BMC Medical Informatics and Decision Making

RESEARCH Open Access

Late feature fusion using neural network with voting classifier for Parkinson's disease detection

Abeer Aljohani[1*]

Abstract
Parkinson's disease (PD) is classified as a neurological, progressive illness brought on by cell death in the posterior midbrain. Early PD detection will assist doctors in reducing the disease's consequences. A collection of skilled models that may be applied to regression as well as classification is known as artificial intelligence (AI). PD can be detected using a variety of dataset formats, including text, speech, and picture datasets. For the purpose of classifying Parkinson's disease, this study suggests merging deep with machine learning recognition approaches. The three primary components of the suggested approach are designed to enhance the accuracy of Parkinson's disease early diagnosis. These sections cover the topics of categorising, combining, and separating. Convolutional Neural Networks (CNN) as well as attention procedures are used to create feature extractors. The related motion signals are fed to a combination of convolutional neural network and long-short-memory model for feature extraction. Besides, for the classification of patients from non-suffers of Parkinson's disease, Random Forest, Logistic Regression, Support Vector Machine, Extreme Boot Classifier, and voting classifier were used. Our result shows that for the PD handwriting and related motion datasets, using the proposed CNN with an attention and voting classifier yields 99.95% accuracy, 99.99% precision, 99.98% sensitivity, and 99.95% F1-score. Based on these results, it is warranted to conclude that the proposed methodology of feature extraction from photos of handwriting and relating motor symptoms, fusing of those features, and following it with a voting classifier yields excellent results for PD classification.

Keywords Parkinson's disease, Fusion, Machine learning, Attention mechanism, Voting classifier

Introduction

Parkinson's Disease (PD) is taken to be a neurodegenerative disorder. Most of the changes occur within the motor system of the brain, which leads to progressive deterioration of the nerve cells. Causes of PD are most likely genetic inheritance combined with certain environmental factors like toxic materials such as pesticides and heavy metals [1]. Further involvement of heritable and non-heritable causes in its pathogenesis has been implicated with genetic mutations associated with PD, including those affecting genes like leucine-rich repeat kinase 2 (LRRK2) and Synuclein Alpha (SNCA). To date, it is estimated that more than 10 million people worldwide are afflicted with PD; thus, it is one of the most common neurodegenerative conditions. Most diagnoses of PD are based on its clinical symptoms, especially motor symptoms. Main manifestations of PD include the following: Tremors-either generally beginning in the hands,

*Correspondence:
Abeer Aljohani
aahjohani@taibahu.edu.sa
[1]Department of Computer Science and Informatics, Taibah University, Medina 42353, Saudi Arabia

arms, jaw or legs or sometimes worsening over time [2]. In addition to tremors, many other signs can also show up when a person has PD. These include difficulties with eating, wherein chewing problems may lead to weight loss and nutritional deficiencies. Other manifestations include slowness of movement that makes daily activities increasingly difficult. Muscle rigidity can also be constant, impeding mobility and adding to discomfort. Besides this, sudden jerking or spasm movements without voluntary control occur very frequently in PD, complicating motor control and increasing the risk of falls. Aggregated, these symptoms grossly affect the quality of life and call for comprehensive medical intervention and support.

The most important complications of PD disease are thinking difficulties, depression, emotional changes, constipation, bladder problems, sleep disorder problems, etc [3]. Due to this problem, constant care for Parkinson's patients is necessary. In the past two years, COVID-19 has affected the social life of all people. The social life of these patients and the need for constant care won't permit them to follow the restrictions of quarantine roles imposed by healthcare systems. Thus, PD patients are a risk group for COVID-19 patients besides the regular complications. Physicians usually track PD symptoms to detect them in patients. One of the standard criteria that can be used for detecting PD is the differences in the dynamic disorder of hand movement while sketching circular shapes. The severity of PD can be specified using H and Y rating scales [4–6]. The physical characteristics of handwriting sketches can also be exploited as biomarkers for PD detection. This approach depend on the time series data typifying handwriting such as the number of strokes and speed of sketching the spiral and wave shapes that have been recorded using digitizing tablets provided with electronic pens [7–10].

Numerous biomedical biomarkers have been employed in the identification of PD, such as protein biomarkers, dopamine metabolites, as well as microRNAs. The modern developments in artificial intelligence, especially in machine learning (ML) as well as deep learning (DL), have led to a notable increase in the use of these algorithms in the healthcare industry. An instance of this is the detection of neurodegenerative disorders such as Parkinson's disease (PD) through picture and electronic healthcare data analysis. This paper tackles the problem of PD identification by utilising a hybrid strategy that combines DL and ML [11–15]. Instead of using only pictures or electronic healthcare datasets for classification, in this research, we fused extracted features from images and healthcare records of healthy and PD subjects—the proposed methods comprised three main steps. The first step belongs to feature extraction. In this step, two different combinations of DL neural networks are used to extract features from the pictures and corresponding electronic healthcare records. Convolutional Neural Networks (CNNs) have garnered a lot of interest from investigators in recent years because of its proven ability to extract features from text, pictures, and videos. In this research, we present a tactic to extract characteristics from both healthy individuals and PD participants by integrating CNNs with attention mechanisms. Relevant features are extracted from electronic health records using a memory-dependent unit and convolutional layer combination [16–20]. The next stage after feature extraction is Fusion. The process of fusing extracted features is done via inner product multiplication. Finally, ML models are used to distinguish between PD and healthy subjects. The newly fused sets of features from pictures and physical characteristics of pen movement propose a new set of discriminator features for PD classification. To evaluate the proposed methods, the Parkinson's Disease Handwriting Database (PaHaw) and spiral-wave pictures are used [7]. The experimental result indicates a state-of-the-art result of 99.95% accuracy for PD versus healthy subjects' classification. The contributions of the proposed model are listed as follows:

- Presenting two novel DL models based on convolutional and attention architectures for extracting features from pictures and electronic healthcare datasets.
- Proposing a new fusion technique based on person correlation to extract distinguishable.
- features from each feature extraction model.
- Explaining the process of extracting features from the individual handwriting pictures for final PD detection.
- Providing a new assistant method for discriminating between PD and healthy subjects to the physicians.

A detailed explanation of the proposed methods and their comparison with simi- lar research is presented in the following.

This study is divided into seven major sections. Section Related works summarizes the historical approaches and current developments regarding distinguishing between individuals with PD and healthy controls. The dataset used in this study is fully described in Section Dataset, along with some statistics about it. The proposed model and the approach for feature extraction and fusion from both signals and images are presented in Section Model. The results on the capability of the proposed model to distinguish between patients with Parkinson's disease and healthy people, and its comparison with other machine learning models, are shown in Section Experimental result. The outcome of each model is also considered. Section Discussion discusses the experimental results where it is crystal clear how much better the suggested model is compared to the rest of the models under investigation. Finally, Section Conclusion concludes by summarizing research contributions and limitations and suggests directions for future work.

Related works

Parkinson's disease is a neurodegenerative disorder that predominantly affects older adults. Around 1 or 2 cases per 1,000 older adults in the general population have the disease and face its challenges. The conventional grounds for diagnosis are symptoms and complications arising from PD. In recent times, machine learning algorithms have been deployed to analyze medical datasets related to PD. Several machine learning and deep learning approaches were already applied for different data types, including handwriting samples, speech recordings, cerebrospinal fluid analysis, and electronic health records of PD patients. Of these, the analysis of handwriting patterns is the most frequent method for differentiating PD patients.

Diaz et al. [21] have proposed a deep convolutional neural network and a voting classifier integrated model to classify PD. They test their model using the PaHaw dataset and present the performance evaluation with other machine learning models such as Support Vector Machines (SVM) and Random Forest (RF). The test results showed that their model yielded an accuracy of 86.67%, with an AUC of 83.33%, a sensitivity of 89.17%, and a specificity of 80.83% in the detection of PD. In contrast, only the following accuracies for the classification of PD were achieved by the models SVM and RF: 75.41% and 71.67%, respectively.

Ali et al. [22] used Naive Bayesian (NB) [23] for PD versus healthy subjects dis- crimination using a handwritten pattern. They proposed a novel cross-validation technique to refer to the issue of overlapping between training and test sets. Also, they converted the imbalance distributions of the target to balance distributions using sampling methods. They reported 71.21% using the PD handwriting database.

Many researchers claim they have proposed new deep learning (DL) methods for identifying Parkinson's disease (PD) in healthy subjects. Loh et al. [24] reviewed these DL techniques for PD classification, exploring models such as Convolutional Neural Networks (CNN) [25], Long Short-Term Memory (LSTM) [26], and hybrid DL models. They reported that the highest accuracy achieved on the PaHaw dataset using CNN was 99.2%. Additionally, they claimed that the best accuracy achieved on a dataset with 19 features (electronic healthcare data) was 98.0% using a CNN-LSTM model.

Deharab et al. [27] developed two novel feature representations from handwriting signatures by fusing analytic signal representation with the area of the second-order difference plot for PD identification. They used a Support Vector Machine (SVM) as the classifier and evaluated their approach using the PaHaw dataset, reporting an accuracy of 86.26% in identifying PD subjects. The primary contribution of their work was the introduction of new feature sets for PD detection.

Basnin et al. [28] applied transfer learning methods to a private handwritten dataset for distinguishing between PD and healthy subjects. They collected 136 handwritten images from PD cases and 36 from healthy patients for evaluation, using VGG 16 [29] as a pre-trained model and adding only two dense layers to the end of the proposed model, achieving an accuracy of 91.36%.

Lamba et al. [30] focused on kinematic features derived from patients' handwriting signatures. They extracted features such as 'Velocity, acceleration, jerk in the horizontal direction' and 'Number of changes in acceleration.' Additionally, they calculated mutual information between the extracted features as a complementary measure. Their evaluation of machine learning models, including SVM, AdaBoost [31], Random Forest (RF), and XGBoost [32], employed sampling methods to balance the distribution of PD and non-PD patients. They reported the following performance metrics using AdaBoost: 96.02% accuracy, 91.93% sensitivity, 100.00% specificity, 100.00% precision, and 95.79% F-measure. They concluded that AdaBoost outperformed RF, XGBoost, and SVM.

In another article by Yousif et al. [33], The authors evaluated the performance of extracting features from speech and handwriting datasets for PD detection. The authors used 8 different pre-trained CNN models, and finally, SVM and K Nearest Neighbors (KNN) [34] were used for classification. The authors concluded that combining VGG16 with SVM can achieve 99.75% accuracy on the NewHandPD dataset. Abdullah et al. [5] proposed the combinations of transfer learning and genetic algorithms for feature selection and Ml models for final classification. The cost function for the genetic algorithm was to improve the accuracy of the final classifier by choosing the best set of features. The author evaluated the proposed method using the NewHandPD dataset. The authors concluded that combinations of ResNet or VGG-16 with KNN with 95.29% accuracy for PD detection.

Zham et al. [35] collected spiral shape drawings from PD and healthy subject candi- dates to check the possibility of distinguishing between healthy and PD subjects. In total, 28 healthy and 27 PD patients participated and were assessed by the Unified Parkinson's Disease Rating Scale (UPDRS). The authors concluded that the speed and pressure of drawing pens can be used to disseminate information about healthy versus PD subjects.

A pattern can be seen from the reviewed researchers. Initially, all the reviewed articles extracted features from participants' handwriting pictures or the dynamic character- istics of the handwriting. Then, the authors used DL or ML models as classifiers for discriminating between healthy and PD subjects [36, 37]. By reviewing similar research, a gap in using complementary information as a new set of features for PD versus healthy subjects discrimination is depicted. Instead of using only pictures, voice, or handwriting physical characteristics in this research, we propose a multimedia approach to use both pictures and related static or dynamic features for discriminating between PD and healthy subjects. The proposed method uses the features from the picture and extracts them using a combination of CNN and the attention layer. Simultaneously, the pro- posed model combines CNN and memory units to extract features from the related dynamic or static features. Combining the Pearson correlation and inner product is used to merge extracted features. This combination is a new set of fused information that can be used for PD versus healthy subjects' discrimination. Finally, a voting classifier was employed to distinguish between PD and healthy patients. Finally, the voting classifier is used to identify the difference between PD and healthy patients. The experimental result indicates 99.85% accuracy, 99.84% precision, 99.86% sensitivity, and 99.85% F1-score using the proposed fusion technique and voting classifier.

Dataset

PaHaW

The PD dataset PaHaW [38] includes handwriting data recorded from 37 PD patients and 38 healthy controls, all right-handed. Data collection was performed with an Intuos 4 M digitizing tablet, at a sampling rate of 200 Hz. This digitizer measured several variables including the xand y-coordinates of the pen position, with corresponding timestamps. All PD subjects were examined during their ON state, and a clinical neurologist evaluated their conditions. Furthermore, healthy subjects were checked for any sign of injuries or movement disorders that could jeopardize their handwriting. The average age of participants is 65.8 years, from 36 to 90 years old. 36 of the contributors were female, and 39 of them were male. The average diagnosis time for positive cases is 8.37 years. Participants' information is shown in Fig. 1 As shown in Fig. 1, the severity of PD is indicated using 6 different stages. Stages 1 to 2.5 can be categorized as early to mild samples. Stages 3 to 5 can be categorized as moderate to severe stages. This research aims to distinguish between healthy and PD subjects. Thus, all stages of PD are considered PD subjects. The dataset contains static features

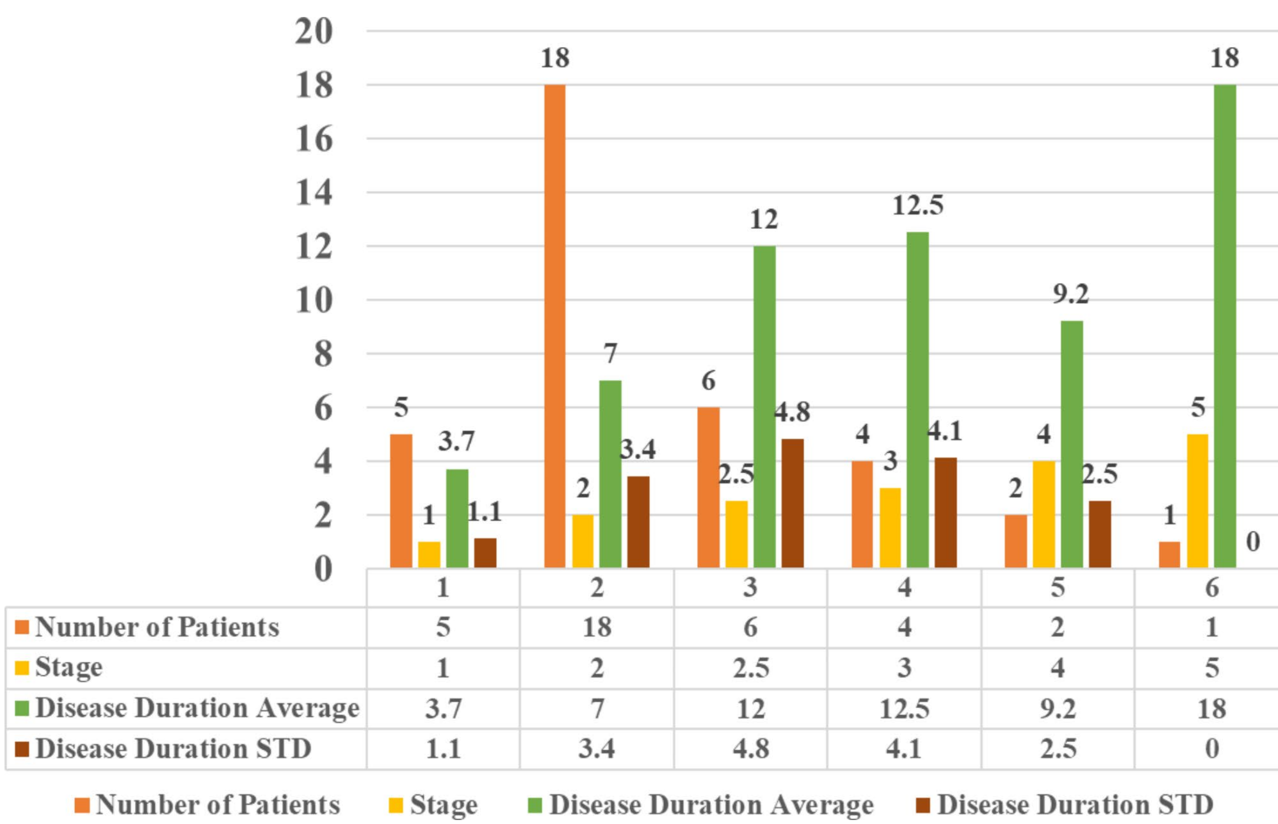


Fig. 1 Severity of diseases based on the disease duration in each subject

like the button state of the pen for creating spiral shapes. We added dynamic features, like speed on the X and Y axes, to augment the dataset. A sample of these drawings and their related features are shown in Fig. 2.

Spiral-wave pictures

The image dataset [37] includes recordings taken from 28 healthy participants and 27 individuals diagnosed with PD. In this dataset, there are a total of 204 images; 102 images were related to spiral drawing and the other 102 images were concerned with wave drawing. Images were obtained from the participants who drew spiral and wave continuous patterns. In this regard, the study obtained ethical clearance from the RMIT University Human Research Ethics Committee, and the work is part of research conducted in compliance with the Declaration of Helsinki guidelines [38]. The recorded handwriting signatures include both static spiral elements and Dynamic Spiral Test (DST) components. The severity rating of PD for the participants is given below in Table 1.

Each patient's severity level in this dataset ranges from 1 to 5, consistent with the previous dataset. Motor assessment severity was evaluated using a modified Hoehn and Yahr (HY) Scale [37]. This dataset assesses motor performance, measures tremors, and aids in diagnosing

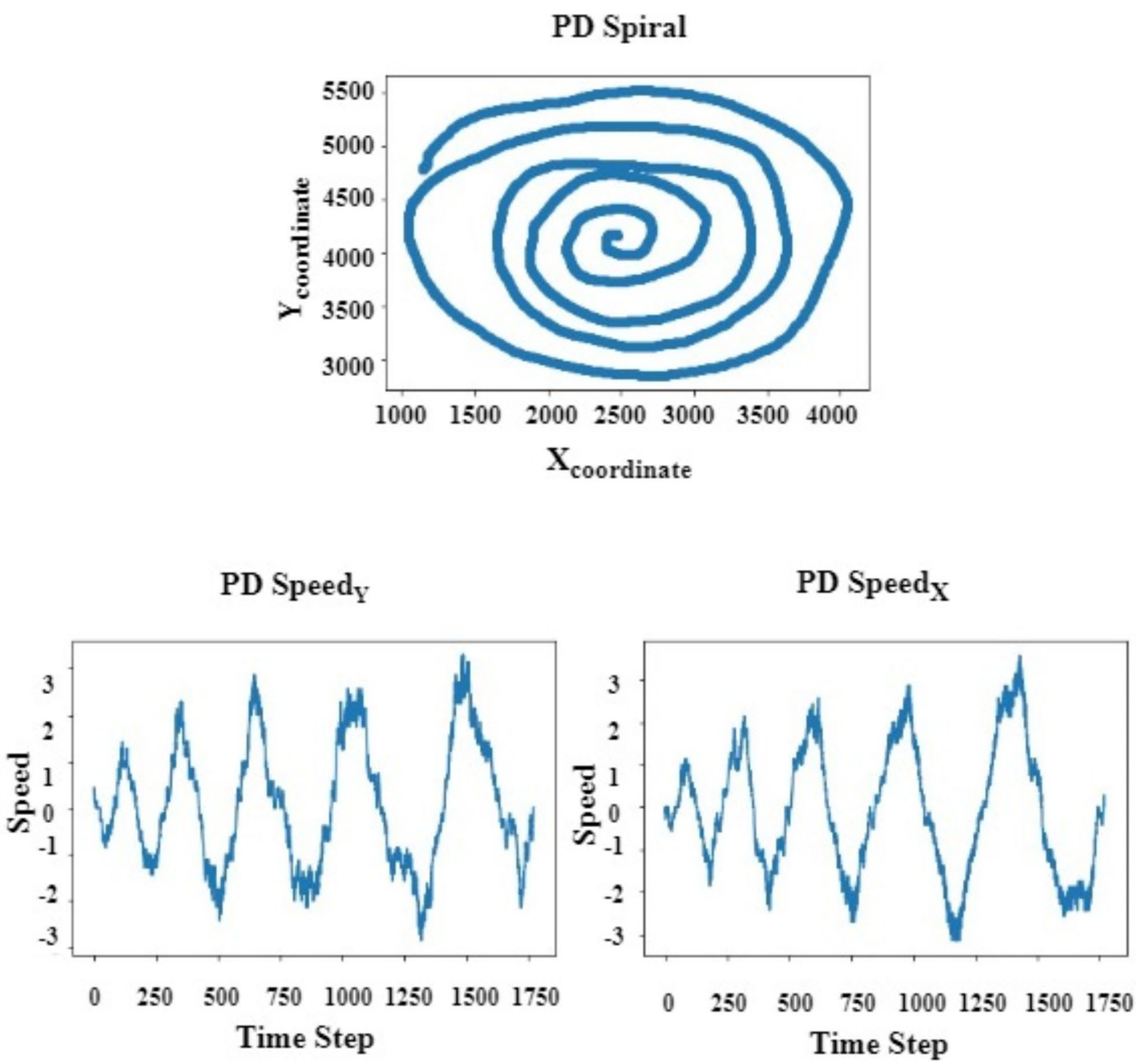


Fig. 2 Samples of the PaHaW drawings. Spiral shapes and speed versus time step

Table 1 The severity group is based on the Unified Parkinson's Disease Rating Scale (UPRDS)

Number of patients	Stage	UPRDS (average)	UPRDS (Std)
12	1-1.5	10.75	2.18
8	2- 2.5	18.38	2.83
7	3–5	28.43	2.63

Parkinson's disease. A sample of these drawings is illustrated in Fig. 3.

As shown in Fig. 3, the difference between the spiral drawing of the healthy participant is more obvious than the wave drawing. This research combines Pictures and dynamic, and kinematic characteristics of subjects' handwriting signatures to differentiate among PD as well as healthy subjects. In total 27 subjects from both datasets with the same level of severity have been chosen for the final dataset. The quantity of healthy subjects in the combined dataset is 28. Two pictures (waves and spiral shapes) per sample are available for training and testing thus, the number of samples from PaHaW is chosen accordingly. The dataset utilized in this research is freely accessible from the Kaggle website and is available at https://www.kaggle.com/datasets/vikasukani/parkinsons-disease-data- set and https://www.kaggle.com/datasets/kmader/parkinsons-drawings/code.

Model

In this section, an explanation of the proposed model is discussed. The proposed model comprises three sections. The first section relates to extracting features from input pictures and speech features. The second section describes the process of merging extracted features. The last section relates to the classification of extracted features.

Feature extractor

Picture

different biological biomarkers can be used to discriminate between healthy and PD subjects. However, gathering biological biomarkers and using them to distinguish between healthy and PD subjects consumes money and time, and it might be considered an invasive process. However, in this research, we proposed a new fused set of features that are extracted using DL models and can be used as a marker for distinguishing between PD and healthy subjects.

As mentioned, we have two sets of information as the input dataset. The first dataset is the handwriting's physical characteristics and the second is the handwriting pictures. The availability of a faster and more powerful Graphical Processor Unit (GPU), solving the gradient vanishing and overfitting problems led to a bright future for the application of DL models especially CNN in the fields of computer vision. In this research, CNN is used to extract feature sets from pictures, static and dynamic related features to the handwriting picture. This development opens up new possibilities for the application of DL models in the diagnosis of Parkinson's disease, potentially leading to more accurate and efficient diagnosis methods. We develop a new model using the combination of attention mechanism, convolutional layer, and transpose convolutions.

To extract suitable features from the picture, a combination of the encoder decoder and attention layer has been developed. The proposed model is comprised of 4 convolutional blocks in the decoder and 4 transpose convolutional blocks with an attention layer on the decoder side. This meticulous approach ensures that all relevant features are extracted and considered in the classification process. Finally, three dense layers are added to the network for final classification. The convolutional block itself breaks down into two convolutional layers with dropout as well as batch normalization layers in between them. After every convolutional block, a max-pooling layer is placed to decrease the width and height of the extracted feature. In the decoder section, trans- pose convolution with the concatenation of the attention layer is used to increase the size of extracted feature maps. The attention mechanism emphasizes the importance of the extracted feature set by calculating the alignment factor to emphasize how much the extracted features should be considered for calculating the output. Concatenating the transpose convolution result with the attention layer, only useful features are extracted from the input dataset.

Figure 4 demonstrates the architecture of the proposed attention layer. As depicted, the attention layer utilizes the feature map extracted from the convolutional transpose along with the feature map from the encoder to compute the attention factors. The context vector for the attention layer is derived using the inner product. The output of the attention layer represents the most significant features extracted from the corresponding encoder layer. Finally, as the classifier, four dense layers are added. Between each layer, a dropout layer is used to mitigate the effect of overfitting. The complete architecture of the proposed model is shown in Fig. 5. As is presented in Fig. 5, the proposed model can be used directly for classification. However, in this research, we extract the feature from each dense layer for concatenation with the next layer.

PaHaW

The original PaHaW comprised only six features. We calculated two extra features (speed in the X and Y directions) and added them to the original dataset. The reason behind adding the kinematic features is the ability of the time series model to extract proper features on PD detection [39]. All of the features including speeds are

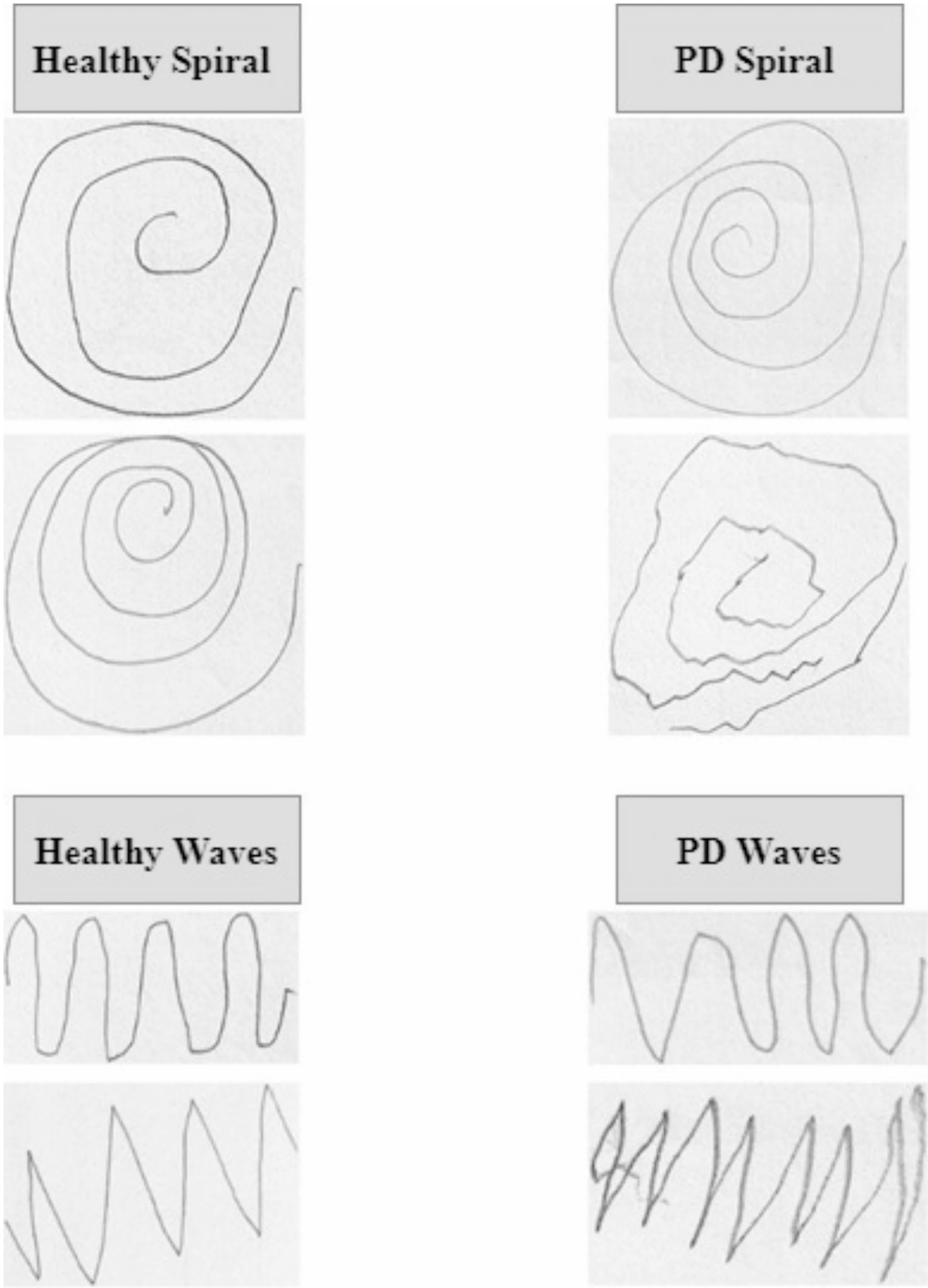


Fig. 3 Samples of the drawings in spiral and wave structures

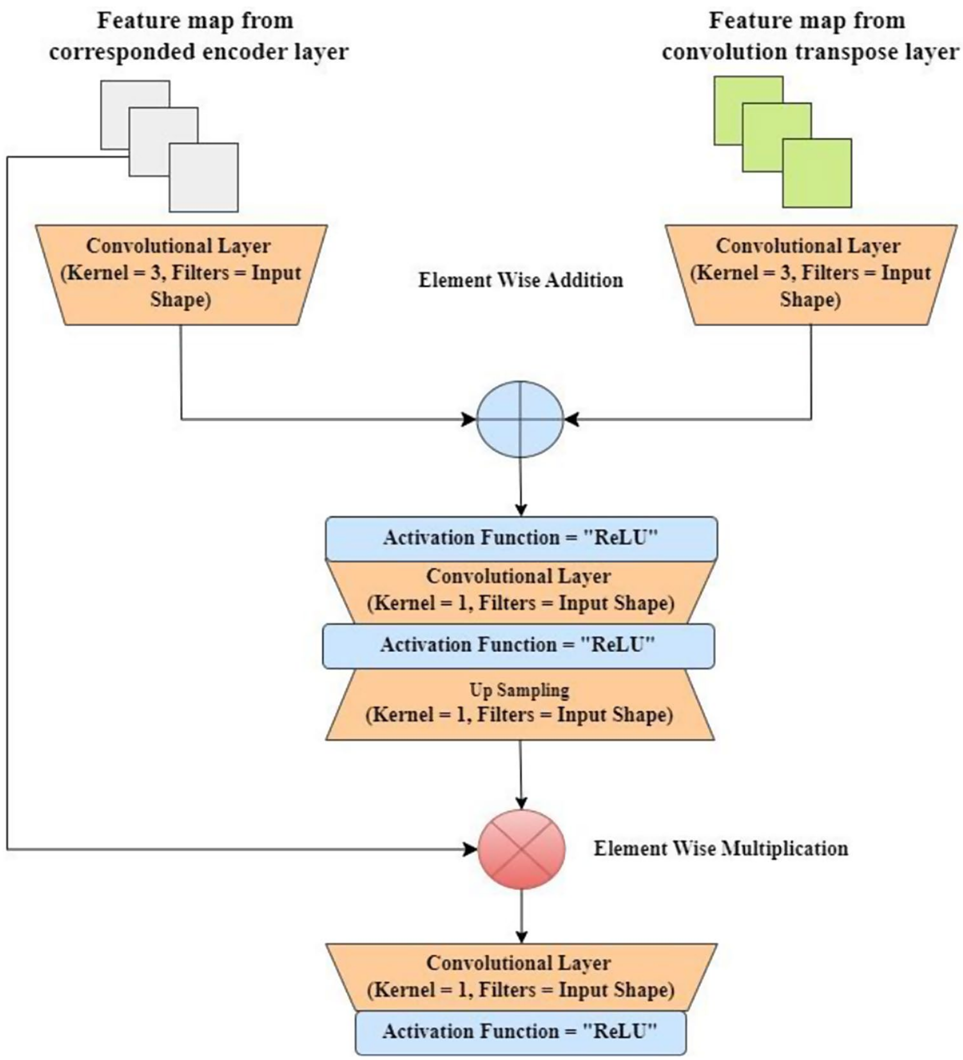


Fig. 4 Architecture of the novel proposed attention layer

fed into the feature extraction model and individual classifiers. The features in the PaHaW are dependent on the time. Memory-type cells are the best options for extracting proper information from time series information. However, in total, only eight features are available in the dataset. To increase the number of features, we used a convolution layer. One-dimension Convolution (Conv 1D) is used in this research. Due to the small feature set of the dataset, we developed a shallow model using only two convolution layers followed by an average pooling layer. After the Conv 1D for feature extraction, two Long-Short-Term-Memory (LSTM) [40] are used. Similar to the first model, four dense layers are used as a classifier. The model's specification is shown in Fig. 6. As shown in Fig. 6, the hybrid CNN-LSTM is used for discriminating between PD and healthy subjects. The number of neurons in each layer of the fully connected layer is 128, 64, 32, and 2. Between each of the fully connected layers, a dropout layer with a 0.2 drop factor is placed. like the previous model, the output of the layer with 32 neurons is extracted for merging.

Merger

Merging extracted features from both the picture set and the handwriting picture's physical tremors will provide a new set of information that can be used for final classification. In the previous sections, the model is first trained to discriminate between PD and healthy instances. However, the result of the last section is not used for the final classification. Give the F_P, F_d as the extracted feature sets from the pictures and electronic health care data [41]. The inner product is used to merge the extracted feature. The process of merging extracted features is shown in Eq. 1.

$$Merged = F_p.F_d \tag{1}$$

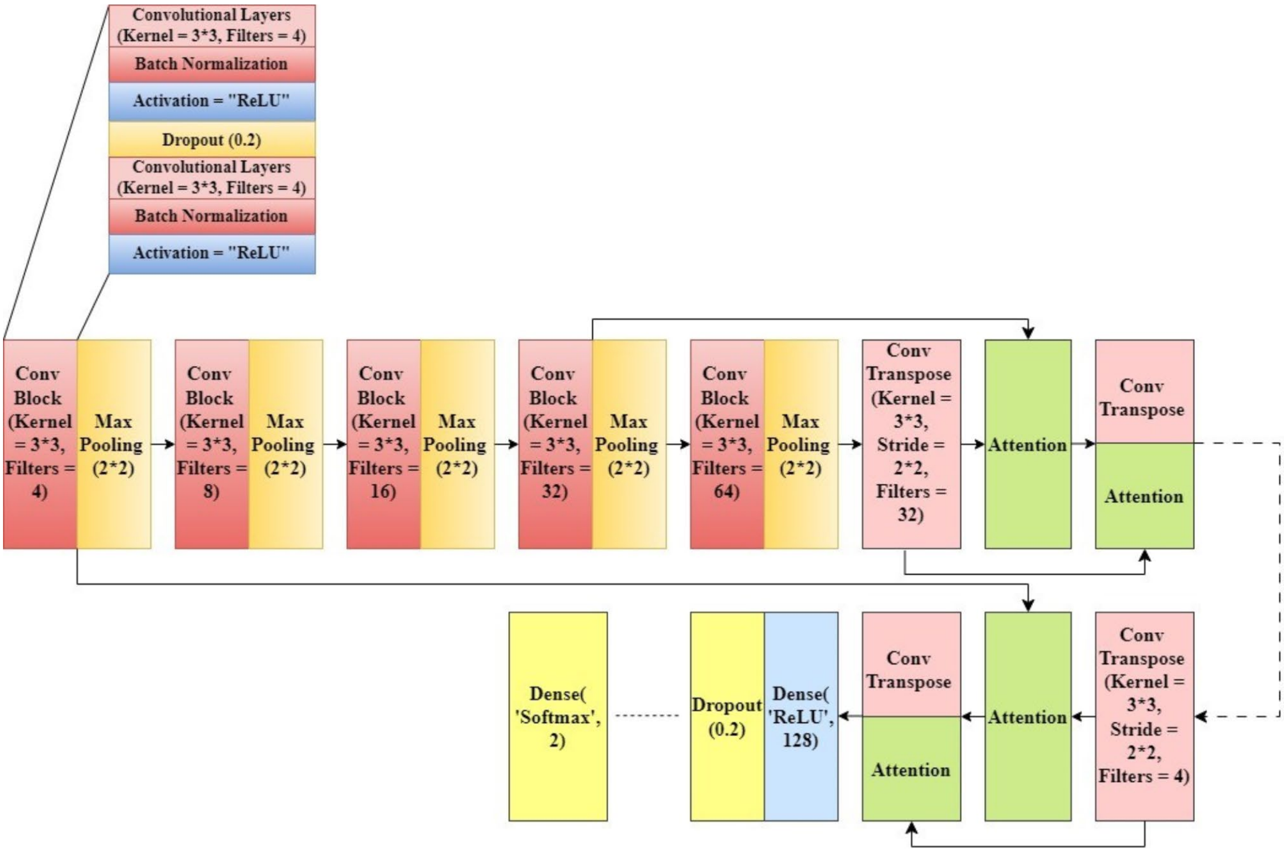


Fig. 5 The architecture of the proposed CNN + attention

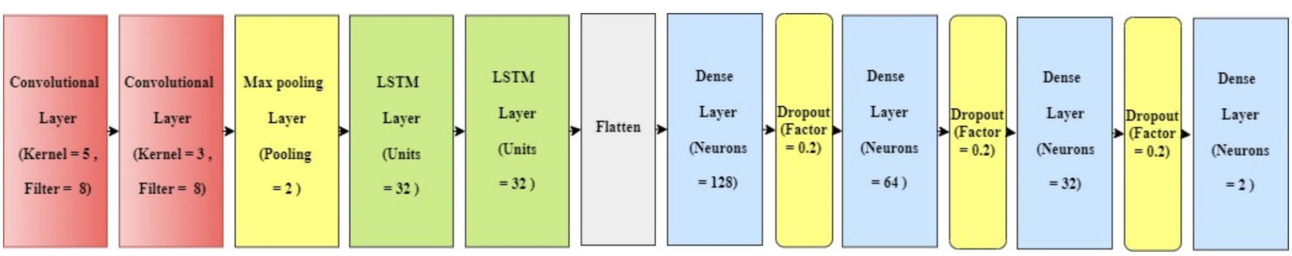


Fig. 6 The proposed shallow combination of CNN (2 convolutional layers) and LSTM (2 LSTM layers) layer

Each extracted feature map from the fully connected layer could enter the merging process Since the number layer and feature vector have been extracted by the same number of neurons. Thus, to choose the best possible outcome of the inner product for merging, first, we calculate the Pearson correlation coefficient to check the linear dependency between each extracted feature vector [42]. The formula for Pearson correlation calculation is shown in Eq. 2.

$$P_r = \frac{\sum_{i=1}^{n}\left(F_{p1} - \bar{F}_{p1}\right) * \left(F_{d1} - \bar{F}_{p1}\right)}{\sqrt{\sum_{i=1}^{n}\left(F_{p1} - \bar{F}_{p1}\right)^2} * \sqrt{\sum_{i=1}^{n}\left(F_{d1} - \bar{F}_{p1}\right)}^2} \quad (2)$$

The least correlated feature vectors are used for the merging process. The whole process of merging tries to investigate similarities between the feature vectors that have shown the least linear relationship to each other. Using the feature representing the various aspects of extracted features from pictures and the electronic health care dataset improves the classifier's performance. Fused features from original pictures, dynamic, and kinematic present a new set of features to identify the difference between PD and healthy subjects. This newly generated feature set can be used more effectively than individual pictures and correspondent handwriting physical features.

Discrimination

The number of instances in both datasets is low; thus, we used the ML models for classification. Motivated by reviewed researchers, we compared the proposed model with already available models such as Logistic Regression (LR), RF, SVM [43], and XGB. Any of these classifiers utilize the features set by different perspectives. LR discriminates between instances by assigning an estimated probability to each instance [16].

RF expands the search area by bootstrap aggregating decision tree estimators [44]. SVM is an optimization classifier that assigns each instance to a specific class by considering a margin of error for classification [45]. XGB employs aggregation of predictions from multiple weak learners to classify instances. Each weak learner contributes valuable information for prediction, enabling the boosting strategy to combine these learners into a more robust model [46].

In this research, rather than using each classifier individually, we implemented a voting aggregation approach with a hard strategy to leverage the best outcomes from all estimators. The voting classifier uses the best possible outcome for classification in the hard strategy. The final classifier aggregates the results of the individual classifiers. The benefit of using each individual and aggregating the result will lead to covering the individual classifier weakness. The main aim of this research is to use the discriminative features and the voting classifier exploit from the discriminative features for final classification. structure of voting classifier with hard strategy is shown in Fig. 7. The proposed algorithm begins by extracting features from handwriting pictures, its physical characteristics and then, by combining the extracted features a new set of distinguishable features is available for final PD versus healthy subjects' identifications.

Experimental result

We utilized a cross-validation approach [47] to evaluate the proposed algorithm. Specifically, we implemented a five-fold cross-validation method, where the dataset is divided into five "folds". In each iteration, four of these folds are used to train the model, while the remaining fold serves as the validation set. Each fold acts as the validation set once across the five iterations, providing a comprehensive performance assessment. The final performance metrics represent the average results from

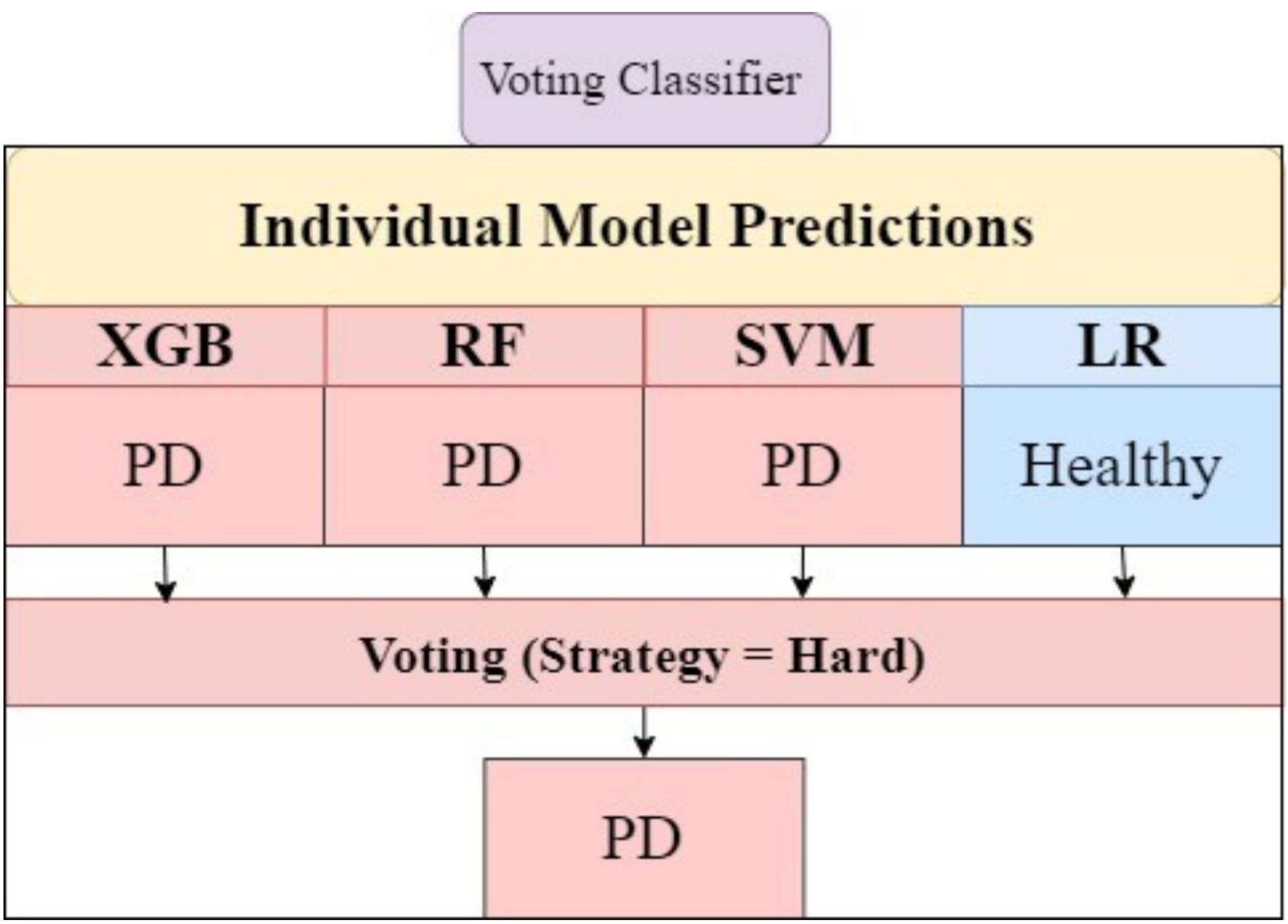


Fig. 7 The results of using the hard voting strategy on the individual classifier

Table 2 Results of PD classification using the separate CNN and CNN-LSTM models

Model	Accuracy	Precision	Sensitivity	F1-score
CNN-Attention	91.24%	91.20%	90.74%	91.01%
CNN-LSTM	89.88%	89.50%	89.66%	89.62%

each of these iterations, offering a robust measure of the model's effectiveness. For feature integration, we combined various types of feature sets image-based features, statistical features, and dynamic features—obtained from patients categorized into either the healthy or PD groups. Each feature set was processed independently before being merged for model training. We trained each deep learning (DL) model separately using the Nesterov Adaptive Moment Estimation (Nadam) optimizer [48], which adjusts the learning rate dynamically during training. The categorical cross-entropy loss function was employed, which is well-suited for classification tasks with multiple classes. This approach ensures that the model benefits from the combined feature sets, thereby enhancing classification accuracy. To avoid overfit- ting, we stopped the training process if the training loss had not been improved after 100 epochs. For both the validation as well as testing stages, we used a batch size of eight and set the training procedure to run for 500 epochs. We employed a number of metrics, including the confusion matrix, accuracy, precision, recall, as well as F1-score, to assess the outcomes. The following is a summary of the formulas used to determine these metrics:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{3}$$

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$Sensitivity = \frac{TP}{TP + FP} \tag{5}$$

$$F1\_score = \frac{2 * TP}{2 * TP + FP + FN} \tag{6}$$

Here, TP, TN, FP, and FN represent True Positive, True Negative, False Positive, and False Negative predictions, respectively. The outcomes for each classifier are summarized in Table 2.

The extracted heatmaps of the presented method on CNN attention are shown in Fig. 8. 7. As shown, the

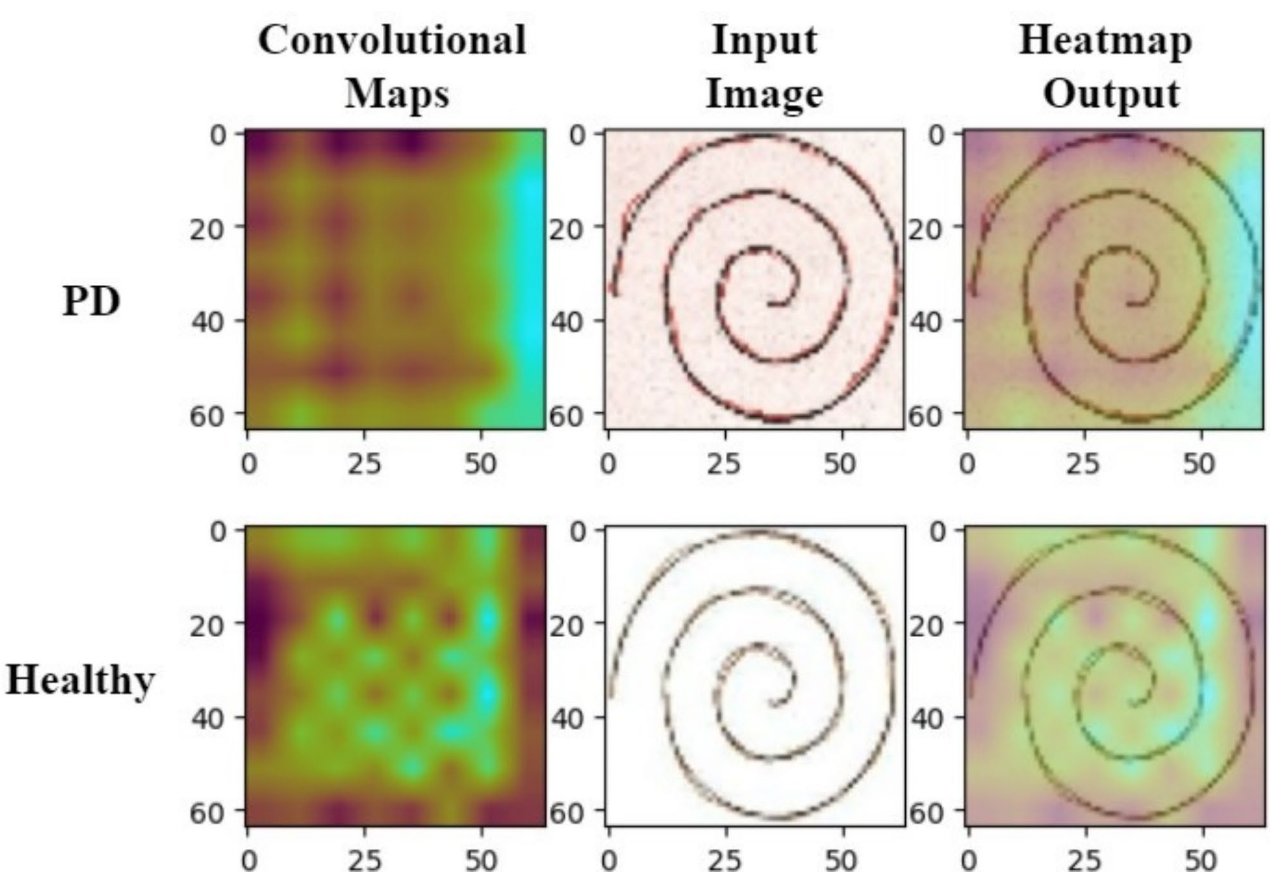
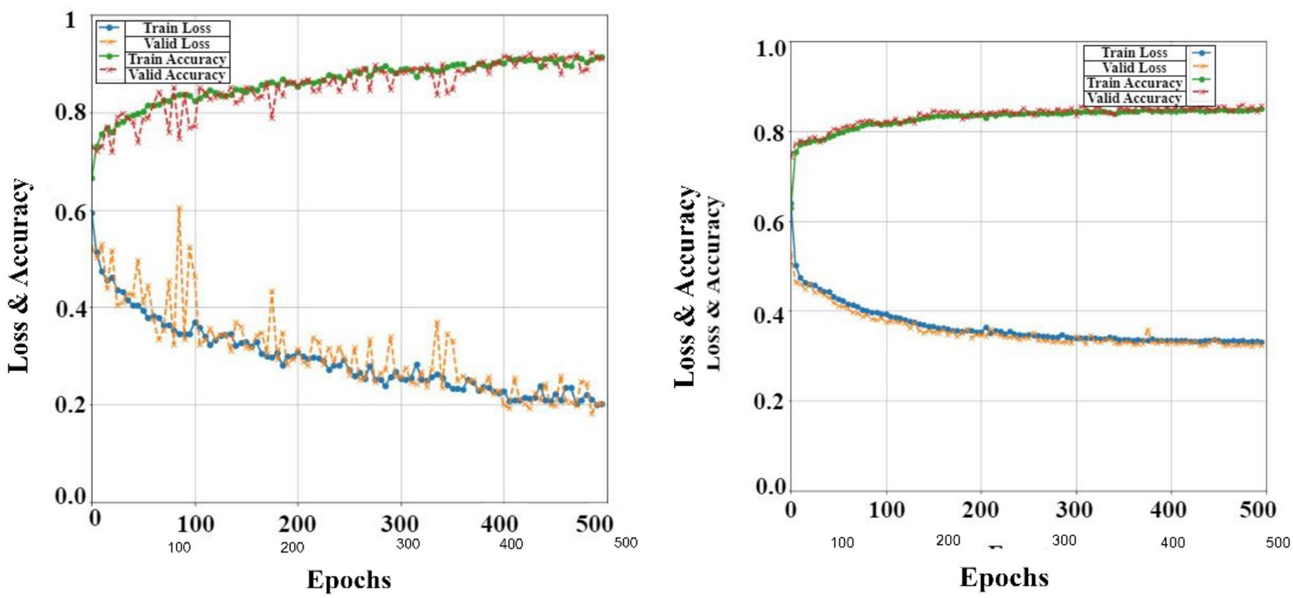


Fig. 8 The extracted feature maps from the input pictures using various convolutional feature maps



(a) (b)

Fig. 9 Learning diagram for feature extraction models using; (**a**) CNN-Attention, (**b**) CNN-LSTM

Table 3 Average results of PD versus healthy subjects' identification using ML models and proposed fusion technique for each class

Model	Classes	Acc	Pre	Sen	F1
Fusion + LR	PD	84.00%	80.76%	84.00%	82.33%
Fusion + LR	Healthy	99.99%	99.98%	99.99%	99.98%
Fusion + XGB	PD	96.20%	80.70%	92.00%	85.98%
Fusion + XGB	Healthy	99.99%	99.98%	99.99%	99.98%
Fusion + RF	PD	87.21%	81.72%	92.00%	85.98%
Fusion + RF	Healthy	99.99%	99.99%	99.99%	99.98%
Fusion + SVM	PD	76.69%	86.71%	70.00%	72.96%
Fusion + SVM	Healthy	99.98%	99.98%	99.98%	99.98%
Fusion + Voting	**PD**	**99.98%**	**99.90%**	**99.91%**	**99.90%**
Fusion + Voting	**Healthy**	**99.99%**	**99.99%**	**99.99%**	**99.99%**

proposed method focused on the curves on the circular dataset to extract the corresponding features. At the same time, the model for healthy subjects focused on the spiral line to distinguish them from the PD subjects. Usually, the models focus on the outlines generated from the original schema.

The results of both models are close to each other. The learning diagrams for both models are shown in Fig. 9.

We evaluated different combinations for CNN-LSTM, and as similar research has shown, using only CNN-LSTM with 3 dense layers and 16, 8, and 2 neurons, respectively, achieved 97.81% accuracy using the PaHaW dataset [48]. However, this architecture must be separate from the CNN-Attention model. We used the ML models as the final classifiers to complete our aim. Due to the limited validation data, the validation loss result is unstable for CNN-LSTM and CNN-Attention models. To identify the optimal hyperparameters, we employed a grid search method to determine the best settings for these machine learning models [49–51]. The best set of hyperparameters is reported as the number of estimators = 300, maximum depth = 50, and learning rate = 0.9 for the XGB classifier. The same hyperparameters, number of estimators = 50, and maximum depth = 100 are reported for the RF. Maximum features = 1 and minimum samples leaf = 1 are also noted for RF. The performance of the SVM and LR didn't improve when utilizing the grid search model. Experimental results on the proposed architecture as a fusion technique and ML models as the classifier are shown in Table 3. As shown in Table 3, the proposed methodology has reached nearly perfect accuracy. For evaluation, five-fold cross-validation is used, and Table 3 represents the average results for each class over five different folds. To rely on the model's performance, we have evaluated the model further on tenfold cross-validation. The accuracy results using 10-fold cross-validation are shown in Table 4. The model's performance is more robust, with an average result of 99.85% accuracy.

Table 3 indicates a rapid improvement in the PD versus healthy subjects' discrimination performance using the voting classifier. To choose the best individual to create a voting classifier, we decided on the RF and XGB. Almost all ML models outperformed the single DL model for PD classification 32 healthy and 34 PD subjects from the original dataset were used as the test set to check the model's performance. The confusion matrix in Fig. 10 explains the investigated result in detail.

As shown in Fig. 10, combining the proposed feature extractor and voting out-performed all the evaluated

Table 4 Results of evaluating the fusion + voting classifier on 10-fold cross-validation

Model	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10	Total
Proposed (Fusion + Voting)	99.92	99.91	99.99	99.98	99.98	99.90	99.94	99.89	99.12	99.93	99.85

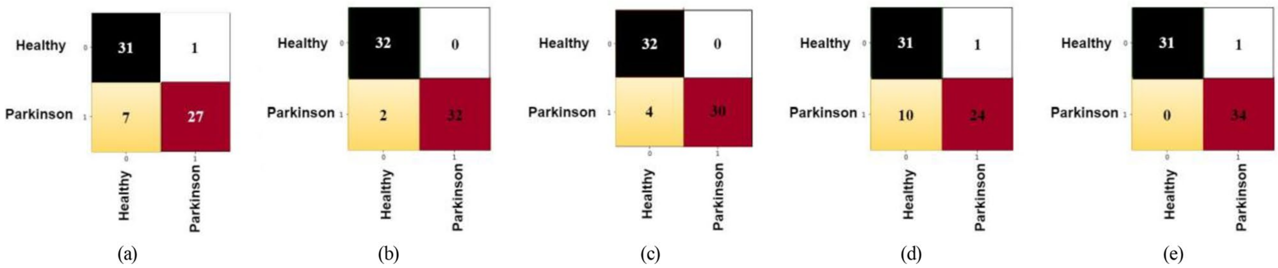


Fig. 10 Instance based confusion matrix; (**a**) Fusion + LR, (**b**) Fusion + XGB, (**c**) Fusion + RF, (**d**) Fusion + SVM, (**e**) Fusion + voting

Table 5 Reported response time and computation complexity by all individual methods

Model	Trainable parameters	Training time (second)	Testing time (second)
CNN-LSTM	38,914	503	0.56
CNN-Attention	702,293	923	0.63
Voting classifier	-	152	1

models for PD versus healthy subjects' classification. The proposed model achieved perfect performance in both PD and healthy discrimination. Using the combined feature set with the voting strategy created a complimentary performance for PD classification. As shown in Table 3, all models have excellent performance for healthy instance discrimination. Furthermore, evaluation is done to prove the robust performance of the methodology using tenfold cross-validation. Table 4 confirms the results' robustness by evaluating the model across ten folds.

However, only XGB and RF demonstrated relatively good performance in PD discrimination. However, using the voting classifier, the voting classifier's complementary performance improved the individual's performance in PD classification. The configuration for running the code is 1xTesla K80, 12GB of virtual randomized accessible memory with one single-core hyperthreaded Xeon Processor @2.3Giga hertz. The environment for training the model is Python 3.7.10 alongside TensorFlow and Scikit learn libraries. All models' training and testing responses are shown in Table 5.

Discussion

PD significantly affects the physical behavior of the patients and has a lasting effect on hand and body movement. With a rough estimation of more than 10 million people worldwide suffering from PD, the urgency of finding effective solutions is clear. Many researchers have proposed various solutions to extract features from available information for discriminating between PD and healthy subjects. Previous researchers have focused on using either ML or DL models to tackle the challenges of PD versus healthy subjects' discrimination. However, in this research, a novel feature fusion procedure is proposed, which could potentially make a significant impact. The proposed algorithms comprised three steps: feature extraction, fusion, and classification. In the first step, a combination of CNN with the attention layer is developed to extract features from pictures. To extract features from the statistical and dynamic features of subjects' handwriting shapes, a combination of CNN and LSTM is used. By evaluating DL models, it is evident that the lack of a large dataset for training can impact their performance for PD versus healthy subject classification. However, training a deep model on the dataset, using the saved weight of the model to extract features, and adding an ensemble ML model showed promising results, instilling hope for its potential in future research. The inner product is used to fuse extracted features. Finally, the fused feature is fed into ML models for classification. Pearson correlation is used to choose the least linearly related features to benefit from the complementary effect of the extracted feature. The proposed approach demonstrates 99.95% accuracy for PD versus healthy subject discrimination. The performance of the proposed algorithms indicates perfect discrimination between positive PD and healthy instances simultaneously. To complete the process of providing a reference to contrast the explored methodology with other reviewed models on PD versus healthy subjects' discrimination, Table 6 is presented.

As shown, the proposed model outperformed previous research. The proposed model has reported a new state-of-the-art performance for PD versus healthy subjects' discrimination with more than 99.85% accuracy.

Table 6 Results of contrasting the explored combination methodology with other methods for PD versus healthy subjects

Author	Dataset	Model's name	Accuracy	Sensitivity	F1-score
Diaz et al. [21]	PaHaw	RF	86.67%	83.33%	-
Ali et al. [22]	PaHaw	NB	71.21%	-	-
Deharab et al. [27]	PaHaw	SVM	86.26%	-	-
Basnin et al. [28]	PaHaw	VGG	91.36%	-	-
Valla et al. [52]	PaHaw	Ensemble	84.86%	75.00%	-
Lamba et al. [53]	PaHaw	Alexnet	93.33%	-	0.96
Loh et al. [24]	PaHaw	CNN-LSTM	99.2%	-	-
Proposed	**PaHaw**	**Fusion + Voting**	**99.85%**	**99.86**	**99.85**

As shown in Table 6, the proposed model outperformed both DL and ML models in the case of PD versus healthy subject classification. In the PaHaW dataset, some patients have shown PD symptoms before their 60th birthdays. Also, 21% of all patients demonstrated their symptoms less than five years after the PD diagnosis. Thus, the proposed model can discriminate PD cases from healthy subjects in the earliest stages and people under 60. Therefore, the investigated merging process can help to provide superior discriminating features for PD versus healthy subject discrimination. The main distinguishable results of the proposed methodology are explained in three facts. First, a novel fusion algorithm is proposed to use dynamic signals and pictures for PD versus discrimination against healthy subjects. Second, the proposed algorithm outperformed similar PD research versus healthy subject classification. Third, the proposed algorithm provides better discriminating features for PD versus healthy subject classification.

The proposed method's performance is further demonstrated in the HandPD Dataset [54, 55]. This dataset comprises 18 male and 17 female healthy samples and 21 male and ten female patient samples. The average age of healthy and PD subjects is 44.05 ± 14.88 and 57.83 ± 7.85 years, respectively. 66% of the HC subjects are women, and only 34% of the HC subjects are men. The subject's age range is between 19 and 79. Finally, the results of testing the proposed methodology on HandPD are demonstrated in Tabale7, underscoring the effectiveness of our proposed method (Table 7).

As shown in Table 6, the performance of the proposed method is consistently using other types of handwriting samples. The proposed model outperformed the base article (83.77% accuracy using the CNN model) by 14.75%.

Table 7 Results of PD detection on HandPD dataset

Model	Acc	Pre	Sen	F1
Fusion + RF	98.52%	98.12%	97.29%	98.37%
Fusion + XGB	97.81%	97.83%	97.82%	97.83%
Fusion + SVM	97.54%	97.50%	97.51%	97.50%
Fusion + voting	98.52%	98.12%	97.29%	98.37%

Limitations and future of the works

The main challenge and limitation of the proposed work is the possibility of gathering datasets from the same patients to fuse their features. Another limitation of the proposed method is using hyperparameter tuning methods for two models simultaneously. Meta-heuristics methods will be investigated to solve this problem in the future. In the future, we plan to further investigate the performance of the proposed model for discriminating between PD subjects in different stages. This will involve the use of additional types of information, such as text, pictures, and speech, to enhance the model's capabilities. The future model offers a discriminator model that can distinguish between various stages of PD. This can significantly impact the estimation of treatment efficacy based on PD progress, instilling hope for better patient outcomes. Since the proposed method provides a new set of fused features for PD classification, the proposed mode will discriminate between Parkinson's patients and those with other neurological diseases with similar symptoms.

Conclusion

This study presents a fusion method to distinguish Parkinson's Disease (PD) patients from healthy individuals using picture and electronic healthcare features. The technique combines Convolutional Neural Networks (CNN) and attention mechanisms to extract features from images and handwriting characteristics. The extracted features are then. They were fused using the inner product with the Pearson correlation factor. Five machine learning models are evaluated as classifiers, with the voting classifier achieving 99.85% accuracy on the PaHaW and corresponding motion datasets. Furthermore, the proposed methodology is evaluated using the HandPD dataset, and it reported 98.52% accuracy, 98.12% precision, 97.29% sensitivity, and 98.37% F1-score. The method outperforms other models and shows potential for further development to discriminate between various stages of PD.

Acknowledgements
Not Applicable.

Author contributions
Abeer aljohani conceived the proposed algorithm and participated in its design helped to draft the manuscript, participated in the proposed solution, and drafted the manuscript. Also Author participated in the design of the study and performed the performance analysis, read and approved the final manuscript.

Findings
This research received no specific findings from any agency, commercial entity, or commercial sources. Abeer Aljohani covered all expenses associated with this study.

Data availability
The dataset used in this study is freely obtained from the Kaggle website and is available: https://www.kaggle.com/datasets/vikasukani/parkinsons-disease-data- set, https://www.kaggle.com/datasets/kmader/parkinsons-drawings/code. Interested researchers can visit the provided links and follow the provided guidelines for data retrieval.

Declarations

Ethics approval and consent to participate
In this self-directed research, carried out independently without oversight from an Institutional Review Board (IRB), ethical considerations were given utmost priority. The authors ensured that participants received comprehensive information about the research objectives, procedures, and potential implications, stressing the importance of voluntary participation and confidentiality. Although there was no formal ethical review board involved, the research adhered to fundamental ethical principles, including respect for autonomy and confidentiality. For questions regarding the study's ethical aspects, please contact the authors at aahjohani@taibahu.edu.sa. The authors committed to maintaining transparency and accountability throughout the research, ensuring that all design and data collection processes strictly followed ethical standards to prioritize participants' well-being and rights. The authors remain dedicated to upholding the integrity of the research process and protecting the welfare of participants.

Consent for publication
Not Applicable.

Competing interests
The authors declare no competing interests.

Received: 7 May 2024 / Accepted: 16 September 2024
Published online: 27 September 2024

References
1. Melissa J, Armstrong MS, Okun. Diagnosis and treatment of Parkinson's disease: a review. JAMA. 2020;323:548–60.
2. Donato Impedovo G, Pirlo G, Vessio. Dynamic handwriting analysis for supporting earlier Parkinson's disease diagnosis. Information. 2018;9(10):247.
3. Alireza T, et al. Source code for optimized parallel inception: a fast COVID-19 screening software. Softw Impacts. 2022;13:100337.
4. Naga Tejaswi M et al. A Hybrid Approach to Parkinson's Disease Detection using Speech Attributes: The Combination of SMOTE and Active Learning. In: 2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC). IEEE. 2023, pp. 350–356.
5. Sura Mahmood A, et al. Deep transfer learning based Parkinson's disease detection using optimized feature selection. In: IEEE Access. 2023;11:3511–24.
6. Alireza T et al. Fast COVID-19 versus H1N1 screening using Optimized Parallel Inception. In: Expert Systems with Applications (2022), p. 117551.
7. Ben Aicha M, et al. Prediction of rheological behavior of self-compacting concrete by multi-variable regression and artificial neural networks. Powder Technol. 2022;401:117345.
8. Barend WF, et al. Perceived autonomy support in individuals with Parkinson's disease requiring emergency care: a cross-sectional pilot study. Neurol Res Pract. 2024;6(1):41.
9. Mahmood Saleh A et al. The role of neural network for the detection of Parkinson's disease: a scoping review. In: Healthcare. Vol. 9. 6. MDPI. 2021, p. 740.
10. Hanff AM, Krüger R, McCrum C, Ley C. Mixed effects models but not t-tests or linear regression detect progression of apathy in Parkinson's disease over seven years in a cohort: a comparative analysis. BMC Med Res Methodol. 2024;24(1):183.
11. Sudip P et al. Bias investigation in artificial intelligence systems for early detection of Parkinson's disease: a narrative review. In: Diagnostics 12.1 (2022), p. 166.
12. Ahmed Shihab A, et al. A systematic review of trustworthy and explainable artificial intelligence in healthcare: Assessment of quality, bias risk, and data fusion. Inform Fusion. 2023;96:156–91.
13. Muhammad J, et al. Explainable machine learning models based on multi-modal time-series data for the early detection of Parkinson's disease. Comput Methods Programs Biomed. 2023;234:107495.
14. Jee-Young L, et al. Multimodal brain and retinal imaging of dopaminergic degeneration in Parkinson disease. Nat Reviews Neurol. 2022;18(4):203–20.
15. Kaushal Kumar and Rajib Ghosh. Parkinson's disease diagnosis using recurrent neural network based deep learning model by analyzing online handwriting. Multimedia Tools Appl. 2024;83(4):11687–715.
16. Alireza T, et al. Hospital readmission and length-of-stay prediction using an optimized hybrid deep model. In: Future Internet. 2023;15:304.
17. Zeyu R, et al. Exploring simple triplet representation learning. Comput Struct Biotechnol J. 2024;23:1510–21.
18. Md Ariful I et al. A review of machine learning and deep learning algorithms for Parkinson's disease detection using handwriting and voice datasets. In: Heliyon (2024).
19. Zeyu Ren S, Wang Y, Zhang. Weakly supervised machine learning. CAAI Trans Intell Technol. 2023;8(3):549–80.
20. Yudong, Zhang, et al. Deep learning in food category recognition. Information Fusion. 2023;98:101859.
21. Moises D, et al. Dynamically enhanced static handwriting representation for Parkinson's disease detection. Pattern Recognit Lett. 2019;128:204–10.
22. Ali L, Zhu C, Zhao H, Zhang Z, Liu Y. An integrated system for unbiased parkinson's disease detection from handwritten drawings. InAdvances in Intelligent Systems and Computing: Proceedings of the 7th Euro-China Conference on Intelligent Data Analysis and Applications, May 29–31, 2021, Hangzhou, China. Singapore: Springer Nature Singapore; 2022. pp. 3–13.
23. Shenglei, Chen, et al. A novel selective nǎive Bayes algorithm. Knowledge-Based Syst. 2020;192:105361.
24. Hui Wen L et al. Application of deep learning models for automated identification of Parkinson's disease: a review (2011–2021). In: Sensors 21.21 (2021), p. 7034.
25. Xinyu Lei H, Pan X, Huang. A dilated CNN model for image classification. IEEE Access. 2019;7:124087–95.
26. Van Greg C, Mosquera G, Nápoles. A review on the long short-term memory model. Artif Intell Rev. 2020;53:5929–55.
27. Elham Dehghanpur Deharab and Peyvand Ghaderyan. Graphical representation and variability quantification of handwriting signals: New tools for Parkinson's disease detection. Biocybernetics Biomedical Eng. 2022;421:158–72.
28. Nanziba B et al. Early detection of Parkinson's disease from micrographic static hand drawings. In: International Conference on Brain Informatics. Springer. 2021, pp. 433–447.
29. Zhong Q, et al. Crack detection of concrete pavement with cross-entropy loss function and improved VGG16 network model. Ieee Access. 2020;8:54564–73.
30. Rohit L, et al. A systematic approach to diagnose Parkinson's disease through kinematic features extracted from handwritten drawings. J Reliable Intell Environ. 2021;7(3):253–62.
31. Mahesh Thyluru R, et al. Homogeneous Adaboost Ensemble Machine Learning Algorithms with reduced Entropy on Balanced Data. Entropy. 2023;25(2):245.
32. Amal Asselman M, Khaldi S, Aammou. Enhancing the prediction of student performance based on the machine learning XGBoost algorithm. Interact Learn Environ. 2023;31:3360–79.
33. Nada RY, et al. A generic optimization and learning framework for Parkinson disease via speech and handwritten records. J Ambient Intell Humaniz Comput. 2023;148:10673–93.
34. Nimish U, et al. Survey on exact knn queries over high-dimensional data space. Sensors. 2023;23(2):629.

35. Zham P, Kumar DK, Dabnichki P, Poosapadi Arjunan S, Raghav S. Distinguishing different stages of Parkinson's disease using composite index of speed and pen-pressure of sketching a spiral. Front Neurol. 2017;8:435.
36. Daniel Valero-Carreras, Alcaraz J, Landete M. Comparing two SVM models through different metrics based on the confusion matrix. Comput Oper Res. 2023;152:106131.
37. Hong Q et al. Generic AI models for mass transfer coefficient prediction in amine-based CO2 absorber, part II: RBFNN and RF model. AIChE J 69.1 (2023), e17904.
38. Muhammed Isenkul B, Sakar O, Kursun et al. Improved spiral test using digitized graphics tablet for monitoring Parkinson's disease. In: Proc. of the Int'l Conf. on e-Health and Telemedicine. 2014, pp. 171–5.
39. Moises D, et al. Sequence-based dynamic handwriting analysis for parkinson's disease detection with one-dimensional convolutions and BiGRUs. Expert Syst Appl. 2021;168:114405.
40. Zhao J, Mao X, Chen L. Speech emotion recognition using deep 1D & 2D CNN LSTM networks. Biomed Signal Process Control. 2019;47:312–23.
41. Helang Lai and Xueming Yan. Multimodal sentiment analysis with asymmetric window multi-attentions. Multimedia Tools Appl. 2022;81:19415–28.
42. Yaqing L, et al. Daily activity feature selection in smart homes based on pearson correlation coefficient. Neural Process Lett. 2020;51(2):1771–87.
43. Jian Z, et al. Predicting TBM penetration rate in hard rock condition: a comparative study among six XGB-based metaheuristic techniques. Geosci Front. 2021;123:101091.
44. Campagner A, Ciucci D, Cabitza F. Aggregation models in ensemble learning: a large-scale comparison. Inform Fusion. 2023;90:241–52.
45. Mokhtar M, et al. A comprehensive survey and taxonomy of the SVM-based intrusion detection systems. J Netw Comput Appl. 2021;178:102983.
46. Hidir Selcuk Nogay and Hojjat Adeli. Diagnostic of autism spectrum disorder based on structural brain MRI images using, grid search optimization, and con- volutional neural networks. Biomed Signal Process Control. 2023;79:104234.
47. Benyamin Ghojogh and Mark Crowley. The theory behind overfitting, cross validation, regularization, bagging, and boosting: tutorial. In: arXiv preprint arXiv:1905.12787 (2019).
48. Sharma J, Soni S, Paliwal P, Saboor S, Chaurasiya PK, Sharifpur M, Khalilpoor N, Afzal A. A novel long term solar photovoltaic power forecasting approach using LSTM with Nadam optimizer: a case study of India. Energy Sci Eng. 2022;10(8):2909–29.
49. Petro Liashchynskyi and Pavlo Liashchynskyi. Grid search, random search, genetic algorithm: a big comparison for NAS. In: arXiv preprint arXiv:1912.06059 (2019).
50. Brage B, et al. A nationwide study of the incidence, prevalence and mortality of Parkinson's disease in the Norwegian population. Parkinson'sDisease. 2022;8(1):19.
51. Aleksa C, et al. Tuning attention based long-short term memory neural-networks for Parkinson's disease detection using modified metaheuristics. In:Scientific Rep. 2024;14(1):4309.
52. Elli V, et al. Tremor-related feature engineering for machine learning based Parkinson's disease diagnostics. Biomed Signal Process Control. 2022;75:103551.
53. Lamba R, Gulati T, Jain A. Automated Parkinson's disease diagnosis system using transfer learning techniques. In: emergent converging technologies and biomedical systems. Springer; 2022. pp. 183–196.
54. Xu S, Pan Z. A novel ensemble of random forest for assisting diagnosis of Parkinson's disease on small handwritten dynamics dataset. Int J Med Informatics. 2020;144:104283.
55. Eghbal H, et al. Meta-heuristics and deep learning for energy applications: review and open research challenges (2018–2023). Energy Strategy Reviews. 2024;53:101409.

Publisher's note

Abeer Aljohani earned her B.Sc. in Computer Science from the College of Computer Science and Engineering at Taibah University, Medina, Saudi Arabia, in 2008. She completed her M.Sc. in Information Technology for Management at Coventry University, U.K., in 2011, and obtained her Ph.D. in Computer Science from Loughborough University in 2019. She is currently an Assistant Professor at the Applied College, Taibah University. Her research interests encompass sensor technology in e-health and smart health management, smart cities, data science, artificial intelligence, evolutionary computation, machine learning, deep learning, pattern recognition, cyber security, computer vision, and image processing.