is fed into the fifth with 256 kernels of dimension 3 × 3. This layer's image size is 13 × 13 × 256. The image is given to a third max pool layer to further reduce the dimensions to 6 × 6 × 256. This layer output is transferred to FC layers.

In the proposed approach, the FC8 layer is a layer of soft-max used for the classification of data.

$$z_j(y) = \frac{e^{a_j(y)}}{\sum_{k=1}^{l} e^{a_k^y}}$$

Here $0 \leq z_j \leq 1$ and $\sum_{k=1}^{l} z_j = 1$ [32].Hence the soft-max activation function is often favored in operation to the nonlinearity of the sigmoid function. A sigmoid function's derivative is σ(a)(1 − σ(a)). When a neuron's output becomes 0 or 1 during back-propagation the gradient becomes 0. Therefore the weights of the neuron do not get updated. These neurons are known as saturated neurons. Not simply this, the neuron weight-related with saturated neurons is also slowly modified. Therefore, if there are many saturated neurons are present, a network with sigmoid activation does not back-propagate. To either end of the sigmoid feature, Y values appear to respond even less to changes in X. This leads to a problem of "vanishing gradients." It lets the changes to the gradients go too far in various directions, making optimization tougher. Thus the Sigmoid activation functions saturate gradients and eliminate them [11]. As in pre-trained Alex-Net (1000), the number of output neurons is not equal to the number of classes in our task (2). This method is called fine-tuning, and it involves the original trained data while training the pre-trained model on data with different representations. It is obtained through alteration of the last fully connected layer accordingly, as described in Table 3.We used a new, randomly initialized fully connected layer with two neurons corresponding to classes HC and PD in our transfer learning scheme.

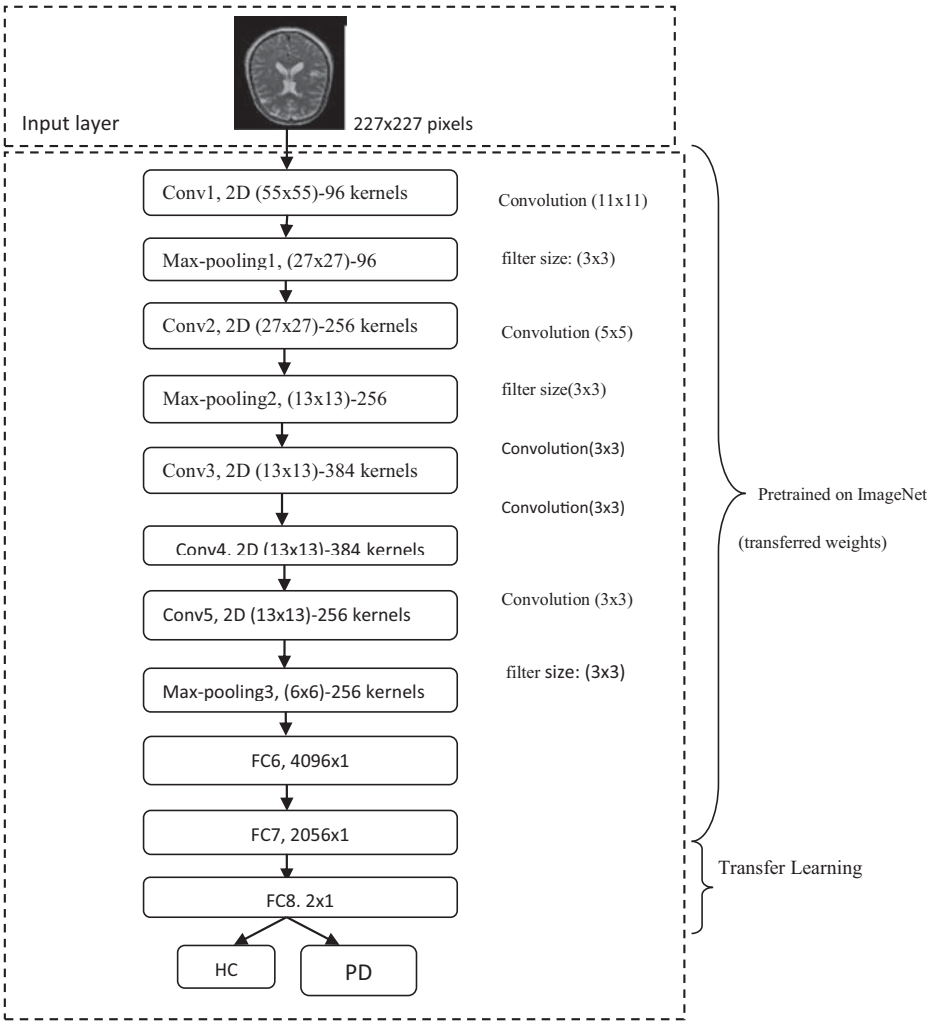

Fig. 7 The proposed Alex-Net CNN architecture

2.3.1 Training and fitting

To improve the performance of image classification, it is essential to use a pre-trained model, also known as transfer learning, as its weights were previously configured with attributes relevant to most computer vision issues. A data set for training, verification, and test was prepared for transfer

Table 3 Modification of last fully connected layer of Alex-Net

Features	Pre-trained Alex-Net	Transfer learned Alex-Net
Weight and Biases	Pre-trained weights and biases	Random Initialization of weights and biases
Activation Function	Soft-max	Soft-max
Number of classes	Classification layer with 1000 classes	The classification layer with two classes corresponds to HC and PD.

learning. The image dataset with 60% of the input data is used for training, 20% for validation and the remaining 20% is used for testing. The images consist of two classes that correspond to HC and PD patients. Training and validation images are used to construct the models, and test images are used to evaluate the performance of the trained model. We used pre-trained CNN models in this study which are available in the Keras Deep learning library (https://www.keras.io/). In this study, it is possible to use the transfer-learning technique, since the dataset used had images similar to those used in ImageNet. If the images are not similar to those used to train the models available, it would have been necessary to retrain some layers. It is performed through the following steps: –.

Output layer training The initial output layer is eliminated. We introduce our unique layer consisting of 2 units with a softmax activation function. The entire network is kept frozen, except for the newly added output layer, i.e. weights are not changed during the training stage. In deep learning, some superior optimizers, such as Adagrad, Adadelta, RMSprop, Adam, have been proposed that can automatically adjust the parameter during training [37]. According to Keras' library, the estimated learning rate for these approaches is 0.01, 1, 0.001, and 0.001 respectively. However, the case is very different when a pre-trained CNN fine-tuning is performed on a small size dataset. Only a part of the model needs to be modified, in addition to the minimal training details which can lead to the testing phase requiring just hundreds of iterations. As we know, the learning rate could be a lower value in the fine-tuning process, but to have the CNN model well trained; it is hard to pick a suitable initial learning rate. To address this problem, the Grid search optimization algorithm is used to search the learning rate due to its high efficiency compared to random search. We conduct multiple experiments for each classification task to determine the optimum number of training epochs, as well as the initial value of the learning rate and its reduction factor and frequency, by adjusting the values of those chosen parameters, using grid search.

Hyperparameter tuning In this stage, we select the values of the hyperparameters, such as initial learning rate, optimization algorithm, and epoch number that affect the network's learning process. Hyperparameter tuning is often performed using grid search, where all possible combinations of the hyperparameters with all of their values form a grid and an algorithm is trained for each combination. Let us assume $h_1,\ldots\ldots\ldots,h_k$ represent hyperparameters of Deep CNN where $\alpha_1,\ldots\ldots\ldots\alpha_n$ being their concerned domains. The training of the model is done with hyperparameter h while training (I_{train}) the MR images of both the healthy and Parkinson's' disease patients. The γ (h, I_{train}, I_{val}) is referred to as the validation accuracy of the Deep CNN Model. The hyperparameter optimization of the Deep CNN Model aims at finding a hyperparameter setting h^* to increase the validation accuracy γ on healthy and Parkinson's disease patients' MR images [44, 35]. The prior configuration of the hyperparameter contexture $h_1,\ldots\ldots\ldots,h_k$ effects the Grid Search Optimization algorithm as the whole hyperparameter contexture is signified by multiple projects there from every single context. Thereafter, the following parameters are assessed on validation data after noting their accuracies. In order to signify the accuracy, the algorithm moves further stereotypically to variate in Deep CNN in continuity. A new hyperparameter setting is recommended by the Grid Search Optimization algorithm with the support of procurement works as there is the determination of the new setup's accuracy on the validation data to develop new correctness [28, 27].

The step-by-step procedure for PD classification using Pre-trained Deep CNN is given in algorithm 3.

Algorithm 3. PD Classification using Pre-trained Deep CNN

Input:-*Pre-trained Alex-Net Model(L_m), Augmented training MR Images I_{train}= (I_R+ I_S), validation images (I_{val}), Iterations per stage Z=<Z_1, …, Z_z>, Total number of stages X, Training data per stage I_{train}= <I_{train}^{1},.., I_{train}^{z}>, Validation data I_{val}, Hyper-parameters $h_{1:k}$, Validation accuracy γ*
// Real MR Images (I_R), Synthetic images (I_S)

Output: - *Output:-Hyper-parameters h*, Predicted labels; 0-for Healthy Controls, 1 for PD patients*

Training:
for I iterations
for each I€ I_{train}:
L← L_m(I) //initialization of Alex-Net framework
L.fc.pop() //removal of last fully connected layer
L.dense(2,Activation_function)
for stage x=1 to X do
for j=1 to k
γ_j=evaluate γ(h_j,I_{train}^{s}, I_{val})
End
for i=k+1 to z_z
g=grid-search$(h_j,\gamma_i)_{i=1}^{j-i}$
h_j=max $args_{h€\alpha}$ a(h, g)
γ_i=evaluate γ (h_i, I_{train}^{s},I_{val})
End
End
End
End
Validation:
for I iterations
for I€ I_{Val}:
Loss ←$f_{cross\ entropy}$ (L)
Updation of L_m with loss.
Reset $h_{1:k}$ = best k configs €(h_1,…… h_{Xz})
// according to validation accuracy γ
End
Return h^*= max args $_{h\,€(}$ h^{X1}, … … .h^{Xz}) γ_i
End

Hyperparameter optimization was achieved in five steps. In the first step, the aim was to randomly test the efficiency of various hyperparameters, in terms of accuracy and loss in the training and validation phases. In the second step, 16 experiments were carried out, as defined in Table 4. Taking into consideration the success obtained in the first phase by AlexNet, certain hyperparameters were chosen as the ones to be used to create the model. Based on the results of the second step, the five top-performing models were selected in terms of accuracy in the validation process and the five best ones in terms of the loss function in the validation phase for the classification of the test dataset. Importantly, random values begin with the values of the initial weights allocated in neural network training. These weights are modified, as the learning process progresses. The initial randomness of these values thus affecting the

Table 4 Grid search results over different values of hyper-parameters

Sr.No	Epochs	Optimizer	Learning rate	Results			
				Loss_train	Acc_train	Loss_val.	Acc_val.
1	200	SGD	0.1	0.56	78.64	0.63	74.9
2	200	RMS-prop	0.1	0.62	77.6	0.78	72.7
3	200	Adagrad	0.1	0.54	78.4	0.61	73.2
4	200	Adam	0.1	0.44	83.6	0.58	79.5
5	200	SGD	0.01	0.53	79.89	0.62	77.3
6	200	RMS-prop	0.01	0.62	78.7	0.76	76.4
7	200	Adagrad	0.01	0.48	82.8	0.60	80.1
8	200	Adam	0.01	0.40	83	0.50	81.1
9	200	SGD	0.001	0.35	88	0.47	84.5
10	200	RMS-prop	0.001	0.36	86.19	0.44	83.6
11	200	Adagrad	0.001	0.33	87	0.38	86.9
12	200	**Adam**	**.001**	**.21**	**91.3**	**.32**	**89.23**
13	200	SGD	0.0001	0.46	85	0.60	80.12
14	200	RMS-prop	0.0001	0.42	84.99	0.56	77.34
15	200	Adagrad	0.0001	0.44	83.23	0.58	79.8
16	200	Adam	0.0001	0.41	86	0.51	83.3

outcome. The fourth step was thus to assess this difference, and the two experiments with the best overall accuracy performance were retrained two more times, after which their new models were re-submitted to the evaluation of the test dataset. In the final step, the batch size for the best sample obtained in the fourth stage varied as 32, 64, 128, and 256. If the algorithm reached maximal epoch, or the validation output declined over a predetermined testing epoch, the training process ended

The training procedure stopped when either the algorithm reached maximum epoch, or the performance of validation decreased over a preset training epoch. We applied the early stoppage in the training phase, whereas in the last six epochs there has been no increase invalidation. The ultimate function at the top level is to decrease the average loss and at the lower level, to optimize the accuracy for each training data. The number of training epochs of the deep CNN model inevitably impacts efficiency and loss. We tuned the hyper-parameters frequently and found the optimum hyper-parameters based on a validation set as shown in Table 5.

After selecting the relevant hyper-parameters we ran the final model for 200 epochs on the test images. With ADAM optimization, we implemented the training with a mini-batch size of 20, a learning rate of 0.001, a weight decay of 0.06, and a momentum factor of 0.9. This phase reflects certain levels of classification. It is at this point, as described earlier, that 2D feature maps are translated to a 1D feature vector and the classification of MR images is performed. In this step, we choose the value of the hyperparameters, such as the number of epochs, initial learning rate, and optimizer, which directly affect the network's learning process. The model is

Table 5 Training relevant hyper-parameters obtained after fine-tuning

Hyper-parameters	Values
Optimizer	ADAM
Learning rate	0.001
Batch size	20
Epochs	200

saved after completing the training and validation processes by running this project. The model stored with the weights learned is tested in the test process after using all the previous programs. At this point, a new dataset is introduced, allowing the evaluation of the generalization power for each class. This is necessary to allow the model to be used with the test dataset and thus to evaluate the network's success in classifying new images. In this analysis, the network's mean training period was 1 h, for 200 epochs.

3 Results and discussion

3.1 Parkinson's disease database

All the experiments are performed using Keras software with Theano as a deep learning backend in python 2.7.0 software running on an Intel ® CoreTMi5-6700 K microprocessor and 6 GB RAM laptop. Data used to prepare this paper are collected from repositories of the Parkinson's Progression Markers Initiative (PPMI). The PPMI database is commonly used by researchers to classify biomarkers of PD progression and to access brain function during Parkinson's disease (https://www.ppmi-info.org). Various datasets of differing modalities are present in the PPMI datasets having an ability to aid the scientists in early diagnosis of Parkinson's disease. The MR images used in this study consist of 162 participants, including 67 healthy individuals and 85 subjects with Parkinson's disease as shown in Table 6.

This work uses the dataset in the form of MR images, including the healthy and Parkinson's disease contents drawn through the data set of PPMI. In order to diagnose early-stage Parkinson's disease, images used are of PD patients at the prodromal stage and early visits. Figure 8a-b represents the T2 weighted MR images to use to differentiate healthy controls and PD subjects.

Thus, there is a preprocessing phase to convert all input images to the size specified by the AlexNet, regardless of their size. From the PPMI dataset, we can find that the MR images consist of a significant amount of noise; therefore, we first reduced the noise from the images by applying modified bilateral filters, and the resulting image is shown in Fig. 9a-b. The Modified bilateral filter decreases noise and thus enhances analytical sensitivity. In PD patients, a significant reduction in grey matter strength is observed compared to Healthy individuals with striatum area modifications.

We have applied data augmentation related to GAN to further increase the size of the training data collection. The image dataset with 60% of the input data is used for training, 20% for validation and the remaining 20% is used for testing. The images consist of two classes that correspond to HC and PD patients. Training and validation images are used to construct the models, and test images are used to assess the performance of the classifiers. Dataset distributions before and after data augmentation are described in Table 7.

Table 6 Description of the dataset used

Subject Description	Parkinson's disease patients	Healthy Individuals
Total number of subjects	67	85
Sex(Male/Female)	64/27	42/32
Age	60.17 ± 10.27	60.33 ± 8.68
Weight	80.16 ± 16.75	81.40 ± 16.28

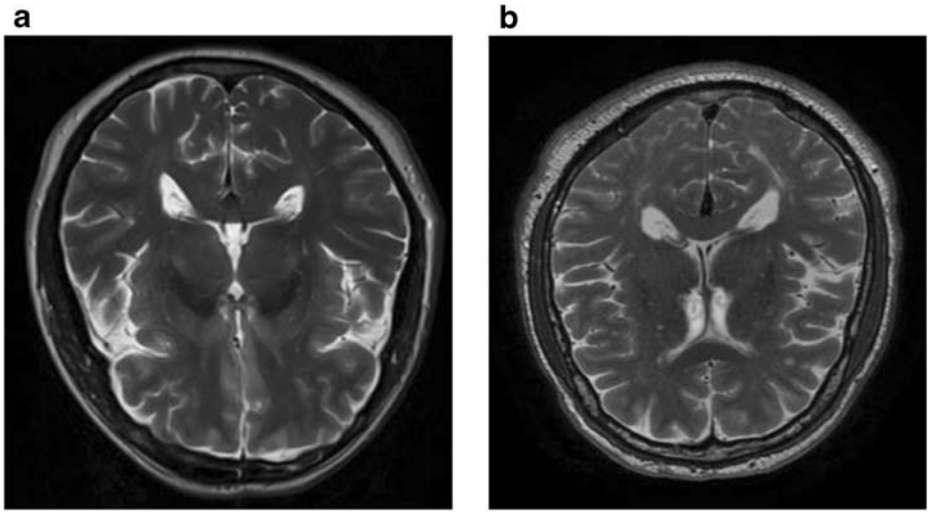

Fig. 8 MR images of (**a**) Parkinson's disease patients (**b**) healthy individuals

With augmented training images, we use the transfer learning technique to train Alex-Net with different model training parameter values. In Addition to replacing and retraining the classifier on the pre-trained Alex-Net, we then use transfer learning to further refine the weights of the pre-trained network by starting backpropagation based on the current training datasets.

As stated in the prior section on the proposed method, the MR images are trained with the appropriate hyper-parameters. The first part of the convolution gets the stored information and carries out the convolution operation of MR images with filters. The extracted features are visualized at each convolution layer as those derived from pixels, separating from the pixels, and acquiring the shapes from the edges. Finally, complicated area as attributes differentiates between the types of the two groups. The

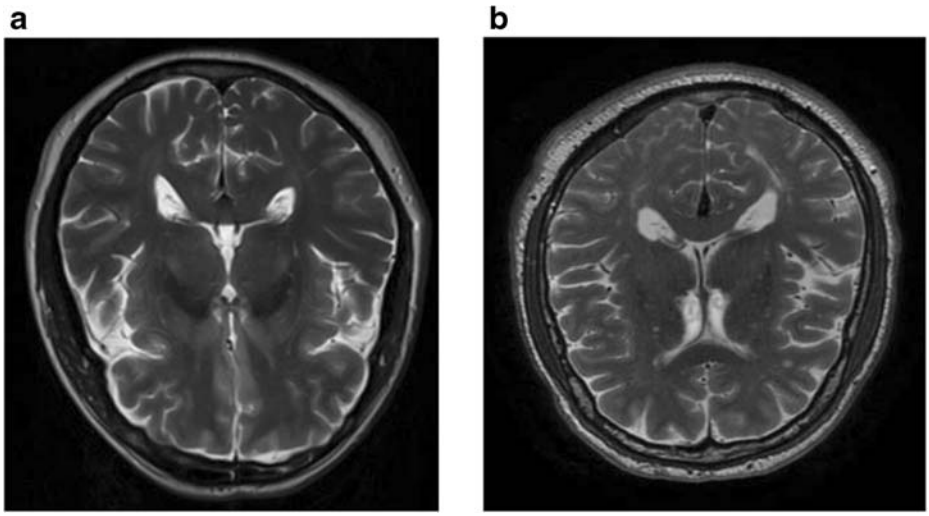

Fig. 9 Improved bilateral filtered images of (**a**) Parkinson's disease patients (**b**) healthy individuals

Table 7 Number of MR images used for training, validation, and testing

Sr. No.	Parkinson's disease patients	Healthy individuals
Original Training images	180	180
Validation images	36	36
Testing images	36	36
Augmented Training Images	2100	2100

first and second convolution layer feature maps are displayed in Fig. 10a, b. The people who suffer from the Parkinson's disease tend to show a considerable amount of reduced grey matter strength in contrast to the normal people with stratum area alterations. The difference in intensity levels and diverse regions the network learns are recorded as Classification process features.

Accordingly, the fully connected layers obtain the high-level mix of features acquired from the previous layers. The CNN approach is interpreted by visualizing the weights. The fully updated layer is compatible with both HC and PD groups, and the feature maps for each group are shown in Fig. 10c . It is found that the final fully connected layer retrieves the discriminatory features that could catch the structural differences between the two classes.

It is observed from the above figures that the first layer acts as a cluster of various edge detectors. As shown in Fig. 10a, the information of the given input image is preserved by the activations. While moving to the higher levels, the activation keeps less information, and in the deeper layers, the information becomes abstract, which is not readable. Figure 10b shows blank filters that represent the nonexistence of filter in the input sample.

To assess the efficiency of the classification network, the accuracy and loss caused during the training and validation stage are tracked. The learning process, as an accuracy chart, is displayed in Fig. 11a. Figure 11b shows the loss occurred during the training and the testing stage. The loss or error rates in the classification of MR images of healthy and PD patients remain constant, close to the learning curve from 150 iterations. The learning process converges from 100 iterations onwards to be stable.

The determination of accuracy value relating to the sum-total of accurate predictions is derived out by analyzing the ratio amongst true positive and false negative. The proportionate value of the true positive is analyzed for giving the sensitivity of the procedure of classification. The specificity is thus measured by the percentage of

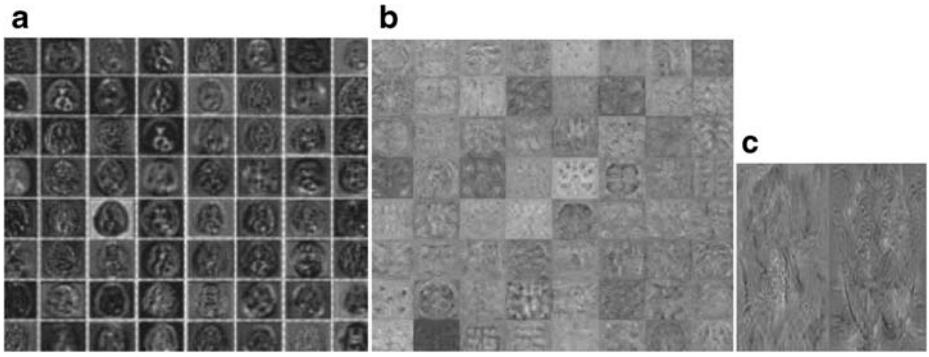

Fig. 10 MR image feature maps corresponding to different layers (**a**) at first convolution layers (**b**) at the second convolution layer (**c**) at the Output layer

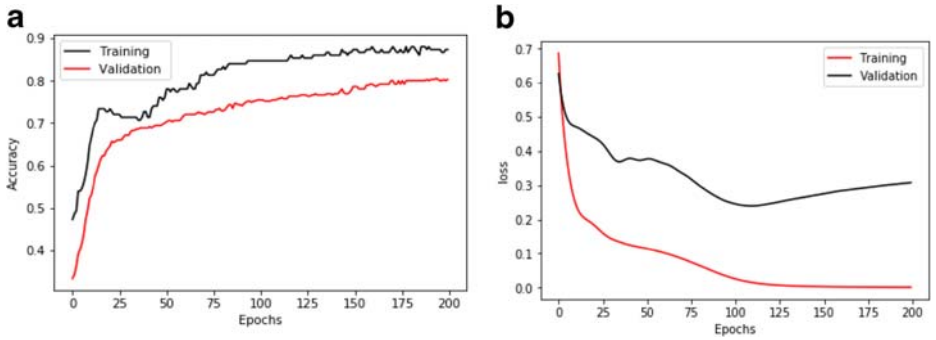

Fig. 11 (**a**) Accuracy (**b**) loss value overtraining and validation MR images during the learning process

true negative where the determination of the character of negative situations is carried out. The proposed approach demonstrates accuracy, sensitivity, and specificity values of 89.23%, 90.27%, and 89.03%, respectively.

The approach proposed is finally analyzed using seventy-two MR images of HC and PD patients with a proportion of 1:1. The test images are only used once for a single said purpose. Every image was evaluated considering the final model to reach a score that represents the chances of an image being classified amongst HC and PD. The score thus achieved is in the continuity of the values ranging from 0 to 1. The receiver operating function (ROC) curve is processed for measuring network performance in both categories. As shown in Fig. 12, the area under the curve (AUC) from the ROC curve is estimated as 0.9723. In this manner, a satisfactory exhibition of the trained model is accomplished that is demonstrated by the ROC curve tending to perform a better characterization for each class.

It can be well ascertained from the higher level of classification accuracy attained from the test set that the model is not over-fitted on the training set. The only case in which the chances of over-fitting may occur is due to a small number of MR images and can be called one of the limitations of this study. We have utilized the standard data augmentation techniques (GAN) for accelerating the network's adaptation. In other words, increasing the samples of the training improves the strength of the classification method. Thus, the proposed approach differentiates HC and PD subjects with a higher degree of accuracy for the objective classification of MR

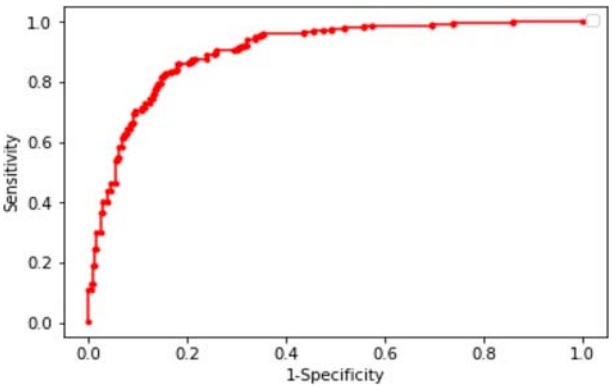

Fig. 12 ROC analysis for Parkinson's disease classification

Table 8 Deep CNN performance with and without data augmentation

Methodology	Accuracy (%)	Specificity (%)	Sensitivity (%)
Deep-CNN(transfer-learned) with data augmentation	89.23	90.27	89.03
Deep-CNN (transfer-learned) without data augmentation	86.63	86.67	84.83

images without the extraction and selection of complicated image features. Furthermore, by adjusting the last layers of the pre-prepared Alex-Net, the transfer of information from the original images to the MR images is conceivable.

The classification results achieved by employed in Deep CNN functioning upon the image synthesis vis-à-vis data augmentation are represented in Table 8. The classification results can be improved with the help of image synthesis models and techniques rather than original images. Further, it is pertinent to mention here that by introducing the synthetic images, the SD outcomes are decreased to a considerable amount in the validation of accuracy, sensitivity and specificity, and AUC. The Deep CNN model using synthetic images helps the CNN classification method to achieve up to 3% higher accuracy compared to the effects of classification using only original images. However, with more digital images, greater accuracies are obtained.

Many studies have explored machine learning techniques to diagnose PD using MR images from the PPMI database. Nevertheless, the preliminary results demonstrated by the existing techniques indicate an accuracy of 40–87%, while our proposed method offers accuracy measure up to 89.23%. These findings reflect our learner's dominance over other related approaches in terms of effectiveness. Our findings indicate that deep learning of magnetic resonance imaging data may enhance assessment of disease development and provide a cost-effective means of potentially encouraging participation in clinical trials with individuals expected to advance in a limited period. The paper proposes a deep learning system for designing imaging biomarkers to predict the progression of Parkinson's disease. The same paradigm can be implemented to predict the development of cognitively healthy subjects to moderate cognitive deterioration that can be helpful in preclinical studies of Parkinson's disease and promote the screening of participants in clinical trials.

4 Conclusions

After Alzheimer's disease, Parkinson's disease is the second predominant neurological disorder. Currently, there is no cure but early detection of PD that provides better and more efficient care for patients. In this study, we have employed a framework to categorize MR images of healthy people and people with Parkinson's disease using a combination of data augmentation and transfer learned Alex-Net. This work attempts at classifying MR images in the HC and PD contents with the application of deep transfer learning. The classification images have been sourced from the PPMI repositories. The MR images are refined by normalizing before being fed with an advanced bilateral filter. For categorization, an Alex-Net model has been preferred as a convolution neural network. To carry out the classification of the HC and PD subjects, the weights of the pre-trained model are used by fine-tuning the last fully connected layer with appropriate hyperparameters. By applying the synthesized images to the training set, the main classification metrics such as accuracy, sensitivity, and specificity are 89.23%, 90.27%, and 89.03%, which are increased by 2.6%, 3.6%, and 4.2%compared to the CNN model respectively. The classified analysis shows that the Transfer-Alex Net performs well in the

classification task, and its capacity would be further strengthened by the suggested GAN-based data augmentation method. The proposed data augmentation approach based on GANs will effectively produce high-quality MR images, contributing to the classification model performance improvement. This work provides a valuable reference in the field of clinical image evaluation based on Deep Learning. The deep learning model proposed here shows an excellent differential fluency by reporting the 0.9723 AUC value from the ROC curve. The scope of the proposed model can be further expanded to bring the AlexNet model's fine-tuning within its ambit for obtaining superior working results. It can be said that with the continuous developments in the deep learning techniques, diagnosing Parkinson's disease objectively will not be a challenging task in the coming future for the clinicians.

References

1. Aarsland D (2016) Cognitive impairment in Parkinson's disease and dementia with Lewy bodies. Parkinsonism Relat Disord. https://doi.org/10.1016/j.parkreldis.2015.09.034
2. Abas MH, Ismail N (2018) VGG16 for plant image classification with transfer learning and data augmentation. Int J Eng Technol 4:90–94
3. Abos A, Baggio HC (2017) Discriminating cognitive status in Parkinson's disease through functional connectomics and machine learning. Sci Rep 7:45347
4. Abu M, Qawaqneh A, Barkana BD (2019) Utilizing CNNs and transfer learning of pre-trained models for age range classification from unconstrained face images. Image Vis Comput 88:41–51. https://doi.org/10.1016/j.imavis.2019.05.001
5. Adeli E, Shi F et al (2016) Joint Feature-Sample Selection and Robust Diagnosis of Parkinson's Disease from MRI Data. NeuroImage. https://doi.org/10.1016/j.neuroimage.2016.05.054
6. Amoroso N, Rocca L (2018) Complex networks reveal early MRI markers of Parkinson's disease. Med Image Anal 48:12–24
7. Babu GS, Suresh S et al (2014) A novel PBL-McRBFN-RFE approach for identification of critical brain regions responsible for Parkinson's disease. Expert Syst Appl 41(2):478–488
8. Chaudhuri KR, Schapira AH et al (2009) Non-motor symptoms of Parkinsons disease: dopaminergic pathophysiology and treatment. Lancet Neurol. https://doi.org/10.1016/S1474-4422(09)70068-7
9. Cigdem O, Yilmaz A et al (2018) Comparing the performances of PDF and PCA on Parkinson's disease classification using structural MRI images. IEEE, Piscataway
10. Dawud AM, Yurtkan K, Oztoprak H (2019) Application of deep learning in neuroradiology: Brain haemorrhage classification using transfer learning. Comput Intell Neurosci. https://doi.org/10.1155/2019/4629859
11. Deepak S, Ameer PM (2019) Brain tumor classification using deep CNN features via transfer learning. Comput Biol Med. https://doi.org/10.1016/j.compbiomed.2019.103345
12. Focke NK, Helms G et al (2011) Individual voxel-base subtype prediction can differentiate progressive supranuclear palsy from idiopathic Parkinson syndrome and healthy controls. Hum Brain Mapp. https://doi.org/10.1002/hbm.21161
13. Fox SH, Katzenschlager R et al (2011) The movement disorder society evidence-based medicine review update: treatments for the motor symptoms of Parkinson's disease. Mov Disord. https://doi.org/10.1002/mds.23829
14. Frid-Adar M, Diamant I, Klang E, Amitai M, Goldberger J, Greenspan H (2018) GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. Neurocomputing 321:321–331. https://doi.org/10.1016/j.neucom.2018.09.013
15. Gavrilov AD, Jordache A, Vasdani M (2019) Preventing model overfitting and underfitting in convolutional neural networks. Int J Softw Sci Comput Intell 10(4):19–28. https://doi.org/10.4018/ijssci.2018100102
16. Ghafoorian M, Karssemeijer N et al (2017) Location sensitive deep convolutional neural networks for segmentation of white matter hyperintensities. Sci Rep 7(1):5110
17. Ghazi MM, Ghazi Yanikoglu B (2017) Plant identification using deep neural networks via optimization of transfer learning parameters. Neurocomputing 235:228–235
18. Gil D et al (2009) Diagnosing Parkinson by using artificial neural networks and support vector machines. Glob J Comput Sci Technol 9:63–71

19. Gu J, Wang Z et al (2018) Recent advances in convolutional neural networks. Pattern Recognit. https://doi.org/10.1016/j.patcog.2017.10.013
20. Hopes L, Grolez G et al (2016) Magnetic resonance imaging features of the nigrostriatal system: biomarkers of Parkinson's disease stages? PLoS One 11(4):e0147947
21. Iqbal T, Ali H (2018) Generative adversarial network for medical images (MI-GAN). J Med Syst 42:11. https://doi.org/10.1007/s10916-018-1072-9
22. Islam J, Zhan Y (2017) A novel deep learning-based multi-class classification method for Alzheimer's disease detection using brain MRI data. https://doi.org/10.1007/978-3-319-70772-3_20
23. Jiang W, Siddiqui S (2018) Hyper-parameter optimization for support vector machines using stochastic gradient descent and dual coordinate descent. EURO J Comput Optim. https://doi.org/10.1007/s13675-019-00115-7
24. Kassani SH, Kassani PH, Wesolowski MJ, Schneider KA, Deters R (2019) Breast cancer diagnosis with transfer learning and global pooling. arXiv preprint arXiv:1909.11839
25. Kaur T, Gandhi TK (2020) Deep convolutional neural networks with transfer learning for automated brain image classification. Mach Vis Appl 31(3). https://doi.org/10.1007/s00138-020-01069-2
26. Kaur T, Gandhi TK (2020) Deep convolutional neural networks with transfer learning for automated brain image classification. Mach Vis Appl 31:20
27. Kaur S, Aggarwal H, Rani R (2019) Diagnosis of Parkinson's disease using principal component analysis and deep learning. J Med Imaging Health Inf 9(3):602–609
28. Kaur S, Aggarwal H, Rani R (2020) Hyper-parameter optimization of a deep learning model for prediction of Parkinson's disease. Springer Mach Vis Appl 31:32. https://doi.org/10.1007/s00138-020-01078-1
29. Kaur S, Aggarwal H, Rani R (2021) MR image synthesis using generative adversarial networks for Parkinson's disease classification. Advances in intelligent systems and computing. 1164. https://doi.org/10.1007/978-981-15-4992-2_30
30. Kokil P, Sudharson S (2019) Automatic detection of renal abnormalities by Off-the-shelf CNN features. IETE J Educ 60:14–23
31. Kolar Z, Chen H, Luo X (2018) Transfer learning and deep convolutional neural networks for safety guardrail detection in 2D images. Autom Constr 89:58–70
32. Kolar Z, Chen H, Luo X (2018) Transfer learning and deep convolutional neural networks for safety guardrail detection in 2D images. Autom Constr 89:8–70. https://doi.org/10.1016/j.autcon.2018.01.003
33. Konidaris F, Tagaris T, Sdraka M, Stafylopatis A (2019) Generative adversarial networks as an advanced data augmentation technique for MRI data. VISIGRAPP 2019 - Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications 5, pp 48–59. https://doi.org/10.5220/0007363900480059
34. Liu S, Tian G, Xu Y (2019) A different scene classification model is combining Res-Net based transfer learning and data augmentation with a filter. Neuro Comput J 338:191–206
35. Mak E, Su L, Williams GB et al (2017) Longitudinal whole-brain atrophy and ventricular enlargement in nondemented Parkinson's disease. Neurobiol Aging 55:78–90
36. Mikołajczyk A, Grochowski M (2018) Data augmentation for improving deep learning in image classification problem. 2018 International Interdisciplinary PhD Workshop, IIPhDW 2018, pp 117–122. https://doi.org/10.1109/IIPHDW.2018.8388338
37. Motta D, Bandeira S, Souza M (2020) Optimization of convolutional neural network hyperparameters for automatic classification of adult mosquitoes. PLoS One 15(7):1–30. https://doi.org/10.1371/journal.pone.0234959
38. Naseer A, Rani M et al (2019) Refining Parkinson's neurological disorder identification through deep transfer learning. Neural Comput Appl. https://doi.org/10.1007/s00521-019-04069-0
39. Pansombut T, Wikaisuksakul S et al (2019) Convolutional neural networks for recognition of lymphoblast cell images. Hindawi Comput Intell Neurosci. https://doi.org/10.1155/2019/7519603
40. Pereira CR, Papa JP et al (2017) Convolutional neural networks applied for Parkinson's disease identification. https://doi.org/10.1007/978-3-319-50478-019
41. Pereira S, Pinto A, Alves V, Silva CA (2016) Brain tumor segmentation using convolutional neural networks in MRI images. IEEE Trans Med Imaging 35(5):1240–1251
42. Petersen K, Nielsen M, Diao P, Karssemeijer N, Lillholm M (2014) Breast tissue segmentation and mammographic risk scoring using deep learning. In an international workshop on digital mammography. Springer, Berlin, pp 88–94
43. Poewe W, Seppi K et al (2017) Parkinson's disease. Nat Rev Dis Primers. https://doi.org/10.1038/nrdp.2017.13
44. Provost JS, Hanganu A, Monchi O (2015) Neuroimaging studies of the striatum in cognition part I: healthy individuals. Front Syst Neurosci 9:140

45. Rana B, Juneja A et al (2015) Graph-theory-based spectral feature selection for computer-aided diagnosis of Parkinson's disease using T1-weighted MRI. Int J Imaging Syst Technol 25(3):245255
46. Rana B, Juneja A et al (2015) Regions-of-interest based automated diagnosis of Parkinson's disease using T1-weighted MRI. Expert Syst Appl 42:4506–4516
47. Sakr GE, Mokbel M et al (2016) Comparing Deep Learning And Support Vector Machines for Autonomous Waste Sorting. IEEE International Multidisciplinary Conference on Engineering Technology
48. Salvatore C, Cerasa A et al (2014) Machine learning on brain MRI data for differential diagnosis of Parkinson's disease and Progressive Supranuclear Palsy. J Neurosci Methods 222:230–237
49. Saranyaraj D, Manikandan M, Maheswari S (2018) A deep convolutional neural network for the early detection of breast carcinoma concerning hyperparameter tuning. Multimed Tools Appl. https://doi.org/10.1007/s11042-018-6560-x
50. Schlegl T, Seeböck P, Waldstein SM, Langs G, Schmidt-Erfurth U (2019) f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks. Med Image Anal 54:30–44. https://doi.org/10.1016/j.media.2019.01.010
51. Shams S, Platania R et al (2018) Deep generative breast cancer screening and diagnosis. Springer Nature, Berlin. https://doi.org/10.1007/978-3-030-00934-2_95
52. Shinde S, Prasad S (2019) Predictive markers for Parkinson's disease using deep neural nets on neuromelanin sensitive MRI. NeuroImage Clin. https://doi.org/10.1016/j.nicl.2019.101748
53. Singh G, Samavedham L (2015) Unsupervised learning-based feature extraction for differential diagnosis of neurodegenerative diseases: A case study on the early-stage diagnosis of Parkinson's disease. J Neurosci Methods 256:30–40
54. Sivaranjini S, Sujatha CM (2019) Deep learning-based diagnosis of Parkinson's disease using a convolutional neural network. Multimed Tools Appl. https://doi.org/10.1007/s11042-019-7469-8
55. Soltaninejad S, Cheng I, Basu A (2018) Towards the identification of Parkinson's disease using only T1 MR images. ArXiv, abs/1806.07489
56. Tabano K, Lee P et al (2016) Brain delivery of drug and MRI contrast agent: Detection and quantitative determination of brain deposition of CPT-Glu using LC-MS/MS and Gd-DTPA using magnetic resonance imaging. Mol Pharm. https://doi.org/10.1021/acs.molpharmaceut.5b00607
57. Talo M, Baloglu UV, Yıldırım O, Acharya UR (2018) Application of deep transfer learning for automated brain abnormality classification using Mr. Images. Cognit Syst Res 54:176–188. https://doi.org/10.1016/j.cogsys.2018.12.007
58. Wang D, Lu Z et al (2019) Cellular structure image classification with small targeted training samples. https://doi.org/10.1101/544130
59. Wang SH, Xie S et al (2019) Alcoholism identification based on an AlexNet transfer learning model. Front Psychiatry. https://doi.org/10.3389/fpsyt.2019.00205
60. Weng Y, Zhou H (2019) Data augmentation computing model based on generative adversarial network. IEEE Access. https://doi.org/10.1109/ACCESS.2019.2917207
61. Wu K, Zhang D (2018) Learning acoustic features to detect Parkinson's disease. Neurocomputing. https://doi.org/10.1016/j.neucom.2018.08.036
62. Xu J, Zhang M (2019) Use of magnetic resonance imaging and artificial intelligence in studies of diagnosis of Parkinson's disease. ACS Chem Neurosci. https://doi.org/10.1021/acschemneuro.9b00207
63. Yamashita R, Nishio M (2018) Convolutional neural networks: an overview and application in radiology. Insights Imaging 9:611–629
64. Yamashita R, Nishio M, Do RKG, Togashi K (2018) Convolutional neural networks: an overview and application in radiology. Insights Imaging 9(4):611–629. https://doi.org/10.1007/s13244-018-0639-9
65. Yoo YJ (2019) Hyperparameter optimization of deep neural network using univariate dynamic encoding algorithm for searches. Knowl Based Syst 178:74–83. https://doi.org/10.1016/j.knosys.2019.04.019
66. Zhen X, Chen J et al (2017) Deep convolutional neural network with transfer learning for rectum toxicity prediction in cervical cancer radiotherapy: a feasibility study. Phys Med Biol. https://doi.org/10.1088/1361-6560/aa8d09

Affiliations

Sukhpal Kaur[1] · **Himanshu Aggarwal**[1] · **Rinkle Rani**[2]

Himanshu Aggarwal
himanshu.pup@gmail.com

Rinkle Rani
raggarwal@thapar.edu

[1] Department of Computer Science and Engineering, Punjabi University, Patiala 147002, India

[2] Computer Science and Engineering Department, Thapar Institute of Engineering and Technology, Patiala 147004, India