IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 3 December 2022, accepted 12 December 2022, date of publication 3 January 2023, date of current version 12 January 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3233969

RESEARCH ARTICLE

Deep Transfer Learning Based Parkinson's Disease Detection Using Optimized Feature Selection

SURA MAHMOOD ABDULLAH[1], THEKRA ABBAS[2], MUNZIR HUBIBA BASHIR[3], ISHFAQ AHMAD KHAJA[3], MUSHEER AHMAD[3], NAGLAA F. SOLIMAN[4], AND WALID EL-SHAFAI[5,6]

[1]Department of Computer Sciences, University of Technology, Baghdad 10066, Iraq
[2]Department of Computer Sciences, College of Science, Mustansiriyah University, Baghdad 14022, Iraq
[3]Department of Computer Engineering, Jamia Millia Islamia, New Delhi 110025, India
[4]Department of Information Technology, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh 11671, Saudi Arabia
[5]Security Engineering Laboratory, Computer Science Department, Prince Sultan University, Riyadh 11586, Saudi Arabia
[6]Department of Electronics and Electrical Communications Engineering, Faculty of Electronic Engineering, Menoufia University, Menouf 32952, Egypt

Corresponding author: Musheer Ahmad (musheer.cse@gmail.com)

This work was supported by the Deanship of Scientific Research, Princess Nourah Bint Abdulrahman University, Riyadh, Saudi Arabia, under Project PNURSP2023R66.

ABSTRACT Parkinson's disease (PD) is one of the chronic neurological diseases whose progression is slow and symptoms have similarities with other diseases. Early detection and diagnosis of PD is crucial to prescribe proper treatment for patient's productive and healthy lives. The disease's symptoms are characterized by tremors, muscle rigidity, slowness in movements, balancing along with other psychiatric symptoms. The dynamics of handwritten records served as one of the dominant mechanisms which support PD detection and assessment. Several machine learning methods have been investigated for the early detection of this disease. But most of these handcrafted feature extraction techniques predominantly suffer from low performance accuracy issues. This cannot be tolerable for dealing with detection of such a chronic ailment. To this end, an efficient deep learning model is proposed which can assist to have early detection of Parkinson's disease. The significant contribution of the proposed model is to select the most optimum features which have the effect of getting the high-performance accuracies. The feature optimization is done through genetic algorithm wherein K-Nearest Neighbour technique. The proposed novel model results into detection accuracy higher than 95%, precision of 98%, area under curve of 0.90 with a loss of 0.12 only. The performance of proposed model is compared with some state-of-the-art machine learning and deep learning-based PD detection approaches to demonstrate the better detection ability of our model.

INDEX TERMS Parkinson's disease, neurological disorder, handwritten records, transfer learning, deep learning.

I. INTRODUCTION

Parkinson's disease (PD) is an incurable neurological disorder that is caused due to the decrement of dopamine levels in a human brain. Dopamine is a neurotransmitter that helps send messages to basal ganglia; the part of the brain which is responsible for movement and coordination control. Dopamine levels decrease when the cells in basal ganglia that are responsible for the synthesis of dopamine, die or become impaired. Parkinson's symptoms may include tremors, restrictive or slowness of movement (Bradykinesia), compromised balance, impaired posture, involuntary movements (dyskinesia), stiff muscles, and speech and writing changes [1]. The Parkinson's disease can prove to be complex

The associate editor coordinating the review of this manuscript and approving it for publication was Mostafa M. Fouda.

to diagnose since there aren't much clinical tests such as blood tests involved. PD is most common in people above 60, but the disease can start off early and not be diagnosed until too late. If the disease is detected at an earlier stage, it becomes easy to manage the symptoms and delay the deterioration caused by the disease [2]. The early onset of PD may result in finger tremors and halts during speech and movement. Finger tremors result in changes in handwriting and thus people with Parkinson's tend to have small and cramped handwriting. This handwriting is termed as ''Mycrographia'' and can prove to be crucial in the early detection of PD. Patients can be diagnosed with PD by finding out the presence of particular patterns in their handwriting, indicating mycrographia or other deformations.

Deep learning methods have excelled at classification problems lately. Deep learning algorithms such as CNNs are proven to have state-of-the-art accuracies in classification tasks. CNNs have been used widely for classification of images, audios, or videos. CNNs extract unique patterns from given data to use for final classification. The ease of availability and usage makes CNNs an excellent choice for classification problems. Prior research has also proven that deep learning algorithms can work more efficiently than the machine learning ones because transfer learning can be applied. Transfer learning makes use of pre-trained CNNs with new use cases, and then one or multiple layers are added at the end [3]. Some of the deep learning architectures include ResNets, EfficientNets, MobileNets etc. Deep learning approaches have been used in medical field for quite some time now. Deep learning models can interpret medical data like X-Ray images, and MRI scans which proves advantageous for diagnosis. With the advancement of AI over the past decade, its application in medical field has also encountered tremendous growth. In medical field, the application of AI is of great potential and is currently being used to diagnose/predict a variety of diseases. Studies indicate that deep learning methods can be far superior in comparison to other high-performing algorithms [4]. Using deep learning approaches for PD detection using handwriting data can prove beneficial as deep learning methods have reached excellent accuracies. A deep learning architecture can be fed with image data comprised of handwritten samples from affected and non-affected people and results can be acquired.

In our proposed model, we proffered to use deep transfer learning models, genetic algorithm, and k-nearest neighbours' technique to develop a system that efficiently detects the patients as healthy or suffering with Parkinson's disease by extracting features from handwritten records.

The rest of the paper is systemized as follows: Section II begins with highlighting the related work that has been done in recent years for Parkinson's detection. Section III provides the description of materials and methods adopted for the proposed work. The dataset used for this study is described followed by the explanation of the proposed framework in Section IV. Whereas the experimental results and analysis of the proposed model including the comparative analysis with existing models is done in section V. Finally section VI recapitulates the proposed work as conclusion.

II. RELATED WORKS

With the advances in deep/machine learning and AI technologies over the past couple of decades these technologies have gained quite a bit of fame in various fields. Of late, the deep learning methodologies have been used extensively in the medical world as well. The rapid growth and research in the AI and deep learning areas have increased its market tremendously in the medical field for diagnoses and prognosis of various diseases. Various studies have been conducted to diagnose PD using a variety of datasets. Also, there are a lot of symptoms that have been studied for detection of PD like olfactory loss, walking patterns, speech patterns, handwriting tests and other motor skill tests.

Recently, Fang [5] proposed improved KNN algorithm entropy for the detection of PD. The UCI dataset was considered for the study. To estimate the efficiency of this improved algorithm, the comparative analysis of already existing approaches was carried out. The KNN (k-nearest neighbors), Random Forest algorithm, and Naïve Bayes algorithm were considered to verify the feasibility of improved algorithm. 5-fold cross validation scheme was used. It was observed that among the existing algorithms when compared with traditional methods, the improved KNN algorithm based on entropy weight showed significant increase in the accuracy.

Kuplan et al. [6] adopted a novel method for the classification of symptoms of PD using MRI scans. The main goal of the study was to explore more clinical data to elaborate the efficacy of artificial intelligence for better detection of the PD disease. Three classification tasks were carried out that focused on stage and major symptoms of Parkinson's. The symptoms included clinical stage, dementia status, and motor skills. After characterizing each and every patient based on their current condition, a novel model was introduced which ran on the combined principles of handcrafted textural feature engineering, multiple feature selectors patch-based learning and IMV. The model showed outstanding performance for all classification tasks.

Gazda et al. [7] also recently proposed an ensemble of deep learning architectures for the detection of PD from offline handwriting. For this purpose, they used 2 datasets namely, PaHaW and NewHandPD. To improve the generality of the model, transfer learning was considered. The ensemble classifier created, consisted of 5 CNN models. Since the PaHaW dataset consists of 8 separate handwritten tasks, the prediction accuracy for all those tasks was calculated. Prediction accuracy was calculated for each specific task via each separate CNN as well as ensemble classifier and then compared. The authors, in their work presented an ensemble of multiple CNNs for the diagnosis of Parkinson's disease. To reduce the computational cost altogether, the approach of multiple fine tuning is adapted. This approach provided competitive results. A detailed study and comparison with other works was provided.

Mohaghegh and Gascon [8] proposed a vision transformer (ViT) for handwritten data to detect Parkinson's disease based on spiral and meander drawings. Their model comprised of three different layers; a pytorch layer (base model), a dropout layer (dropout value=0.1) and finally a linear layer as the classifier. They made use of DeiT; pre-trained on ImageNet and self-supervised with DINO, as the base model. DeiT refers to data-efficient image transformer which is a type of vision transformer used for image classification jobs. Using 5-fold validation schemes, an accuracy of 92.37% was achieved together with a standard deviation of 0.013.

Fratello et al. [9], carried out a study in which data was collected from 9 PD patients and 22 healthy people. The data was collected in collaboration with Casa di Cura Le Terrazze institute where every subject, but one was right-handed. The participants ranged from 25 to 60 in age. An application was created in this study that helped the authors to record data from subjects via a tablet for the recognition of patterns in handwritten data in order to diagnose PD post which; for the classification purposes, three models were proposed for three different handwritten data. Highly discriminate features were extracted using the Mann–Whitney test [10]. For the first 2 models, a linear SVM approach was considered while as for the third model a medium KNN was used. It was observed that the first 2 models provided an accuracy of 71.6% and 75.5% respectively. The third model achieved an accuracy of 77.5%.

Loh et al. [11] used the Electroencephalography (EEG) signals for the detection of Parkinson's disease. The approach was considerably different from other approaches that generally used the handwritten data for the PD analysis. Gabor transforms were used to obtain the spectrograms from EEG signals after splitting them in half. Thus, from each of the EEG signals, two spectrograms were collected. A 2D-CNN based architecture was proposed to classify the spectrogram signals of control group, PD patients with medication and PD patients without medication. The study included 4 kinds of experiments, first to classify all the three groups, and the other three for a binary classification of the three groups in different combinations e.g., control group and PD with medication, control group and PD without medication and lastly PD with medication and PD without medication. The last layer of the proposed model differed according to the output of each experiment. For the first experiment where the last layer could have 3 outputs, a SoftMax function was used while as for the binary classification a sigmoid activation was used. The results were examined by applying a 10-fold cross-validation scheme. For experiment 1 the accuracy was the highest, with a value of 99.46%. The 2nd, 3rd and 4th experiments had the accuracy values of 99.44%, 98.84% and 92.60% respectively.

Chakraborty et al. [12] developed a system to observe some patterns in spirals and waves sketched by patients with PD and eventually detect the disease. For this system, voting ensemble classifiers were used along with the two-dimensional CNNs were considered for analyzing the patterns in sketches and detecting the Parkinson's disease. To validate and train their proposed system; a dataset with 204 images was considered. This dataset had 102 images of spiral sketches and an equal number of wave sketches. The complete study comprised of 3 different sections. Section 1 comprises of a generator section which served the spiral and wave images of a specific patient. Section 2 defines the CNN architecture that is in charge for producing feature representations. After the features are extracted, they are sent through the final dense layer for getting the predictions for every image. Section 3 defines the type of meta-classifiers that predict the probabilities and make final predictions. Logistic regression (LR) and Random Forest classifier (RF) are used for the development of the meta-classifiers. Both classifiers take input as prediction probabilities of wave and spiral CNNs and produce the outcome as Parkinson's disease or healthy. The model achieved an accuracy of 93.3%, an average recall of 94% and average precision of 93.5%. The work successfully achieved a multistage classifier system for the detected of the Parkinson's disease with the help of spiral and meander sketches. They leveraged two systems namely ensemble voting classifier and CNNs which resulted in an average F1 score of 93.94%.

Nõmm et al. [13], presented a research which was based on a group of 34 people equally divided into sub-groups of PD patients and healthy people, therefore having 17 subjects per group. The groups have a mean of 69 years along with the standard deviation of 4 years. For this research, special software was constructed for ipad Pro equipped with stylus to collect different data from the patients. The data was in the form of writing and drawing tests. AlexNet architecture was proposed after necessary data enhancement and augmentation. The final accuracy was observed to be 93%. The experimental results of their research included the application of deep learning-based networks in the area of Parkinson's diagnosis.

Tuncer et al. [14] used voice signals to detect PD. A fusion of SVD (singular value decomposition) and minimum average maximum tree (MAMa) is proposed to find out unique features from the voice signals. In the pre-processing phase the authors developed a new feature signal from three levels of MAMa tree. After the feature signal was generated, SVD was applied to it for the extraction of features. Relief feature selection method id implemented to extract about 50 different features. For classification purposes, the KNN algorithm and 10-fold cross validation are used. The experimental results show that with the KNN classifier, an accuracy of 92.46% was achieved. The proposed algorithm in the study can be used for distinct signals like ECG, EEG, PCG and EMG as well and detect several other diseases.

Das et al. [15], in their research, tested the performance of multiple CNNs on 2 datasets. The datasets in consideration were taken from Kaggle's repository. The second dataset in question was provided by the authors of [16]. The images consisted of spiral drawings, wave drawings, in the first dataset and hand-drawn cube and triangle images in dataset 2. Two approaches were considered for the study. In the first approach, the CNNs like

VGG19, ResNet50, MobileNet-v2, Inceptionv3, Xception, and Inception-ResNet-v2 were trained from scratch on both the datasets whereas in the second approach Transfer learning was applied. The authors investigated the usage of deep convolutional neural networks in the detection of Parkinson's with the help of hand drawn images. The fine tuning gave better results as compared to approach 1 and 2 which comprised of training from scratch and using two shallow neural networks respectively. The models ResNet50 and MobileNet-V2 were explored that outperform other leading CNNs.

Johri and Tripathi et al. [17], developed a classifier made up of two modules, for cheap and efficient diagnosis of Parkinson's disease. Two separate datasets namely VGFR dataset and voice impairment dataset were used. The VGFR dataset is composed of the signals recorded for the reactions of subjects to the vertical ground force. The voice impairment dataset was contributed by the Max little university of Oxford and contains voice measurements of 91 people. The deep learning method that is proposed is used to detect two Parkinson's symptoms i.e., Gait and speech impairment. The proposed model has 2 modules; the first one being the VGFR spectrogram-detector which works on distorted walking styles and the second one is the voice impairment classifier which is established on the basis of speech distortion of the PD subjects. For the first module, the sensor readings per patient are converted to a spectrogram. This spectrogram shows a pattern which is acquired with the help of these signal values. The 2D spectrogram images are used as the inputs for the CNN. It was observed that for the VGFR module the accuracy came out to be 88.17% and for the voice impairment module accuracy was 89.15%. In their study, the authors proposed a novel system based on the principles of bi-directional GRUs and 1D convolution. This approach was to detect the distinguishing patterns in the handwritten material acquired from people with and without PD. Promising performance values were achieved by the authors.

Tuncer and Dogan [18] implemented a novel multi pooling technique for classification that used 8 pooling methods, commonly called and octopus-based method to solve 3 classification problems. These were Gender, PD and gender + PD classification problems. The aim of using the octopus-based method is to achieve a lightweight nature since this method doesn't use any algorithm to optimize or update any weights during training. Concepts like SVD, NCA have been utilized by the authors for feature extraction and selection respectfully. Several other algorithms like SVM, KNN, logistic regression, and decision tress have been exploited for the classification phase. The results showed that the KNN solved the PD problem the best with a high accuracy. The authors also managed to solve all the three problems with only 32 features.

Khatamino et al. [19] classified HW dataset with the help of CNN-based approach. The dataset was split into separate sections of spiral and spiral images. Spiral drawings were drawn by using 2D features of the dataset. Rescaling along with normalizing was done on these signals, for square transformation. Data was converted into square matrix and then fed to the proposed CNN model. Finally, the normalized square matrices were sketched. Two datasets were used; SST and DST and both were made of 72 images, where 57 entries belonged to PD patients and 15 entries belonged to normal subjects. Early stopping was used to reduce the value of loss to prevent the drop in validation accuracy. A CNN based on LOOVC and K- fold cross validation was proposed that efficiently responds to the features, extracted. Another achievement of the research included getting higher accuracies with fewer features. The model achieved over 88% accuracy.

A. LITERATURE GAP

It is important to address that PD is a life altering disease and can cause long term suffering. It's crucial to detect it with higher precision and accuracy. While the above-mentioned papers present numerous techniques and methods to collect data from control subjects and PD subjects, further research is required to find out the best working algorithms to finally diagnose PD with utmost precision and accuracy. Majority of the research is diverted towards the working of various leading deep learning models. In addition, common and generally used algorithms like K-nearest neighbours, Naïve bayes and Random forests have been exploited for the problem until now. However, there is a scope for research in the area of adaptive heuristic algorithms such as the genetic algorithm. The working of heuristics algorithms for PD detection can be elucidated.

B. OUR CONTRIBUTIONS

Based on the literature reviewed and existing research gaps, a novel transfer learning-based model is proposed for automatic detection of Parkinson's disease which explores the merits of genetic algorithm and K-nearest neighbor for efficient detection performance. In this study, the objective function of GA is based on KNN, which is a distance-based algorithm. The model never learns a discriminative function, during the process but only calculates distances between two vectors. By reducing the features, the complexity of traditional KNN is further decreased. Thus, the training complexity is highly reduced as compared to other traditional CNN based models. This paper has following main contributions compared to the existing models investigated for Parkinson's disease detection.

- ⋆ Instead of employing handcrafted feature extraction technique, an automatic feature extraction model based on transfer learning networks is suggested.
- ⋆ Multiple transfer learning neural networks are employed to eliminate possible bias from any single TL network for precise and bonafide detection.
- ⋆ Optimum features selection out of the extracted stacked features from TL networks is performed through genetic algorithm and K-nearest neighbor procedure to achieve more accurate detection unlike traditional CNN based models.

- ⋆ The performance is compared with many existing PD detection methods to demonstrate the better performance of the proposed model.

III. MATERIALS AND METHODS

The different materials and methods adopted to develop the proposed PD detection model are described in this section. Primarily, the transfer learning models, K-nearest neighbor classifier, and genetic algorithm optimizer are explored in this proposed model.

A. TRANSFER LEARNING

Transfer learning makes it possible for pre-trained networks to be used for new use cases which might prove beneficial in saving up resources and providing improved efficiency. The general idea of transfer learning is to use the previously gained knowledge and apply it for a newer problem with different data. Transfer learning also saves a lot of time of training since a new model doesn't need to be trained from scratch. This approach is also fruitful when it comes to the absence of enough data. It allows a user to apply an entirely new dataset to solve completely different problems. It allows the user to specify the dimensions of last layers according to will. Also, not only does the transfer learning approach allows users to change the dimensions of output layer, it allows the users to fine tune other hyper-parameters as well as weights in the other layers of the pre-trained model. Typically, in transfer learning the starting layers are fixed or locked and resistant to any change, while as last layers are adjustable.

B. KNN CLASSIFIER

The k-nearest neighbor or k-NN algorithm is considered to be one of the most straightforward and uncomplicated machine learning algorithms. The simple nature of the algorithm is achieved due to the fact that is doesn't consider any parameters because of which it can also be called a non-parametric algorithm. The action on data is performed during the last stages of this algorithm and often called as a lazy learning technique. The algorithm can be used for classification as well as regression purposes, but most prominent application of the algorithm can be observed in classification problems. The concepts of this algorithm are easy to understand and apply. The k represents the neighboring points of data surrounding the new data point. The algorithm compares the new data point with its neighbors (k) and then groups it with the most similar neighbors [20]. The value of the K is generated randomly at the beginning of the algorithm which is usually taken within the range of 3-5. The similarity among the data point is found out by the means of distances between them. To be more specific, Euclidean distances are calculated between new point and its neighbors. The new data point is appointed to the group comprised of neighbors with least Euclidean distances.

There hasn't been a specified way to determine an optimum number of k, so some trial and error is always expected. Still, the most commonly used value of k is taken as 5; however different problems might require changes according to the need. Very small values of k e.g., 1 or 2 can prove to be noisy and eventually lead to misleading interpretations due to outliers. The KNN is utilized as the objective function during the feature optimization phase, and it also applied to do the classification task after getting the optimized feature vector.

C. GENETIC ALGORITHM

Genetic algorithm (GA) is an adaptive heuristic algorithm whose working is governed on the principles of genes and natural selection. Genetic algorithms being closely related to the evolution theory mimic the concept of natural selection. Natural selection refers to the survival of the species that are able to adapt or mutate according to the changes in their habitat and surroundings. The core idea behind the natural selection is "Survival of the fittest." Every generation is comprised of a different set of individuals commonly known as a population. Each and every individual of a particular populations acts as a point in the search space [21]. There are generally 5 phases of the GA such as (1) *Initial population*, (2) *Fitness Function*, (3) *Crossover*, (4) *Mutation*, and (5) *Selection*. The basic procedure of a GA is shown in Figure 1.

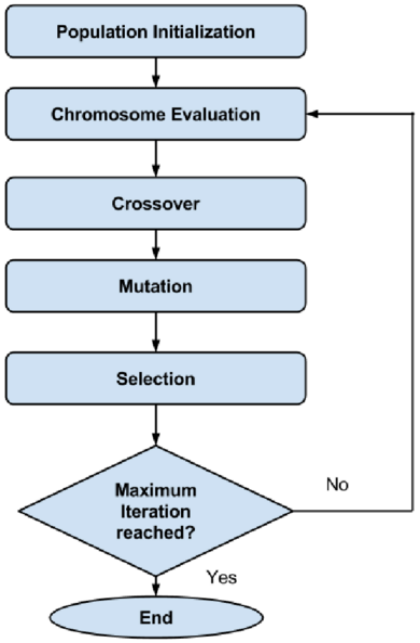


FIGURE 1. Genetic algorithm procedure.

Starting with the initial phase called the initial population phase; this phase consists of the population in question. Each individual is represented using a unique string. Here, each individual acts as a solution to any problem that has to be solved. Once the population is considered the next phase known as the fitness function starts. Fitness function, as the name suggests, displays the fitness level of an individual. An individual needs to be fit in order to survive in its habitat by being competitive against other individuals. The fitness function provides a score which represents the fitness level. The selection of the individuals is dependent on the fitness score. The third phase is the selection phase where the fittest

individuals are selected to pass on their genes to younger ones. In this phase a couple of individuals are selected on the basis of their fitness scores. These individuals can also be called parents. Parents with high fitness scores are more likely to be considered for reproducing the off-springs. After the selection phase, a very significant phase takes place which is known as the crossover phase. Here random genes/ set of genes are selected from the chromosomes of both the parents and are swapped with each other. Thus, the offspring contains half the genes of both the parents. This new individual is then added to the existing population. The last phase of the genetic algorithm is the mutation phase. This mutation refers to the slight changes in the genes of a newly created offspring. These changes are subject to the changing patterns of the environment. Mutation occurs so that the new population is able to deal with the changes and thrive, instead of not coping up with the changes which could lead to extinction. The algorithm is said to terminate when no new offspring contains any kind of mutation.

IV. PROPOSED METHODOLOGY

In this section the proposed methodology for PD detection using handwritten data will be discussed. This section begins with a brief dataset description followed by the elaborated discussion of proposed methodology.

A. DATASET DESCRIPTION

For the experimental part, NewHandPD [22] dataset has been taken into consideration. The dataset has been published and made available to the public for research purposes. The dataset is entirely dedicated to the handwritten specimens which prove to be beneficial for our work. The NewHandPD contains images collected via a smart pen and a tablet respectively. The NewHandPD dataset was introduced by Pereira et al. [22]. This dataset is the extension of the HandPD dataset [23]. There are 594 total images in the data set, 160 of which are male and 104 of which are female. The Healthy Group and Patient Group are the two different sorts of groups that make up the data set. There are 315 samples overall in the healthy group and 279 samples total in the patient group. In each category, samples are drawn from both males and females. Depending on the sort of drawing that individual receives; the data set is split into three categories: Circle, Meader, and Spiral. Each group has a depiction of the form it stands for; for example, in the Circle group, a patient and a member who is in good health are asked to draw over the provided circle. The images of all groups are combined in our study into two categories: healthy and patient. The images from all three groups that are healthy make up the healthy group. Images from all groups obtained from patients make up the patient group. Some example images from the Patient group and healthy group are shown in Figure 2 and 3, respectively.

B. PHASE-1: DATA PREPROCESSING

Each group's images are read, and if they fall under the same category, they are preserved with that label (Healthy or Patient). The images are resized to 256 × 256 for use in Transfer Learning (TL) models. Because certain TL models have requirements for image sizes, this is done in advance to prevent issues in the future. The images are read as RGB (3 channel images) since TL models normally work on colored images only, as they were trained on colored images. Additionally, image pixels gray values are scaled down from 0 to 1 by dividing them by a factor of 256. Reason being, the deep learning models work better on values which are in the range of 0 and 1.

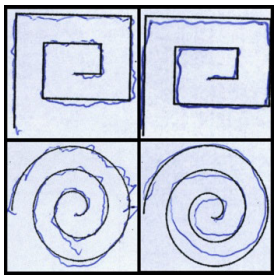


FIGURE 2. Examples from spiral and meander classes of the patient group.

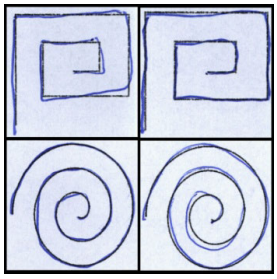


FIGURE 3. Examples from the spiral and meander classes of the control group.

C. PHASE-2: FEATURE EXTRACTION

In this work, feature extraction is carried out using three transfer learning models, specifically the following ones [20].

1) RESNET50

The ImageNet dataset was used to train the 50-layer deep neural network known as ResNet50. The network has learned a wide range of attributes as a result of being trained on more than a million images. The network's input shape (image input size) is 224 × 224 × 3. The network was first used for computer vision tasks, but as it advanced, it has also shown promising performance in non-computer vision applications.

2) VGG19

Another transfer learning model that was educated on ImageNet is VGG19. The network has 19 layers and was trained using several different characteristics. Although this network has also been utilized for computer vision applications, it also performs well in other image-related tasks.

3) INCEPTION-V3

The Inception model is based on the depth of neural network that is employed. This model consists of 48 symmetric and asymmetric layers, including convolution, pooling, dropouts, etc., make up the model. In terms of accuracy and computational cost, the v3 model outperforms its antecedents (v1 and v2).

By freezing their training and removing their tops, these transfer learning models are utilized in our proposed work. These models' training has been stopped, and features are now obtained using the weights from earlier training (ImageNet). These models' top include upwards of 20,000 classes, which are not necessary for our study. Therefore, we deleted the top from each of the TL models and placed our own neural network (top) on top of each of them in order to effectively employ TL models. The top is composed of a thick layer on top of a flattened layer. The number of features we extract from each model individually is the number of nodes given to the dense layer. The structure of individual transfer learning model extracting features from input images is shown in Figure 4. Where, the whole feature extraction process collectively through all three TL models is depicted in Figure 5.

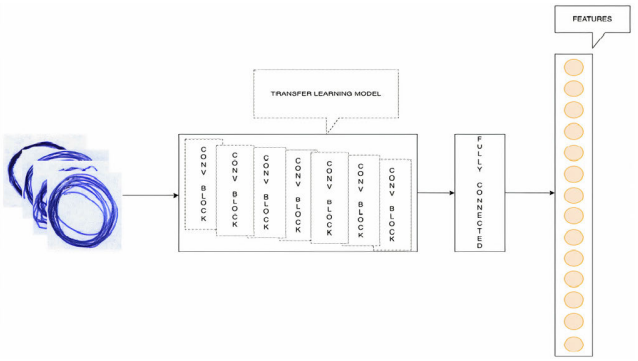


FIGURE 4. Structure of each TL model.

Each model has had 100 features taken out of it. After passing through each TL model, these characteristics are then layered horizontally on top of one another, giving each input image a final form of 300 (100+100+100). Following the feature extraction procedure, each image is divided into 300 features, resulting in the form of our entire dataset being (594 × 300), which is composed of 594 images.

The rationale behind employing three models is to eliminate bias for any one model in particular. Our algorithm model primarily focuses on the optimization process and is input independent due to a variety of features. By using three models we ensure that our study doesn't produce biased results.

D. PHASE-3: FEATURE OPTIMIZATION

The features that were extracted in the earlier step are then fed into this stage. The critical stage of the study occurs here, when a collection of features is chosen using an optimization approach and sent to the machine learning algorithm to assess performance. The results of the optimization process are the

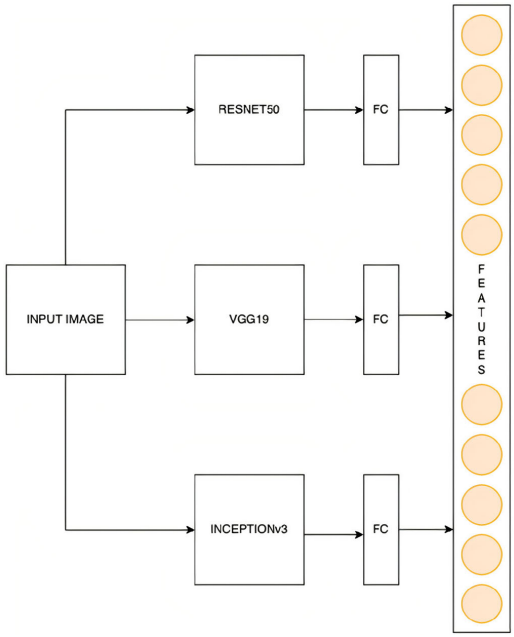


FIGURE 5. Feature extraction process.

characteristics that were chosen. This phase needs to choose a suitable machine learning (ML) algorithm, and a proper optimization method. In our study, the accuracy metric serves as the cornerstone for the optimization process, with the machine learning approach acting as an objective function. Since, the KNN has low computational cost and it is perfectly suited for small sized dataset, we choose the k-Nearest Neighbors as objective function during feature optimization phase. Due to the durable performance of genetic algorithm, it is chosen as the optimization algorithm.

The feature vector optimization takes place as per the following GA process:

Step 1. Initialize the GA parameters
Step 2. Generate random population (Initial Population).
Step 3. Calculate the fitness of each member of initial population
// calculate accuracy of each feature generated
Step 4. While *iteration* < *Max_itr*
Step 4.1: Choose two parents at random from the population and perform crossover over the parents. This process is continued until the crossover ratio from the entire population has been reached.
Step 4.2: Choose a parent from population and perform mutation. This step is also repeated until the mutation ratio from total population is attained.
Step 4.3: Calculate the fitness of newly generated children.
Step 4.4: Select the top candidates from the extended population and forward them as population for the next iteration.
Step 4.5: Go to Step 4.
Step 5. Stop and output the best vector produced.

In step 1 genetic optimization algorithm parameters are initialized, the various parameters are crossover ratio,

mutation ratio, population size, total iterations etc. A population of binary vectors, or vectors that exclusively include 0s and 1s, is generated in step two. Using the chaotic logistic map approach, the population is produced at random. This approach generates random numbers by using a dynamic key. The same pseudo-random number vectors are calculated every time using the same key. In order to retain the influence of the complete feature set on our algorithm, an additional vector made up entirely of 1s is introduced to the population. Accuracy metric serves as a parameter for fitness evaluation, in step 3. The algorithm estimates accuracy by converting the feature matrix in accordance with each candidate in the population.

The algorithm is then executed through a number of specified iterations. In each iteration, step 4 randomly chooses two parents, performs crossover on these two parents, and produces two children. The initial crossover rate is what determines the final crossover rate. The total population generated then, contains both parents and children. The algorithm then moves on to the next stage of *mutation*. The candidate is chosen at random from the entire population, and it is subjected to mutation. Up until a population-wide mutation rate is attained, the mutation process is repeated as well. The algorithmic population size is unaffected by this phase. Usually, the mutation is carried out to prevent being trapped in local minima. Following these two methods, each created child's fitness is determined, and only those children who score highest on the fitness scale are kept. The hyper parameters, (n), which represent the algorithm's whole population, are used to choose candidates for subsequent generations. Step 4 is repeated till the stopping condition is not met. We go on to step 5 after exiting step 4, where the best candidate found in step 4 is output and frozen. The process then comes to an end with the optimized feature vector. The hyperparameter of the mentioned GA process are as follows.

Population size:	20
Total Iteration:	200
Crossover rate:	70%
Mutation rate:	30 %
Selection process:	20 candidates with best accuracies
Random Seed:	13

A total of 300 features are input to our GA algorithm, which generates a matrix of size 20*300, where 20 being population size and 300 being vector size (size of individual candidate in population). The resultant vector also has a size of 300 which is then multiplied by feature matrix, in order to evaluate results. The GA generates a total of 20 vectors as output, among them only best (with highest objective score) is used in following steps.

The model begins with reading the data from memory and shaping and scaling down the data. We have converted the images into $256 \times 256 \times 3$ input vectors and have scaled them down by dividing 256. The intuition behind this is that our model learns better when the data is scaled down. The input vectors are fed to three different networks namely ResNet, VGG19 and Inception. The three different networks give the different feature vectors. The resultant feature vectors are stacked on top of each other to get a single output vector which we feed to the optimization algorithm, which in our case is Genetic optimization algorithm (GA). As shown in the workflow diagram, GA comprises of three phases: crossover, mutation, and selection, which give us the final output and performance evaluation metric. The complete procedure of proposed methodology is presented below in stepwise form.

Step A. Load the NewHandPD Image dataset.

//Data pre-processing

Step B. Scale down the images by dividing each image's values by 256. (Scaling creates floating point values which show better results).

//Feature extraction

Step C. Pass each image through three Neural Nets such as ResNet50, VGG19 and Inception_v3. Each model outputs a total of n features, thus each image is transformed into the 3*n features. The net result of this stage is a stack of feature maps, each of length 3*n.

//Feature optimization

Step D. After obtaining feature maps from step C, the optimization algorithm is applied. The optimization algorithm applied in this study is Genetic Algorithm (GA), which has three stages i.e., Crossover, Mutation and Selection. The initial population 'm' (population of binary vectors where 1 represents the presence of feature and 0 represents absence of the feature) is chosen. This step is followed by objective function evaluation. The objective function chosen here is KNN. Each binary vector is multiplied by feature matrix and the KNN algorithm is applied and evaluated on this matrix. Accuracy is calculated and is stored against each vector in population. The objective is to maximize accuracy (objective function). The anticipated GA process for feature optimization presented in Section D is enforced.

Step E. After optimization phase, the algorithm moves to next step, which is returning the optimized vector. This vector is then used to test and evaluate the test set.

The schematic diagram of the proposed transfer learning and optimization-based Parkinson's disease detection framework is shown in Figure 6.

V. PERFORMANCE RESULTS

In this part, we present the findings of a series of tests designed to evaluate the performance of the suggested model. The experiment assessed the prediction capability of the proposed methodology on the NewHandPD dataset and determined how each feature subset contributed to the total classification accuracy. Various analyses of the results are carried out based on how the performance of the algorithm is assessed. The metrics used to assess the detection performance of the proposed model includes Accuracy,

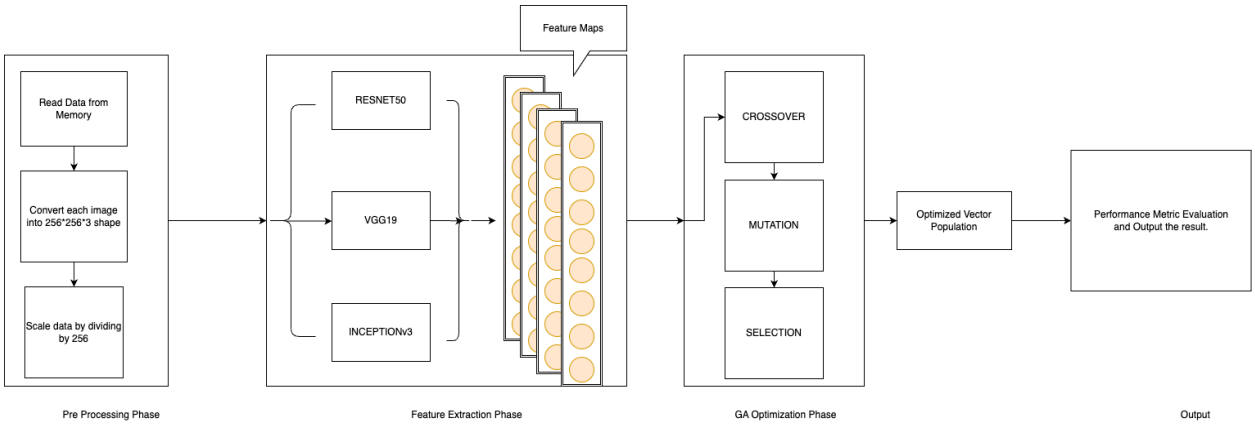


FIGURE 6. Schematic diagram of the proposed framework for PD detection.

Recall, Precision, and AUC which are briefly formulated as follows.

Accuracy is expressed as:

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

Recall is formulated as:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

Mathematically, the precision is accounted as:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

where,

TP stands for True Positives
TN stands for True Negatives
FP stands for False Positives
FN stands for False Negatives

A. BEHAVIOUR BASED ON POPULATION SIZE

In this analysis, the population size is varied from 10 to 50 with an increment of 5 and iterations are set to 200. The peak accuracy at each population size is recorded. The obtained results are presented in graphical form shown in Figure 7.

At population sizes of 20 and 40, the highest accuracy of 95.29% is attained. The algorithm is frozen at population size 20 for future analyses in this study since the loss was lowest at that level for the population.

B. BEHAVIOUR BASED ON ITERATION COUNT

Here, the iteration varied from 50 to 500 with an increment of 50, and the population size is fixed at 20. The number 20 was chosen since it produced the best results in the prior analysis (95.29%). The highest accuracy achieved in each iteration is noted. Graphical plots shown in Figure 8 are plotted to present the behavior and findings. The algorithm shows no signs of improvement beyond iteration 200, where the greatest average was attained (flat line in graph). The loss seen throughout this procedure dropped up to

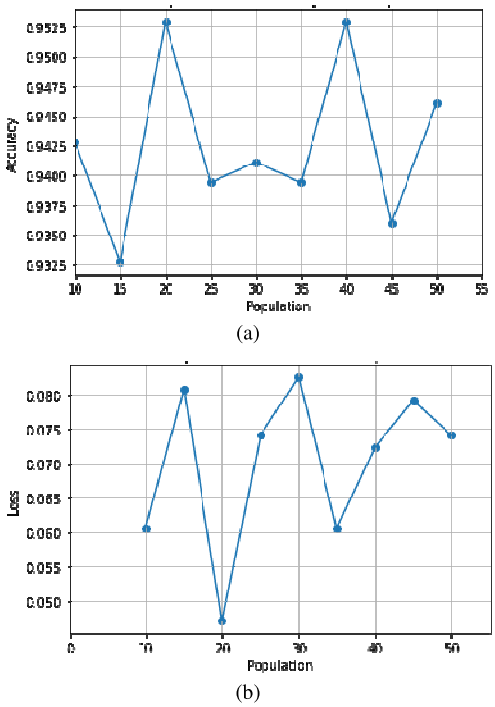


FIGURE 7. Performance of proposed model versus population size (a) accuracy, (b) loss.

iteration 100, following which it began to increase until it reached iteration 200. The loss exhibits a sharp rise after the 200th cycle. For this reason, 200 is selected as the ideal iteration size in our study. Thus, the algorithmic parameters for further performance assessments in this study are 20 for the population and 200 for the iteration count.

C. PERFORMANCE ANALYSIS

The input feature matrix is split into a train and a test set, and the best vector acquired from the optimization procedure is multiplied with the feature map to produce the optimized feature map. The test set is used to calculate the KNN algorithm's score, while the train set is used to fit the KNN algorithm. The confusion matrix is described

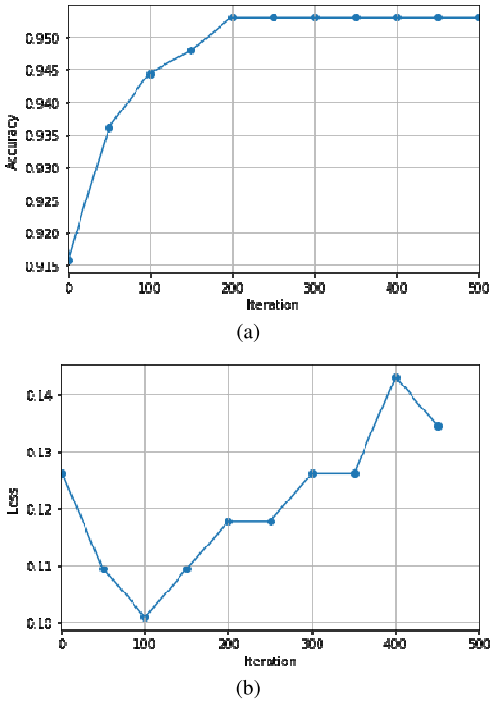


FIGURE 8. Performance of proposed model versus iteration count (a) accuracy, (b) loss.

as follows. The dataset is split into a train set and a test set for assessment. 20% of the total data is maintained for testing while the remaining 80% is provided to the train set. There are 119 samples in total which are stored for analysis. The confusion matrix shown in Figure 9 is attained for the proposed model for population size of 20 and iteration count of 200. Total values successfully predicted are 105 (42+63), while incorrect predictions are 14 (12 + 2).

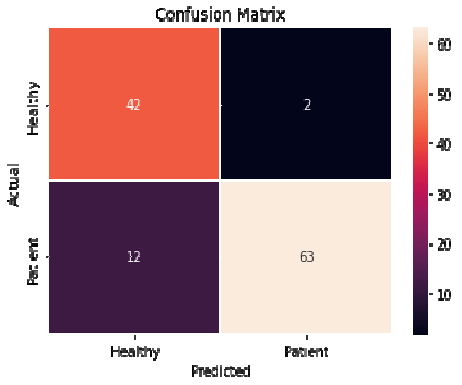


FIGURE 9. Confusion matrix for population=20 and iteration=200.

At a population size of 20 and iteration count 200, the algorithm is frozen, and performance is evaluated in each iteration. Accuracy, train loss, test loss, and area under the curve (AUC) are the different metrics that are evaluated.

1) ACCURACY

The proposed algorithm achieves a maximal accuracy of 95.29% after that, the algorithm overfits and stops training altogether. In each iteration of the algorithm, accuracy gains are seen as visible in Figure 10. The algorithm achieves its peak at and after iteration 200 and then exhibits steady behavior after that.

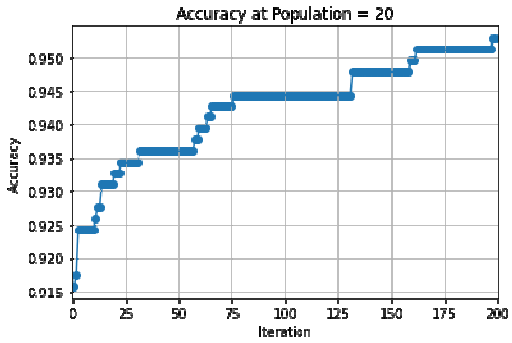


FIGURE 10. Accuracy of proposed model.

2) LOSS

In order to create an optimum feature map (× train) at each iteration, the best vector generated for each population is multiplied by × train to determine the training loss. This feature map is used to fit the KNN algorithm, which is subsequently tested on test data. Using the × test, the test loss is computed. The train and test loss's variation with iteration are shown in Figure 11 and 12, respectively. The minimal test loss obtained is 0.09, while the minimum training loss is 0.01.

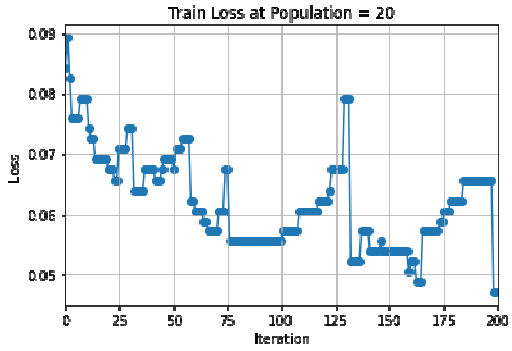


FIGURE 11. Training loss plot.

3) AREA UNDER CURVE (AUC)

The level or measurement of separability is represented by Area under curve (AUC). It reveals how well the anticipated model can differentiate across classes. The greater the AUC, the better the model is in correctly classifying the classes i.e., Healthy classes as Healthy and Patient classes as patient. The AUC behavior for the proposed algorithm is shown in Figure 13. The highest AUC value obtained is 0.928, while the final AUC value i.e., after the 200th iteration is 0.9010.

4) RECALL

The recall acts as a measure of how well our algorithm detects True Positives. Recall reveals how many people we

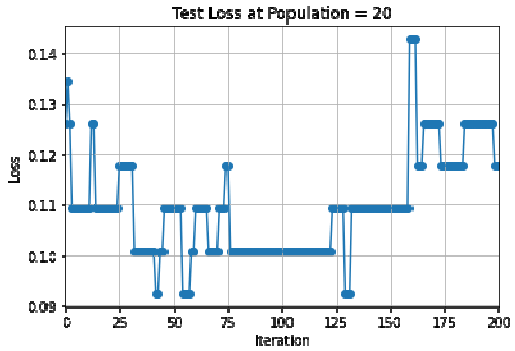


FIGURE 12. Test loss plot.

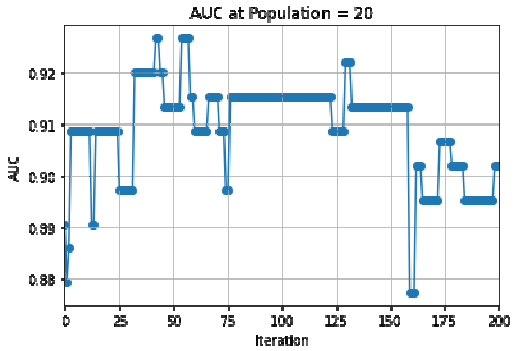


FIGURE 13. Area under curve (AUC) plot.

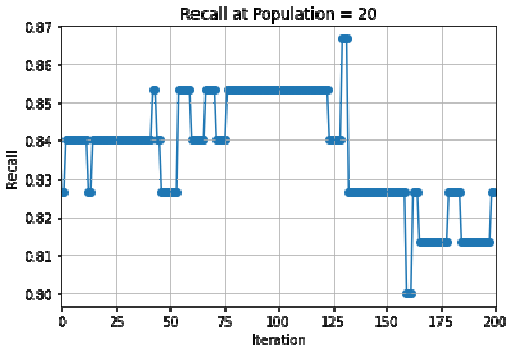


FIGURE 14. Recall plot.

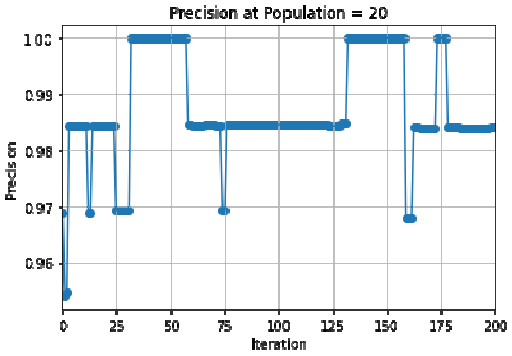


FIGURE 15. Precision plot.

accurately recognized as patients out of all those who truly are Patients. Our model's peak recall is 0.8690, and the final recall measure is derived as 0.829. Recall serves as an indicator of how well our model can locate the pertinent facts. It is also known as the True Positive Rate or Sensitivity.

5) PRECISION

The proportion of True Positives to all Positives is known as precision. In terms of our problem statement, that would be the proportion of patients with Parkinson's disease that we are able to identify accurately out of all those who truly have it. The proposed model achieves a peak precision score of 1.00 as can be seen in Figure 15. However, it really ends up being 0.985 at the end of last iteration. Furthermore, precision provides us with a count of the pertinent data points.

D. COMPARISON ANALYSIS

In order to fairly assess the performance of the proposed framework, we need to compare the obtained results against some recently investigated PD detection schemes suggested in [5, 7-9, 12, 13, 17, 19, 25-29]. We prepared Table 1 to present various performance parameters scores and compare with various networks like CNN, Random Forest, Linear SVM, AlexNet, LSTM and ESN etc., with that of the proposed model. As depicted in the Table, the accuracy obtained in the conventional models varies from 88.0% to 93.88% and our model achieves 95.29% accuracy, which is fairly better than all the listed recent schemes. The comparison of accuracies is also graphically shown in Figure 16. The table gives the performance comparison of every metric in varying architectures, which shows the better working of the proposed model and the evaluation parameters. In addition to accuracy, the Table 1 also compares the performance metrics such as recall, precision, Loss, and AUC. It shows that the AUC, recall, precision and Loss parameters are also comparable. The Loss, which is only 0.12, along with high accuracy is an indicator of better detection ability of the proposed model as compared to the mentioned existing detection schemes. Hence, the comparison analysis helps to make a vivid deduction about the better performance of the proposed model.

E. DISCUSSION

We investigated the detection of Parkinson's disease using an enhanced feature extraction procedure. The dataset is made up of drawings taken from both healthy and affected persons. The study creates feature maps for each input image using the well-known transfer learning (TL) models. To obtain features, several TL (ResNet, VGG, and Inception) models are used. The stacked output from all three models is used to extract the features. The fundamental goal of employing three models is to eliminate bias for any one model in particular. Our algorithm model primarily focuses on the optimization process and is input independent due to a variety of features.

The feature extraction is followed by the optimization process, which in our work uses a genetic algorithm. Given the objective function (KNN), the genetic algorithm produces

TABLE 1. Performance comparison of proposed model.

Model	Accuracy	Recall	Precision	Loss	AUC
Deep Learning CNN [19]	88.0	-	-	-	-
Random Forest [25]	89.4	-	-	-	-
CNN Based Visual Features [26]	83	-	-	-	-
Linear SVM [9]	71.6	-	0.7	-	-
Alex Net [13]	88.23	0.8232	0.93	-	-
Decision Tree, Random Forest and SVM [27]	82.45	-	-	0.6	-
Multi Model Approach [8]	90	-	-	-	-
Improved KNN [5]	93.88	0.97	-	-	-
Multiple Fined Tuned CNN [7]	92.7				
HOG Descriptors [28]	89.6	-	-	-	-
Deep Neural Networks [17]	89.15	-	-	-	-
Quantum ReLU Activation [29]	83	0.83	0.92	-	0.84
CNN Multistage Classifier [12]	93.3	0.94	0.935	-	
Proposed Model	**95.29**	**0.86**	**0.98**	**0.12**	**0.90**

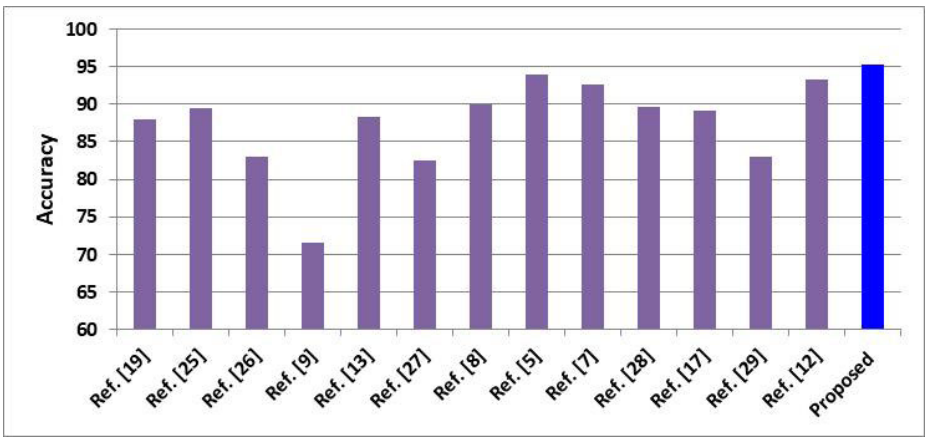


FIGURE 16. Comparison of accuracy from proposed model and other recent schemes.

a population of binary vectors and, over the course of its run, produces the best binary vectors. The GA methods, namely, crossover, mutation and selection are employed during its course. The algorithm produces the optimized vector, which is then used to evaluate the test set. The objective function of algorithm is based on KNN, which is a distance-based algorithm. The algorithm never learns a discriminative function, during the process but only calculates distances between two vectors (lazy learner).By reducing features the complexity of traditional KNN is further decreased. Thus, the training complexity is highly reduced as compared to other traditional CNN based models.

The algorithm was evaluated on a number of dimensions, and the outcomes were interpreted in several ways. Population Size against Accuracy, Iterations vs Accuracy, Population Size vs Loss, and Iteration size vs Loss are some of the several metrics used to assess performance.

The population's size is first set to 10 with a subsequent increase of 5 up to 50, in each run, in order to assess the performance of the population's size vs accuracy and population size against loss. The iteration size was set to 200 and was kept constant during the course of this evaluation. It was seen that the best accuracy was obtained at population size 20 and 40. The loss was recorded as 0.04 and 0.06 at 20 and 40 population sizes respectively. The population size of 20 was chosen for further analysis since it demonstrated the best accuracy and loss.

The second evaluation includes the investigation of accuracy and loss against iterations. Initially iterations were set at 50 and increased by 50-fold up to a maximum of 500. The population size being set at 20, the accuracy and loss are measured at each run. The algorithm shows best convergence at iteration size 200. The accuracy measured at iteration 200 was found out 95.29% while as the loss was minimum i.e., 0.1 at iteration number 100. After 200th iteration the model overfits and shows a flat accuracy curve. While as loss increases after 100th iterations and shows a spike after 200th iteration.

We thus conclude that the method shows the best convergence at population sizes of 20 and the 200th iteration. The algorithm's final findings show accuracy of 95.29 percent and a loss of 0.12.

VI. CONCLUSION

This paper proffered to present a novel framework for the accurate detection of Parkinson's disease through handwritten records available from a standard NewHandPD dataset. The proposed framework is based on the transfer learning models such as ResNet, VGG19, and InceptionV3 so as to reduce the burden of the training time. The collective features from the TL models are fed to the optimization process using genetic algorithm to get optimized feature vector for better classification results. The optimization phase considers the accuracy as the fitness value and KNN as the objective function. The classification using the optimized features is done with the help of KNN which is computationally less intensive. The performance of the proposed model is studied and assessed through various analyses. The proposed model found to possess better classification accuracy than many recently investigated schemes. The Loss is very negligible and has good precision, and other performance lineaments. The experimental and performance comparison analysis validated the better performance of the proposed model in accurately detecting the Parkinson's disease.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGMENT

The authors would like to acknowledge the Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R66), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

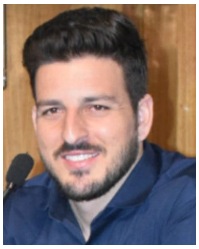
REFERENCES

[1] B. R. Bloem, M. S. Okun, and C. Klein, "Parkinson's disease," *Lancet*, vol. 397, no. 10291, pp. 2284–2303, 2021.
[2] H. Li, C.-M. Pun, F. Xu, L. Pan, R. Zong, H. Gao, and H. Lu, "A hybrid feature selection algorithm based on a discrete artificial bee colony for Parkinson's diagnosis," *ACM Trans. Internet Technol.*, vol. 21, no. 3, pp. 1–22, Aug. 2021.
[3] M. Kim, J. Yun, Y. Cho, K. Shin, R. Jang, H. J. Bae, and N. Kim, "Deep learning in medical imaging," *Neurospine*, vol. 16, no. 4, p. 657, 2019.
[4] M. Bakator and D. Radosav, "Deep learning and medical diagnosis: A review of literature," *Multimodal Technol. Interact.*, vol. 2, no. 3, p. 47, Aug. 2018.
[5] Z. Fang, "Improved KNN algorithm with information entropy for the diagnosis of Parkinson's disease," in *Proc. Int. Conf. Mach. Learn. Knowl. Eng. (MLKE)*, Feb. 2022, pp. 98–101.
[6] E. Kaplan, E. Altunisik, Y. E. Firat, P. D. Barua, S. Dogan, M. Baygin, F. B. Demir, T. Tuncer, E. Palmer, R.-S. Tan, P. Yu, J. Soar, H. Fujita, and U. R. Acharya, "Novel nested patch-based feature extraction model for automated Parkinson's disease symptom classification using MRI images," *Comput. Methods Programs Biomed.*, vol. 224, Sep. 2022, Art. no. 107030.
[7] M. Gazda, M. Hires, and P. Drotar, "Ensemble of convolutional neural networks for Parkinson's disease diagnosis from offline handwriting," Dept. Comput. Inform., Intell. Inf. Syst. Lab, Tech. Univ. Kosice, Košice, Slovakia, Tech. Rep. 9, 2022.
[8] M. Mohaghegh and J. Gascon, "Identifying Parkinson's disease using multimodal approach and deep learning," in *Proc. 6th Int. Conf. Innov. Technol. Intell. Syst. Ind. Appl. (CITISIA)*, Nov. 2021, pp. 1–6.
[9] M. Fratello, F. Cordella, G. Albani, G. Veneziano, G. Marano, A. Paffi, and A. Pallotti, "Classification-based screening of Parkinson's disease patients through graph and handwriting signals," *Eng. Proc.*, vol. 11, no. 1, p. 49, 2021.
[10] A. Gold, "Understanding the mann-whitney test," *J. Property Tax Assessment Admin.*, vol. 4, no. 3, pp. 55–57, 2007.
[11] H. W. Loh, C. P. Ooi, E. Palmer, P. D. Barua, S. Dogan, T. Tuncer, M. Baygin, and U. R. Acharya, "GaborPDNet: Gabor transformation and deep neural network for Parkinson's disease detection using EEG signals," *Electronics*, vol. 10, no. 14, p. 1740, Jul. 2021.
[12] S. Chakraborty, S. Aich, J. Seong-Sim, E. Han, J. Park, and H.-C. Kim, "Parkinson's disease detection from spiral and wave drawings using convolutional neural networks: A multistage classifier approach," in *Proc. 22nd Int. Conf. Adv. Commun. Technol. (ICACT)*, Feb. 2020, pp. 298–303.
[13] S. Nõmm, S. Zarembo, K. Medijainen, P. Taba, and A. Toomela, "Deep CNN based classification of the archimedes spiral drawing tests to support diagnostics of the Parkinson's disease," *IFAC-PapersOnLine*, vol. 53, no. 5, pp. 260–264, 2020.
[14] T. Tuncer, S. Dogan, and U. R. Acharya, "Automated detection of Parkinson's disease using minimum average maximum tree and singular value decomposition method with vowels," *Biocybernetics Biomed. Eng.*, vol. 40, no. 1, pp. 211–220, Jan. 2020.
[15] A. Das, H. S. Das, A. Choudhury, A. Neog, and S. Mazumdar, "Detection of Parkinson's disease from hand-drawn images using deep transfer learning," in *Proc. Congr. Intell. Syst.* Singapore: Springer, Sep. 2020, pp. 67–84.
[16] L. S. Bernardo, A. Quezada, R. Munoz, F. M. Maia, C. R. Pereira, W. Wu, and V. H. C. De Albuquerque, "Handwritten pattern recognition for early Parkinson's disease diagnosis," *Pattern Recognit. Lett.*, vol. 125, pp. 78–84, Jul. 2019.
[17] A. Johri and A. Tripathi, "Parkinson disease detection using deep neural networks," in *Proc. 12th Int. Conf. Contemp. Comput. (IC)*, Aug. 2019, pp. 1–4.
[18] T. Tuncer and S. Dogan, "A novel octopus based Parkinson's disease and gender recognition method using vowels," *Appl. Acoust.*, vol. 155, pp. 75–83, Dec. 2019.
[19] P. Khatamino, I. Canturk, and L. Ozyilmaz, "A deep learning-CNN based system for medical diagnosis: An application on Parkinson's disease handwriting drawings," in *Proc. 6th Int. Conf. Control Eng. Inf. Technol. (CEIT)*, Oct. 2018, pp. 1–6.
[20] C.-L. Liu, C.-H. Lee, and P.-M. Lin, "A fall detection system using K-nearest neighbor classifier," *Exp. Syst. Appl.*, vol. 37, no. 10, pp. 7174–7181, Oct. 2010.
[21] S. Katoch, S. S. Chauhan, and V. Kumar, "A review on genetic algorithm: Past, present, and future," *Multimedia Tools Appl.*, vol. 80, no. 5, pp. 8091–8126, 2021.
[22] C. R. Pereira, S. A. Weber, C. Hook, G. H. Rosa, and J. P. Papa, "Deep learning-aided Parkinson's disease diagnosis from handwritten dynamics," in *Proc. 29th Conf. Graph., Patterns Images (SIBGRAPI)*, Oct. 2016, pp. 340–346.
[23] C. R. Pereira, D. R. Pereira, F. A. Silva, J. P. Masieiro, S. A. Weber, C. Hook, and J. P. Papa, "A new computer vision-based approach to aid the diagnosis of Parkinson's disease," *Comput. Methods Programs Biomed.*, vol. 136, pp. 79–88, Nov. 2016.
[24] R. Ribani and M. Marengoni, "A survey of transfer learning for convolutional neural networks," in *Proc. 32nd SIBGRAPI Conf. Graph., Patterns Images Tuts. (SIBGRAPI-T)*, 2019, pp. 47–57.
[25] S. Xu and Z. Pan, "A novel ensemble of random forest for assisting diagnosis of Parkinson's disease on small handwritten dynamics dataset," *Int. J. Med. Informat.*, vol. 144, Dec. 2020, Art. no. 104283.
[26] M. Moetesum, I. Siddiqi, N. Vincent, and F. Cloppet, "Assessing visual attributes of handwriting for prediction of neurological disorders—A case study on Parkinson's disease," *Pattern Recognit. Lett.*, vol. 121, pp. 19–27, Apr. 2019.
[27] A. Parziale, C. A. Della, R. Senatore, and A. Marcelli, "A decision tree for automatic diagnosis of Parkinson's disease from offline drawing samples: Experiments and findings," in *Proc. Int. Conf. Image Anal. Process.* Cham, Switzerland: Springer, 2019, pp. 196–206.
[28] J. P. Folador, M. C. S. Santos, L. M. D. Luiz, L. A. P. De Souza, M. F. Vieira, A. A. Pereira, and A. De Oliveira Andrade, "On the use of histograms of oriented gradients for tremor detection from sinusoidal and spiral handwritten drawings of people with Parkinson's disease," *Med. Biol. Eng. Comput.*, vol. 59, no. 1, pp. 195–214, Jan. 2021.
[29] L. Parisi, D. Neagu, R. Ma, and F. Campean, "Quantum ReLU activation for convolutional neural networks to improve diagnosis of Parkinson's disease and COVID-19," *Exp. Syst. Appl.*, vol. 187, Jan. 2022, Art. no. 115892.

SURA MAHMOOD ABDULLAH received the B.Sc. degree in computer science (software) from the University of Technology, Iraq, in 2006, and the M.Sc. degree in software engineering from the Institute of Informatics of Higher Studies, Iraqi Commission for Computers and Informatics, Baghdad, Iraq, in 2013. She is a Lecturer with the Department of Computer Sciences, University of Technology, Baghdad. Her research interest includes artificial intelligence.

THEKRA ABBAS received the bachelor's degree in computer science from the Department of Computer Sciences, University of Technology, Baghdad, Iraq, in 1987, the M.Sc. degree in computer science from Mustansiriyah University, Baghdad, and the Ph.D. degree in computer science from Central South University, Hunan, China. She is an Assistant Professor with the Department of Computer Sciences, Mustansiriyah University, where she is the Head of the Computer Science Department. She also leads and teaches modules in computer science. She has more than 25 years of experience including extensive project management, supervised researches in the related area. Her research interests include information technology, big data, and multimedia.

MUNZIR HUBIBA BASHIR received the B.Tech. degree in computer engineering and technology from Amity University, Noida, India, in 2020, and the M.Tech. degree in computer engineering from Jamia Milia Islamia, New Delhi, India, in 2022. Her research interests include (but not limited to) neural networks, pattern recognition and classification, and deep learning.



ISHFAQ AHMAD KHAJA received the B.Tech. degree from the University of Kashmir, Srinagar, Jammu and Kashmir, India, in 2017, and the M.Tech. degree from the Department of Computer Engineering, Jamia Millia Islamia, New Delhi, in 2020, where he is currently pursuing the Ph.D. degree with the Department of Computer Engineering. His areas of research interests include machine learning, deep learning, cryptography, optimization techniques, and image processing.

MUSHEER AHMAD received the B.Tech. and M.Tech. degrees from the Department of Computer Engineering, Aligarh Muslim University, India, in 2004 and 2008, respectively, and the Ph.D. degree in chaos-based cryptography from the Department of Computer Engineering, Jamia Millia Islamia, New Delhi, India. From 2007 to 2010, he was at the Department of Computer Engineering, Aligarh Muslim University. Since 2011, he has been an Assistant Professor with the Department of Computer Engineering, Jamia Millia Islamia. He has published over 100 research papers in internationally reputed refereed journals and conference proceedings of the IEEE/Springer/Elsevier. He has more than 2500 citations of his research works with an H-index of 30, i-10 index of 70, and cumulative impact factor of more than 200. He is listed among World's Top 2% Scientists in studies conducted by Elsevier BV and Stanford University, in 2021 and 2022. His research interests include multimedia security, chaos-based cryptography, cryptanalysis, machine learning for security, image processing, and optimization techniques. He has served as a reviewer and a technical program committee member of many international conferences. He was a Referee of some renowned journals, such as *Information Sciences*, *Signal Processing*, *Expert Systems With Applications*, *Journal of Information Security and Applications*, IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS (JSAC), IEEE TRANSACTIONS ON CYBERNETICS (TCYB), IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE (TPAMI), IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS (TII), IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY (TCSVT), IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS (TCAS), IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS (TNNLS), IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS, IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS (TITS), IEEE TRANSACTIONS ON RELIABILITY (TR), IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING (TNSE), IEEE TRANSACTIONS ON NANOBIOSCIENCE, IEEE SYSJ, IEEE MULTIMEDIA, IEEE ACCESS, *Wireless Personal Communications*, *Neural Computing and Applications*, *International Journal of Bifurcation and Chaos in Applied Sciences and Engineering*, *Chaos Solitons & Fractals*, *Physica A Statistical Mechanics and its Applications*, *Signal Processing: Image Communication*, *Neurocomputing*, *IET Information Security*, *IET Image Processing*, *Security and Communication Networks*, *Optik*, *Optics and Laser Technology*, *Complexity*, *Computers in Biology and Medicine*, *Computational and Applied Mathematics*, and *Concurrency and Computation*.

NAGLAA F. SOLIMAN received the B.Sc., M.Sc., and Ph.D. degrees from the Faculty of Engineering, Zagazig University, Egypt, in 1999, 2004, and 2011, respectively. She has been with the Faculty of Computer Science, PNU, KSA, since 2015. She is a Teaching Staff Member with the Department of Electronics and Communications Engineering, Faculty of Engineering, Zagazig University, Egypt. Her current research interests include digital image processing, information security, multimedia communications, medical image processing, optical signal processing, big data, and cloud computing.

WALID EL-SHAFAI was born in Alexandria, Egypt. He received the B.Sc. degree (Hons.) in electronics and electrical communication engineering from the Faculty of Electronic Engineering (FEE), Menoufia University, Menouf, Egypt, in 2008, the M.Sc. degree from the Egypt-Japan University of Science and Technology (E-JUST), in 2012, and the Ph.D. degree from the Faculty of Electronic Engineering, Menoufia University, in 2019. Since January 2021, he has been a Postdoctoral Research Fellow with the Security Engineering Laboratory (SEL), Prince Sultan University (PSU), Riyadh, Saudi Arabia. He is currently working as a Lecturer and an Assistant Professor with the Electronics and Communication Engineering (ECE) Department, FEE, Menoufia University. He has several publications in his research areas in several reputable international and local journals and conferences. His research interests include wireless mobile and multimedia communications systems, image and video signal processing, efficient 2-D video/3-D multi-view video coding, multi-view video plus depth coding, 3-D multi-view video coding and transmission, quality of service and experience, digital communication techniques, cognitive radio networks, adaptive filters design, 3-D video watermarking, steganography, and encryption, error resilience and concealment algorithms for H.264/AVC, H.264/MVC, and H.265/HEVC video codecs standards, cognitive cryptography, medical image processing, speech processing, security algorithms, software defined networks, the Internet of Things, medical diagnoses applications, FPGA implementations for signal processing algorithms and communication systems, cancellable biometrics and pattern recognition, image and video magnification, artificial intelligence for signal processing algorithms and communication systems, modulation identification and classification, image and video super-resolution and denoising, cybersecurity applications, malware and ransomware detection and analysis, deep learning in signal processing, and communication systems applications. He serves as a reviewer for several international journals.