Received: 22 April 2021 | Revised: 19 June 2021 | Accepted: 6 July 2021

DOI: 10.1111/exsy.12787

ORIGINAL ARTICLE

Expert Systems WILEY

Efficient detection of Parkinson's disease using deep learning techniques over medical data

Lipsita Sahu[1] | Rohit Sharma[2] | Ipsita Sahu[1] | Manoja Das[1] | Bandita Sahu[3] | Raghvendra Kumar[3]

[1]Department of Biotechnology, Gandhi Institute of Engineering and Technology, Gunupur, Odisha, India

[2]Department of Electronics & Communication Engineering, Faculty of Engineering and Technology, SRM Institute of Science and Technology, Delhi—NCR Campus, Ghaziabad, Uttar Pradesh, India

[3]Department of Computer Science and Engineering, Gandhi Institute of Engineering and Technology, Gunupur, Odisha, India

Correspondence
Rohit Sharma, Department of Electronics & Communication Engineering, Faculty of Engineering and Technology, SRM Institute of Science and Technology, Delhi—NCR Campus, Delhi-Meerut Road, Modinagar, Ghaziabad, Uttar Pradesh 201204, India.
Email: rohitr@srmist.edu.in

Abstract

Parkinson's disease is a degenerative disease that leads to brain disorder and nonfunctioning of different body parts. Deep learning tools like artificial neural network (ANN), convolution neural network (CNN), regression Analysis (RA), and so on, has been considered to a great extent in recent days. Several data sets based on the motor and nonmotor symptoms are applied to different classifier for correct identification of Parkinson's patient from healthy people. In this paper, hybridization of two deep learning tools such as, RA and ANN are done for effective diagnosis of the disease by probability estimation. The communal merits of individual approaches of the existing approaches are realized in this context for accurate probability estimation. Data preprocessing and probability estimation of preprocessed data is done in RA. The second existing approach is used to identify the PD patient by comparing with a predefined threshold value of a neuron. The estimation is performed on the data set of speech recognition, iron content, and pulse rate among a group of people. The proposed approach is compared with the existing approaches like, SVM and k-NN classifier. The computed result reveals the superiority of the proposed algorithm with 93.46% accuracy.

KEYWORDS
accuracy, artificial neural network, Parkinson's disease, regression analysis, specificity

1 | INTRODUCTION

Parkinson's disease is a brain disorder, degenerative long-term disease that leads to nonfunctioning (Little et al., 2008; Wingate et al., 2020) or nonresponse of some organs. It is caused by early nonresponse or death of the neurons (Shahid & Singh, 2020; Sivaranjini & Sujatha, 2020). These are dopaminergic neurons, found the nigral region (Fu et al., 2016). It affects the dopamine messengers which are a communication media for signal transmission between the brain and other body parts. Once the dopamine production is affected, the regulatory aspects and cognitive functions are degraded. The dopamine metabolism produces greater number of oxygen species and thereby increasing the iron content. It damages the cell components and thereby the neuronal response and function in this disease. It becomes a common disease in recent days commonly found in the age group of 65 years and above. It causes a mortality rate of approximately 1.4% throughout the globe (Dorsey et al., 2018). Dopaminergic neurons are the neurotransmitter that is used to transmit the signal from the brain to different parts of the body. Hence, dopamine depletion reduces the cognitive abilities, movement capabilities, and reproduction. Depletion of the dopaminergic neurons causes two types of symptoms, such as motor and non-motor. The motor symptoms are movement based and include stiffness, rigidity, shaking movement slowness and also movement difficulty. The non-motor symptoms are mental based and includes psychosis, sleep disturbance, depression, genitourinary problem. When the dopaminergic neurons depletion reaches to 60%, the symptoms appear in the person such as difficulty walking, tremors, and

other movement issues. Thereby, deposition or incrementing the dopamine by required amount enhances the motor skills like movement capabilities and nonmotor symptoms like cognitive abilities. Both of the symptoms are related to the aging factor (Tysnes & Storstein, 2017).

Plethora of algorithms and approaches are developed by the researchers for the early detection of the disease. Artificial neural network (ANN; Al Imran et al., 2018; Baby et al., 2017; Loconsole et al., 2018) simulates the neuronal behaviour to identify the symptoms. Nerve impulses are transmitted between the neurons through axons. Expert systems (ES) and fuzzy logic (FL; Liu et al., 2014; Pepa et al., 2020) are also used for pattern recognition of the symptoms. Several neuroimaging modalities such as single photon emission computed tomography (SPECT), magnetic resonance brain imaging (MRI), transcranial sonography, positron emission tomography (PET), functional magnetic resonance imaging (FMRI) is used for diagnosis of PD (Sivaranjini & Sujatha, 2020). SPECT and PET imaging modalities are used for visualization of the striatal region to show the dopaminergic deficiency (Mo et al., 2018). But using this modality the disease can only be identified if 80% neurons damages. Hence early diagnosis of PD is not possible for early management to allow for better treatment procedure to be followed. Recently, MRI produces promising results in with better characterization of the disease. Deep transfer learning is used to identify the PD patient based on handwriting data set (Kamran et al., 2021). Handwriting samples are combined to produce a larger data set. This is applicable to the pre-learned dataset only. Convolutional neural network (CNN) is implemented in (Oh et al., 2020) on EEG signal for identification of the PD. The architecture is a 13-layered CNN that produces the performance with 88.25% accuracy.

It detects the iron deposition in substantia nigra to facilitate the recognition of the disease progression. But, a comprehensive analysis of this approach is needed to investigate the pathological changes. Increases iron levels causes oxidative stress and thereby the neuronal death. Deep learning and neural network are the emerging technology to understand the spatial structure of nigral region. It is used for lesion detection, registration, shape modelling and disease classification. Deep neural network has an exceptional generalization capacity and capable of extracting higher level features that provide better accuracies in the disease classification. Convolutional neural network (CNN) is an ongoing research area to identify the Parkinson's disease (Martinez-Murcia et al., 2018; Shahid & Singh, 2020). The authors in Pereira et al. (2016) have used this approach for early detection of PD based on handwritten skills of the affected person. Classification of PD patient based on vocal tests has done for individual alphabets only. But other independent variables like pulse rate, iron content, and so on, are over looked (Behroozi & Sami, 2016). Using SAS (Das, 2010) regression analysis is used as a classifier to identify the PD patient. It achieved the accuracy of 92.9%. Baye's probability estimation can be implemented to identify the PD patient using relative probability. However, it is not suitable for random sample. As most of the approaches are applied with random sample Bayesian probability may not be used to compute the accurate probability. Several hybrid approaches are suggested by the researchers to achieve the combined benefits of existing approaches. Hybridization of random forest tree and synthetic minority oversampling techniques (SMOTE) is done for classification of Parkinson's patient (Polat, 2019). The approach produces an accurate result but suffers from computational complexity in terms of space. The main pitfall of this approach is its reduced efficiency in processing the larger data set.

The main emphasis of the paper includes the identification of the disease by classifying the PD and HP through the attributes value as mention in the subsequent sections. The probability has been computed for the PD patient based on the reduced pitch period entropy (PPE), the varying values of detruded fluctuation analysis (DFA), recurrence period density entropy (RPDE), increased pulse rate, and iron deposition in the substantia nigra. As iron deposition plays an important role in changing the physical and mental behaviour of a person, the present context only computes the iron content and does not classify the variants of substantial nigra.

In this paper, we have proposed a hybridized approach of regression analysis (RA; Harrell Jr, 2015) and artificial neural network (ANN) for early detection of the PD. The collected data set (Marín Méndez, 2019) is initially undergone with regression analysis. Further, the regressed data set is applied to the ANN for finding out the probability of PD symptoms. The probability of PD and healthy people (HP) have considered as 1 in total. We have applied data set for the speech to identify the PD patients. The main contribution of the paper includes (i) identification of PD probability using regression analysis as the data preprocessing technique, (ii) Computation of the probability through supervised ANN classifier, (iii) Hybridization of regression analysis and ANN for finding the probability of PD, (iv) Computer simulation using programming in C language based on dopamine production and iron content using SCILAB software and (v) performance analysis as compared to other existing approach like CNN and ANN approach.

Rest of the paper is organized into five sections. Section 2 briefly introduces the basic knowledge about regression analysis and ANN. The details of the new hybrid approach are explained in Section 3. Section 4 represents the computer simulation of the proposed approach. The performance analysis of different approaches is done in Section 5. Some conclusion and future scope are provided in Section 6.

2 | MATERIALS AND METHODS

2.1 | Regression analysis

Regression analysis is a relationship estimation method based on statistical approach. It is used to establish a one-to-one or one-to-many relationship between the dependent and independent variables. It has three variants (Chatterjee & Hadi, 2015) such as, simple linear regression, multiple

linear regressions and nonlinear regression. First two variants are used for simple dataset. But complex data set required nonlinear regression. Simple regression is used to estimate the relationship between a single dependent and single independent variable. Multiple linear regressions establish relationship between one dependent and two or more independent variables. Nonlinear regression establishes the relationship by considering the data as a function. The variants are represented using Equation (1) (simple linear regression), Equation (2) (multiple linear regression) and Equation (3) (nonlinear regression).

$$y = a + a_1 x_1 + \epsilon \tag{1}$$

$$y = a + a_1 x_1 + a_2 x_2 + a_3 x_3 + \epsilon \tag{2}$$

$$y = f(x_1, \alpha) = \left(\frac{\alpha_1 x_1}{\alpha_2 + x_1}\right) \tag{3}$$

where a is a constant, ϵ is the residual error, y is the dependent variable, x_1, x_2, x_3 are the independent variables, f is the nonlinear function and α_1 is the parameter vector.

2.2 | Artificial neural network

It is a deep learning tool also called as neural network (Kukreja et al., 2016). It is a network of nodes called as neurons responsible for sensing and transmitting the sensed signal to the brain. The connecting edges and the neurons are labelled with some value called as weights. At each neurons the received signals are processed linearly or nonlinearly to produce the output signal. The weights are adjusted to affect the strength of signal. Figure 1 describes a basic ANN.The dependent variable y is computed using an activation function as shown in Equation (4).

$$y = f(x, \alpha) = a + \sum_{i=1}^{n} x_i \alpha_i \tag{4}$$

where x and α represents the input and the weights respectively.

3 | PROBLEM FORMULATION

Regression analysis is used to estimate the relationship between dependent and independent variable. Multiple linear regression technique takes the benefits of computing the value of an attributes which is dependent on several other independent variables or attribute linearly. It is able to model the potentially important variables into a single one. It leads to produce more accurate result and clear understanding of the association between the dependent and independent variables. As the problem addresses the identification of PD from a randomly

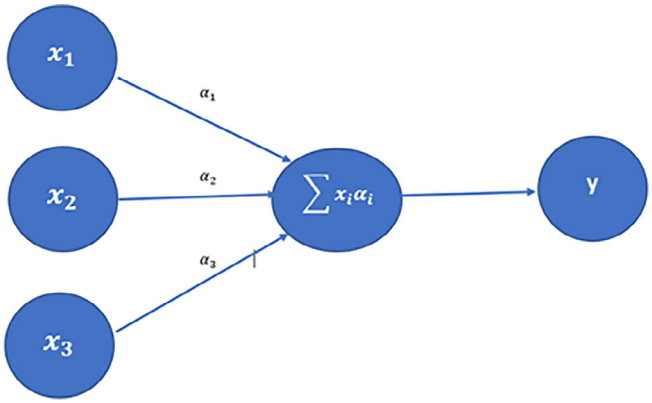

FIGURE 1 ANN components

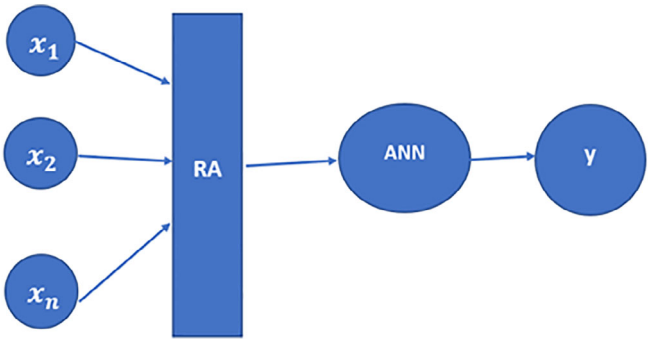

FIGURE 2 Proposed model

collected data set, nonlinear regression may not be beneficial and becomes expensive due to more number of variables. The data set undergone with the multiple linear regression followed by the activation function of the neural network. One standard value for each attribute of the data set is assumed as the threshold value. The probability of PD patient is computed using regression analysis which is having a mark able mean square error. Hence, ANN is used in second step for the computation of the probability. The proposed model is as shown in Figure 2.

The independent variables undergo the process of regression analysis and then ANN is applied. The probability of PD and HP are computed using the following equation.

$$\ln\left(\frac{p_{PD}}{p_{HP}}\right) = a + a_1x_1 + a_2x_2 + a_3x_3 + \epsilon \tag{5}$$

where p_{PD},p_{HP} represent the probability of PD patient and healthy people such that

$$p_{PD} + p_{HP} = 1 \tag{6}$$

Solution to the Equation (5) can be written as follows

$$\left(\frac{p_{PD}}{p_{HP}}\right) = e^{a+a_1x_1+a_2x_2+a_3x_3+\epsilon} \Rightarrow \frac{p_{PD}}{1-p_{PD}} = e^{a+a_1x_1+a_2x_2+a_3x_3+\epsilon}$$

$$\Rightarrow p_{PD} = e^{a+a_1x_1+a_2x_2+a_3x_3+\epsilon} - e^{a+a_1x_1+a_2x_2+a_3x_3+\epsilon} * (p_{PD})$$

$$\Rightarrow p_{PD} = \frac{e^{a+a_1x_1+a_2x_2+a_3x_3+\epsilon}}{1-e^{a+a_1x_1+a_2x_2+a_3x_3+\epsilon}} \tag{7}$$

Equation (7) is obtained by applying the multi linear regression analysis. Now, the ANN is implemented to produce the probability of PD patient.

3.1 | Proposed algorithm

Based on the proposed model shown in previous section the computational steps are identified and listed in Algorithm 1. It is a two-step process. In the first step the input values are fed to the regression analysis components. Data preprocessing is achieved in this step. Multiple linear regressions are applied using Equation (7) to the preprocessed data for finding the probability of PD patient. In the second step, the probability obtained for different input variables are passed through an ANN. To achieve the combine benefits of individual approach the existing algorithms are hybridized. The regression analysis algorithm is used to preprocess the data. The second algorithm is used to compute the probability of the preprocessed data. Regression coefficients (Austin & Merlo, 2017) are used as the weight parameters in the ANN. The values are compared with the threshold value at the neuron and the corresponding output is generated.

Algorithm 1. $RA(x,y,a,\epsilon)$

Input : independent variable, residual error,
Output : probability of PD in terms of the independent variable

1. Read the independent variable as the data set
2. Apply regression analysis using Equationn (5)
3. Find p_PD using Equation (7)
4. Stop

ANN approach and computation of the probability is enlisted in Algorithm 2. It describes the input, output and computational steps. Threshold value is initialized at the neuron to produce the PD probability.

Algorithm 2. $\mathrm{ANN}(x,y,\alpha)$

Input : independent variable, weights,
Output : probability of PD in terms of the independent variable

1. Read the independent variable as the data set
2. Read the coefficient vector
3. Initialize the threshold value at the neuron.
4. Apply ANN formula using Equation (4)
5. Find p_{PD} by comparing with the threshold value.
6. Stop

Once the probability of the PD patients with respect to different values of the independent variable is computed, it is fed to the ANN. The ANN computes the PD probability by incorporating weights with it. Different types of parameters we have considered for the identification of disease such as pitch period entropy (PPE), detrended fluctuation analysis (DFA), Recurrence period density entropy (RPDE), pulse and iron content. Hybridization of RA and ANN is done in Algorithm 3.

Algorithm 3. $\mathrm{RA}-\mathrm{ANN}(x,y,a,\epsilon)$

Input : independent variable, residual error, weights, m,n and threshold
Output : probability of PD in terms of the independent variable

1. Read the independent variable as the data set
2. Read the coefficient vector
3. Initialize the threshold value at the neuron.
4. For $i=1$ to m
5. For $j=1$ to n
6. Apply regression analysis using Equation (5)
7. Find p_PD $[i]$ using Equation(7)
8. Apply ANN using Equation (4)
9. Find $p_{PD}[i]$ by comparing with the threshold value.
10. Stop

Using the parameter as mentioned above, we have computed for 10 values of each attributes from the data set. Probability estimation is done as per algorithms. The probability obtained using RA, ANN and RA-ANN is shown in Table 1. Each of the attribute is associated with a threshold

TABLE 1 PD probability based on parameters values

Sl. No.	PPE	DFA	RPDE	PULSE	IRON	p_{PD} RA	p_{PD} ANN	p_{PD} RA-ANN
1	0.85247	0.71826	0.57227	240	158.75	1	0	0
2	0.76686	0.69481	0.53966	234	108.23	1	1	1
3	0.85083	0.67604	0.58982	232	138.28	0	0	0
4	0.41121	0.79672	0.59257	178	145.7	1	1	1
5	0.3279	0.79782	0.53028	236	149.56	1	1	1
6	0.5078	0.78744	0.65451	226	109.12	1	1	1
7	0.76095	0.62145	0.54543	322	147.98	1	1	1
8	0.83671	0.62079	0.51179	318	107.10	0	1	0
9	0.80826	0.61766	0.50447	318	138.25	1	1	1
10	0.85302	0.62247	0.54855	493	142.72	0	0	0

TABLE 2 Varying attributes value for PD and HP

Attribute	PD	HP
PPE	Reduced frequency less than the range (100–250) Hz	(100–250) Hz
DFA	Strides interval range 0.95–1.15	Increased strides
RPDE	Unstable with deviation 0.05	stable
PULSE	Increased pulse rate	72
IRON	$107.21 \pm 34.25\,\mu g/dl$	$110.21 \pm 48.75\,\mu g/dl$

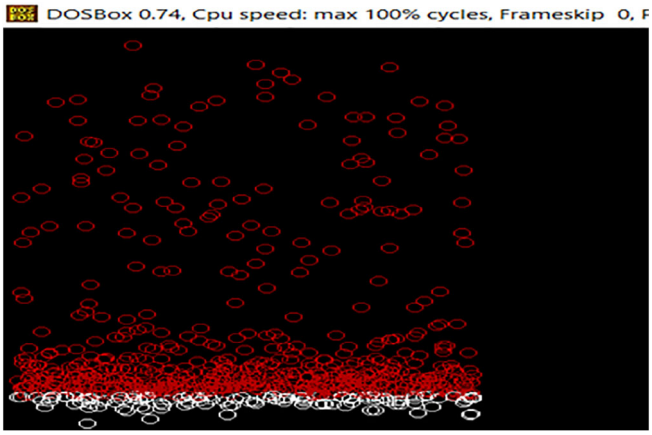

FIGURE 3 Probability-based classification based on PPE

value and it is not same for the PD and HP as shown in Table 2. If the count of a parameter goes beyond the value (Oh et al., 2020), PD probability increases. Otherwise, it is assumed to be zero.

4 | EXPERIMENTAL SETUP

The proposed approach is simulated using programming in C language with SCILAB software for plotting. Five attributes such as PPE, DFA, RPDE, pulse rate, and iron content are considered for classification of the HP and PD people. Each attributes with 741 values are fed to the RA components and the computed probabilities are stored in a file. Again, the probability values are operated in with ANN components to produce the final probability of PD patient. Standard value (Hinnell et al., 2012; Martínez-Martín et al., 1994) for each attribute is identified. Based on those values, the classification is done. Data set is fed to the RA component and the classification of PD and HP based on PPE is as shown in Figure 3. By considering the DFA as the independent variable, the data set is fed to the computer and the classification is recorded in Figure 4. RPDE is an

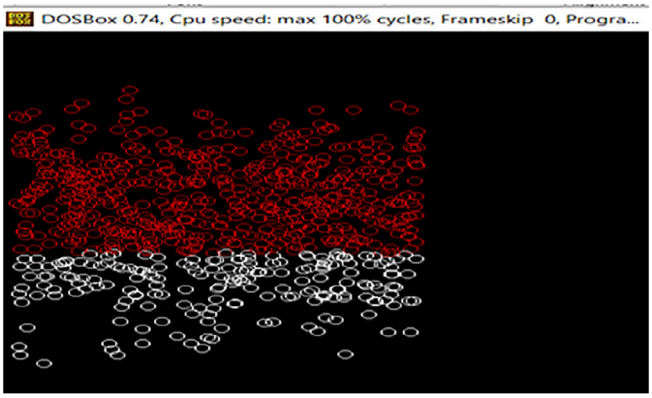

FIGURE 4 Probability-based classification based on DFA

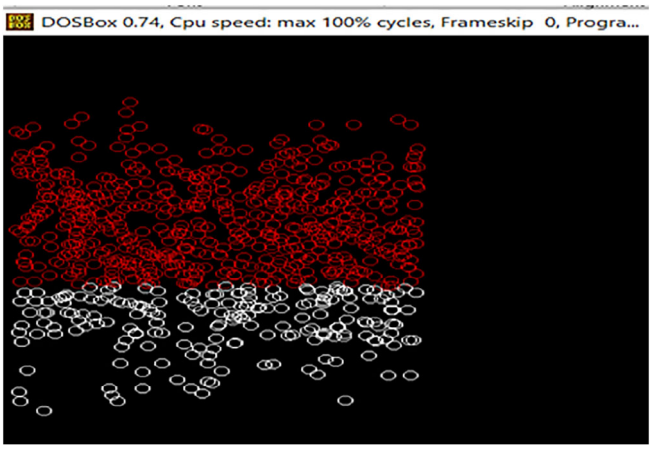

FIGURE 5 Probability-based classification based on RPDE

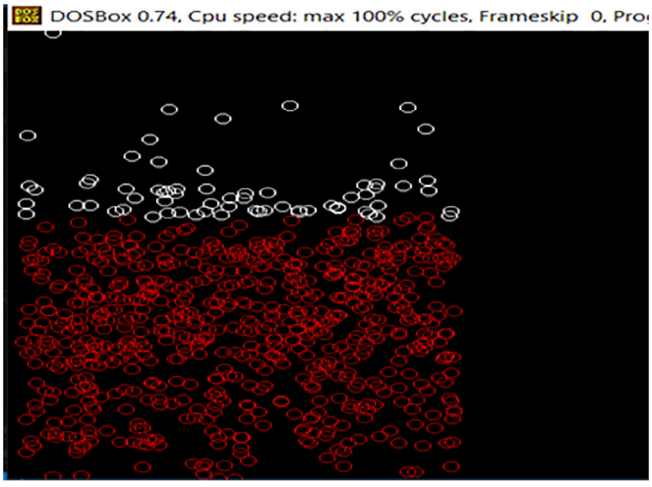

FIGURE 6 Probability-based classification based on pulse

important parameter used to identify the probability of PD. Approximately 741 from the dataset is applied to the programming set up and obtained the classification as shown in Figure 5. Probability is computed by taking the value from the data set for pulse rate of different people and the estimation-based classification is represented in Figure 6. The iron content increases in a PD patient. Based on the iron content (Chen et al., 2019; Mostile et al., 2017), the probability is estimated to identify the class of a person belong to PD or HP class. The iron-content based classification is shown in Figure 7. The estimated probability for each classification is compared with the desired probability. It is obtained from the computation that each classification hardly deviates from the desired value.

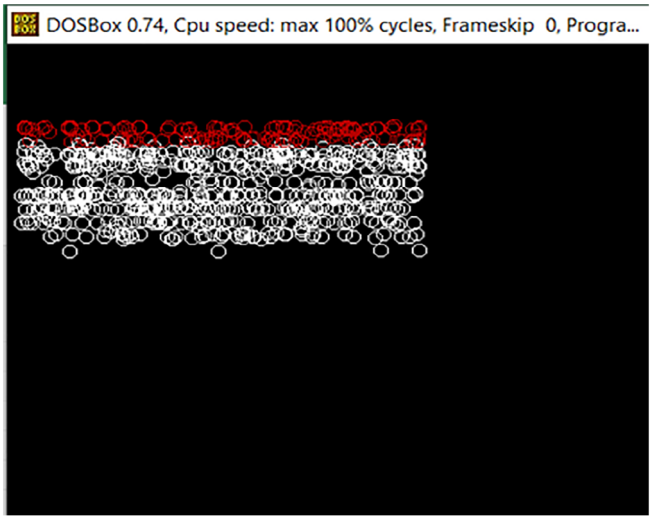

FIGURE 7 Probability-based classification based on iron content

TABLE 3 RME and SD

Approach	RME	*SD*
RA	0.18819	0.936
ANN	0.18813	0.934
RA-ANN	0.18801	0.375

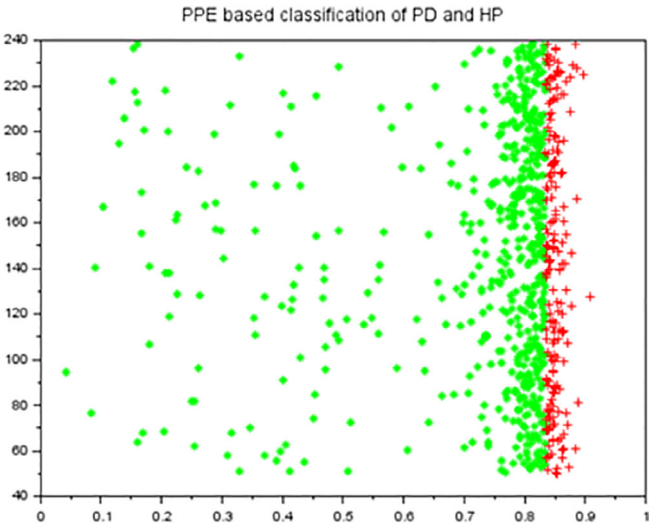

FIGURE 8 PPE based classification

5 | PERFORMANCE ANALYSIS

Performance analysis of the proposed approach is done based on the applied variables. Probability estimation for RA, ANN and RA-ANN is done in the algorithms. The values obtained are compared with the desired probability and the classification group as well. The root means square error (RME) and standard deviation (*SD*) computed for each of the independent variables (Kriegel, 2018) as mentioned in Table 1. The hybridized approach produces the result with less error due to its multilinearity neurons. The values for these three approaches are as shown in Table 3.

Data set is separated into different attributes based on the independent variables. The separated data set is applied to the proposed approach. It is applied to RA, ANN, RA-ANN, k-NN (k-nearest neighbour), SVM (linear) and SVM (RFB (radial basis kernels) model; Lahmiri et al., 2018). PPE data set as an independent variable is fed to the proposed model and the classification obtained is recorded in Figure 8. The

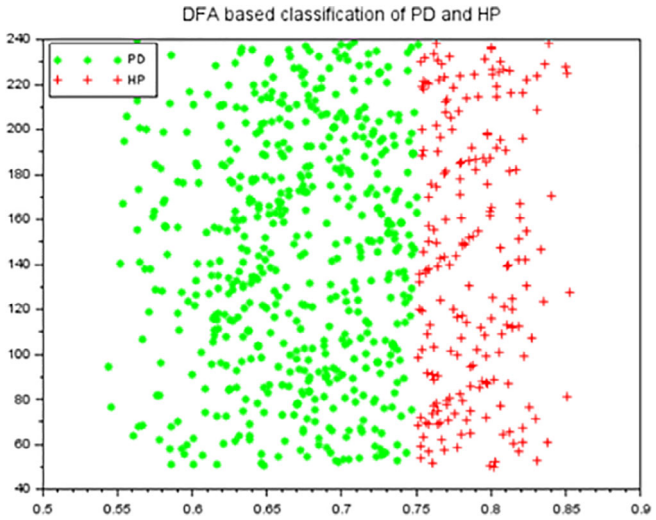

FIGURE 9 DFA based classification

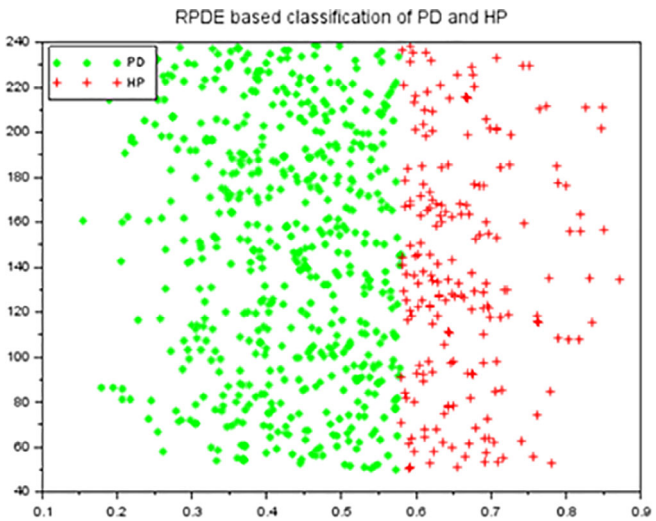

FIGURE 10 RPDE based classification

probability is computed based on the values of DFA attributes associated with the RA coefficient. The classification is done as per the probability and pictorially shown in Figure 9. The classification of PD and HP based on RPDE component is computed according to the hybridized algorithm and the classification is represented in Figure 10. Pulse rate may vary for each individual. It is quite different between the PD and HP group of people. The pulse rate from the data set is fed to the proposed model to classify the group of people based on values. The classification is recorded in Figure 11. The classification of PD and HP based on the independent variable iron is fed to the above-mentioned model and computed group is as shown in Figure 12.

For the close investigation of the results, sensitivity, accuracy and specificity obtained for each method. There exist some other terms that describes these parameters are true positive, true negative, false positive, and false negative. The total number of subjects with positive tests, the total number of subjects with negative tests, and the total subjects of study used to compute the sensitivity, specificity, and accuracy of different approaches. With reference to Table 3, the approaches with these above parameters' values are shown in Table 3. Sensitivity, specificity and accuracy (Beach & Adler, 2018; Grover et al., 2018; Quattrone et al., 2018) is computed using Equation (8), Equation (9) and Equation (10) respectively

$$\text{Sensitivity} = \frac{T_P}{T_P + F_N} \tag{8}$$

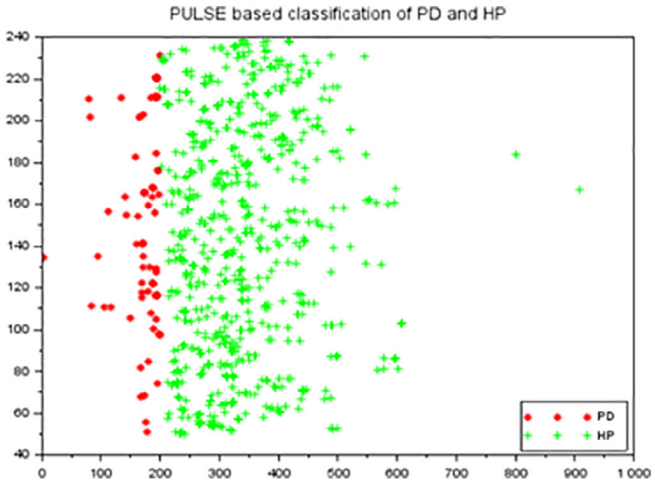

FIGURE 11 Pulse based classification

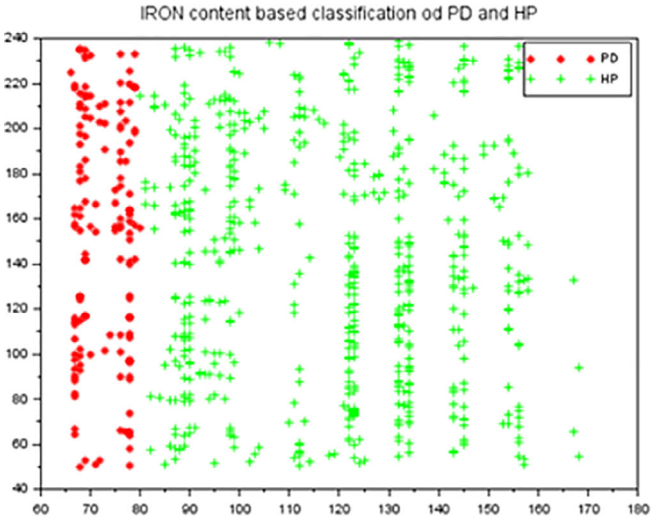

FIGURE 12 Iron based classification

$$\text{Specificity} = \frac{T_N}{T_N + F_P} \tag{9}$$

$$\text{Accuracy} = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \tag{10}$$

Number of patient identified correctly is represented by T_P. Number of patient identified incorrectly is represented by F_N. Number of healthy people identified correctly is represented by T_N. Number of healthy people incorrectly is represented by F_P. Values for each independent variables are fed to RA, ANN and the proposed model. By applying the regression analysis to the parameters along with the RA coefficient, the computed values are recorded in RA column of Table 4. The RA computed value is again fed to the ANN model and the obtained values are shown in ANN column of that table. Finally, the probability is computed for the proposed model and compared with the desired probability. Only five value we have shown in the table. However it is computed for 741values of each independent variable. The overall accuracy obtained in the proposed approach is 93.46 which is superior to all other approaches as shown in Table 5. It also depicts the specificity and sensitivity of the proposed approach. Results obtained from simulation and computation of the proposed model reveals the supremacy of RA-ANN.

Based on the Parkinson's dataset several studies have reported with comparable performance. Using MR and SPECT images in CNN (Sivaranjini & Sujatha, 2020) the classification of PD and HP done with an accuracy of 88.9% with the use of several dimensionality reduction (Salvatore et al., 2014) and machine learning approaches (Long et al., 2012; Nemmi et al., 2015) produces the result with an accuracy of

TABLE 4 Probability estimation and classification

Sl. No.	Parameter	Coefficient	RA	ANN	RA-ANN	Probability desired
1	PPE	1.362	−1.4915	1.160	1	1
2		−0.003	433	−0.0023	0	0
3		7.856	−1.0012	6.6841	1	1
4		−0.021	115.77	−0.0086	0	0
5		0.005	−770.46	−0.0016	0	0
6	DFA	1.362	−4.3196	0.97827	1	1
7		−0.003	0.00010	−0.002084	0	0
8		7.856	−1.0049	5.31097	1	1
9		−0.021	59.2415	−0.01673	0	0
10		0.005	−252.2582	0.00398	**1**	**0**
11	RPDE	1.362	1.2601	0.7794	1	1
12		−0.003	−0.0087	0.0016	0	0
13		7.856	1.0832	4.3336	1	1
14		−0.021	−0.0025	−0.0124	0	0
15		0.005	28.4511	0.0026	0	0
16	PULSE	1.362	−1.122	326.88	1	1
17		−0.003	1.0051	−0.702	0	0
18		7.856	−0.0010	1822.592	1	1
19		−0.021	0.00437	−3.738	0	0
20		0.005	0.000437	1.18	**1**	**0**
21	IRON	1.362	0.10212	216.2175	1	1
22		−0.003	−0.0011	−0.3246	0	0
23		7.856	12.5643	1086.3276	1	1
24		−0.021	−0.0062	−3.0597	0	0
25		0.005	−0.0101	0.7478	0	0

Note: Bold values are the dissimilarity in result.

TABLE 5 Result comparison of different approaches

Approach	Specificity	Sensitivity	Accuracy
RA	52.3	70	70
ANN	65.00	82.87	80
RA ANN	**67.34**	**95.64**	**93.46**
K-NN	80	75	77.50
SVM (linear)	85	85	85
SVM (RFB)	85	87	90

Note: Bold values are the dissimilarity in result.

approximately 74%. k-NN classifier is applied to the data set with different distance values and the accuracy obtained is 77.50% on an average. SVM classifier with linear and radial basis function is applied to the data set. Considering the scaling factor to be 3 and penalty factor as 1 we obtained the accuracy of 85% for linear and 90% for radial basis function. Using SAS as a software needs two phases. The first one is data preprocessing and the second one for data analysing. It achieves 92.9% accuracy.

Further, the proposed approach is compared with some other existing approaches such as deep transfer learning (DTL) and CNN with different datasets as depicted in Table 6. The DTL approach uses handwriting data set and produces 99.22% accuracy. However, it is implemented with pre-learned data and the accuracy may decreases while processing the raw data. Similarly, 13-layered CNN is used with EEG and MR preprocessed dataset and produce 88.25% and 88.9% accuracy, respectively. Table 6 reveals the superiority of the proposed approach

implemented on speech dataset with five attributes each with 741 values that produces 93.46% accuracy. It is feasible and efficient as it is capable to process the raw data values.

Figure 13 depicts the accuracy comparison of the above-mentioned approaches along with the use of SAS. It shows the supremacy of the RA-ANN approach among all. We have considered several values of the independent parameters ranging from 10 to 741. It shows the accuracy is 100% for less values and decreases with increases with the increased value. The specificity of the classification approaches compared in Figure 14. Based on different independent variables with a set of 741 values the specificity is computed. True negative, true positive, false negative and false positive values are identified on applying the data set in several approaches. The sensitivity computed based on these attributes are recorded in Figure 15.

TABLE 6 Comparative study of deep learning tools applied approaches

Approach	Data set used	Accuracy	Types of data
DTL	Handwriting	99.22%	Pre-learned
Thirteen-Layered CNN	EEG	88.25%	Preprocessed
CNN	MR	88.9%	Preprocessed
PROPOSED	Speech	93.46%	Raw, preprocessed

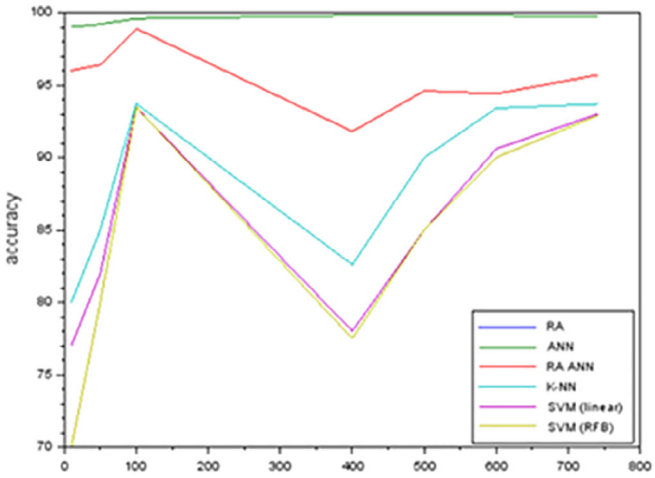

FIGURE 13 Accuracy comparison of prediction approaches

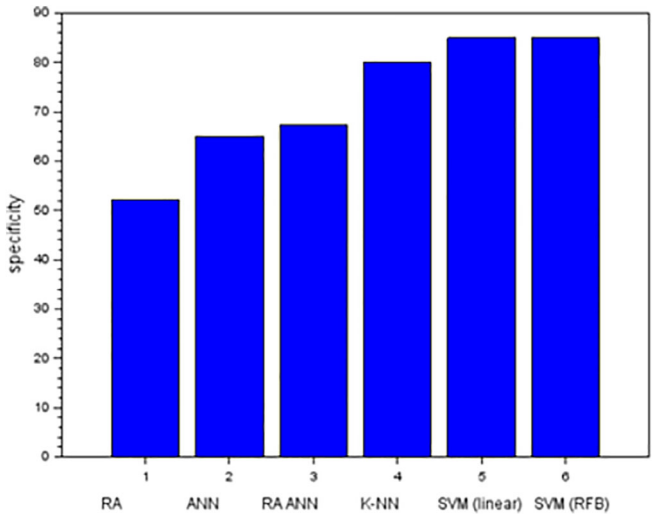

FIGURE 14 Specificity comparison

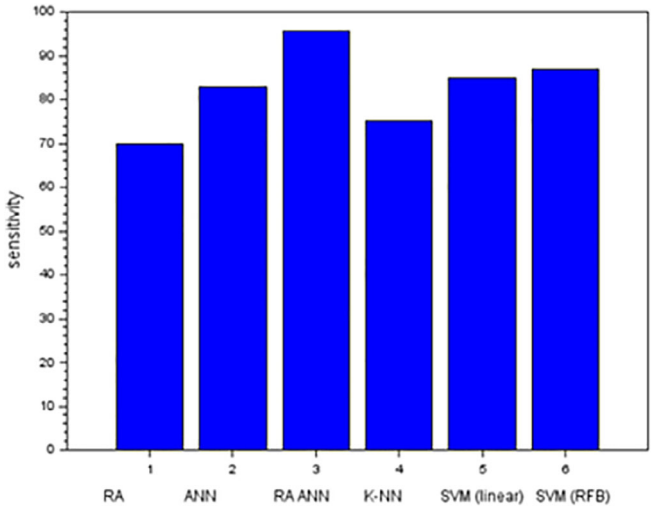

FIGURE 15 Sensitivity comparison

6 | CONCLUSION

Early detection of Parkinson disease has become essential to our society. Deep learning has received greater attention in recent years. The progress of the disease and probability of getting this disease can be assessed by speech and some other motor and nonmotor symptoms. This paper has implemented a hybrid deep learning approach. A single approach lacks in accuracy. And the bottleneck of that single approach of RA is resolved by hybridizing it with ANN model. Classification is done based on the data set with 741 values of independent parameters as mentioned in literature. An accuracy of 93.46% is obtained for PD and HP class. The proposed approach can be extended with more deep learning tools and application.

CONFLICT OF INTEREST

None.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

REFERENCES

Al Imran, A., Rahman, A., Kabir, M. H., & Rahim, M. S. (2018). The impact of feature selection techniques on the performance of predicting Parkinson's disease. *International Journal of Information Technology and Computer Science*, *10*(7), 40.

Austin, P. C., & Merlo, J. (2017). Intermediate and advanced topics in multilevel logistic regression analysis. *Statistics in Medicine*, *36*(20), 3257–3277.

Beach, T. G., & Adler, C. H. (2018). Importance of low diagnostic accuracy for early Parkinson's disease. *Movement Disorders*, *33*(10), 1551–1554.

Baby, M. S., Saji, A. J., & Kumar, C. S. (2017). Parkinson's disease classification using wavelet transform based feature extraction of gait data. In 2017 international conference on circuit, power and computing technologies (ICCPCT). IEEE (pp. 1–6).

Behroozi, M., & Sami, A. (2016). A multiple-classifier framework for Parkinson's disease detection based on various vocal tests. *International Journal of Telemedicine and Applications*, *2016*, 6837498.

Chatterjee, S., & Hadi, A. S. (2015). *Regression analysis by example* (pp. 1–423). John Wiley & Sons.

Chen, Q., Chen, Y., Zhang, Y., Wang, F., Yu, H., Zhang, C., & Luo, W. (2019). Iron deposition in Parkinson's disease by quantitative susceptibility mapping. *BMC Neuroscience*, *20*(1), 1–8.

Das, R. (2010). A comparison of multiple classification methods for diagnosis of Parkinson disease. *Expert Systems with Applications*, *37*(2), 1568–1572.

Dorsey, E. R., Elbaz, A., Nichols, E., Abd-Allah, F., Abdelalim, A., Adsuar, J. C., & Murray, C. J. (2018). Global, regional, and national burden of Parkinson's disease, 1990–2016: A systematic analysis for the global burden of disease study 2016. *The Lancet Neurology*, *17*(11), 939–953.

Fu, Y., Paxinos, G., Watson, C., & Halliday, G. M. (2016). The substantia nigra and ventral tegmental dopaminergic neurons from development to degeneration. *Journal of Chemical Neuroanatomy*, *76*, 98–107.

Grover, S., Bhartia, S., Yadav, A., & Seeja, K. R. (2018). Predicting severity of Parkinson's disease using deep learning. *Procedia Computer Science*, *132*, 1788–1794.

Hinnell, C., Hurt, C. S., Landau, S., Brown, R. G., Samuel, M., & PROMS-PD Study Group. (2012). Nonmotor versus motor symptoms: How much do they matter to health status in Parkinson's disease. *Movement Disorders*, *27*(2), 236–241.

Harrell, F. E., Jr. (2015). *Regression modeling strategies: With applications to linear models, logistic and ordinal regression, and survival analysis* (pp. 1–553). Springer.

Kriegel, Z. (2018). Effect of Parkinson disease on concurrent speech and manual task performance (Doctoral dissertation, Bowling Green State University) (pp. 1–113).

Kukreja, H., Bharath, N., Siddesh, C. S., & Kuldeep, S. (2016). An introduction to artificial neural network. *International Journal of Advance Research and Innovative Ideas in Education*, *1*, 27–30.

Kamran, I., Naz, S., Razzak, I., & Imran, M. (2021). Handwriting dynamics assessment using deep neural network for early identification of Parkinson's disease. *Future Generation Computer Systems*, *117*, 234–244.

Lahmiri, S., Dawson, D. A., & Shmuel, A. (2018). Performance of machine learning methods in diagnosing Parkinson's disease based on dysphonia measures. *Biomedical Engineering Letters*, *8*(1), 29–39.

Little, M., McSharry, P., Hunter, E., Spielman, J., & Ramig, L. (2008). Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *Nature Precedings*, *56*(4), 1015.

Loconsole, C., Cascarano, G. D., Lattarulo, A., Brunetti, A., Trotta, G. F., Buongiorno, D., & Di Sciascio, E. (2018). A comparison between ANN and SVM classifiers for Parkinson's disease by using a model-free computer-assisted handwriting analysis based on biometric signals. In 2018 international joint conference on neural networks (IJCNN). IEEE (pp. 1–8).

Long, D., Wang, J., Xuan, M., Gu, Q., Xu, X., Kong, D., & Zhang, M. (2012). Automatic classification of early Parkinson's disease with multi-modal MR imaging. *PLoS One*, *7*(11), e47714.

Liu, S., Shen, Z., McKeown, M. J., Leung, C., & Miao, C. (2014). A fuzzy logic based Parkinson's disease risk predictor. In 2014 IEEE international conference on fuzzy systems (FUZZ-IEEE). IEEE (pp. 1624–1631).

Marín Méndez, J. J. (2019). Desarrollo de un algoritmo con machine learning para la clasificación de pacientes con Parkinson (Master's thesis) (pp. 1–175).

Mo, S. J., Axelsson, J., Jonasson, L., Larsson, A., Ögren, M. J., Ögren, M., & Riklund, K. (2018). Dopamine transporter imaging with [18 f] fe-pe2i pet and [123 i] fp-cit spect—A clinical comparison. *EJNMMI Research*, *8*(1), 1–13.

Martinez-Murcia, F. J., Górriz, J. M., Ramírez, J., & Ortiz, A. (2018). Convolutional neural networks for neuroimaging in Parkinson's disease: Is preprocessing needed? *International Journal of Neural Systems*, *28*(10), 1850035.

Martínez-Martín, P., Gil-Nagel, A., Gracia, L. M., Gómez, J. B., Martinez-Sarries, J., Bermejo, F., & Cooperative Multicentric Group. (1994). Unified Parkinson's disease rating scale characteristics and structure. *Movement Disorders*, *9*(1), 76–83.

Mostile, G., Cicero, C. E., Giuliano, L., Zappia, M., & Nicoletti, A. (2017). Iron and Parkinson's disease: A systematic review and meta-analysis. *Molecular Medicine Reports*, *15*(5), 3383–3389.

Nemmi, F., Sabatini, U., Rascol, O., & Péran, P. (2015). Parkinson's disease and local atrophy in subcortical nuclei: Insight from shape analysis. *Neurobiology of Aging*, *36*(1), 424–433.

Oh, S. L., Hagiwara, Y., Raghavendra, U., Yuvaraj, R., Arunkumar, N., Murugappan, M., & Acharya, U. R. (2020). A deep learning approach for Parkinson's disease diagnosis from EEG signals. *Neural Computing and Applications*, *32*(15), 10927–10933.

Pereira, C. R., Weber, S. A., Hook, C., Rosa, G. H., & Papa, J. P. (2016). Deep learning-aided Parkinson's disease diagnosis from handwritten dynamics. In 2016 29th SIBGRAPI conference on graphics, patterns and images (SIBGRAPI). IEEE (pp. 340–346).

Polat, K. (2019). A hybrid approach to Parkinson disease classification using speech signal: The combination of smote and random forests. In 2019 scientific meeting on electrical-electronics & biomedical engineering and computer science (EBBT). IEEE (pp. 1–3).

Pepa, L., Capecci, M., Andrenelli, E., Ciabattoni, L., Spalazzi, L., & Ceravolo, M. G. (2020). A fuzzy logic system for the home assessment of freezing of gait in subjects with Parkinsons disease. *Expert Systems with Applications*, *147*, 113197.

Quattrone, A., Morelli, M., Nigro, S., Quattrone, A., Vescio, B., Arabia, G., & Caracciolo, M. (2018). A new MR imaging index for differentiation of progressive supranuclear palsy-parkinsonism from Parkinson's disease. *Parkinsonism & Related Disorders*, *54*, 3–8.

Salvatore, C., Cerasa, A., Castiglioni, I., Gallivanone, F., Augimeri, A., Lopez, M., & Quattrone, A. (2014). Machine learning on brain MRI data for differential diagnosis of Parkinson's disease and progressive supranuclear palsy. *Journal of Neuroscience Methods*, *222*, 230–237.

Shahid, A. H., & Singh, M. P. (2020). A deep learning approach for prediction of Parkinson's disease progression. *Biomedical Engineering Letters*, *10*(2), 227–239.

Sivaranjini, S., & Sujatha, C. M. (2020). Deep learning based diagnosis of Parkinson's disease using convolutional neural network. *Multimedia Tools and Applications*, *79*(21), 15467–15479.

Tysnes, O. B., & Storstein, A. (2017). Epidemiology of Parkinson's disease. *Journal of Neural Transmission*, *124*(8), 901–905.

Wingate, J., Kollia, I., Bidaut, L., & Kollias, S. (2020). Unified deep learning approach for prediction of Parkinson's disease. *IET Image Processing*, *14*(10), 1980–1989.

AUTHOR BIOGRAPHIES

Lipsita Sahu is currently persuing her m.tech in department of biotechnology, GIET University, Bunupur, odisga, India. She has completed her MSC in Biotechnology from, GIBS, Gunuour under Berhampur university in the year 2019.

Rohit Sharma is working as an Assistant Professor in the Department of Electronics and Communication Engineering, SRM Institute of Science and Technology, Delhi NCR Campus Ghaziabad, India. He is an active member of ISTE, IEEE, ICS, IAENG, and IACSIT. He is an editorial board member and reviewer of more than 12 international journals and conferences, including the topmost journal *IEEE Access and IEEE Internet of Things Journal*. He serves as a Book Editor for seven different titles to be published by CRC Press, Taylor & Francis Group, USA and Apple Academic Press, CRC Press, Taylor & Francis Group, USA, Springer, and so on. He has received the Young Researcher Award in '2nd Global Outreach Research and Education Summit & Awards 2019' hosted by Global Outreach Research & Education Association (GOREA).

Ipsita Sahu is currently pursuing her MTech in Department of Biotechnology, GIET University, Gunupur, Odisha, India. She has completed her MSc in Biotechnology from, GIBS, Gunupur under Berhampur University in the year 2019.

Manoja Das is now working as an Associate Professor in Biotechnology at the Department of Biotechnology, GIET University, Gunupur, Odisha. Dr. Das has got 20 years of teaching experience at the graduate and post graduate level. His area of interest lays in the field of microbiology, plant biotechnology and Nano biotechnology. He has published around 30 research papers in different repute journals. He is also a popular science article writer in local languages. He is the recipient of young scientist award 2002 of Odisha Bigyan Congress. Currently around six PhD scholars are working under him for their PhD work. He has guided many projects of UG and PG students.

Bandita Sahu is currently pursuing her PhD in the Department of Computer Science and Engineering Veer Surendra Sai University of Technology (VSSUT; A Unitary Technical University, Established by Govt. of Odisha), Burla, Odisha. She has received her MTech from National Institute of Technology (NIT), Rourkela, India in 2015. She has 16 years of teaching experience. Her current research includes intelligent robotics, machine intelligence and cluster computing. She has published more than 12 papers in top international journals and conference proceedings. She has guided five MTech scholars.

Raghvendra Kumar is working as Associate Professor in Computer Science and Engineering Department at GIET University, India. He received BTech, MTech and PhD in Computer Science and Engineering, India, and Postdoc Fellow from Institute of Information Technology, Virtual Reality and Multimedia, Vietnam. He serves as Series Editor Internet of Everything (IOE): Security and Privacy Paradigm, Green Engineering and Technology: Concepts and Applications, publishes by CRC Press, Taylor & Francis Group, USA, and Bio-Medical Engineering: Techniques and Applications, Publishes by Apple Academic Press, CRC Press, Taylor & Francis Group, USA.

How to cite this article: Sahu, L., Sharma, R., Sahu, I., Das, M., Sahu, B., & Kumar, R. (2022). Efficient detection of Parkinson's disease using deep learning techniques over medical data. *Expert Systems*, *39*(3), e12787. https://doi.org/10.1111/exsy.12787