Machine Vision and Applications (2020) 31:32
https://doi.org/10.1007/s00138-020-01078-1

SPECIAL ISSUE PAPER

Hyper-parameter optimization of deep learning model for prediction of Parkinson's disease

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

Received: 16 October 2019 / Revised: 15 March 2020 / Accepted: 31 March 2020 / Published online: 9 May 2020

Abstract
Neurodegenerative disorder such as Parkinson's disease (PD) is among the severe health problems in our aging society. It is a neural disorder that affects people socially as well as economically. It occurs due to the failure of the brain's dopamine-producing cells to produce enough dopamine to enable the motor movement of the body. This disease primarily affects vision, speech, movement problems, and excretion activity, followed by depression, nervousness, sleeping problems, and panic attacks. The onset of Parkinson's disease is diagnosed with the help of speech disorders, which are the earliest symptoms of it. The essential goal of this paper is to build up a viable clinical decision-making system that helps the doctor in diagnosing the PD influenced patients. In this paper, a specific framework based on grid search optimization is proposed to develop an optimized deep learning Model to predict the early onset of Parkinson's disease whereby multiple hyperparameters are to be set and tuned for evaluation of the deep learning model. The grid search optimization consists of three main stages, i.e., the optimization of the deep learning model topology, the hyperparameters, and its performance. An evaluation of the proposed approach is done on the speech samples of PD patients and healthy individuals. The results of the approach proposed are finally analyzed, which shows that the fine-tuning of the deep learning model parameters result in the overall test accuracy of 89.23% and the average classification accuracy of 91.69%.

Keywords Parkinson's disease · Hyperparameter optimization · Deep learning model · Grid search optimization

1 Introduction

Parkinson's disease (PD) is the most common neurological disorder which affects the central nervous system [1]. There is a considerable rise in the number of its sufferers, mainly in the developing nations. The prior symptoms of Parkinson's disease are trembling, impaired mental response, and improper posture [2]. It is a severe medical condition that is prevalent in advanced as well as progressing nations, where around 10 million people have been diagnosed [3]. The leading cause of the disease is not traceable yet, but from the signs and symptoms related to it, this disease can be cured if discovered at the initial stages. It is uncertain to predict whether PD is a genetic or hereditary disease, and there is no treatment for its control and prevention yet [4]. The customary determination of PD includes a doctor taking a neurological history of the patient and carrying out an assessment of an assortment of motor skills. Several types of blood or clinical studies have been shown to assist with the detection of PD. Parkinson's disorder can be challenging to diagnose correctly, especially in the early stages. Sometimes, doctors can ask for brain scans or laboratory tests to eliminate other diseases. The use of blood tests, cognitive imaging methods such as MRI and PET scan may be used to avoid medical conditions such as a coma or depression that are close to Parkinson's disease [5]. Since there is no complete analytic test, the task is frequently troublesome, especially in the beginning periods, when motor symptoms are not extreme. Signs can be so unobtrusive in these first stages that they go unnoticed, leaving the disease undiagnosed for expanded periods. Clinical

✉ Sukhpal Kaur
sukhpal91@gmail.com

Himanshu Aggarwal
himanshu.pup@gmail.com

Rinkle Rani
raggarwal@thapar.edu

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

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

conditions prompting misdiagnosis or undiscovered are one of the broadest areas where medical expert frameworks get increasing interest.

The investigation of real-life datasets in a clinical setting by utilizing deep learning and machine learning strategies, techniques, and tools help in developing an appropriate and informative framework that can help clinicians in decision making. The new developments in deep learning have made it applicable very quickly in the medical fields as it proves very advantageous in preparing higher dimensional data by capturing essential features [6]. The models based on deep learning are prepared appropriately for recognition tasks of medical images, dermatology, and visual disorders. Up until this point, very few models based on deep learning have been utilized for determining cerebrum issues like Alzheimer's disease, mental disorder, and Parkinson's disease [3, 6–10]. Even though these models have recorded high accuracy for separating brain disorders from healthy individuals, their medical applicability has not yet been set up because of a few reasons. One of the basic confinements in the present deep learning-based demonstrative models is that a large number of parameters will be produced during the initialization of the deep learning model [10, 11], these parameters must be optimized to achieve a higher rating of accuracies [12, 13]. In this paper, a multi-stage optimization procedure using grid search is being proposed to develop a deep learning model to predict the early onset of Parkinson's disease. While most of the previous related works [14–19] consider prediction accuracy as the sole objective; however, the proposed approach optimizes the deep learning model for both accuracy and complexity in multiple stages. In the first stage, the topology of the network will be optimized to determine the number of layers, number of neurons in every layer, and the type of activation function for each layer. For each optimization algorithm, it is the goal of the second stage to optimize the learning rate. Ultimately, the model output with the optimized topology and learning rate is evaluated in the third stage with different optimization algorithms utilizing various epochs and batch sizes. At last, the results are compared and validated with three publicly available real-life datasets and some other machine learning approaches. The main contributions of the proposed approach are as follows:

1.1 Proposed an efficient approach based on hyperparameter tuning of the deep learning model for the prediction of Parkinson's disease.
1.2 Comparative analysis of different hyperparameter values is performed, and the optimal parameters are selected.
1.3 Experimental results are performed on three different real-life data sets. It is shown that the proposed approach is efficient in terms of processing time and can find optimal hyperparameters with high precision.
1.4 A significant reduction in the search time by 30 s and an improvement of 5% in the classification accuracy is achieved by the proposed approach when compared with the other state-of-the-art methods.
1.5 An optimized deep learning model proves its ability to be used as an early prediction tool for Parkinson's disease, thereby providing an opportunity for early therapeutic intervention.

The rest of the paper is organized as follows. Section 2 represents the contribution of previous studies on the classification of Parkinson's disease and the method of grid search optimization discussed in previous articles. Section 3 describes our primary grid search optimization technique to optimize and refine the deep learning model. Section 4 provides the results and analysis of the work done. The paper ended with our conclusion and suggested future work, as discussed in Section 5.

2 Related works

Some previously completed research has aided the progress of this paper in directing the proposed methods and augmenting the understanding of the deep learning model.

The kernel-based support vector machine for the treatment of Parkinson's disease was proposed by Little et al. [20] to identify dysphonia. For experimentation, the investigators used continuous phonations from 23 individuals with Parkinson's disease and eight control persons. The new measure of dysphonia, such as pitch frequency and another ten steps, are found to boost classification accuracy, as suggested in many telemonitoring applications. Harel et al. [4] reported that PD signs are visible for up to 5 years before professional diagnosis, including reduced loudness, elevated voice tremor, and breathability. The author uses a data set of 263 phonations from 43 subjects (17 females, 26 males, and ten healthy controls, 33 identified with Parkinson's disease) from the local voice and speech center. A comparative analysis of different classification algorithms, including decision tree, neural networks, DMneural, and regression in the diagnosis of Parkinson's disease, was carried out by Das [17]. The authors have used experimental voice tests of Parkinson's disease patients who are suffering from speech disorder for research. The analysis results show that the neural network exceeds other classifiers with regards to its accuracy in classification. Yadav et al. [18] said that only in the middle and late mid-aged Parkinson's disease signs emerged, making it difficult for researchers to foresee this. There are a variety of recommendations for PD. The study focused on the articulation of the voice and attempted to establish a standard through three methods of data mining. The three methods for data mining are derived from three distinct data mining

environments, i.e., from the mathematical classifier, tree, and SVM classifier. The three performance metrics, i.e., accuracy, specificity, and sensitivity, are used in measuring the output performances of the three classifiers. The primary purpose of this study [18] is to create the best network for Parkinson's disease individuals. Nevertheless, the only condition treated was the vocal sample, and other symptoms such as environmental and age factors, difficulties with speech and development and trembling arms, legs, hands were not taken into account. However, still, the cases are reported with an inappropriate determination. The authors of this paper have achieved 82.051% accuracy. In essence, another investigator Ramani et al. [21] have taken a telemonitor for calculating six features of significance algorithms and a total output of thirteen classification algorithms to overcome the above constraints.

Khemphila et al. [21] have used artificial neural network with back backpropagation to segregate healthy individuals from those with PD. Information gain was used to pick those important characteristics dependent on entropy values. Nonetheless, incidents with the wrong diagnosis are still identified. The authors of this article also received a precision of 82.05%. Rustempasic et al. [22] has proposed the Artificial Neural Network to distinguish healthy individuals and persons with the disorder of Parkinson. The principal component analysis is used to select relevant features before classification. The technology proposed resulted in a classification accuracy of 81.33%.

When assessing the presence of PD, Sriram et al. [10] used the patient voice. The work included the method of machine learning, which in the last decade has been designed to improve the understanding of the PD data set. Research has led to more considerable heterogeneity in Parkinson's disease samples in the parallel coordinates. In contrast to most k-NN algorithms, SVM has achieved good accuracy (88.9%). The algorithm of classification, like Random Forest, has been more precise (90.26%) with less precision (69.23%) shown by Naïve Bayes. It is therefore evident that several studies have been conducted which have used disordered voice samples to identify Parkinson's disease.

Shahbakhi et al. [23] also picked four tailored attributes by using a genetic algorithm. Classification is carried out with the aid of a vector supporting machine that generated 83.66% accuracy with four optimized characteristics. Suganya et al. [24] suggested a different optimization algorithm for the classification of PD by taking the adverse consequences of making the adverse consequences of artificial neural networks into account. The study carried out the algorithm for metaheuristic information mining to detect and classify Parkinson's diseases (Table 1).

As it is seen, several techniques of diagnosing Parkinson's disease are now being used in medical research and public health. Deep learning models generally play an ever more critical role in the field of healthcare. The objective of this paper is to create and implement an end-to-end deep learning model that can enable the doctors and clinicians to have faster and more accurate predictions.

Table 1 A summary of existing Parkinson's disease identification methods

Authors	Methods	Accuracy (%)
Little et al. [20]	SVM	81
Das et al. [17]	Decision tree, neural networks, DMneural, and regression	82.9
Yadav et al. [18]	Statistical classifier, tree classifier, support vector machine	82.56
Khemphila et al. [21]	Artificial neural network	82.05
Sriram et al. [10]	SVM, k-NN, Random forest, Naïve Bayes	84.26
Shahbakhi et al. [23]	Genetic algorithm, SVM	86.23
Suganya et al. [24]	Ant-colony Optimization, Artificial Bee-optimization, Particle Spam optimization,	86.77

2.1 Preliminaries and background

2.1.1 Hyperparameters optimization of deep learning models

Let $h_1,\ldots, h_k$ denotes the hyperparameters of the deep learning model and $\alpha_1,\ldots\alpha_n$ be their respective domains. The deep learning model is trained with hyperparameter h on training voice samples (D_{train}) of PD patients and healthy individuals. The validation accuracy of the deep learning model is signified as λ (h, D_{train}, D_{val}). The main aim of the hyperparameter optimization of the deep learning model is to find a hyperparameter setting h^* to maximize the validation accuracy λ on voice samples of PD patients and healthy individuals [25–27].

2.1.2 GSO-based parameter optimization model

The grid search optimization algorithm takes into consideration the hyperparameter optimization in the respective stages with training data in increased amounts. In the first stage, a small subset of training data is used by applying sequential optimization for quick identification of an initial set of promising hyperparameters settings. These settings are further utilized to initialize the grid search optimization algorithm on the next stages to operate with a better knowledge to merge to the most favorable solution [1] rapidly.

The initial arrangement of hyperparameter settings $h_1,\ldots,$ h_k impact the grid search optimization algorithm, where all the hyperparameter settings are indicated with many assignments by each context. From there on, the assessment of the underlying parameters on the validation data is done, followed by recording their accuracies. To mark the correctness, the algorithm at that point advances in rounds to continually change in deep learning model. The grid search optimization algorithm recommends another hyperparameter setup with the assistance of procurement work, whereby the exactness of the new setting is determined on validation data bringing about new accuracy [11, 28].

Algorithm 1 (Grid Search Optimization)
Input:-*Hyper-parameters* $h_{1:k}$,
Iterations per stage $X=<X_1,\ldots,X_z>$,
Total number of stages Z,
Training data per stage $D_{train}=<D_{train}^{1},\ldots, D_{train}^{z}>$,
Validation data D_{val},
Validation accuracy λ
Output:-Hyper-parameters h^*
for stage z=1 to Z do
for i=1 to ℓ
λ_i=*evaluate* $\lambda(h_i, D^{s}_{train}, D_{val})$
end
for j=ℓ*+1 to* X_z
$g=grid\text{-}search(h_i,\lambda_i)_{i=1}^{j-i}$
$h_j=max\ args_{h\epsilon\alpha}\ a(h, g)$
λ_i=*evaluate* $\lambda\ (h_i, D^{s}_{train}, D_{val})$
End
Reset $h_{1:k}$ = *best k configs* €*(* $h_1,\ldots\ldots h_{Xz}$*)*
// according to validation accuracy λ
End
Return $h^*= max\ args_{\ h\ \epsilon(}\ h^{X1},\ldots\ldots h^{Xz}\ )\ \lambda_j$

During each stage z, the k best arrangements dependent on validation accuracy passed from the antecedent stage ℓ are first assessed on the current stage's training data D_{train}. After that, the grid search algorithm is introduced with these k settings and connected for Xs –L iterations on D_{train}, where X_{s} are all number of repetitions for stage z. At that point, the top configurations dependent on validation accuracy are utilized to instate the subsequent stage's run. In the wake of running, all S arranges the calculation ends and outputs the configuration with the most exceptional validation accuracy from all hyperparameters investigated by all stages. Toward the end of the considerable number of steps, the algorithm ends with output the configuration with the most astounding validation accuracy [29]. Algorithm 1 above describes the GSO based parameter determination technique for deep learning models. The calculation is introduced for Parkinson's infection forecast utilizing voice tests of PD patients and healthy individuals. The situation will be wholly investigated in the following section [3].

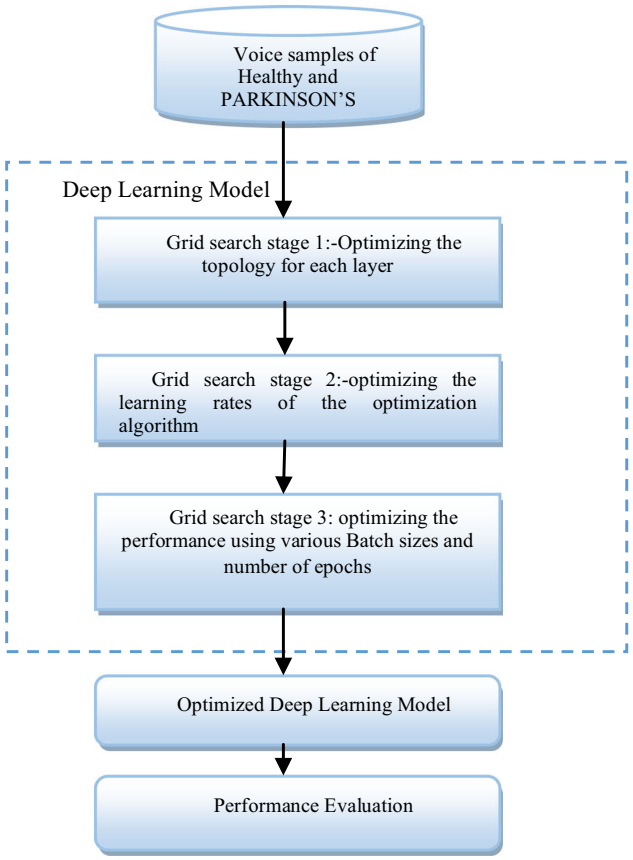

Fig. 1 The overall workflow of our approach

3 The proposed approach

Figure 1 illustrates the overall workflow of the proposed approach. Inputs are used for speech recordings of people with stable and Parkinson's disease that are obtained from the UCI machine learning center. The classification of the Parkinson's disease is achieved by using optimized deep learning model. Grid search is an optimization technique for hyperparameters. It can be used in different areas such as the number of layers, the number of neurons of the single layer, the types of activation function, the different learning rates, the batch sizes, the number of epochs, and finally, the optimization algorithm type. In compliance with the unusual circumstances, the following concept is split into three stages [12].

Grid search optimization is done in stages because each stage of the network requires a vast number of operations due to a large number of tuned hyperparameters, and costs in the other levels would further compound this. It indicates that the code should be developed with the use of concurrent algorithms and high-performance computing equipment and GPU-capable computers to address this obstacle. In the first stage, the topology of the network will be optimized to determine the layer numbers, neuron numbers in every layer, and

the type of activation function for every segment. For each optimization algorithm, it is the goal of the second stage to optimize the learning rate. Ultimately, the model output with the optimized topology and learning rate is evaluated in the third stage with different optimization algorithms utilizing various epochs and batch sizes. A cross-validation (CV) method is applied to deal with the problem of overfitting of the trained model at each stage. The trained classification model is used in CDMS to diagnose Parkinson's disease patients. To achieve the best model with higher validation accuracy, a tenfold cross-validation method is applied to the training. In this way, accuracy and loss values are used as performance metrics.

3.1 Grid search optimization stage 1:Optimizing the deep learning model topology

The proposed approach includes the conventional multilayer perception (MLP) of an underlying deep learning model architecture, which usually comprises one input layer, and two or more processing layers and one output layer. The input layers' neurons are of equal size to the input features; neurons in processing layers usually can be any number and neurons of the output layer similar to those of the output layer. The goal is to develop the topology of the deep learning model by optimizing the number of layers, neurons per layer, and the type of activation function per layer by loading with the training voice samples. An important part to efficiently enhance the classification accuracy is the number of hidden layers and the number of neurons in each layer.

To sharpen the number of neurons in shrouded layers, it is considered that the number of neurons in the top layers should be more than a number of neurons in lower layers [30]. Furthermore, such requirements are known to speed up the grid search process as well as to ensure the stability of the network. For example, if layer 2 has 30 neurons, there should be less than 30 neurons in layer 3. This dramatically reduces the number of operations (*O(n^2)). Besides, there were 1872 qualified criteria using the batch of 50 to 200 iterations for the configured model, which took 97 s. However, provided that the metric of accuracy and loss is required in this problem and the corresponding values are normalized to have values between 0 and 1, in the final layer, a neuron with Sigmoid always was called the activation feature. The similar values between 1 and 40 can be used in steps of 5 as several neurons. The resulting arbitrary values from 1 to 40 here mean 5, 10, 15, 20, 25, 30, 35, and 40. Thus, despite the suggestions from different studies, there is no exact parameter to figure out these values. To order to do so, there must be a minimal and precise number of measures rather than a large number, which can prove overwhelming. Nevertheless, the collection starts with the trial and error approach not beyond a specific class, and therefore there's no optimal value for this purpose.

Table 2 Values of various parameters used at grid search stage I

Hyperparameter values	
Layers (Number)	2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13
Neuron in different layer (Number)	1, 2, 4, 5, 10, 12, 15, 20, 25, 30, 35, 40
Epochs	100
Batch sizes	10
Activation function	Softmax, Softplus, Relu, Tanh, Sigmoid, linear
Optimizers	SGD
Learning rates	0.001
Performance matrices	Classification accuracy, loss function
Cross-validation	Tenfold

To evaluate the results comparing different activation functions ReLU, Sigmoid, Tanh, and Softplus, the varying hidden layered deep learning model with differing neurons are used for testing, keeping the rest of the hyperparameters are fixed.

In Algorithm 2, grid search stage-1 is explained. On the Python GridSearch CV repository, after the creation of the deep learning model, hyperparameter optimization of the grid search stage-1 is performed. The input parameters for the grid search stage1 algorithm are the speech samples PD patients and healthy individuals, the different neuronal values, and the hidden layers of various functions of activation. The optimized hyperparameters h_1 is the resulting output of the proposed approach.

Algorithm 2. Grid Search Stage-1

1. Input: - Voice samples of PD patients and healthy individuals, values of neurons, hidden layers, different activation function.
2. Output:-Hyper-parameter h_1
3. create_model ()
   //initialization of Deep learning model with basic parameters
4. ml.add (layers, neurons, activation function)
   //define grid search parameters
5. gd=GdSearchCV(par_gd=par_gd,est=ml)
6. gd_res=gd.fit(voice samples of healthy and PD patients)
7. print("best hidden layers, number of neurons,activation function %(gd_res.score, gd_res.best_par))

Table 2 shows the various hyperparameter details set for the advancement of the topology in the GSO stage-1.

Figure 2 demonstrates the optimized configuration of the deep learning model after the grid search stage-1. The method suggested thus utilizes 22 neurons in the input layer as the speech samples of PD patients and healthy individuals have 22 features. After grid search stage-1, the optimized deep

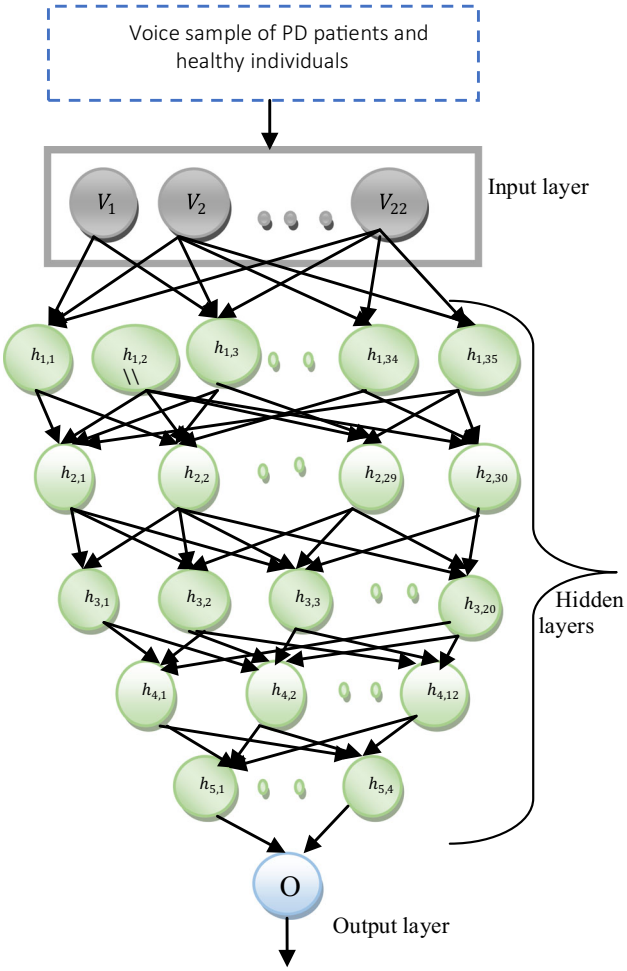

Fig. 2 Optimized deep learning model after grid search optimization stage-1

learning model consists of seven layers of 22, 35, 30, 20, 12, 4, and 1 neuron, respectively.

3.2 Grid search optimization stage 2: Optimizing the learning rates of optimization algorithms

The hyperparameter of the optimization algorithm, such as the learning rate, is tuned after designing the topology of the deep learning model. The calculated number of layers and neurons used in constructing the deep model architecture were used to optimize the learning rate for each of the optimization algorithms.

As can be seen, the grid search for the learning rate and optimization algorithms are carried out at this point. The importance of the learning rate (μ) in the deep learning model is a significant consideration. When there is too little learning pace, algorithm conversion is very slow, and the network does not converge [31] on the other hand, when the value is too big, the deep learning model fails to converge.

Table 3 Values of various parameters used at grid search stage III

Hyperparameter	Values
Layers (number) neuron in a different layer	7, 22, 35, 30, 20, 12, 4, 1
Epochs	100
Batch sizes	10
Activation Function	ReLU, Sigmoid,
Optimizers	SGD, RMSprop, Adagrad, Adam, Nadam
Learning rates	1.0, 0.1, 0.01, 0.001, 0.0001, 1e−05
Losses	Classification accuracy
Cross-validation	Tenfolds

In algorithm 3, grid search optimization stage-2 is explained. Once grid search stage-1 has been applied to the deep learning model, hyperparameter optimization of the grid search stage-2 is performed. The GSO stage-2 algorithm input parameters are the optimized hyperparameter values of the stage-1; different values of learning rates, and optimization algorithms. The optimized hyperparameters h_2 is the corresponding output of the grid search stage-2 algorithm.

Algorithm 3. Grid Search Optimization Stage-2

1. Input: -Optimized hyper-parameter values obtained at stage-1 h_1, different values of learning rates and optimization algorithm.
2. Output:-Hyper-parameter h_2
3. create_model ()
   //initialization of deep learning model with optimized hyper-parameter obtained at stage1 h_1
4. ml.add(values of learning rates, optimization algorithm)
5. gd=GdSearchCV(par_gd=par_gd,est=ml)
6. gd_res=gd.fit(voice samples of healthy and PD patients)
7. Print ("best learning rate, optimization params))

Table 3 represents the hyperparameter details set for the optimization of the learning rates and optimization algorithm in the grid search stage-2.

The framework of the tuned deep learning model after the GSO second stage contains 0.01 learning rate and Adam as an optimizer.

3.3 Grid search optimization stage 3: Optimizing the performance using different epochs and batch size

The GSO stage-3 with tenfold cross-validation for four distinctive group sizes is used to monitor the attainment of the optimized deep learning model with tuned hyperparameters for optimizing algorithms. The stage-3 optimization of grid search is described in Algorithm 4. The GSO-3 algorithm's input parameters are the optimized hyperparameter values

Table 4 Values of various parameters used at grid search stage III

Hyperparameter	Values
Layers (number) number of neuron in a different layer	7, 22, 35, 30, 20, 12, 4, 1
Number of epochs	25, 50, 75, 100, 150, 175, 200
Batch sizes	20, 50, 100, 150, 200
Activation function	ReLU, Sigmoid
Optimizers	SGD, RMSprop, Adagrad, Adam, Nadam Adam
Learning rates	0.1
Losses	MSE, MAE
Cross-validation	Tenfold

of stages 1 and 2, the different values of batch sizes, and the epochs. The corresponding performance of the proposed approach is the optimized hyperparameter h3.

Algorithm 4.Grid Search Optimization Stage-3

1. Input: -Optimized hyper-parameter values obtained at stage-1 and 2, different value of Batch sizes and number of epochs.
2. Output:-Hyper-parameter h_3
3. create_ model ()
   //initialization of deep learning model with optimized hyper-parameter obtained at stage1 and 2 i.e. h_1 and h_2
4. ml.add(values of batch sizes , number of epochs)
   //define Grid search parameters
5. gd=GdSearchCV(par_gd=par_gd,est=ml)
6. gd res=gd.fit(voice samples of healthy and PD patients)
7. print("best value of Batch size and Epochs" %(%(gd_res.score, gd_res.best_par)))

Table 4 represents the hyperparameter details set for the different batch size and epochs in the grid search stage 3.

4 Results and discussion

Two significant aspects of the experimental method were observed. To investigate their impact on the performance of feed-forward deep learning models, arbitrary yet realistic configurations of selected hyperparameters had been explored. A contrast of optimized deep learning model output with the state of artwork was explored in the second part of the study. In this article, all the experiments are carried out using the 2.7.12 edition of Python. Using Theano at the backend, the Keras library is used to construct the deep-learning model. The experiments are performed by using 196 speech recordings of 31 individuals, 23 of which are impaired by PD. Voice samples are obtained from an Irvin (UCI) University of California machine learning repository. Every section in the informational collection is the voice representation of people and each line of 196 voices from those people. The sole motive behind this research is to discriminate PD patients and healthy individuals since 0 reflect healthy individuals, and 1 reflects the PD patient [15, 20], according to the standard size scale. Validation is done using three different types of datasets–telemonitoring data set from Parkinson, Breast cancer diagnosis of Wisconsin dataset, and Pima Diabetes dataset.

The findings of the method outlined in Sect. 4 are discussed in this section. The first analysis will be carried out using the grid search optimization to different values, which further suggests the parameters which will possibly produce the best results. The hyperparameters used for analysis are (i) the number of hidden layers (ii) by layer neurons (iii) activation functions (iv) learning rate, (v) method of optimization, (vi) batch size and (vii) the number of periods. The configuration of the deep learning model is used for testing, which is trained at a constant learning rate for 200 epochs without previous stops. K-fold cross-validation was used to reduce the effects of chosen preparing information and test information on the model assessment. This involves the division of training data into unrepeated subsets. In training, $k - 1$ subsets are used, and the majority of the subsets are used for testing. In this document, a tenfold cross-validation approach is used to select training and testing subsets. The total informational collection is separated into ten different subgroups with ten times Cross-validation, of which nine subsets are used for training, while one subset tests the trained deep learning model. The entire process is repeated ten times, with various training and test subsets. About 174 samples are used as experiments in tenfold cross-validation, while other examples are used for testing from 196. During the training, the validation loss is observed. When there comes stability in validation loss, the training stops to prevent over-fitting. At the top level, the objective function reduces the average loss over the k validation folds.

At the lower level, for each training data k, we are maximizing the accuracy. The Validation accuracy is measured using an average of 10 folds in the optimized deep learning model. The number of training epochs of the deep learning model impacts performance and loss eventually. The early stoppage is a technique to determine an arbitrary number of training epochs and avoid training once the output of the model stops improving with a validation dataset. Stop the algorithm, for example, when the accuracy approaches a predetermined level or if the loss value hits a minimum, or the maximum number of iterations is reached. We used the second stop test to equate our approach. As the number of iterations rises, accuracy and loss are stable when the network reaches the convergence level.

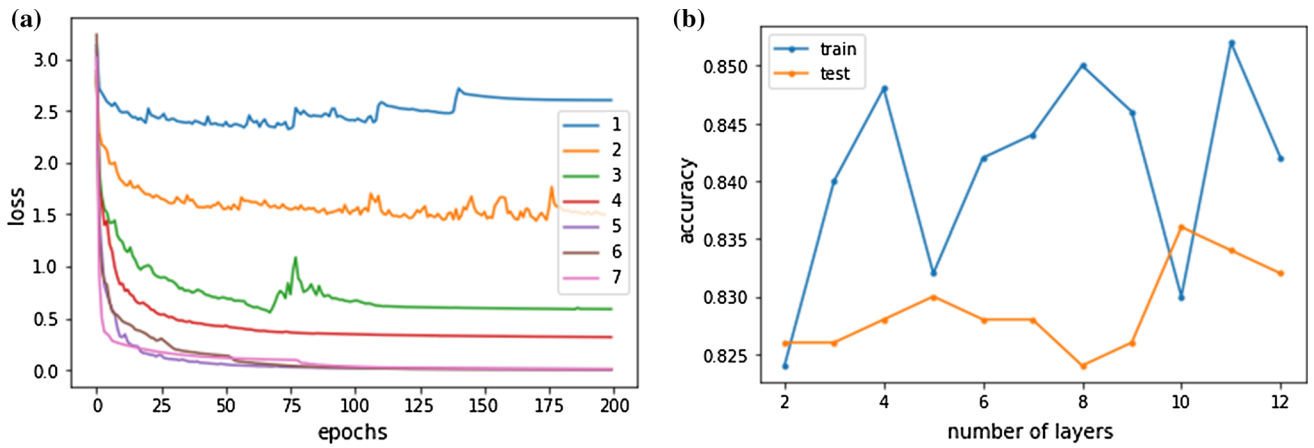

Fig. 3 **a** The loss value **b** Accuracy with varying layers on voice samples of PD patients and healthy individuals across different epochs

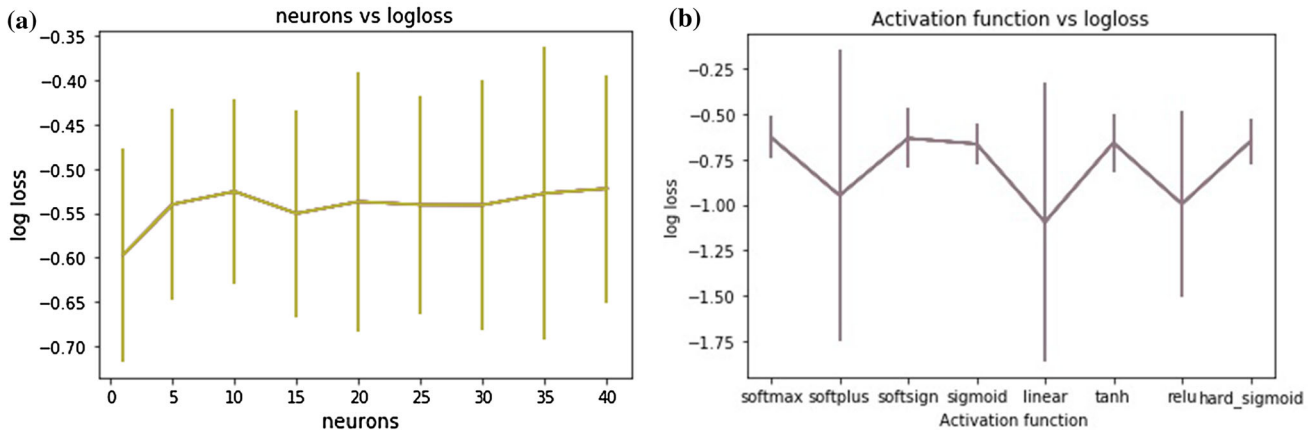

Fig. 4 Loss values on voice samples of PD patients and healthy individuals on deep learning model with **a** number of layers **b** activation function

4.1 Comparing activation function, number of hidden layers and neurons per layers

Figures 3a, b and 4a, b show the results of numerous hidden layers, multiple neurons, and various activation functions. For measuring the performance over the grid search stage 1, we use the loss of tenfold CV with constant values for batch size, epochs, learning rate, and then optimization algorithm. Figure 3a and b shows the results of numerous hidden layers on voice samples of healthy and PD patients as a new addition to the deep learning model layers. The accuracy of the deep learning model on voice samples of PD patients and healthy individuals rises gradually and loss decline, as shown in Fig. 3a, b. The deep learning model tends to overfit the training voice samples more readily. Here the plot shows a little over-fitting on training voice samples due to the large capacity of the network, i.e., due to a large number of layers and the number of nodes per layer. The model shows better performance on test samples of healthy and PD patients with five hidden layers. Therefore five hidden layers are selected as a balance between allowing for sufficient complexity and avoiding over-fitting. Figure 4a indicates a line diagram reflecting a loss over training voice samples for both healthy and PD patients with a single model structure (one to 40 nodes per hidden layer). It implies that five hidden layers are used as a bridge between having adequate flexibility and minimizing overfitting. The chart above indicates that the loss declines as the number of nodes grow, which helps to understand the results. Through achieving a balance between complexity and overlap, a right, first hidden layer size of 35 neurons is obtained when the size of subsequent layers decreases. The layer activation of the hidden layers of the deep learning model is shown in Fig. 4b.

For various hidden layers and the number of neurons, the deep learning model is trained using the different activation functions such as logical sigmoid, tangent, soft plus, and linear rectifier modules, across 100 iterations with cross-validation (CV) by ten times. The above analysis demonstrates the log loss when comparing different activation functions over voice samples of healthy and Parkinson's disease patients. The chart shows ReLU's highest performance over various activation functions. The deep learning

model is a stable model on PD patients and healthy individuals ' voice samples that can be seen from the layer activation graph stability. This indicates that the layer's weights are correctly configured. In combination with ReLU activation, the deep learning model performs better by obtaining a higher degree of accuracy than Sigm or Tanh. The neurons with the ReLU activation function are more natural to optimize, faster to converge, better in generalization, and give quick computation. From these outcomes, it tends to be seen that the increase in neurons and hidden layers, when combined with activation function, substantially affects the performance of the deep learning model.

4.2 Different learning rates and optimization algorithm

The effect of optimization algorithms and learning rates is shown in Fig. 5a–c at various epochs with the remaining parameters fixed. The deep learning model performance is measured according to classification accuracy and loss value in the context of visually averaged evaluation metrics. Different learning rates are implemented to maintain stability in the deep learning model, as shown in Fig. 5a.

Figure 5b demonstrates train and test accuracy with varying rates of learning on speech samples of PD patients and healthy individuals across 200 epochs with optimized hyperparameters obtained in stage-1. The six-line plots display six separate measured levels of learning. The plot illustrates behavioral variations with elevated learning rates, i.e., 1.0 and the model's incapacity to know anything with a minimal learning rate, i.e., 1E−05. It might be seen that the deep learning model could learn the voice samples of healthy and Parkinson's disease patients adequately with the learning rates of 0.1, 0.01, and 0.001. With the assistance of the adopted deep learning model configuration obtained in GSO stage 1, the results propose a balanced learning rate of 0.1, thus resulting in better performance of the model. The above figure shows that the selected learning rate of 0.1 worked well for convergence, which proves that 0.1 is well-suited LR.

Figure 5c demonstrates the comparisons among various gradient descent optimization algorithms on voice samples of PD patients and healthy individuals in combination with the ReLU activation function for 200 epochs with tenfold cross-validation. The above figure proves as an example for ranking the performance of the optimization algorithms on voice samples of PD patients and healthy individuals. Therefore, a variant of the Adam algorithm outperformed the other algorithms. The SGD algorithm gave the worst performance, and it also requires around all the 200 epochs, which results in volatile accuracy on the train and test voice samples. This could be a direct result of the truancy of bias adjustment, which decreases SGD performance in terms of sparsing gradients to the end of the optimization. While the model converged in 200 iterations, the efficiency of the model has to be checked for several epochs to display the stability of the deep learning model. There is a similarity in the performance of both RMS-prop and Adam having effectively learned the problem within the 50 training epochs, thereby making minute weight updates without converging. Another approach to tackle this issue is by hyperparameterizing the number of training and multiplying the training models with different values and then selecting how many epochs better fit on the train or a holdout test dataset. In the voice samples of PD patients and healthy individuals, two-loss measures, MAE, and MSE are used to measure the performance of different batch sizes using various optimization algorithms. An evaluation of the concert with batch size ranging from 20 to 200 is used to study the impacts of the batch size on the presentation of the fine-tuned deep learning model.

To study the impact of the batch size on the efficiency of the deep learning model, an output evaluation with a batch size range of 20–200 is carried out. It can be observed here that convergence of the algorithms takes place with an increase in Batch size. To demonstrate the strength of the deep learning model, the exhibition of the model for an enormous number of Batch sizes is necessary. As observed in Fig. 6a and b, as we expanded the batch sizes, the losses increased expectedly. Figure 6a, b is an ideal example of how optimization algorithms work in different batch sizes. Other algorithms have surpassed the varieties of Adam's calculation, eminently Nadam. For the Adagrad algorithm, the worst result is reported. The Nadam algorithm, which has shown better performance among the optimizations algorithms on voice samples of PD patients and healthy individuals, also gives the closest explained variance to 1.0.

The explained variance represents the gap between the actual data and the model. The more the level of discrepancy among them represents the stronger bond, which shows good predictions. To the results, there is an improvement in the overall accuracy of the deep learning model from 86.06 to 91.69% when there is an increase in the batch size from 20 to 50. The deep learning model is trained to forecast the early onset of the condition of Parkinson on the proposed hyperparameters. The classification accuracy, sensitivity, and specificity are common standards, with the assistance of various parameters such as FP, FN, TP, and TN, to evaluate the functioning of the optimized deep learning model. Where "TP" refers to the true positive, few instances that fall into the grouping of Parkinson's disease display the people affected with PD in this manner. "TN" is true negatives, which indicates that few occurrences in the lively class group are healthy persons. "FP" is false positive, showing that the condition sufferers of Parkinson experience a few events that fell into the categories of whole classes. In the end, "FN" often talks of a false negative, which means that few cases occurring in

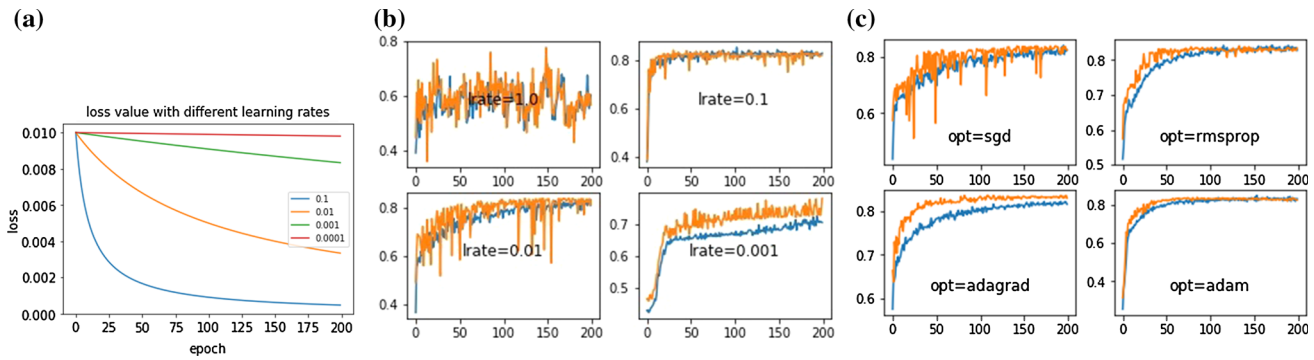

Fig. 5 **a** Loss value with different learning rates. **b** Accuracy with varying proportions of learning. **c** Accuracy with optimization algorithms

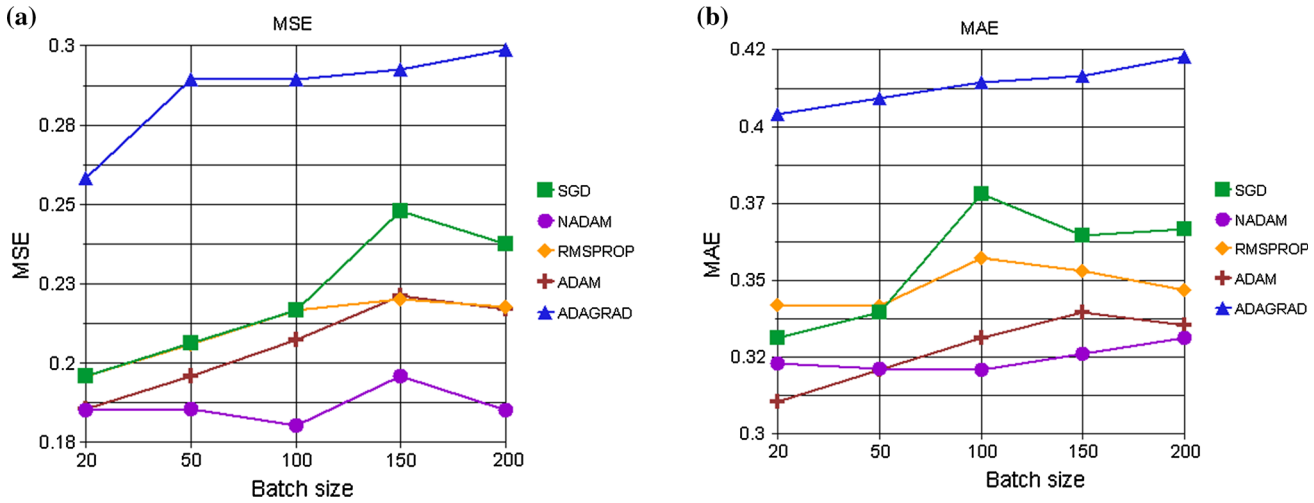

Fig. 6 **a** MSE **b** MAE values for various batch sizes utilizing diverse optimization algorithms on voice samples of PD patients and healthy individuals

the disease class of Parkinson represent stable individuals. The basic equations are described as follows:-

Accuracy: The accuracy is characterized as the capacity to recognize PD patients and healthy individuals accurately.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \times 100 \tag{1}$$

Mean Square Error: Mean square error is used to measure the nearness of the proposed strategy during the preparation stage to accomplish the lowest error.

$$\text{M.S.E} = \frac{1}{n}\sum_{n}^{1}\left(\hat{X}_i - X_i\right)^2 \tag{2}$$

The trained deep learning system is evaluated on the proposed parameters with the difference in characterization of the performance of the model, accuracy, and loss characterization. Figure 7a, b indicates the loss function value and accuracy of trained deep learning model on varying iterations. The deep learning model expands to maximum efficiency on the test set just 89.23%, and at last, the loss function reaches up to zero values, as shown in Fig. 7b, after a training accuracy of 91.69%. The deep learning model also culminated in 89.23% test accuracy in the study range of 19 tests.

4.3 Dataset description

The findings from the experiment are explained in this section after using publicly available data on the proposed framework. The tests are performed with four different datasets acquired from UCI machine learning repository [15]. The first is Parkinson's voice dataset; the second is the Pima Diabetes dataset, and the third is

Parkinson's telemonitoring Dataset and the last one is the Wisconsin Breast Cancer dataset. The description details of the employed data sets are given below.

4.3.1 Parkinson's voice dataset

The Parkinson's voice dataset is accessed from the University of California, Irvine Machine Learning storehouse. This dataset is possessed of the scope of biomedical voice estima-

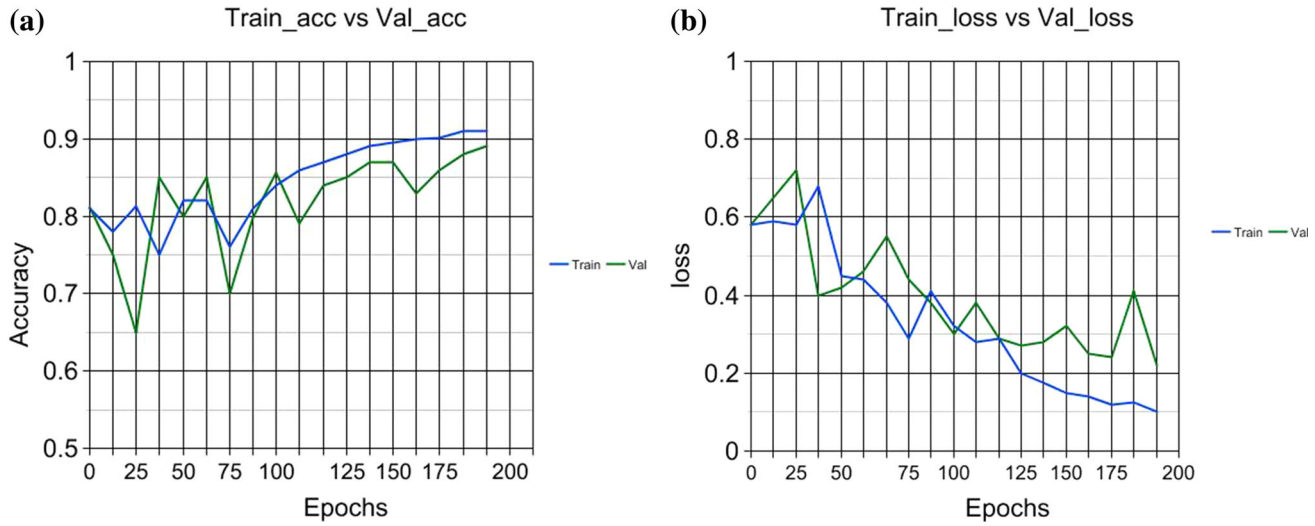

Fig. 7 **a** Accuracy and **b** Loss value of the training and test voice samples during the training process

Table 5 Description of Parkinson's voice dataset

Sr. No.	Attribute name	Description
1	Fo (Hz): MDVP	Average fundamental vocal frequency
2	Fhi (Hz): MDVP	Highest fundamental vocal frequency
3	Flo (Hz): MDVP	Lowest fundamental vocal frequency
4	jitter (Abs): MDVP	
5	jitter (%):MDVP	Variation in fundamental frequency
6	PPQ: MDVP	Fundamental frequency variation
7	RAP: MDVP	Fundamental frequency variation
8	DDP: Jitter	Fundamental frequency variation
9	Shimmer (dB):MDVP:)	Amplitude variation measures
10	DDA: Shimmer	Amplitude variation Measures
11	NHR	Noise-tonal components
12	HNR	Noise-tonal components ratio
13	Status	1-PD patients and, 0-Healthy individuals
14	D2	Complex dynamic measurement
15	Spread1	Frequency nonlinear measures
16	Spread2	Frequency nonlinear measures
17	PPE	Frequency nonlinear measures
18	DFA	Scaling fractal signal exponent
19	RPDE	Complex dynamic measures
20	APQ: MDVP	Amplitude variation measures
21	Shimmer: APQ3	Amplitude variation measures
22	Shimmer: APQ5	Amplitude variation measures
23	Shimmer: MDVP	Amplitude variation measures

tions from 31 individuals, 23 with Parkinson's disease (PD). Every section in the table is a specific voice measure, and each line corresponding to one of 196 voice recordings from these people. The principle point of the information is to segregate healthy individuals from those with PD, as indicated by the "status" section, which is set to 0 for healthy individuals and 1 for PD patients. The detailed description of The Parkinson's voice dataset is given in Table 5.

4.3.2 PIMA diabetes dataset

The PIMA Diabetes dataset is acquired from UCI machine learning repositories. This dataset contains a clinical descrip-

Table 6 Description of PIMA Diabetes dataset

Sr. No.	Attributes	Description
1	preg	Number of times pregnant
2	pres	Diastolic blood pressure (mm Hg)
3	plas	Plasma glucose (2-h)
4	BMI	body mass index (mass in kg/(height in m)^2)
5	age	Age (years)
6	Pedi	Pedigree Diabetes function
7	insu	Serum insulin 2 hours (μ U/ml)

tion of 768 female patients. This dataset likewise includes numeric eight features values where an evaluation of label '0' corresponds to healthy individuals, and another label '1' corresponds to diabetes patients. The PIMA Diabetes dataset description details are given in Table 6.

4.3.3 Parkinson's telemonitoring dataset

The Parkinson's telemonitoring Dataset is acquired from UCI machine learning repositories. The experiments are performed with the vocal chronicle samples of Parkinson's disease patients and healthy individuals. The last form of the dataset contains 756 cases with 26 characteristics. The dataset includes 188 patient records with 107 men and 81 female members. The age gathering of the individuals in the dataset is extending from 33 to 87. The detailed description of the dataset is given in Table 7.

4.3.4 Wisconsin breast cancer dataset

Wisconsin Breast Cancer data set is taken from UCI ML Repository. The samples in the database occasionally collected by Dr. Walberg from Wisconsin Hospitals, Madison. This dataset contains an aggregate of 699 instances, with 241 fatal and 458 benign cases. Each instance has ten features with relegated integer values running from 1 to 10, and one class attribute with a twofold estimation of either 2 or 4, corresponds to malignant or benign breast cancer diagnoses, separately. Each feature value lies in the range 1–10, where 1 means the normal state and ten speaks to the most abnormal state. The detailed description of the Wisconsin Breast Cancer Dataset is given in Table 8.

All data set characteristics are pre-processed and scaled up to $[-1, 1]$. The labels in class have been marked as y to $\{-1, 1\}$. To ensure that each fold has the same Distribution, we have generated ten folds using a stratified sample for tenfold cross-validation. The Validation accuracy is prediction accuracy over validity folds using tenfold cross-validation. Running time is the average CPU time spent on training the deep learning model (Intel Core i5-7700HQ). The results shown in Table 9 are obtained through each of the 200 iterations. It reveals that the Optimized DNN approach has

Table 7 Description of Parkinson's telemonitoring dataset

Sr. No.	Attributes	Description
1	Jitter (local)	Frequency parameters
2	Jitter (local, absolute)	
3	Jitter (rap)	
4	Jitter (ppq5)	
5	Jitter (DDP)	
6	Autocorrelation	Harmonicity parameters
7	Harmonic-to-noise	
8	Noise-to-harmonic	
9	Number of pulses	Pulse parameters
10	Mean period	
11	The standard deviation of the period	
12	Number of periods	
13	Shimmer (local)	Amplitude parameters
14	Shimmer (dda)	
15	Shimmer (local, dB)	
16	Shimmer (apq11)	
17	Shimmer (apq5)	
18	Shimmer (apq3)	
19	Median pitch	Pitch parameters
20	Mean pitch	
21	Minimum pitch	
22	Maximum pitch	
23	Standard deviation	
24	The fraction of locally unvoiced frames	Voicing parameters
25	Degree of voice breaks	
26	Number of voice breaks	

Table 8 Data set description of Wisconsin breast cancer

Sr. No.	Features	Description
1	Clump thickness	1–10 value
2	Uniformity in cell size	1–10 value
3	Uniformity in cell shape	1–10 value
4	Cell size single epithelial	1–10 value
5	Nuclei bare	1–10 value
6	Marginal adhesion	1–10 value
7	Nucleoli normal	1–10 value
8	Mitoses	1–10 value
9	Bland chromatin	1–10 value
10	Class	2-benign, 4-malignant

enhanced the validation accuracies on all data sets. The accuracy results are almost the same in Parkinson's telemonitoring dataset, cancer Wisconsin dataset, and Pima Diabetes data sets. The accuracies of the Optimized DNN approach are 5% greater than that of the traditional DNN technique

Table 9 Numerical results of optimized DNN and conventional DNN during the validation process

Dataset	Optimized DNN		Conventional DNN	
	Accuracy (%)	Running time (sec.)	Accuracy (%)	Running time (sec.)
Pima diabetes dataset	86.50	0.64	83.50	0.97
Parkinson's telemonitoring Dataset	86.56	0.49	82.72	0.67
Wisconsin breast cancer dataset	85.23	0.33	75.23	0.78

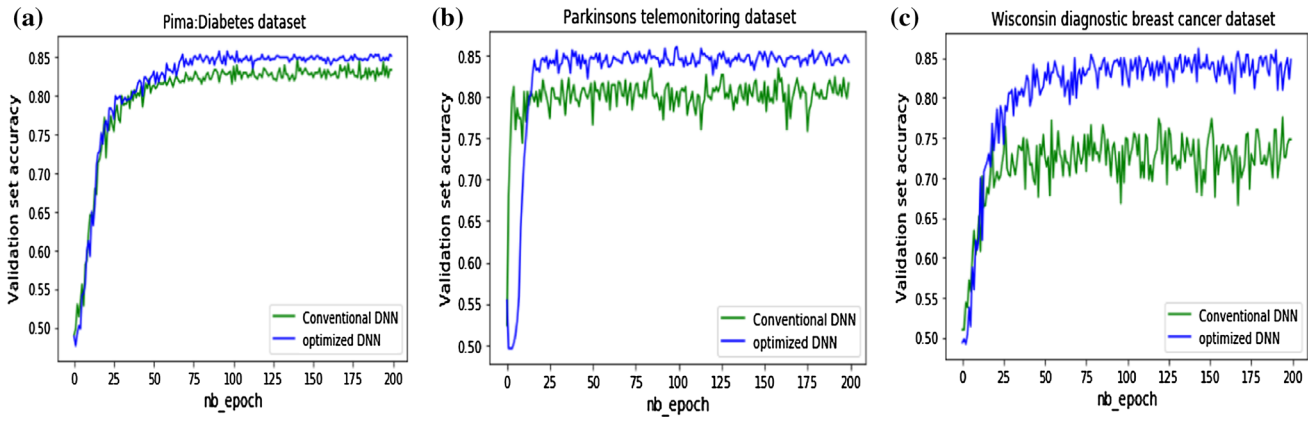

Fig. 8 Validation set accuracy **a** Pima: Diabetes dataset **b** Parkinson's telemonitoring dataset **c** Wisconsin breast cancer dataset

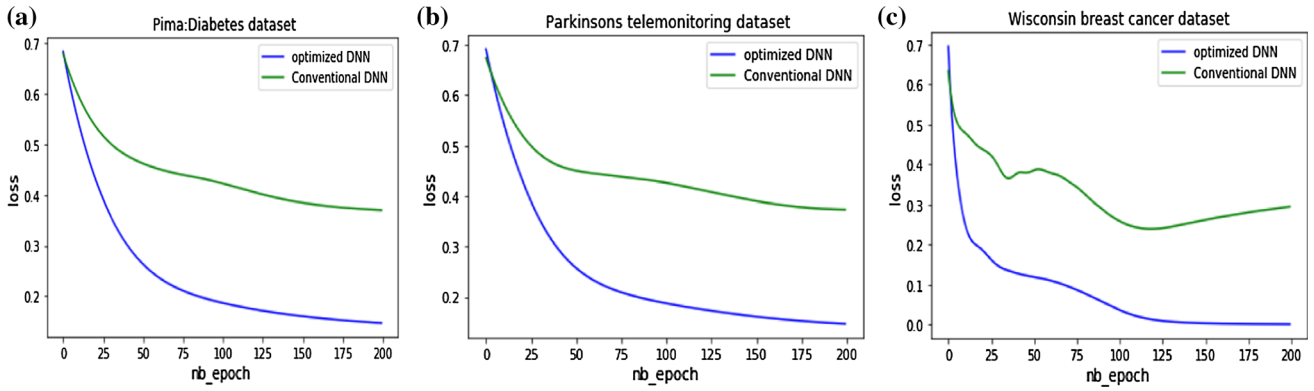

Fig. 9 Loss value **a** Pima: Diabetes dataset **b** Parkinson's telemonitoring dataset **c** Wisconsin breast cancer dataset

in all other datasets. The optimized deep learning model is relatively quick for the high-dimensional datasets. We computed the accuracy and loss value on the three datasets as appeared in Figs. 8a–c and 9a–c to further contrast the convergence effects. We can see that the Optimized DNN method is much faster than the conventional DNN method and has less degradation in the validation fold. Although the Optimized DNN converges very rapidly to an optimal solution, it is more costly computationally than the conventional DNN. We consider that the proposed model is stable, with our streamlined hyperparameter optimization process. Results show that our approach converges rapidly and achieves energetic generalization efficiency on several validation data sets. We have compared our approach to another state of the art approaches, using the same dataset, to test the efficacy of our proposed method, and the findings are given in Table 1. The previous prediction approaches have produced excellent results, with accuracy varying from 78 to 85%. As can be seen, our system has achieved maximum efficiency. With 89.23% overall accuracy on voice samples of PD patients and healthy individuals, our proposed approach produced better prediction performance in comparison to previous studies.

It is observed that the optimized deep learning model accomplishes the best classification precision of 89.23% by means of tenfold cross-validation analysis. The auspicious achievement acquired on voice samples of PD patients,

and healthy individuals have demonstrated that the method proposed can appropriately differentiate PD patients from healthy individuals. In light of the experimental investigation, it tends to be securely surmised that the advanced diagnosis method can help the doctors in making a correct demonstrative decision.

In this work, multiple implementations of the architecture have been carried out in order to select the appropriate values of parameters concerning the deep neural Network. Therefore it carries a higher cost of computation for training the system in comparison with other approaches. The trained classification model is competent enough to diagnose the patients of PD due to which its practicability is recommended and applied. The model can be well implemented in the clinical process of diagnosis by the clinicians. In future studies, an emphasis will be made on the applicability of the proposed method on diagnosing the other medical problems.

5 Conclusions and future work

Parkinson's disease is the second prevalent neurological disease after Alzheimer's disease. It affects different parts of the human body, the speech being the most susceptible. The speech study was therefore used as the foundation for diagnosing Parkinson's disease utilizing various methods for optimum numbers of trials. Right now, we have executed the idea of optimized deep learning to identify precisely PD patients to enable medical staff to make better and faster decisions. To increase the network training pace, the hyperparameter tuning of the deep learning model has been done, which has improved the precision to 91.69% with the learning rate 0.1, the ReLU activation function, ADAM as an optimizer, and seven layers. To develop prediction models with an accuracy of a high level in a relatively short period, the proposed method can analyze Parkinson's disease results automatically. It is concluded that when the deep learning model is utilized with the grid search hyperparameter tuning, the overall and mean classification precision is improved to 89.23% and 91.69%, respectively. Our experimental results show that we boost our system considerably in comparison with a state of the art automated selection method, including search performance and search results. The approach proposed here is meant to detect the early onset of Parkinson's disease. For future studies, our intention is to enhance the deep learning model to multiple levels of achieving a much better diagnosis of Parkinson's disease. In the future, a corresponding variant of grid search optimization on GPU's is under consideration.

Compliance with ethical standards

Disclosure There are no money related or different relations that could prompt an adverse circumstance.

References

1. Cai, Z., Gu, J., Wen, C., Zhao, D., Huang, C.: An intelligent Parkinson's disease diagnostic system based on a chaotic bacterial foraging optimization enhanced fuzzy KNN approach, Hindawi Computational and Mathematical Methods in Medicine, pp. 1–24 (2018)
2. Rinaldi, S., Trompetto, C.: Balance dysfunction in Parkinson's Disease. In: Hindawi BioMed Research Journal, pp. 1–10 (2015)
3. Colony, B., Maabreh, M., AlFuqaha, A.: Parameters optimization of deep learning models using Particle swarm optimization. In: International Conference of Wireless Communications, and Mobile Computing (2017)
4. Harel, B., Cannizzaro, M., Snyder, P.: Variability in fundamental frequency during a speech in prodromal and incipient Parkinson's disease: a longitudinal case study. Brain Cogn. **56**, 24–29 (2004)
5. Wang, X., Gong, G.: Detection analysis of epileptic EEG using a novel random forest model combined with grid search optimization. Front. Hum. Neurosci. **13**, 1–13 (2019)
6. Zhang, Y. N.: Can a smartphone diagnose Parkinson disease? A Deep Learning Model Method and Telediagnosis System Implementation, Hindawi Parkinson's disease, pp. 1–11 (2017)
7. Bind, S., Tiwari, A., Sahani, A.: A survey of machine learning-based approaches for Parkinson disease prediction. Int. J. Inf. Technol. Comput. Sci. **6**, 1648–1655 (2015)
8. Singh, R., Srivastava, S.: Stock prediction using deep learning. Multimed. Tools Appl. **76**, 18569–18584 (2016)
9. Fatawa, A. H., Jahardi, M. H., Ling, S. H.: Efficient diagnosis system for Parkinson's disease using deep belief network, pp. 1324–1330 (2016)
10. Sairam, N., Mandal, I.: New machine-learning algorithms for prediction of Parkinson's disease. Taylor&Francis Int. J. Syst. Sci. **45**, 647–666 (2014)
11. Tahmassebi, A., Gandomi, A., Fong, S.: Multi-stage optimization of a deep model: a case study on ground motion modeling. Plos One. **3**, 1–23 (2018)
12. Koutsoukas, A., Monaghan, K.J., Li, X., Jun, H.: Deep-learning: investigating Deep Learning Models hyperparameters and performance comparison of shallow methods for modeling bioactivity data, Springer. J. Cheminf. **3**, 1–13 (2017)
13. Lee, S., Zokhirova, M., Nguyen, T.: Effect of hyper-parameters on deep learning networks in structural engineering, pp. 537–544 (2018)
14. Yaseen, M., Anjum, A., Rana, O., Antonopoulos, N.: Hyper-parameter optimization of deep learning for video analytics in clouds. In: IEEE Transactions on Systems, Man, and Cybernetics: Systems, pp. 2168–2216 (2018)
15. www.uci-repository.com
16. Madrigal, F., Maurice, C., Lerasle, F.: Hyper-parameter optimization tools comparison for multiple object tracking applications. J. Mach. Vis. Appl. **3**, 1–12 (2018)
17. Das, R.: A comparison of multiple classification methods for the diagnosis of Parkinson's disease. Expert Syst. Appl. **37**, 568 (2010)
18. Yadav, G., Kumar, Y., Sahoo, G.: Prediction of Parkinson's disease using data mining methods: a comparative analysis of tree, statistical, and support vector machine classifiers. Indian J. Med. Sci. **65**(6), 231–242 (2011). https://doi.org/10.4103/0019-5359.107023

19. Lorenzo, P. R., Nalepa, J.: Hyper-parameter selection in deep learning models using parallel particle swarm optimization. In: GECCO Companion, pp. 1864–1887 (2017)
20. Little, M.A., McSharry, P.E., Roberts, S.J., Costello, D.A., Moroz, I.M.: Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. J. Biomed. Eng. **6**, 1–23 (2007)
21. Khemphila, A., Boonjing, V.: Parkinson's disease classification using neural network and feature selection, world academy of science, engineering, and technology. Int. J. Math. Comput. Phys. Electr. Comput. Eng. **6**, 377 (2012)
22. Rustempasic, I., Can, M.: Diagnosis of Parkinson's disease using principal component analysis and boosting committee machines. Southeast Eur. J. Soft Comput. **102**, 10 (2013)
23. Shahbakhi, M., Taheri, D.: Speech analysis for diagnosis of Parkinson's disease using genetic algorithm and support vector machine. J. Biomed. Sci. Eng. Sci. Res. **7**, 147 (2014)
24. Suganya, P., Sumathi, C.P.: A novel metaheuristic data mining algorithm for the detection and classification of Parkinson's disease. Indian J. Sci. Technol. **8**, 1 (2015)
25. Hassan, M., Sabar, N., Song, A.: Optimizing deep learning by hyper-heuristic approach for classifying good quality images, pp. 528–539 (2018)
26. Wang, L., Zhou, B., Xiang, B., Mahadevan, S.: Efficient hyperparameter optimization for NLP applications. In: International Conference on Natural Language Processing in Empirical Methods, pp. 2112–2117 (2015)
27. Gracia, N., Castaneda, A. L.: Effect of dataset sampling on the hyperparameter optimization phase over the Efficiency of a machine learning algorithm, Wiley Hindawi, pp. 1–16 (2019)
28. Zeng, X., Luo, G.: Progressive sampling-based Bayesian optimization for efficient and automatic machine learning model selection. Springer Health Inf. Sci. Syst. **3**, 1–21 (2017)
29. Yaseen, M., Anjum, A., Rana, O., Antonopoulos, N.: Deep Learning hyperparameter optimization for video analytics in clouds. IEEE Trans. Man Cybern. Syst. Oper. **3**, 2168–2216 (2018)
30. Gnana, S. K., Deepa, S. N.: Review on methods to fix number of hidden neurons in neural networks. In: Hindawi Mathematical Problems in Engineering, pp. 1–12 (2013)
31. Wang, P., Ge, R., Xiao, X., Cai, X., Wang, G.: Rectified linear unit based deep learning for biomedical multi-label data. J. Comput. Life Sci. **9**, 419–422 (2016)

Sukhpal Kaur is currently pursuing her Ph.D. from Department of Computer Engineering, Punjabi University, Patiala,India. She has done M.Tech. at Punjabi University, Patiala. She has contributed 10 articles in various research journals. Her areas of interest are bioinformatics, machine learning and data mining.

Dr. Himanshu Aggarwal, Ph.D., is currently serving as Professor in Department of Computer Scinece & Engineering at Punjabi University, Patiala. He has more than 27 years of teaching experience and served academic institutions such as Thapar Institute of Engineering & Technology, Patiala, Guru Nanak Dev Engineering College, Ludhiana, and Technical Teachers Training Institute, Chandigarh. He is an active researcher who has supervised more than 40 M.Tech. dissertations and contributed 100 articles in various research journals. He has guided 12 Ph.D. scholars and is currently guiding 8 Ph.D. scholars. He is on the editorial and review boards of several journals of repute. His areas of interest are software engineering, computer networks, information systems, ERP and parallel computing.

Dr. Rinkle Rani is working as Associate Professor in Computer Science and Engineering Department, Thapar University, Patiala. She has done her Post graduation from BITS, Pilani and Ph.D. from Punjabi University, Patiala. She has more than 22 years of teaching experience. She has over 120 research papers, out of which 56 are in international journals and 64 are in conferences. She has guided 07 PhD and 46 ME theses. Her areas of interest are Big Data Analytics and Machine Learning. She is member of professional bodies: ACM, IEEE and CSI.