A Multi-agent Feature Selection and Hybrid Classification Model for Parkinson's Disease Diagnosis

MAZIN ABED MOHAMMED, College of Computer Science and Information Technology, University of Anbar, Ramadi, Anbar, 11, Iraq
MOHAMED ELHOSENY, Department of Computer Science, College of Computer Information Technology, American University in the Emirates, Dubai, 503000, United Arab Emirates and Faculty of Computers and Information, Mansoura University, Egypt
KARRAR HAMEED ABDULKAREEM, College of Agriculture, Al-Muthanna University, Samawah 66001, Iraq
SALAMA A. MOSTAFA, Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, Batu Pahat, Johor, 86400, Malaysia
MASHAEL S. MAASHI, Software Engineering Department, King Saud University, Riyadh 11451, Saudi Arabia

Parkinson's disease (PD) diagnostics includes numerous analyses related to the neurological, physical, and psychical status of the patient. Medical teams analyze multiple symptoms and patient history considering verified genetic influences. The proposed method investigates the voice symptoms of this disease. The voice files are processed, and the feature extraction is conducted. Several machine learning techniques are used to recognize Parkinson's and healthy patients. This study focuses on examining PD diagnosis through voice data features. A new multi-agent feature filter (MAFT) algorithm is proposed to select the best features from the voice dataset. The MAFT algorithm is designed to select a set of features to improve the overall performance of prediction models and prevent over-fitting possibly due to extreme reduction to the features. Moreover, this algorithm aims to reduce the complexity of the prediction, accelerate the training phase, and build a robust training model. Ten different machine learning methods are then integrated with the MAFT algorithm to form a powerful voice-based PD diagnosis model. Recorded test results of the PD prediction model using the actual and filtered features yielded 86.38% and 86.67% accuracies on average, respectively. With the aid of the MAFT feature selection, the test results are improved by 3.2% considering the hybrid model (HM) and 3.1% considering the Naïve Bayesian and random forest. Subsequently, an HM, which comprises a binary convolutional neural network and three feature selection algorithms (namely, genetic algorithm, Adam optimizer, and mini-batch gradient descent), is proposed to improve the classification accuracy of the PD. The results reveal that PD achieves an overall accuracy of 93.7%. The HM is integrated

Authors' address: M. A. Mohammed (corresponding author), College of Computer Science and Information Technology, University of Anbar, Ramadi, Anbar, 11, Iraq; email: mazinalshujeary@uoanbar.edu.iq; M. Elhoseny, Department of Computer Science, College of Computer Information Technology, American University in the 7 Emirates, Dubai, 503000, United Arab Emirate and Faculty of Computers and Information, Mansoura University, Egypt; email: melhoseny@ieee.org; K. H. Abdulkareem, College of Agriculture, Al-Muthanna University, Samawah 66001, Iraq; email: khak9784@mu.edu.iq; S. A. Mostafa, Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, Batu Pahat, 13 Johor, 86400, Malaysia; email: salama@uthm.edu.my; M. S. Maashi; email: mmaashi@ksu.edu.sa.

1551-6857/2021/05-ART74 $15.00
https://doi.org/10.1145/3433180

with the MAFT, and the combination realizes an overall accuracy of 96.9%. These results demonstrate that the combination of the MAFT algorithm and the HM model significantly enhances the PD diagnosis outcomes.

CCS Concepts: • **Computing methodologies → Supervised learning by classification**; **Classification and regression trees**;

Additional Key Words and Phrases: Parkinson's disease, machine learning, feature evaluation, voice feature, multi-agent system, multi-agent feature filter, Hybrid Classification Model, Convolutional Neural Network

ACM Reference format:
Mazin Abed Mohammed, Mohamed Elhoseny, Karrar Hameed Abdulkareem, Salama A. Mostafa, and Mashael S. Maashi. 2021. A Multi-agent Feature Selection and Hybrid Classification Model for Parkinson's Disease Diagnosis. *ACM Trans. Multimedia Comput. Commun. Appl.* 17, 2s, Article 74 (May 2021), 22 pages.
https://doi.org/10.1145/3433180

1 INTRODUCTION

Parkinson's disease (PD) is a neurodegenerative disease that affects millions of people worldwide and has no known cure. The primary symptoms associated with this disease include tremors, freezing of gait, and alterations in gait and speech. This kind of disease usually affects the daily activities of affected people and reduces the quality of life [1, 2, 3]. Despite the absence of a cure, PD can be managed using a few available drugs. However, the use of these drugs over a long time leads to rapid neurodegeneration [4]. The detection of PD in the early stage is one of the significant problems associated with the disease due to its unknown leading cause. Therefore, researchers from different academic fields are conducting joint discussions to gain a comprehensive insight into the disease.

PD is a degenerative disease that progresses gradually, and its leading cause remains unknown [5]. The origin of PD is yet to be determined. However, researchers have noted several factors, such as environmental and hereditary factors, which possibly contribute to the development of the disease. Research has shown that in every 100,000 people, approximately 100 to 250 are affected by PD [6]. Even though the disease is widespread among people aged 50, the main and early symptoms of PD have also been detected in people aged 30 to 50. The part of the human body that is often affected by this disease is the central nervous system, thus causing the degeneration of dopamine-producing neuron cells. Dopamines are chemicals created through considerable nigra (basal ganglia), which is responsible for moving signals inside the brain part. PD patients experience movement disorders when these cells are lost. Two significant classes of PD symptoms are as follows: motor and non-motor symptoms. The motor symptom effects are related to movements and are progressively discernible when contrasted with non-motor effects [7]. Motor symptoms are observed in patients with slow movements (bradykinesia), tremor, postural instability, and rigidity. Non-motor symptoms can be observed at a specific interval, and some of the manifestations of these symptoms include olfactory disorder, speech problems, sleep disorder, and swallowing problems [1, 6, 7]. PD has the following three effects on speech: prosody, articulation, and phonation. The phonation is the utilization of vocal folds for discourse (as a speech), and enunciation alludes to the output of speech using exceptional tissues. Meanwhile, prosody is related to the pitch, loudness, and amplitude used for the production of sound. PD has been treated at the early stages using a wide range of chemical methods because of the absence of a permanent cure. One of the commonly used chemical methods is referred to as Levodopa (L-dopa). DBS is another effective therapy of PD. The DBS involves the electrical stimulation of certain parts of the brain using an implantation device that is almost the same as a cardiac pacemaker [8]. Presently, the finding made by specialists to evaluate the seriousness level of PD is led by utilizing different strategies dependent on many research areas, including speech disorders, cognitive deficits, gait cycle, and other

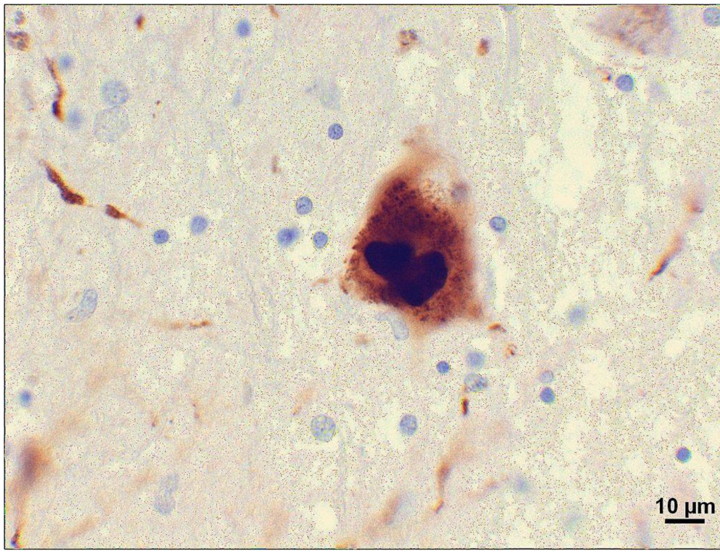

Fig. 1. Lewy body of the substantia nigra [10].

studies. Unmistakably, the absence of an explicit test for PD contributes to the difficulty in identifying PD subjects. Furthermore, symptoms and signs similar to those of PD may have different causes; for example, dementia with Lewy bodies, dynamic supranuclear paralysis, and particular sorts of stroke as shown in Figure 1. Overall, the accurate diagnosis of PD is a challenging task involving many physical, psychological, and neurological examinations. Furthermore, according to Reference [9], the main challenges in computer-assisted PD diagnosis include the following:

- Different data sources
- Detection of PD at the early stages of the disease
- Patient monitoring at home
- Collection of digitized versions of old exams

Machine learning is a part of computational insights committed to the advancement of methods that empower computer programs to enhance its presentation dependent on early (learned) data. Since the introduction of the earliest reference point of "perceptron," new numerical models of the working component of the brain have been sought after daily. Such an exceptional study has motivated various studies that planned to utilize artificial intelligence methods and machine learning techniques to help PD detection. In the context of the learning process, the problem of PD prediction is considered to be a clustering classification problem. However, a framework is built to handle different sets of data availabilities. This framework can classify only one type of class with a restricted class set of the PD samples. Thus, the detection of the correct class is easy, which leads to high accuracy [11–13].

A **multi-agent feature filter (MAFT)** algorithm for an automated voice-based PD diagnosis model is proposed in this article. The PD diagnosis model includes **neural network (NN), random forests (RFs), logistic regression (LR)**, support vector machine, **k-nearest neighbor (k-NN), Naïve Bayesian (NB), decision tree (DT)**, AdaBoost, **stochastic gradient descent (SGD)**, CN2 rule inducer, and **hybrid model (HM)**. The HM, which comprises a binary **convolutional neural network (CNN)** and three feature selection algorithms (namely, **genetic algorithm (GA), Adam optimizer (AO)**, and **mini-batch gradient descent (M-BGD)**), is proposed to improve the classification accuracy of the PD. By contrast, the MAFT algorithm has been applied for feature ranking and selection. The MAFT algorithm mainly aims to reduce the complexity of the prediction model, build a robust training model, and accelerate the training phase of the model to prevent over-fitting and improve the overall performance of the prediction model. This algorithm also contains multiple feature selection algorithms, which are controlled by a multi-agent system, to determine the score of the attributes and select the most significant features to be used in the classification phase.

The rest of the research article is organized as follows. Section 2 presents related work. Section 3 describes the materials and explains the methods. Section 4 presents details of the experimental setup, datasets, simulations, and results. Section 5 provides the required comparison for the proposed method with state-of-the-art studies. Finally, Section 6 presents the conclusion and future work.

2 RELATED WORK

Efforts are continuously provided by researchers in different fields to find effective management ways for the disease and increase the quality of life of PD patients. Such works, which range from chemical- and behavioral-based studies to computer-assisted diagnosis, can be found in the existing literature. The main contribution of this work is in computer-assisted diagnosis, wherein computer tools are provided to help researchers diagnose PD effectively and rapidly. Efforts were made in this article to review the most important works related to the automatic identification and diagnosis of PD using a computer method. In most works, the use of artificial learning is employed for the sole purpose of learning the most relevant features that can be used for diagnosis. For instance, Gupta et al. [14] introduced the **optimum-path forest (OPF)** for the automatic identification of PD. Subsequently, these authors also proposed an evolutionary-based approach for the selection of a unique set of features that can be used to obtain effective PD detection rates [15]. The OPF classifier is a suitable tool due to its easy management and absence of parameters.

Conclusively, some researchers that focused on the diagnosis of PD by analyzing voice disorders have applied classification approaches to the original PD dataset. Thus, these researchers have achieved high classification results. Meanwhile, some other researchers have applied feature evaluation methods before reducing the features of the dataset and proceeding to classification. Their results revealed that the accuracy of classifiers increases with the reduction of features. Nevertheless, the features may be overweight or over-reduced when a wide range of feature evaluation methods are used together. Therefore, the classification generalization may be negatively affected, thereby exposing the classifiers to over-fitting.

Mostafa et al. [5] evaluated the performance of three conventional classifier techniques, which include NN, NB, and DT, to diagnose the PD. Their study aimed to solve the problem of PD diagnosis by applying these classification methods on different Parkinson's datasets, which include human voice disorder detection features. The evaluation results demonstrate that DT achieves the best result among the other classification methods with an accuracy rate of 91.63%. However, the diagnostic accuracy is still low compared with that of other studies despite the significant contribution of this study. The author of Reference [16] proposed using sequential forward and backward selection methods to reduce features. Different numbers of features have been fed into numerous classifiers. The best diagnosis accuracy is achieved at 90.6% when the number of selected features is 4, 5, 6, and 7. However, the low number of selected features is questionable, because there is no guarantee if the classification results cannot present any over-fitting scenario. The authors of Reference [17] proposed an approach involving a combination of the k-means clustering-based feature weighting method and a complex-valued artificial neural network. A significant diagnosis accuracy of 99.52% as an accuracy ratio has been presented in this study. However, the authors have not revealed the most significant features. Furthermore, the feature selection approach based on the k-means technique is quite easy to understand, robust, and fast. This technique depends highly on expert knowledge despite its popularity. Thus, scholars should be aware of some weaknesses. For instance, the user has to specify k (the number of clusters) in the beginning. Therefore, the quality of the selected features is extremely dependent on the first selection of the initial centroid of the clusters. The authors of Reference [18] proposed a new feature selection approach based on the integration of the relief and ant-colony optimization algorithms. A significant diagnosis accuracy of 99.5% as an accuracy ratio has been presented in this study. However, if the same number of features is used despite different subset features, then the diagnosis results may be enhanced.

A large portion of previously mentioned studies is predominantly founded on the utilization of voice disorder and frequency-domain feature to PD diagnosis. However, these attributes could not be easily connected to a clinical pointer. This study primarily aims to establish a valuable approach to help PD diagnosis by utilizing features extracted from voice disorders. This approach is mostly committed to its utilization in a clinical situation to help physiotherapists in PD identification. Most studies in the PD diagnosis focus on using many machine learning methods and a proper feature selection technique to obtain good accuracy. The issue of finding the quality of an attribute is comprehensive in inductive machine learning calculations, thus resulting in difficulties.

The contributions in this study can be summarized as follows.

- An MAFT algorithm is proposed for an automated voice-based PD diagnosis model.
- An HM, which comprises a binary CNN and three feature selection algorithms (namely, GA, AO, and M-BGD), is proposed to improve the classification accuracy of the PD.
- A comprehensive performance investigation is conducted for different PD diagnosis models based on the overall feature set, and the selected features are chosen following the feature selection of the MAFT model.

3 MATERIALS AND METHODS

This study primarily aims to identify the best combination of features that can be used in improving the PD classification accuracy outcomes while maintaining a balanced feature selection. Herein, multiple feature evaluation methods and classification approaches are applied to improve the accuracy of PD diagnosis. The main activities conducted in this study include data preparation, feature evaluation, method setting, classification, and evaluation of classifiers. The original Parkinson's dataset is processed in this study to obtain a new filtered dataset. Figure 2 shows a representation of different research activities conducted in this study.

PD datasets contain different features, which are indicative of healthy people and are influenced by PD. Several preprocessing steps were taken to prepare and improve dataset quality. The next step, which is the evaluation of features, involves the application of the MAFT containing several algorithms for the evaluation and ranking of features. The MAFT aims to determine the value of the attributes and select the most significant features. The Parkinson's original dataset is the outcome of the evaluation, and the classes of the dataset contain 22 features. The setting activities involve the implementation of 11 commonly used classification methods: KNN, RFs, NN, NB, LR, SGD, CN2 rule inducer, SVM, AdaBoost, and the proposed model. The 10-fold cross-validation techniques are used to evaluate the classifiers after feature selection. The subsequent sections provide details of the significant methods and activities of the work.

3.1 PD Data Preparation

The PD data preparation, which is the first step that was performed, involved the normalization of data, in which the importance of the attributes of the PD database acquired from **University**

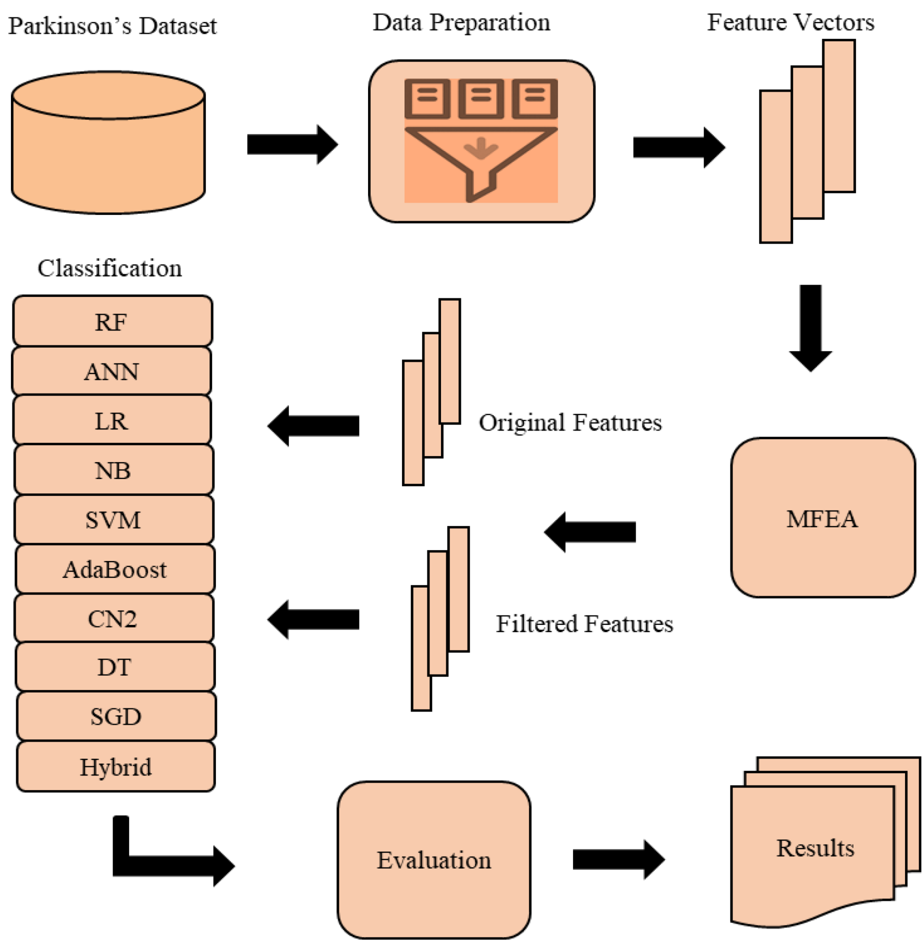

Fig. 2. Main research activities for machine learning-based identification approach to improve the accuracy of automated PD diagnosis.

of California-Irvine (UCI) had an assortment of information ranges. Upon completion of the normalization process, the feature selection began. This process involved the removal of any attributes that are incapable of improving classification results based on each feature selection algorithm. Two schemes were used in this stage: the technique with no feature selection and that with the feature selection process. The MAFT was employed for the technique for feature determination and selection to improve accuracy. In the classification phase, 11 classifiers, including KNN, RF, LR, NN, AdaBoost, NB, CN2 rule inducer, SVM, SGD, DT, and the proposed model, were used. Figure 3 shows the flowchart of the methods used for the preparation of the PD data used in this study.

According to Reference [19], the dataset for PD can be extracted using many available methods. The authors obtained the physical signals, which were then analyzed and filtered into 5,923 signals [20]. The main contribution of their work lies in the updating of functions to facilitate normal data distribution further. Therefore, stable UPDRS features can be manifested for the regression schemes. Moreover, the authors applied a wide range of linear signal preparation techniques to obtain the signal characteristics; for example, short-time autocorrelation and nonlinear techniques,

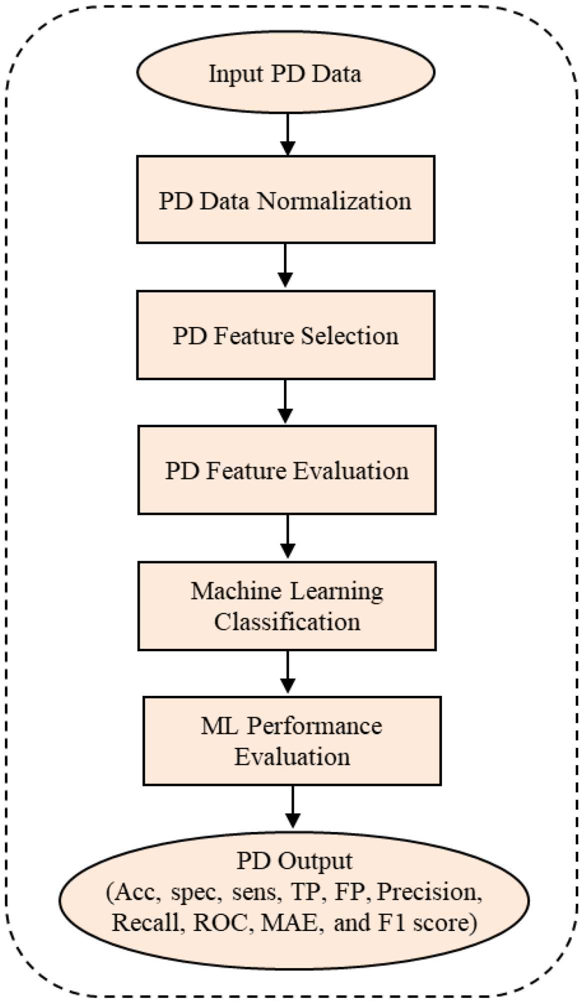

Fig. 3. Flowchart of the methods of PD data preparation.

such as regression tree and classification method, in vowel phonation (dysphonia) measurements. The attributes of the feature extraction task were obtained from the PD dataset. The measures were compared with UPDRS using non-parametric statistical tests, such as Gaussian kernels and Spearman correlation coefficients, to ensure that the assessment is conducted accurately. Afterward, the authors used regression methods to map the extracted features to obtain the final **Unified PD Rating Scale (UPRDS)**. A statistical procedure was used in the regression investigation to examine the dependence of features and analyze their correlations. The class model was constructed during the process of feature selection by applying Bayesian and Akaike data measures [21]. The methods are capable of handling a wide range of features with diverse categorization while measuring the best fit of the typical. Any input feature that is unrelated to the target output is rejected or passed to the following circle of the testing process (PD Pathology, PD Stages and Life Expectancy, and PD Symptoms [10]).

According to the study by the authors of Reference [22], the technique of the feature selection process was used by the least absolute shrinkage with selection operator to choose the best

Table 1. Dataset of PD Features [24]

PD Features	Attribute Information
DFA	Signal fractal scaling exponent
MDVP: Fo(Hz), Fhi(Hz), and Flo(Hz)	Minimum, maximum, and average vocal essential frequencies
HNR and NHR	Measurements of the ratio of noise to tonal modules in the voice
MDVP: Shimmer (dB), Shimmer, APQ5, APQ3, DDA, and APQ	Measurements of the difference in amplitude
D2 and RPDE	Non-linear dynamical difficulty measurements
MDVP: Jitter (Abs), Jitter (%), PPQ, RAP, and DDP	Measurements of the difference in essential frequencies
PPE, Spread 1, and Spread2	Non-linear measurements of essential frequencies difference

attributes, and the same results were obtained for the two methods; this approach did not affect the number and dimensions of the features. This process facilitated the acquisition of the PD dataset, which contains 23 attributes of numerical numbers. The procedure yielded a PD dataset containing 23 attributes of actual numerical sums. The name of the patient and the recording number were the first two features; however, these features were not included in the classification. The major features of the PD dataset and their description are presented in Table 1. The model generalization was performed using the cross-validation technique. The process of typical generalization involved the application of idle numbers in the parameter set of the model. For the classification result calculation of the **mean absolute error (MAE)**, the authors divided the dataset into two to minimize the required computation time for the tests (PD Symptoms) [23]. Overall, the test results showed that the prediction accuracy is acceptable. The Parkinson's datasets reveal the following: (1) all the attributes are numerical of actual statistics, (2) complete without loud or non-numeric information, and (3) without patterns. Nevertheless, an imbalance was observed in the PD database, in which 75% fall under PD cases and 25% are healthy cases. The classification task becomes complex due to this imbalance, while the robustness of the classification results reduces [23].

3.2 Multi-agent Feature Filter (MAFT) Algorithm

The MAFT algorithm performs feature selection to reduce the features of the PD dataset in a manner that does not affect the learning of the classification models but rather improve their prediction performance. The algorithm mainly aims to reduce the complexity of the prediction models, accelerate the training phase, and build a robust training model by preventing over-fitting and improving the overall performance of the prediction models. The review of the literature introduced several deployment models of software agents in feature selection and optimization. The study conducted by the authors of Reference [25] proposed a dynamic feature selection model in text mining. The feature selection is performed by deploying software agents to observe and compare the scoring of the subset features with pre-defined thresholds and then select the best subset of features accordingly. Another work made by the authors of Reference [26] proposed a multi-agent soccer model. The agents perform the feature selection based on C4.5 DT as a feature selection operator. The agents then perform feature analysis based on the scoring variation of some data mining methods. In both studies, reducing the dimensions of the features results is crucial in improving the training and prediction performance of the tested classification models.

The MAFT operates via software agents that interact collaboratively to solve the feature selection problem. These agents contribute to the flexibility of the feature selection process by segregating the selection functionalities. The agents then interact and reason over the input, process, and output of each function and make the necessary amendment during runtime. The agents run in the three phases of input, process, and output. The input phase represents fetching the input parameters of the training data, initial subset, default setting, and processing criteria. The process phase represents performing a filtering operation, which includes searching, selection, evaluation, and ranking processes, to generate different subsets of features. The output phase is divided into two parts: an individual subset of features for each agent and a collaborative subset of features for multi-agent.

Let a PD dataset D with a feature vector and V of an n length, in which the $V = \{F, C\}$. The F represents the input features, in which $F = \{f_1, f_2, \ldots, f_m\}$, and C represents the corresponding output class labels, in which $C = \{c_1, c_2, \ldots, c_n\}$. The ultimate representation for the D in the MAFT has a general format of Equation (1):

$$D = \begin{array}{c} v_1 \\ v_2 \\ : \\ . \\ v_n \end{array} \begin{bmatrix} f_1 & f_2 & .. & f_m \\ f_1 & f_2 & .. & f_m \\ f_1 & f_2 & .. & f_m \\ f_1 & f_2 & .. & f_m \end{bmatrix}, \begin{bmatrix} c_1 \\ c_2 \\ : \\ . \\ c_n \end{bmatrix} \tag{1}$$

The basic procedure of the MAFT is the deployment of five feature selection agents $\boldsymbol{\Omega}$ on the PD dataset to find the best $v_i \in$ V to ensure the processing of high prediction results by the classification models. First, each agent, $\alpha_i \in \Omega$, randomly generates a subset s_i of features. Then, the $\boldsymbol{\Omega}$ apply independent measures μ_i in every round of feature optimization according to their represented operators to evaluate the generated subsets of features S, in which $S = \{s_1, s_2, \ldots, s_5\}$. The agents rank the features according to the evaluation scores to $R = \{r_1, r_2, \ldots, r_m\}$. Subsequently, in the following round, the Ω applies a search strategy on the generated s to add or remove certain features from the corresponding s. During this process, the $\boldsymbol{\Omega}$ share the feature evaluation matrix $\mathcal{E}$, which includes the evaluation results for each evaluation round, to influence the search operation through the search space S. Based on this procedure, different preliminary copies of subsets are produced by the $\boldsymbol{\Omega}$ during the feature selection optimization process. The evaluation operators are included for wrapping and filtering methods along with the classification and search operators. Symmetrical uncertainty and information gain are examples of evaluation operators. Each operator used a suitable type of search method (for example, greedy search as the best search method). The variation process of the search strategies in the MAFT ensures a generalized search space is generalized without constraints from the local optimums. Each agent is examined in different locations of the search space, and different evaluation criteria are applied. Thus, the variation of the independent evaluation criteria and the application of different mining operators prevent inherited bias of a particular mining operator. The $\boldsymbol{\Omega}$ generate different evaluated patterns of features, in which each $\alpha_1 : \mathcal{S} \times \mathcal{E} \rightarrow f_1 \gg \ldots \gg f_m$ during the F permutation process (referred to as $\gg$). The production of the same pattern of features s by $\boldsymbol{\Omega}$ is unlikely. Hence, the $\boldsymbol{\Omega}$ comprises a collaborative filter mechanism \$, which receives the s_i of each α_i and determines the final best subset based on the ranking and frequency of appearance criteria $\{s_{best}\} \rightarrow \$(a_{1,\ 2,\ \ldots}, s_{1,2\ldots})$

3.3 Machine Learning Classification Methods

Machine learning is a branch of computational intelligence that is focused on the development of a method through which the performance of a computer program can be improved on the basis of prior information. Since the introduction of "perceptron," researchers have continued to gear efforts toward new mathematical modeling of the working mechanism of the brain [27, 28]. These efforts have inspired several other works related to the use of machine learning-oriented techniques that are capable of facilitating PD recognition. The classification process involved the use of four kinds of classifiers, including RFs, KNNs, artificial neural networks (ANNs), LR, NB, SVM, AdaBoost, CN2 rule inducer, DT, and SGD, and the proposed automated voice-based PD diagnosis HM was utilized.

- **Random Forests (RFs):** The RF is an ensemble technique typically utilized in the process of classification, wherein the use of different DTs is employed in data classification [28]. Herein, bootstrap templates are built from the RF main numbers, and a raw classification process or regression tree is established in every bootstrap pattern. Moreover, consideration is given to each node in this classifier rather than selecting only the best disclosure from all predictors.
- **K-Nearest Neighbors (KNNs):** The central idea behind the KNN is that the distance between each dataset and data with similar properties is minimal. If the data have no classes, then the classification of such data can be conducted by observing the closest neighbor [29]. The distance can be calculated using the Euclidean measurement to find equality among data, which is commonly used as presented in Equation (2):

$$d(x,y) = \sqrt{\sum_{n}^{i=1} (a_i(x) - a_i(y))^2}. \tag{2}$$

- **Artificial Neural Network (ANN):** A multilayer perceptron is an NN technique often used in image, speech, and vision recognition systems. Recent studies conducted have extensively employed multilayer perceptron in the area of medical diagnosis [30]. The multilayer perceptron is a feedforward NN comprising the following three layers: input, hidden, and output layers. The input values are fed into the input layer, and the information is sent by the hidden layer to the output layer.
- **Logistic Regression (LR):** LR is a type of parametric classification pattern in machine learning techniques despite the word "regression" in the primary definition and concept [30]. Therefore, LR schemes are those with a characterized number of parameters considering the number of information attributes, and the output of this model is often categorical prediction. LR is also one of the most fundamental and common algorithms used in solving the classification problem. The logistic function is a Sigmoid function, which takes any real value between zero and one. The presented LR model uses a ridge (L2) regulation, and the constant is 1. This model is defined as

$$\sigma(t) = \frac{e^t}{e^t + 1} = \frac{1}{1 + e^{-t}} \tag{3}$$

- **Naive Bayes (NB):** The NB is another basic machine learning technique dependent on the Bayes hypothesis with freedom presumptions between perception information. Moreover, the NB is advantageous due to its simple learning model and the absence of any complex iterative parameter estimation [31].
- **Support Vector Machine (SVM):** This popular supervised machine learning model is used in addressing problems associated with the binary prediction [32]. The main concept of this method is rooted in the theories of the hyper-plane and the margin. The linear separator capable of separating the training data is observed for the learning process. Moreover, the margin between the linear separator and the training dataset is maximized using this separator. The linear separation among the statistics in its creative demonstration cannot

be directly identified by the SVM. Thus, Vapnik [32] suggested that the training data should be transformed from the original space to another space of high dimension.

- **AdaBoost** is often used as the short form of Adaptive Boosting, a machine learning meta-algorithm that was developed by Yoav Freund and Robert Schapire, who won the 2003 Gödel Prize for their work. Performance can be improved through this meta-algorithm if used together with other kinds of learning algorithms. The final output of the boosted classifier is obtained by combining the output of the other learning algorithms ("weak learners") into a weighted sum. AdaBoost is adaptive, because it allows the tweaking of subsequent weak learners in favor of misclassified instances by previous classifiers. AdaBoost also demonstrates sensitivity to outliers and noisy data [34].
- **CN2 rule inducer:** This learning algorithm is used for rule induction, and its main idea is rooted in the concepts of the ID3 and QA algorithms. The design of the CN2 is created such that it works with imperfect data [35]. CN2 is rooted in the concepts of the two aforementioned algorithms. Thus, this algorithm can create a similar set of rules to the AQ and deal with noises similar to ID3.
- **Decision Tree (DT):** DT is an effective and simple supervised classification method that can be easily interpreted. The existing non-linear correlations between the input and output of the system are identified by DT. The DT is an iterative method through which parameters are separated into nodes and branches, and these nodes comprise a single foundation and inertial and leave nodes [35].
- **Stochastic Gradient Descent (SGD):** SGD is an approach characterized by its simplicity and high level of efficiency. This approach is often used in discriminative learning of linear classifiers under convex loss functions, such as (linear) SVMs and LR [36]. SGD only recently received attention in the area of large-scale learning despite its long existence in the field of machine learning.

3.3.1 Automated Voice-based PD Diagnosis Hybrid Model. This section describes the proposed HM, which comprises a binary CNN and three feature selection algorithms (GA, AO, and M-BGD). The CNN is used to classify the PD based on the patients' voice, while the three other algorithms are used to perform a global search and adjust the weights of the CNN. Figure 4 presents a description of the proposed HM along with the new MAFT algorithm.

As shown in Figure 4, all the initial feature selection processes in the convolution layer include GA, AO, and M-BGO algorithms, which cooperatively generate an initial set of features. Subsequently, the final feature selection process in the pooling layer is performed by the MAFT model. Evaluations of the solutions are iteratively repeated until the best solutions to the CNN weights are obtained and the minimum classification error rate is achieved. In Figure 4, the optimization process of CNN using various feature selection methods of initial GA, AO, M-BGD, and final MAFT is described. The HM contains five fully connected layers as described below.

As a result of convolution and pooling layers, the full matrix of features is used for conversion to a one-dimensional vector using the attending process. This output is inserted into the fully connected layer as an input to the so-called neurons. Therefore, the fully connected layer comprises two layers: one for feature vector produced by convolutional and pooling layers, and the other is a 2×2 size vector representing the PD classification number of the classes. Once all blocks of CNN are created, the training process for the entire model can be applied. In the training process, the weights of CNN are continuously updated until the best solutions are obtained with high accuracy. Three optimization methods, including GA, AO, and M-BGD algorithms, are utilized in the proposed model. These methods are compared considering the classification accuracy. In the

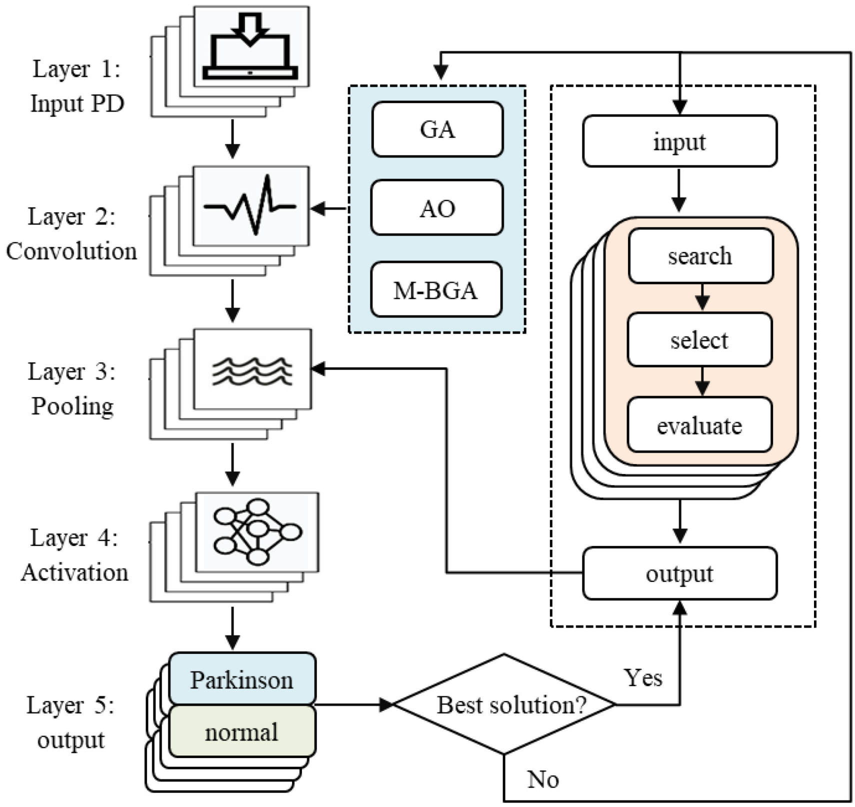

Fig. 4. Automated Voice-based PD Diagnosis Hybrid Model.

context of GA, the population is continuously updated during the extraction time, and several numbers of generations to improve the quality of solutions. The quality of the CNN classification for accuracy relies on the weights of CNN layers.

(1) Input layer: This layer receives the input of voice-based PD and feeds it to the next layer, which is called the convolutional layer. In the proposed model, the input is obtained from the database of voice-based PD. This database belongs to a repository of machine learning of the UCI. The dimension of the voice-based PD dataset is $128 \times 128 \times 2$.
(2) Convolutional layers: This layer facilitates the creation of the feature by using voice PD input in convolution operations, such as filters (kernels). The CNN design comprises six or (eight) convolutional layers, and 2×2 size of several kernels is distributed as follows: 6 kernels in the first and second convolutional layers, 12 kernels for the third and fourth layers, 28 kernels for the fifth and sixth layers, and 36 kernels for the last two layers. The normal/uniform distribution is used to initiate the weights of kernels randomly. Matrix form is used to store all the CNN weights, including the weight of fully connected layers and the convolutional kernels to use them for any further operation. However, a one-dimensional vector is used to store the initial GA population. In matrix form, every weight matrix is transformed to the one-dimensional vectors to act as the initial GA population. Additional details of this operation will be explained later.

(3) Pooling layer: This layer decreases the dimension of the feature map through the calculation of the average of kernels in the convolutional layers using the average pooling function.
(4) Activation layer: This layer task enhances the convergence rate for the learning process. An activation function, namely, ReLU, is located after the convolutional and pooling layers.
(5) Output layer: This layer presents the output of Parkinson or normal class.

Considering the initial feature selection, a weight matrix is used to store the initial weights of the kernels and the fully connected layer [32]. One solution (individual) comprises 5,426 weights in the proposed model. A population of at least 10 ten solutions is obtained in each generation. Therefore, the total number of parameters is 50,424 as a minimum. The inputs of this function are the weight matrix and the size of the output vector. The size of the one-dimensional vector is equal to the number of elements in the weight matrix. The proposed model comprises 5,426 elements for the signal dimensional vector, which acts as the initial population of GA. The genes are constructed in real number-based representation, which reflects the vector of CNN weights. The uniform distribution is used to generate the initial population in real number-based representation randomly. Each weight represents a gene, and the single solution represents the single chromosome (individual). In each generation, the fitness function is used to evaluate the quality of each solution to find the best solution, which reflects the best set of CNN weights. Figure 4 shows the evaluation block. The corresponding NN is constructed by a set of weights obtained by the GA. The following parameters for the GA are used in the proposed model.

- Solution number in each generation ranges from 10 to 40.
- Generation number ranges from 50 to 500.
- Mutation rate set ranges from 0.1 to 0.01.
- Normal and uniform distributions are used to initialize the network weights.

The solution fitness is measured using the error rate of classification formulated in Equation (4). The value of 0 refers to the best error rate, while the value of 1 refers to the worst error rate. Thus, the predictions are inaccurate:

$$\textit{Classification error rate} = \frac{\textit{False positives} + \textit{False negatives}}{\textit{Total number of samples}} \tag{4}$$

The evaluated solutions are sorted in accordance with their fitness values. The fit solutions are possibly chosen in the next generation to create new children (solutions). Genetic operators, including mutation and crossover, are utilized on the chosen individuals. One-point crossover is utilized on the chosen parent to generate new children for the next generation. The mutation operator is applied to one gene of the children using random uniform distribution to diversify the population.

4 RESULTS AND DISCUSSION

The classification of PD subjects was conducted through the implementation of the aforementioned classification techniques. A total of 11 classification techniques, including KNN, SVM, RF, LR, NN, NB, DT, AdaBoost, SGD, CN2 rule inducer, and the proposed HM, were used in this study. Standard evaluation metrics were also used in comparing the classification techniques. The extracted and selected features from the raw data are fed into the supervised and unsupervised approaches as inputs.

Table 2. PD Database Details [19]

Instances Number	197
Dataset Features	Multivariate
Attribute Characteristics	Real
Number of Attributes	23
Missing Values	0
Associated Tasks	Classification

4.1 PD Dataset

The PD database is a real database comprising human voice recordings. The dataset also contains features of voice disorder prediction used for PD prediction. Data collection was conducted through the use of the Intel At-Home Testing Device. The data were collected from the patients who volunteered from the comfort of their homes [19]. Some erroneous recordings were removed from the recorded data. The signal analysis was simplified using the sustained vowels method, which was also used in minimizing the ambiguous influences of the speech [39]. Afterward, the data were conducted online through the Internet to a clinic for the detection of the UPDRS of the patients. The database composition includes different biomedical voice quantities of 31 people, of which 23 of them were identified with PD [37]. A copy of the PD database was obtained from the machine-learning repository of the UCI [19]. The explanation of the PD database details is given in Table 2 below.

4.2 Performance Evaluation

This phase involves the evaluation of each method to determine which of the methods is capable of producing the best outcome. Certain parameters were used at this phase to evaluate the different methods, and these parameters include F1 score, Accuracy, true positive, false positive, Precision, Recall, MAE, and ROC from the confusion matrix. The constituents of the confusion matrix include True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN).

- **Accuracy** is defined as the closeness factor of metrics for data value reading near the actual data outcomes:
$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \times 100. \tag{5}$$
- The **Precision** measurement is utilized to calculate the quantity of the matters that are of good significance. The capability of the method to discard insignificant matters was calculated through this measurement. The formula of this metric is as follows:
$$Precision = \frac{TP}{TP + FP}. \tag{6}$$
- **The Recall** measurement is utilized to calculate the quantity of the matters that are evaluated. The capability of the method to discard insignificant matters was calculated through this measurement. The formula of this metric is presented as follows:
$$Recall = \frac{TP}{TP + FN}. \tag{7}$$
- **ROC curve** is used in measuring the performance of the classification problem at a wide range of threshold settings. ROC is a probability curve, and the degree of separability is represented by AUC, which also shows the capability of the model to distinguish between classes. A high AUC is indicative of the good performance of the model considering the

prediction of 0s as 0s and 1s as 1s. Considering the analogy, a high AUC indicates the capability of the model to distinguish between patients that are affected by PD and those that are not.

- **The F1 metric** is identified as a harmony measurement for recall and precision values. This metric reaches its best ratio in 1 (best value recall and precision). The F1 score can be calculated in the following formula:

$$F1 = \frac{Precision * Recall}{Precision + Recall} \tag{8}$$

4.3 k-Fold Cross-validation

The k-fold cross-validation, which produces satisfactory classification results through the division of all features into k sets with approximately similar numbers of individuals in every set, was employed in the training and testing data schema [38]. The features are divided into k sets, which are the testing and training groups. This method is conducted repeatedly for k times to ensure that each instance is used only once for testing. The present study used k = 10, implying that the training data is 90% of the data, while the remaining 10% is used for data testing in each iteration.

4.4 PD Classification Results and Evaluation

The initial outcomes achieved are for the MAFT of the PD database. The MAFT algorithm computes the average number of times that all the threshold t and PD features appeared in the demand value of the x^{in}. A maximum of 14 physical group features is required for the threshold process, while the 8 PD features are excluded (except for the status feature). The worth scores of the features for five agents are presented in Table 3 based on MAFT. The table reveals different results for feature selection and ranking from the evaluator agents. The last column of Table 3 presented the last ranked features by the **Ω** after deliberating on the combination of the subsets of feature sequence and the mean rank scores calculated via the MAFT approach.

The feature vector of the filtered database comprises 14 of the 22 features of the primary PD dataset. At the initial stage, the Zero R classifier is applied to the original and filtered Parkinson's datasets to obtain the initial accuracy result for both datasets. Similar results were obtained for the two datasets considering the error rates and accuracy of the Zero R, with a slight improvement of 4% accuracy observed in the filtered dataset. The Zero R classifier achieved an initial accuracy rate of 79.02% for the original dataset, while that of the filtered dataset was 74.64%. Then, using the MAFT benchmark, PD was diagnosed by utilizing KNN, RF, NN, NB, LR, SVM, SGD, AdaBoost, DT, CN2 rule inducer, and the proposed HM. The classification learning methods for the tests were evaluated using 10-fold cross-validation. A total of 100 tests are performed involving the use of ML methods, PD datasets, and 10-fold cross-validation. The overall results for the original dataset are presented in Table 4 based on the following parameters: Accuracy, Precision, Recall, TP, FP, AUC, MAE, and F1 scores.

The results for the original Parkinson's dataset revealed that the highest accuracy of 93.7% was achieved by the HM but yielded the lowest accuracy of 75.4% was achieved by the NB. An intermediate average accuracy of 86.35% was realized by other methods, including ANN, KNN, SVM, LR, AdaBoost, SGD, CN2 rule inducer, and DT.

By contrast, the test results of the filtered PD database revealed that the highest diagnostic accuracy of 96.9 % was achieved for the proposed model but yielded the lowest accuracy of 78.5% was achieved by the NB. Average accuracy of 86.35 % was achieved for the other methods, such as ANN, KNN, SVM, AdaBoost, SGD, DT, CN2 rule inducer, and Logistic Regression. The overall results for the filtered dataset are shown in Table 5.

Table 3. Twenty-two Feature Evaluations of the PD Dataset

Feature No.	**Feature Name**	**PD Feature Evaluation**					
		α_1	α_2	α_3	α_4	α_5	Ω
1	MDVP:Fo (Hz)	13	1	13	3	15	6
2	MDVP:Fhi (Hz)	11	2	7	10	19	7
3	MDVP:Flo (Hz)	15	3	1	6	14	4
4	MDVP:Jitter (%)	19	20	15	17	7	19
5	MDVP:Jitter (Abs)	1	15	14	4	4	8
6	MDVP:RAP	22	18	6	12	5	15
7	MDVP:PPQ	14	17	20	21	1	21
8	Jitter:DDP	2	4	8	9	6	2
9	MDVP:Shimmer	21	19	10	7	10	14
10	MDVP:Shimmer (dB)	16	22	12	13	8	18
11	Shimmer:APQ3	20	16	17	18	3	20
12	Shimmer:APQ5	18	14	11	8	9	12
13	MDVP:APQII	3	5	3	5	12	1
14	Shimmer:DDA	4	11	21	14	2	11
15	NHR	5	6	5	20	11	9
16	HNR	6	21	16	15	18	16
17	RPDE	7	13	22	22	20	18
18	D2	9	9	18	16	21	13
19	DFA	8	12	19	19	13	17
20	spreadI	12	7	2	2	17	3
21	spread2	17	8	9	11	22	10
22	PPE	10	10	4	1	16	5

Table 4. Overall Results of the Original PD Dataset

ML Model	**Accuracy**	**AUC**	**Precision**	**Recall**	**F1 Score**
HM	93.7	0.91	92.4	91.5	92.2
RF	90.3	94.2	90	90.3	90
LR	85.6	89	85	85.6	84.9
NB	75.4	87.6	83	75.4	77
SVM	87.7	89.3	88.8	87.7	86.3
AdaBoost	90.3	87.2	90.3	90.3	90.3
NN	84.6	87.1	84.6	84.6	82.9
KNN	85.1	85.4	84.5	85.1	84.3
DT	87.7	84.1	87.5	87.7	87.6
CN2	82.1	83.5	83.1	82.1	82.4
SGD	87.7	78.5	87.5	87.7	86.9

Remarkable improvements were generally found in the diagnosis results obtained for the 11 classifiers during the implementation of the filtered dataset. The HM achieved the highest accuracy improvement of 3.2%, while the lowest accuracy improvement was recorded for the NN and KNN (0.5%). The average development in the diagnostic accuracy outcomes reached 4%. The differences in the diagnostic accuracy outcomes obtained from the ML models for the two databases are illustrated in Figure 5.

Table 5. Results of the Filtered PD Dataset

ML Model	**Accuracy**	**AUC**	**Precision**	**Recall**	**F1 Score**
HM	96.9	0.93	93.4	92.2	94.3
RF	87.2	93.9	86.8	87.2	86.9
LR	86.2	89.1	85.7	86.2	85.4
NB	78.5	89.4	83.7	78.5	79.7
SVM	88.2	88	89.2	88.2	86.9
AdaBoost	91.3	89.3	91.5	91.3	91.4
NN	85.1	87.9	86.7	85.1	82.8
KNN	85.6	84.5	85.2	85.6	84.6
DT	86.2	81.9	86.1	86.2	86.1
CN2	81.5	84.8	81.5	81.5	81.5
SGD	86.7	77.1	86.3	86.7	85.8

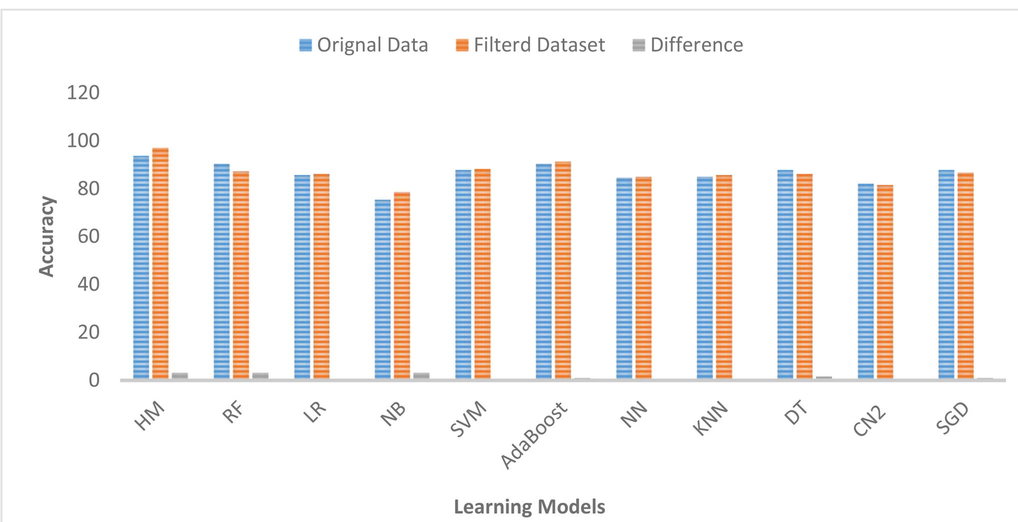

Fig. 5. Difference between the outcomes of the original and filtered PD datasets.

Moreover, the computation of the AUC was conducted in accordance with the ROC generated by sweeping through all possible discriminating thresholds in the scores output by a given classifier. Thus, the AUC is an excellent metric for measuring the overall performance of a classifier. Figures 6 and 7 present the ROC curve of the filtered dataset for all the classifiers considering healthy and PD patients, respectively.

Primarily, classification can be accurately performed by investigating the following: (i) the best method of predicting or diagnosing a given disease or problem, (ii) the best algorithm for the evaluation and selection of features, and (iii) the best parameter setting for the ML techniques based on the selected features. Herein, the most suitable and efficient classifiers have identified healthy people and persons affected by the PD through the application of a wide range of classification approaches to the PD database. The related works reveal that when features are reduced, the feature vector of the dataset is enhanced, the classifier complexity is minimized, and the diagnostic accuracy results are improved. For instance, NNs, which are classifiers that can accurately deal

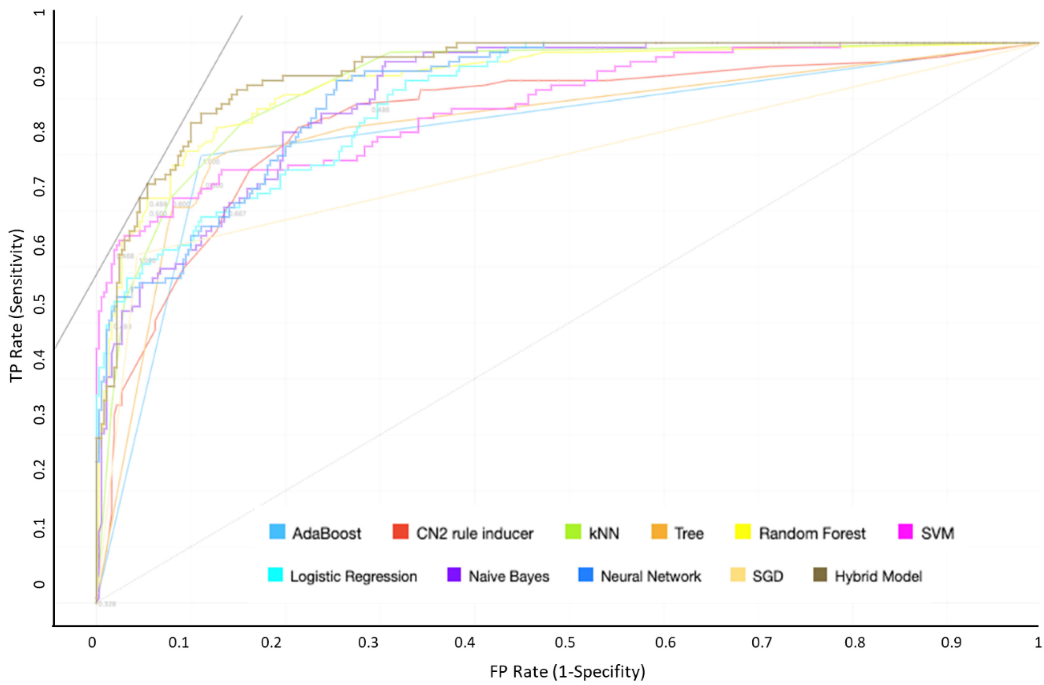

Fig. 6. ROC curve of the filtered dataset considering the records of healthy patients.

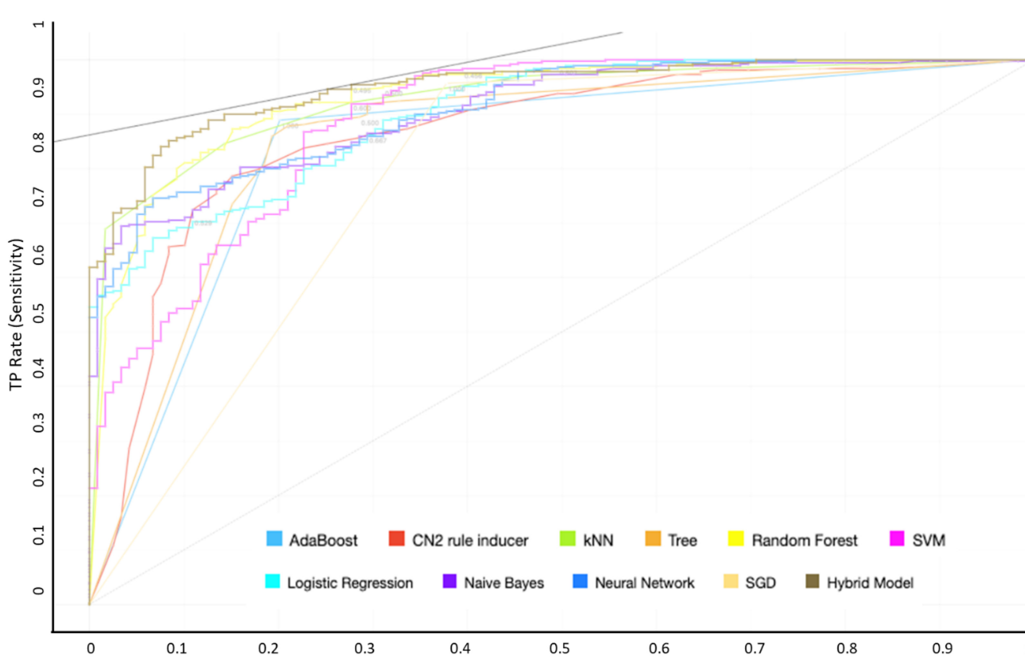

Fig. 7. ROC curve of the filtered dataset considering the records of PD patients.

Table 6. Comparison with State-of-the-art Methods

Classifier	All features (22)	Benchmark study (14 features)	Proposed study (14 features)
RF	90.3	**92.3**	87.2
LR	85.6	83.6	**86.2**
NB	75.4	75.4	**78.5**
KNN	85.1	84.1	**85.6**
DT	87.7	86.7	86.2

with complex decision boundaries, often demonstrate sensitivity to the feature selection. Thus, such classifiers are also likely to experience over-fitting.

Nevertheless, the evaluation, selection, and ranking of the PD features in the classification issues are not guided by a fixed rule. This process is executed on the basis of the attributes possessed by the PD data training, the kind of classifiers, and the complexity of the decision boundaries. The main contribution of the current study is a MAFT approach of the multi-agent system to obtain a satisfactory evaluation, selection, and ranking of the PD features. The critical role of feature selection in the achievement of diagnosis accuracies for the classification scheme is highlighted by executing the MAFT in the PD disease diagnosis. One of the key strengths of the MAFT is that it can adopt a general feature choice done the application of a combination of collaborative methods of the feature evaluation approach. Furthermore, the MAFT does not over-reduce and strictly avoids overweight features, thus preventing over-fitting of classifiers. The present work is aimed at identifying the best mix of features that can be used to achieve a balanced feature selection and improved PD classification precision outcomes.

This study is limited by the runtime, which is not considered during the evaluation process. Moreover, one way of minimizing the effect of the implanted data is through the redundancy of the filtered features. Thus, the classification results of the NN can be a robust indicator, because these results process features of independent correlations. However, the results of the SGD classification improved progressively without any oscillation. Moreover, the use of additional voice data can help the classifier distinguish other degenerative diseases that can lead to a voice disorder.

5 COMPARISON WITH STATE-OF-THE-ART METHODS

The most related work to the current study is found in Reference [18], wherein 14 features were selected as the best according to the proposed hybrid selection method. This study has the same number of selected features as the current study but with different subsets, of which 7 of 14 features (5, 8, 9, 12, 13, 15, 20) were selected by the proposed method and ignored by the benchmark study. Moreover, in the current study the experiment was executed using the Orange Python tool, and the hyperparameters of the selected classifiers were kept default. Furthermore, the benchmark study was reimplemented following the experimental settings of the current study. The benchmark study selected five classifiers (RF, LR, NB, KNN, and DT) as a baseline for comparison with other studies. The same strategy was also followed as a cornerstone for the comparison with state-of-the-art studies. However, the comparison was conducted in accordance with the number of improvements in the accuracy per each classifier as shown in Table 6.

As shown in the above table, the benchmark and the proposed study have different effects on the accuracy improvement per each classifier. According to the group of features selected by the benchmark study, the increased accuracy is found only in RF classifiers, where the accuracy ratio improved by 2%. On the one hand, the same group has negative accuracy improvement on LR, KNN, and DT. On the other hand, the group of selected features by the proposed method revealed a significant enhancement for LR, NB, and KNN with respective scores of 0.6%, 3.1%, and 0.5% as accuracy improvement ratios. However, the negative impact of the proposed method is observed considering RF and DT classifiers. According to Reference [17], important issues, such as medical diagnostics and diagnostic systems, even a 0.1% increase is crucial. Consequently, the proposed method is expected to provide an important contribution to this field.

6 CONCLUSION

A variety of classification methods for the recognition of PD based on voice disorder features were implemented in this study. The entire construction of the PD detection process was also discussed, starting from the initial stage of data acquisition, followed by the feature extraction and the selection process. In addition, the various methods of classification were described using a theoretical background. Finally, the 11 classification methods used in this study were compared based on rate Accuracy, True Positive, False Positive, Precision, Recall ROC, MAE, and F1 score results. The inputs for the classifiers were the 14 selected features. Subsequently, the leave-one-out cross-validation was employed in generalizing the performances of the classifiers. Multiple feature evaluation and classification approaches were investigated in the present study to improve the PD diagnosis. The testing of the MAFT involved the use of 11 independent classifier methods: NB, KNN, RF, NN, DT, LR, SGD, SVM, AdaBoost, CN2 rule inducer, and the proposed HM. A total of 100 tests were performed for the two datasets (original and filtered). The results showed that of all the classifiers, the highest diagnostic accuracy of 96.6% was achieved by the HM. Remarkable improvements were generally observed in the diagnosis accuracy results of the filtered dataset when the MAFT is used. The HM and NB achieved the highest accuracy improvement of 3 %, whereas RF had the negative improvement. Most of the tested classifiers yielded an average improvement of 3.90%. Moreover, this study presents a comparative review of multiple machine learning methods and aims to provide relevant insights for future research initiatives. The proposed review summary of the application of numerous techniques in identifying and detecting PD based on voice feature can support future studies and provide a solid overall state-of-the-art background. Nevertheless, this study also has some limitations. The runtime period has not been evaluated, and even the computational effort for each method is not analyzed. The authors aim to study the previously mentioned parameters in the future. Furthermore, the properties of the dataset with feature evaluation in the MAFT and its results will be tested with databases possessing a variety of properties. In the future, the proposed method will be investigated with different medical cases and additional benchmark datasets. Furthermore, the performance of the proposed method will be validated in accordance with the computational load and time complexity criteria to measure the capability of the proposed methods to be adopted in the real-time scenario.

REFERENCES

[1] S. A. Mostafa, A. Mustapha, M. A. Mohammed, R. I. Hamed, N. Arunkumar, M. K. A. Ghani, M. M. Jaber, and S. H. Khaleefah. 2019. Examining multiple feature evaluation and classification methods for improving the diagnosis of Parkinson's disease. *Cogn. Syst. Res.* 54 (2019), 90–99.

[2] K. J. Kubota, J. A. Chen, and M. A. Little. 2016. Machine learning for large-scale wearable sensor data in Parkinson's disease: Concepts, promises, pitfalls, and futures. *Movement Disorders* 31, 9 (2016), 1314–1326.

[3] D. Avci and A. Dogantekin. 2016. An expert diagnosis system for Parkinson's disease based on genetic algorithm-wavelet kernel-extreme learning machine. *Parkinson's Disease.*

[4] H. L. Chen, G. Wang, C. Ma, Z. N. Cai, W. B. Liu, and S. J. Wang. 2016. An efficient hybrid kernel extreme learning machine approach for early diagnosis of Parkinson′ s disease. *Neurocomputing* 184 (2016), 131–144.

[5] S. A. Mostafa, A. Mustapha, S. H. Khaleefah, M. S. Ahmad, and M. A. Mohammed. 2018. Evaluating the performance of three classification methods in diagnosis of Parkinson's disease. In *Proceedings of the International Conference on Soft Computing and Data Mining*. Springer, Cham, 43–52.

[6] K. Mueller, R. Jech, and M. L. Schroeter. 2013. Deep-brain stimulation for Parkinson's disease. *N. Engl. J. Med.* 368, 5 (2013), 482–483.

[7] D. Georgiev, M. Domellof, K. Hamberg, L. Forsgren, and G. M. Hariz. 2019. Sex differences, quality of life and non-motor symptoms in Parkinson's disease.

[8] A. Rueda, J. C. Vásquez-Correa, C. D. Rios-Urrego, J. R. Orozco-Arroyave, S. Krishnan, and E. Nöth. 2019. Feature representation of pathophysiology of Parkinsonian dysarthria. In *Proceedings of the Annual Conference of the International Speech Communication Association (INTERSPEECH'19).* 1–5.

[9] C. R. Pereira, D. R. Pereira, J. P. Papa, G. H. Rosa, and X. S. Yang. 2016. Convolutional neural networks applied for Parkinson's disease identification. In *Machine Learning for Health Informatics.* Springer, Cham, 377–390.

[10] 2018. Parkinson's Disease-Symptoms, Stages and Life Expectancy, Pathology, Lecturio. Retrieved from https://www.lecturio.com/magazine/parkinsons-disease/.

[11] K. H. Abdulkareem, M. A. Mohammed, S. S. Gunasekaran, M. N. Al-Mhiqani, A. A. Mutlag, S. A. Mostafa, N. S. Ali, and D. A. Ibrahim. 2019. A review of fog computing and machine learning: Concepts, applications, challenges, and open issues. *IEEE Access* 7 (2019), 153123–153140.

[12] M. K. Abd Ghani, M. A. Mohammed, N. Arunkumar, S. A. Mostafa, D. A. Ibrahim, M. K. Abdullah, M. M. Jaber, E. Abdulhay, G. Ramirez-Gonzalez, and M. A. Burhanuddin. 2020. Decision-level fusion scheme for nasopharyngeal carcinoma identification using machine learning techniques. *Neural Comput. Appl.* 32, 3 (2020), 625–638.

[13] M. A. Mohammed, K. H. Abdulkareem, S. A. Mostafa, M. K. A. Ghani, M. S. Maashi, B. Garcia-Zapirain, I. Oleagordia, H. Alhakami, and F. T. AL-Dhief. 2020. Voice pathology detection and classification using convolutional neural network model. *Appl. Sciences* 10, 11 (2020), 3723.

[14] D. Gupta, S. Sundaram, A. Khanna, A. E. Hassanien, and V. H. C. De Albuquerque. 2018. Improved diagnosis of Parkinson's disease using optimized crow search algorithm. *Comput. Electric. Eng.* 68 (2018), 412–424.

[15] D. Gupta, A. Julka, S. Jain, T. Aggarwal, A. Khanna, N. Arunkumar, and V. H. C. de Albuquerque. 2018. Optimized cuttlefish algorithm for diagnosis of Parkinson's disease. *Cogn. Syst. Res.* 52 (2018), 36–48.

[16] K. Wrobel. 2019. Diagnosing Parkinson's disease with the use of a reduced set of patients' voice features samples. In *Proceedings of the IFIP International Conference on Computer Information Systems and Industrial Management.* Springer, Cham, 84–95.

[17] H. Gürüler. 2017. A novel diagnosis system for Parkinson's disease using complex-valued artificial neural network with k-means clustering feature weighting method. *Neural Comput. Applications* 28, 7 (2017), 1657–1666.

[18] A. Ul Haq, J. Li, Z. Ali, M. H. Memon, M. Abbas, and S. Nazir. Recognition of the Parkinson's disease using a hybrid feature selection approach. *J. Intell. Fuzzy Syst.* 1–21.

[19] M. Little, P. McSharry, E. Hunter, J. Spielman, and L. Ramig. 2008. Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *IEEE Trans. Biomed. Eng.* 56, 4 (2009).

[20] A. G. Ramayya, A. Misra, G. H. Baltuch, and M. J. Kahana. 2014. Microstimulation of the human substantia nigra alters reinforcement learning. *J. Neurosci.* 34, 20 (2014), 6887–6895.

[21] M. Can. 2013. Neural networks to diagnose the Parkinson's disease. *Southeast Eur. J. Soft Comput.* 2, 1 (2013).

[22] A. Tsanas, M. A. Little, P. E. McSharry, and L. O. Ramig. 2009. Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests. *IEEE Trans. Biomed. Eng.* 57, 4 (2009), 884–893.

[23] E. Kaya, O. Findik, I. Babaoglu, and A. Arslan. 2011. Effect of discretization method on the diagnosis of Parkinson's disease. *Int. J. Innov. Comput. Info.* 7 (2011), 4669–4678.

[24] M. Hariharan, K. Polat, and R. Sindhu. 2014. A new hybrid intelligent system for accurate detection of Parkinson's disease. *Comput. Methods Programs Biomed.* 113, 3 (2014), 904–913.

[25] S. Doan and S. Horiguchi. 2004. An agent-based approach to feature selection in text categorization. In *Proceedings of 2nd International Conference on Autonomous Robot and Agent.* 362–366.

[26] F. Farahnakian and N. Mozayani. 2009. Evaluating feature selection techniques in simulated soccer multi-agents system. In *Proceedings of the International Conference on Advanced Computer Control.* IEEE, 107–110.

[27] M. S. P. Subathra, M. A. Mohammed, M. S. Maashi, B. Garcia-Zapirain, N. J. Sairamya, and S. T. George. 2020. Detection of focal and non-focal electroencephalogram signals using fast Walsh-Hadamard transform and artificial neural network. *Sensors* 20, 17 (2020), 4952.

[28] M. A. Mohammed, K. H. Abdulkareem, A. S. Al-Waisy, S. A. Mostafa, S. Al-Fahdawi, A. M. Dinar, W. Alhakami, A. Baz, M. N. Al-Mhiqani, H. Alhakami, and N. Arbaiy. 2020. Benchmarking methodology for selection of optimal COVID-19 diagnostic model based on entropy and TOPSIS methods. *IEEE Access.*

[29] M. A. Mohammed, M. K. A. Ghani, R. I. Hamed, S. A. Mostafa, D. A. Ibrahim, H. K. Jameel, and A. H. Alallah. 2017. Solving vehicle routing problem by using improved K-nearest-neighbor algorithm for best solution. *J. Comput. Sci.* 21 (2017), 232–240.

[30] M. A. Mohammed, B. Al-Khateeb, A. N. Rashid, D. A. Ibrahim, M. K. A. Ghani, and S. A. Mostafa. 2018. Neural network and multi-fractal dimension features for breast cancer classification from ultrasound images. *Comput. Electric. Eng.* 70 (2018), 871–882.

[31] C. I. Sánchez, R. Hornero, A. Mayo, and M. García. 2009. Mixture model-based clustering and logistic regression for automatic detection of microaneurysms in retinal images. In *Medical Imaging 2009: Computer-Aided Diagnosis, Vol. 7260.* International Society for Optics and Photonics, 72601M.

[32] N. Arunkumar, M. A. Mohammed, S. A. Mostafa, D. A. Ibrahim, J. J. Rodrigues, and V. H. C. de Albuquerque. 2020. Fully automatic model-based segmentation and classification approach for MRI brain tumor using artificial neural networks. *Concurr. Comput.: Pract. Exper.* 32, 1 (2020), e4962.

[33] N. Arunkumar, M. A. Mohammed, M. K. A. Ghani, D. A. Ibrahim, E. Abdulhay, G. Ramirez-Gonzalez, and V. H. C. de Albuquerque. 2019. K-means clustering and neural network for object detecting and identifying abnormality of brain tumor. *Soft Comput.* 23, 19 (2019), 9083–9096.

[34] C. Ying, M. Qi-Guang, L. Jia-Chen, and G. Lin. 2013. Advance and prospects of AdaBoost algorithm. *Acta Automatica Sinica* 39, 6 (2013), 745–758.

[35] J. Van Zyl and I. Cloete. 2004. FuzzConRI—A fuzzy conjunctive rule inducer. In *Proceedings of the Workshop on Advances in Inductive Rule Learning (ECML'04).* 194–203.

[36] M. Zinkevich, M. Weimer, L. Li, and A. J. Smola. 2010. Parallelized stochastic gradient descent. In *Advances in Neural Information Processing Systems.* MIT Press, 2595–2603.

[37] J. Mekyska, Z. Galaz, Z. Mzourek, Z. Smekal, I. Rektorova, I. Eliasova, M. Kostalova, M. Mrackova, D. Berankova, M. Faundez-Zanuy, and K. López-de-Ipina. 2015. Assessing progress of Parkinson's disease using acoustic analysis of phonation. In *Proceedings of the 4th International Work Conference on Bioinspired Intelligence (IWOBI'15).* IEEE, 111–118.

[38] T. T. Wong. 2015. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recogn.* 48, 9 (2015), 2839–2846.

Received June 2020; revised October 2020; accepted November 2020