Early Parkinson Detection Using Fully-Connected Deep Neural Network based on Vocal Features

Article *in* International Journal of Network Security · June 2020

4 authors:

Ahlem Kehili
University of Tunis El Manar

Karim Dabbabi
University of Tunis El Manar

Chrif Adnen

Atssee Fst
University of Tunis El Manar

Early Parkinson Detection Using Fully-Connected Deep Neural Network based on Vocal Features

Ahlem Kehili[1], dabbabi Karim[2], Cherif Adnen[3]

[1,2,3]Research Unite of Processing and Analysis of Electrical and Energetic Systems,
Faculty of Sciences of Tunis, University Tunis El-Manar, 2092 Tunis -TUNISIA

Summary
Parkinson's disease (PD) is considered to be a growing neurodegenerative disease characterized by many motors and non-motor specifications. During the early stages of this disease, vocal impairments are usually the disorders that can appear for patients with PD. Thus, diagnosis systems based on vocal perturbations have become at the head of recent studies for the detection of PD. In our study, a PD detection system based on vocal features was proposed using Full-Connected Deep Neural Network (FC-DNN) as a classifier, and Jitter and shimmer variants plus mean fundamental frequency (meanF0), harmonic to-noise ratio (HNR) and duration as vocal features. The experimental tests were performed on Spanish dataset and the results have shown the superiority of FC-DNN in terms of the evaluated performances (Accuracy=100%, precision=99%, Recall=99, F-measure=99.1, and Matthew Correlation (MCC=0.95) with comparison to other tested classifiers of machine learning and to classic approaches.

Keywords:
Deep learning algorithm, machine learning algorithm, early Parkinson detection, vocal features, PD classification, FC-DNN

1. Introduction

Parkinson's disease (PD) remains a coherent line of research in the world due to the fact that the outgoing diagnoses still inaccurate in terms of time, cost and biology. Also, thoughts of doctors are responsible for the confusion of PD syndromes with those of other Parkinsonian diseases [1] by considering that particular diseases have precise diagnoses based on biochemical responses of patients, whereas other recurrent diseases are difficult to be detected using biochemical data and composition [2]. In general, the diagnoses of widespread diseases like PD do not rely on firm biochemical data to judge whether the individual is affected by a neurodegenerative disease or not. However, medical professionals rely solely on physical tests by focusing on diminished hand-eye and impaired movement coordination for early detection. Although, SPECT and DatSCANs are such analyzes that can be used to determine whether a patient has PD or not, whereas they were considered relatively inaccurate and not very profitable [3]. In addition, indirect diagnosis of PD can be performed using PET, CT and IMR scans which can be mainly applied for the elimination of other possible confounding diseases [3]. Doctors actually use a probability process to determine whether a patient has PD or not by first eliminating the possibility that he is affected by another disease. With the presence of PD, the results of these examinations have shown very weak correlations because the same symptoms are common to many other diseases. As instance, impaired movement and dizziness may be symptoms engendered by hypoglycemia and anemia. Some unexpected outcomes of these coping medications include mood swings, confusion, delusions, psychological changes and hallucination [4]. Thus, an early detection system for Parkinson's disease is necessary to allow doctors to start rehabilitation treatment early so that the patient does not have to consume medication for an extended period of time. Additionally, machine learning was explored in previous studies as a mean to develop an effective diagnosis for the identification of cancer [5] and heart failure [6] among other diseases. The analysis of human brain images [7] was also carried out by machine learning technology via the identification of brain tumors and the detection of retinal diseases in the laboratory [8]. In some other organizations, machine learning was used for the diagnostic of disease and there were warnings that inaccurate results were delivered due to the fact that these contain some inherent pitfalls [9]. At the same level, machine learning techniques fail in general to provide clinically validated diagnostic approaches when determining the possibility of existence of a disease. Both stochastic back propagation and classification are such machine learning techniques that suffer from over-fitting. Thus, a justified diagnosis of disease cannot be provided [2]. On the other hand, the integration of new approaches in machine learning techniques has a good impact on the verification of certain unexpected phenomena in contemporary medicine, such as the relationship between the prevalence of disease and demography [10]. Exploring a modern machine learning diagnosis for the treatment of diseases will not be only limited to reduce the medical pecuniary losses, but it will also proffer cost-effective and rapid solutions with a wide range of applications in the society.
Eventually, the clinician constructs his overall decision from the interpretation of the precedent results. Moreover, machine learning algorithms are such techniques that can be helped to make such a decision. Indeed, they are based

Manuscript received June 6, 2020
Manuscript revised June 20, 2020

on the combination of many input variables reflecting different characteristics and therefore they produce a single value that can assist the clinician. Also, these methods are researched on the basis of both statistical differences and statistical learning between PD and Control (NC) groups [11] [12] [13] [14] [15] [16]. Furthermore, we can find other approaches like logistic lasso [17], Naïve-Bayes [18], plus the general trends based on the most commonly used method which is the Support Vector Machine (SVM) with radial or linear basic function kernel. ANN-based methods have gained popularity in the last years in virtue of the fast increasing innovations that engender the construction of new architectures and the development of effective training algorithms. Thus, a wide range of applications [19] was performed using such algorithms like deep neural networks and multi-layer neural networks. Among the applications of these algorithms, we can cite those of speech recognition [20], genomics [21] and drug discovery [22]. In the field of image classification and computer vision, deep learning has recorded its presence via the particularly use of conventional neural networks (CNNs) which have made a real revolution of the-state-of-the-art [19]. In object recognition, these architectures have practically led to reach the human-level performance, and even they have surpassed it [23]. In fact, the more the networks are deeper the higher is the number of abstraction levels. This allows the computation of more complex features, but with the resulting of a more much complexity for training. Therefore, the performances can be degraded due to some limitations existing in both training algorithms and architectures, which can engender their over-fitting [24].

The success achieved in PD classification is strongly related to the choice of suitable feature extraction and artificial learning models. Many studies were performed in the literature using the same publicly available dataset [25] composed of 95 sound recordings with 31 instances each one (8 healthy individuals and 23 PD patients). In [26], another PD database was proposed and it was composed of 40 instances from multiple speech recordings (20 healthy individuals and 20 PD patients). Indeed, common features were extracted from both databases, such as the measure of amplitude and fundamental-frequency variations, vocal fundamental frequency etc. In PD detection studies, the features explored to carry out the experiments were extracted from these databases and they were referred to as baseline features. Also, there were other features based on signal processing techniques, which were explored for PD detection. Furthermore, the extraction of the relevant features in PD classification can be performed via important tools, such as Mel-frequency cepstral coefficients (MFCCs), Tunable Q-Factor Wavelet Transform (TQWT) and signal-to-noise ratio (SNR) [27]. In fact, to perform the classification task, most studies were relied on the combination of individual feature types rather than exploring those ones separately from a training model. In these studies, feature selection methods were applied in order to reduce the extended feature space [28]. Despite of the existing of a lot of symptoms (e.g. posture and balance dependencies and slowed movement) identifying a subject as a PD patient, however dysphonia which can be denoted as the changes in articulation and speech, is considered as the most meaningful forerunner of PD. For this reason, many studies were relied on speech for PD classification task.

There are many symptoms characterizing PD patients and some of them are manifested in the existing of vocal deflections (like vocal loudness, frequency and instability abnormality), while the other impairments make appearing impaired vocal quality and vocal breaks to PD patients. Speech processing techniques were considered as the most commonly used tools for the detection of speaking anomalies and they were preferred for automatically extracting the PD-related vocal features. In fact, there were several studies performed during the last decade using machine learning algorithms for the task of vocal features-based PD detection. In [27], authors have presented a novel approach for PD detection using vocal features and many feature selection techniques at the aim to feed the inputs of such model by the top 10 features which have the high relevant scores in the selection. Indeed, Relief and Local Learning-base Features Selection (LLBFS), Least Absolute Shrinkage and Selection Operation (LASSO), and Minimum Redundancy Maximum Relevance (mRmR) were used as feature selection methods, whereas SVM and Random Forest (RF) were explored as classifiers to evaluate the performances on the selected features. The precision rate was reached up to 98.6% using these classifiers combined with different features consisting of HNR, vocal fold excitation and shimmer. For the lowest classification error in that study, it was obtained using Relief selection. In [29], the proposed model for PD detection was based on voice signals and its inputs was composed of HNR, Pitch, jitter, fundamental frequency, shimmer, plus other statistical measures based on these parameters. For the selection of the informative features from the whole feature set, many feature selection techniques were utilized, such as ROC curves, t-test, correlation rates, Fisher’s Discriminant Ratio. Concerning the determination of the optimal number of features, it was performed using a wrapper approach employing SVM as classifier in order to construct a feature-performance curve. After the specification of the optimal features, the training of KNN, SVM and Discrimination-Function-Based classifiers was effectuated. In that study, the best performance was achieved using KNN classifier by reaching an accuracy value of 93.82%. In [30], authors have used Praat Software to extract different features from voice signals in order to distinguish between healthy individuals and those ones holding PD. Comparing MFCC, shimmer, pitch, jitter to the individual’s glottal pulse, it was mentioned that this latter and MFCC coefficients were different in their characteristics and therefore they had higher fluctuations when comparing healthy individuals to PD patients. As regards to the examinations performed using shimmer and jitter features, it was concluded that healthy individuals had lower values of features with comparison to PD patients. In [31], a model relied on a novel hybrid Artificial Intelligence-based classifier was suggested for early examination of PD. In this study, the data was explored from the University of California-Irvine (UCI) Machine Learning repository and it was composed of 68 instances with clinical scores and dysphonic measures. A training of Multi-Layer Perceptron (MLP) with custom cost function was firstly carried out in order to assign the essential scores of features. After that, the 20 features with high essential scores were inputted to the Lagrangian Support Vector Machine (LSVM) for classification. The overall performance of the hybrid MLP-LSVM classifier has shown an excellent value of accuracy rate (100%) with comparison to the available similar studies. A tunable Q-factor wavelet transform (TQWT) was recently applied for the diagnoses of PD using vocal signals from different individuals [32]. In that study, experiments were carried out using different types of feature sets constituting of multiple voice instances from 252 individuals. In fact, the extracted features were inputted to numerous classifiers that were also combined to the majority voting classifier. As summary of that study, TQWT features have succeeded to achieve better performance and even have surpassed the state-of-the-art voice features that are frequently employed in PD classification. Moreover, it was concluded that the combination of MFCC with TQWT has led to boost up the classification performance when mRmR selection was used. It is worth to mention from the above-mentioned works that the related PD studies were generally performed on the basis of voice-based features combined with machine-based learning algorithms. In fact, there were some other studies which have used different data sources like wearable sensors [33], Electroencephalogram (EEG) [34] and smart pens [35] to extract features, and they were not only limited to vocal-based features to perform the PD classification.

In PD studies, deep learning which is a subdivision of machine learning was also successfully implemented besides common machine learning algorithms. For example in [35], a smart pen was well-designed to capture hand-written dynamics from both PD patients and healthy individuals. The CNN was used in that work as a classifier and its inputs were fed by a time series data which has the role of modeling the handwritten dynamics. Indeed, the proposed CNNs were constructed on prior-trained deep-learning architectures, such as ImageNet, LetNet and Cifar10. Over all experiments, CNN has shown better results than Open Path Forest (OPF) classifier in terms of the evaluated performances. This comes back to its proficiency in learning important features to discriminate between healthy individuals from PD patients. In a study in [36], Deep Neural Network (DNN) consisting of a stack auto-encoder (SAE) and a softmax layer as a classifier was proposed. Indeed, the extraction of the intrinsic information from speech features was carried out via SAE, while the softmax layer was explored to interpret the encoded features for the classification of patients. The experimental tests were performed on two databases and the results of the evaluated performances have shown that DNN was a convenient classifier for the diagnoses of PD in comparison to the state-of-art machine learning models. In [37], another study for PD diagnoses was suggested on the basis of the effectiveness of DNN. The data explored in that study was composed of speech records and digital bio-markers collected from PD and non-PD individuals using mobile application. In fact, two types of feature sets were extracted from the preprocessed speech signals using the open-source tool (OpenSmile). For the AVEC-first feature set that has a dimension up to 2200, Minimum Redundancy Maximum Relevance (mRMR) was applied as a feature selection approach to these features. Concerning the second feature set, it was composed of 60 features with MFCC. Both feature sets were inputted to many artificial learning classifiers comprising 3-layered DNN. The classification results have shown that DNN has achieved better result in term of accuracy (85%) in comparison to all tested models, and even it has surpassed the average clinical diagnosis accuracy of non-experts (73.8%). In [34], a study was built on the hypothesis that there is a direct relationship between PD and the brain abnormality EEG signals. These latter represent the principal indicators in early diagnosis of PD. Indeed, the EEG signals were explored from 20 healthy individuals and 20 PD patients and then inputted into 13-layer CNN architecture for further detection of PD. In this study, CNN model has reached acceptable results in terms of the measured performances from which it has achieved the values of 71%, 91.77% and 88.25% for sensitivity, accuracy and specificity, respectively. The resulted development in wearable sensors offers the possibility to capture with minimum cost the disorders of the impairments of motor functions for individuals, which themselves a main reason for arising the PD. At this regard, a study in [33] was performed for the purpose of classification the bradykenesia which has the characteristic of an impaired ability when the body is moved. The data used in this study was collected from 10 patients with idiopathic PD. For the preprocessed and labeled feature vectors, they were used as inputs to deep learning and machine learning pipelines. The experimental results have shown the superiority of CNN-based classifier with comparison to the traditional machine learning models in terms of accuracy rates. In [38], the CNN model was utilized as a classifier and it has combined in its inputs different types of vocal features at feature- and model-level (first framework) at the aim to differentiate healthy individuals from PD patients. Also, 9-layered CNN was used in this study to perform another feature-level-based combination (second framework). Indeed, the combination of TQWT+MFCC has reached the best accuracy rate (84.5%) with comparison to all binary combinations. In fact, there were no improvements in terms of accuracy using the first framework. However, there were an increase in both MCC and accuracy rates for all feature combinations using the second framework. Moreover, the highest overall classification success was achieved using the model-level CNN combination approach with comparison to that one based on SVM. Also, the proposed CNN was a power alternative to the study suggested in [28] by reaching good values of accuracy, F-Measure and MCC (86.9%, 91.7% and 63.2%, respectively).

In regards to RNN models, they are considered to be naturally adapted to temporal sequences data and there were many other variants which were developed for sequenced features. Long-short-term memory is such variant (LSTM) of RNN, which first was proposed by Hochreiter and Schmidhuber. In fact, it has achieved impressive performances in numerous sequence-based tasks, such as language translation [39], acoustic modeling of speech [40], handwriting recognition [41], and language modeling [42]. In [43], authors have suggested the gated recurrent unit (GRU) model which is structurally similar to and simpler than LSTM, and has showed comparable, and even better, performance. In [44], this model was used to predict the initial diagnosis of heart failure (HF) by modeling the temporal relations among health data from electronic health records (EHRs) of patients. In [45], a deep recurrent neural network referred to as TimeNet was explored to firstly extract features from time series and then classify them. This model has shown better performances than the dynamic time warping and domain-specific recurrent neural network [45]. To overcome the problem of forecasting for the spatiotemporal sequences, a conventional LSTM (ConvLSTM) model was integrated in [46]. It was concluded that this model has performed better than the fully connected LSTM since it has a good power in capture the spatiotemporal correlations from the sequential data for precipitation now casting.

In this paper, we present the problem of PD detection as a features-based classification task. Indeed, the proposed model is composed of vocal features constituting of jitter and shimmer variants plus fundamental frequency (F0) and duration as features, and a Fully-connected Deep Neural Network (FC-DNN) model as a classifier for PD classification. In fact, the selection of DNN model comes back to its ability and potentiality for modeling nonlinear and complex relationships from data. Also, we have taken the courage to test this model in virtue of the satisfied results for PD classification task achieved in [33-35].

The remaining sections of this paper are presented as follows: In the next section, we exhibit the proposed method. The database and the evaluation metrics are presented in section 3 and 4, respectively.The data analysis and the interpretation of the experimental results are given in section 5. Conclusion and further work are presented in section 6.

2. The Proposed Method

The proposed model is composed of four stages: feature extraction, feature selection, data partition and classification. These stages are presented in the following subsections.

2.1 Feature Extraction

To assess and track the evolution of Parkinson disease (PD) after pharmacological and surgical treatment, speech features can be explored. This comes back from the fact that the PD can engender the affection of speech even at an early stage. In our study, we have used jitter and shimmer variants, mean F0, HNR and duration as features. Indeed, 14 acoustic features were extracted from .wav files and then saved in .CSV format using Parselmouth [47] and Praat softwares. Each column in the CSV file represents the status which is set to 1 for PD and 0 for healthy control (HC). However, each row in this file contains only one instance which corresponds to one voice recording. More details about the extracted features are shown in Table 1.

2.2 Feature Selection

As a preprocessing stage for classification, the features of the voice samples were selected and then linearly mapped into the interval ranging from -1 to 1. As regards to feature selection stage, it was carried out using Principal Component Analysis (PCA). This latter is defined as a well-known statistical procedure which is consecrated to extract features and reduce their dimensions. In fact, the conversion of a set of observations that have correlated variables into a smaller set of values of linearly uncorrelated variables is performed using PCA via an orthogonal transformation. Also, PCA is relied on the assumption that only the features with the most variance contain the information about certain classes. Indeed, a smaller set of n dimensions which are described by n leading eigenvectors of a global covariance matrix is explored to present the p-dimensional dataset. In this paper, the selected features which include all the principal components that have more than 0.1%, 0.5%, 1%, 5%, and 10% of the total variance were tested.

2.3 Data Partition

The input data were randomly split into train and test data sets using data partition component. In our study, we have explored the Cross-Validation (CV) method since the number of samples in the dataset is small. Indeed, a 10-fold cross-validation was explored, where each fold is constituted of 80% of samples for training and the remaining 20% were employed for testing purposes.

After the preprocessing and data partition stages, the selected features are then inputted into the selected classifier to perform the classification task.

Table 1: Detailed explanations of feature sets

Attribute Name	***Measure***	***Attribute description***
F0 (Hz)	Mean and standard deviation	Measure of vocal fundamental frequency
Jitter variants	Local_jitter, Local absolute_jitter, Jitter: RAP, Jitter: PPQ, Jitter: DDP Jitter: PCA	Several measures of variation in fundamental frequencies in order to detect the cycle-to-cycle changes from them
Shimmer variants	Local_Shimmer, Local_Shimmer(dB),Shimmer:APQ3,Shimmer: APQ5, APQ11, Shimmer: DDA, Shimmer: PCA	Several measures of variation in amplitudes in order to detect the cycle-to-cycle changes from them
HNR	Mean and standard deviation	Second measures of ratio of noise to tonal components in the voice
Duration	Mean and standard deviation	This feature is self-explanatory
Status		0 for HC and 1 for PD

2.4 Fully-Connected Deep Neural Network (FC-DNN)

The basic form of the Neural Network (NN) is the perceptron from which weights are used to weight the input signals and then feed them directly to the output neurons without the use of any hidden layer. For this the perceptron has a simple architecture and a computational efficiency even with very large databases [48]. Despite of these advantages, it nevertheless has the disadvantage of being limited to linearly-separable functions for learning. To resolve this problem, a multilayer perceptron called Deep Neural Network (DNN) was proposed in [49] at the aim to reach more powerful learning mechanics. In fact, the wide adoption of DNNs was the important factor for the emergence of Deep Learning (DL) as a new field.

Also, the NN is considered as a simple process made up of several hidden layers including nonlinear processing units which are responsible for the extraction of features by passing the input data from one layer to another until a desirable output will be produced [48]. One of the most used algorithms in DL is the Fully-Connected Deep Neural Network (FC-DNN). This designation comes back from the fact that the layers are fully connected (dense) via the neurons of a network layer. Indeed, all neurons of the previous layer transmit the input to each neuron of the next layer. For this they are called densely connected. An instance of the representation of the dense neural network (DNN) on Tensor Flow Playground is given in Fig.1.

On the other hand, we have used a densely connected layer instead of that one of convolution. This is justified by the ability of the first one to learn features from all the combinations of those from the previous layer. However, the second one is limited with consistent features from a small repetitive field.

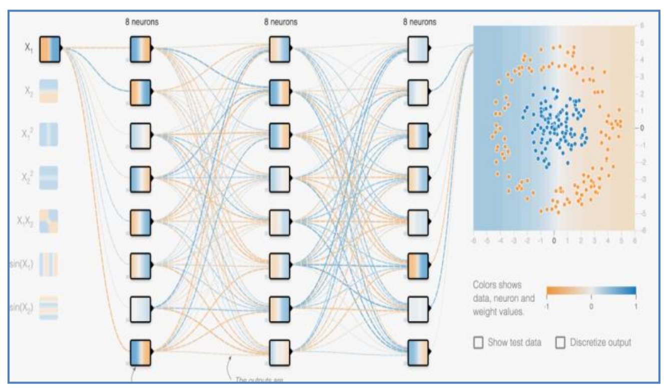


Fig. 1 Dense Neural Network (DNN) representation on Tensor Flow Playground [50].

Let's consider $x \in \mathbb{R}\, m$ the input to a fully connected layer and $y_i \in \mathbb{R}$ is the $i-th$ output from the fully-connected layer. The expression of $y_i \in \mathbb{R}$ is given such that

$$y_i = \sigma\,(w_1 x_1 + \cdots + w_m x_m) \quad (1)$$

Where σ denotes a nonlinear function and w_i represents the learnable parameters in the network.

Thus, the full output y is given as follows:

$$y = \sigma\left(w_{1,1}x_1 + \cdots + w_{1,m}x_m\right) : \sigma\left(w_{n,1}x_1 + \cdots + w_{n,m}x_m\right) \quad (2)$$

In our work, we have used the FC-DNN composed of two hidden layers with 50 fully connected neurons for the first

layer and 25 for the second one. Indeed, the first hidden layer takes the vocal features as input, and then passes them through the subsequent layers. After that, the performance of the network is evaluated using a loss function. To improve the knowledge of the network, an optimizer is explored as a metric to estimate the performance during the learning phase by increasing the network weights and therefore decreasing the loss. In our model, we have used the mean square error (MSE) as a loss function and the Stochastic Gradient Descendent (SGD) as an optimizer. We have also explored the classical sigmoid function σ as an activation function.
To avoid the over-fitting of the gradient, adding constraints to the network weights is the standard technique for dealing with this problem. Indeed, the size of the network will be forced by these constraints to take only small values and the loss function error will be therefore calculated by adding these values. In this model, we have also used the kind of regularization called “Lasso” for which the cost was calculated as the absolute value of the weight coefficients. Furthermore, the dropout technique was applied in order to randomly set some weights to zero. For the dropout rate, it was set to 0.05.
Concerning the training of the fully connected deep neural network (FC-DNN), it was effectuated using Tensor Flow. In fact, the computation of each gradient was performed on a small chunk of data referred to as mini batch (typically 50-500 data points) because it is not feasible to compute the gradients on a large dataset entirely at each step. So, the mini batches were employed in order to speed up the gradient descendent. For the learning rate and the number of iterations, they were set to 0.01 and 100000, respectively. As regards to the size of batch, it was set to 50.
Indeed, the general architecture of the proposed classifier (FC-DNN) is shown in Fig.2.

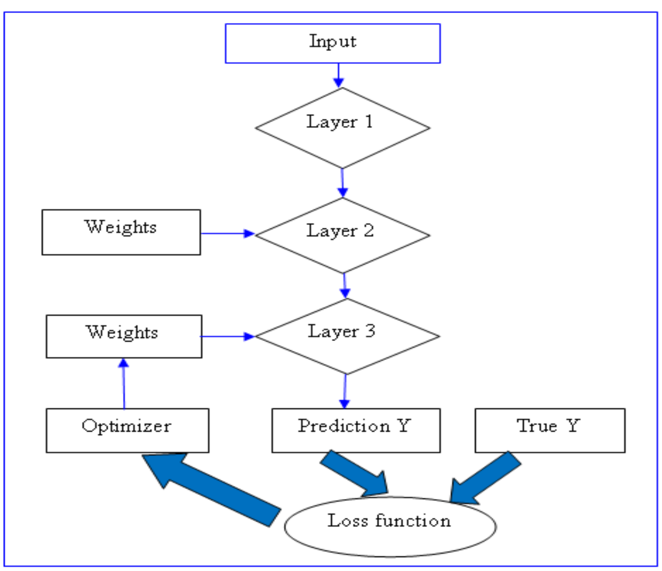


Fig. 2 The general architecture of FC-DNN classifier.

To compile our model, we have explored Google Colab framework with a GPU and an 8GB-RAM plus a processor of 1.2 GHz.

3. Database

3.1 Spanish Corpus

This dataset is composed of several voice recordings from 100 people: 50 healthy people and 50 people with PD. Indeed, it is divided between two groups of men and women each one includes 25 individuals. All participants were Colombian Spanish native speakers. Concerning the age range of each group, it is varied from 33 to 73 (mean 62.2 +_11.2) years old for men and from 44 to 75 (mean 60.1 +_7.8) years old for women. As regards to healthy controls, the age of men is ranged from 31 to 86 (mean 61.2 +_11.3) years old, while that one for women is ranged from 43 to 76 (mean 60.7 +_7.7) years old. It should be additionally mentioned that the participants in healthy controls do not suffer from any other neurological disease. Moreover, all sounds in this database are recorded under noise controlled-conditions using a dynamic Omni-directional microphone at a sampling rate of 44100 Hz and a 16-bits resolution. Furthermore, the diagnosis of all patients is established by neurology experts on the basis of UPDRS and H&Y scales. More details about the values of these scales and the age of each intervenient can be found in [51].

4. Evaluation Metrics

To evaluate the predictability of different performances for the explored classifiers, assessment metrics are required. Indeed, the accuracy is considered as one of the most frequently used metric, but it can produce misleading results when using data from unbalanced class distribution. Also, there are other evaluation metrics, such as F-measure and Matthews Correlation (MC) that can be used to measure how well a class can discriminate among different classes. In fact, these metrics can be applied even in the case of class unbalance.
In our experiments, we have evaluated the performance of the tested classifiers using recognition accuracy, specificity and sensitivity, which are obtained from the confusion matrix. This latter was explored to predict the success via four parameters -True Positive (tp), True Negative (tn), False Positive (fp) and False Negative (fn). Indeed, the recognition accuracy, sensitivity and sensibility are determined using the above-mentioned parameters and they are expressed respectively such that:

$$Accuracy = \frac{tp + tn}{fn + fp + tp + tn} * 100\% \qquad (3)$$

$$Specificity = \frac{tn}{tn + fp} * 100\% \qquad (4)$$

$$Sensitivity = \frac{tp}{tp + fn} * 100\% \qquad (5)$$

For F-measure, it is computed such that:

$$F - measure = \frac{2 \times precision \times recall}{precision + recall} \qquad (6)$$

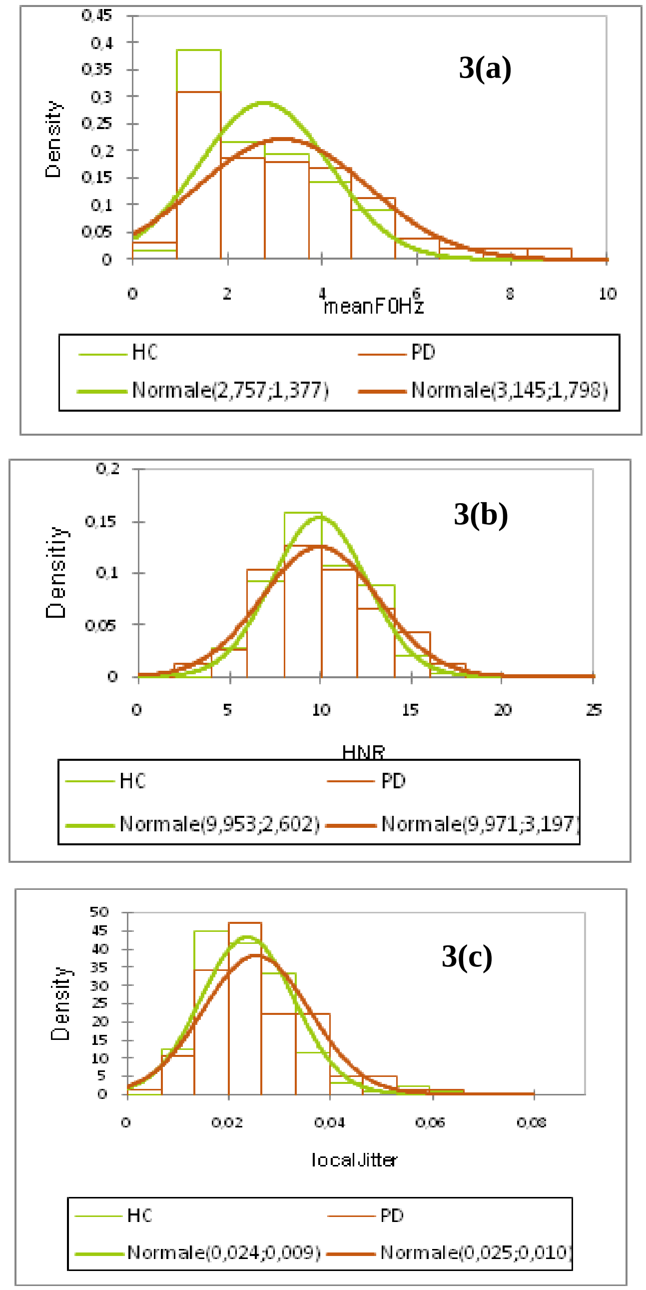
In order to quantify the quality of the binary classification, Matthews Correlation Coefficient (MCC) was also used as another metric. In general, this metric was considered as a balanced measure which can be employed even in the case of unbalanced class distribution. Fundamentally, MCC can be defined as a correlation coefficient between the predicted and the actual instances and it can take a value between +1 and -1. Indeed, a value of +1 denotes a perfect prediction, whereas that one of -1 indicates the disagreement between actual and predicted labels.
Indeed, the expression of MCC is given such that:

$$MCC = \frac{tp * tn - fn * fp}{\sqrt{(fn + tp)(fp + tn)(fp + tp)(fn + tn)}} \qquad (7)$$

5. Experimental Results

5.1 Data Analysis

The histograms of the extracted features for control and Parkinson speakers are shown in Fig.3 (a), (b), (c), (d) and (e). We can observe from these figures that the values of means and standard deviations (Std) of the visualized features are closed for both control and Parkinson people. Furthermore, we can see from Fig.3(c) that the local jitter has the best normal distribution for control and Parkinson people with comparison to other features. Indeed, this feature has achieved the best values of means and the minimum ones of standard deviations (Std) for both PD and HC (HC: means=0.024 and Std=0.009; PD: means=0.025, Std=0.010). For the second best normal distribution, it was reached by local shimmer (as shown in Fig.3(d)) by achieving the second best values of means and the second lowest ones of standard deviations (Std) for both PD and HC people (HC: means=1.241 and Std=0.229; PD: means=1.290, Std=0.262). As regards to the worst normal distribution for both PD and HC, it was reached using duration feature by achieving the highest values of means (HC: means=187.658, PD: means=160.788) and the maximum ones of standard deviations (HC: Std=42.968, PD: Std=30.499) as shown in Fig.3(e).



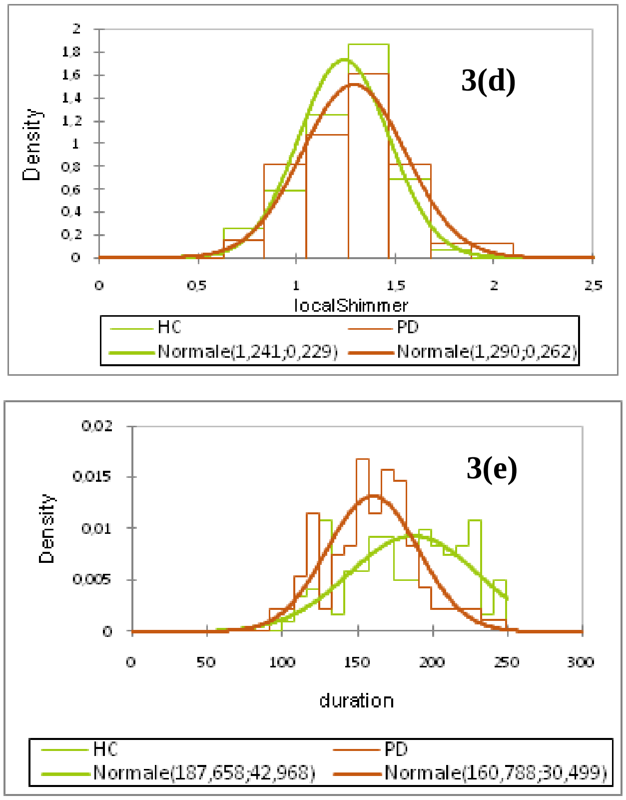


Fig. 3 Histograms of some extracted features for control and Parkinson speakers.

For the quartiles of the extracted features, they are reported in Fig.4. Indeed, a two-tailed t-test shows that the averages of the values of the extracted features for control and Parkinson speakers do not differ significantly. Also, we can say that both duration and HNR features were the main acoustic indices which were sufficiently sensitive for the early separation of PD from HC for the task of DDK evaluation. However, the other acoustic indices (like Mean F0, Local jitter and Local shimmer) were less sensitive for the same task because of their weakness to capture deficits in syllable repetitions and the reduction in their ratios due to PD.

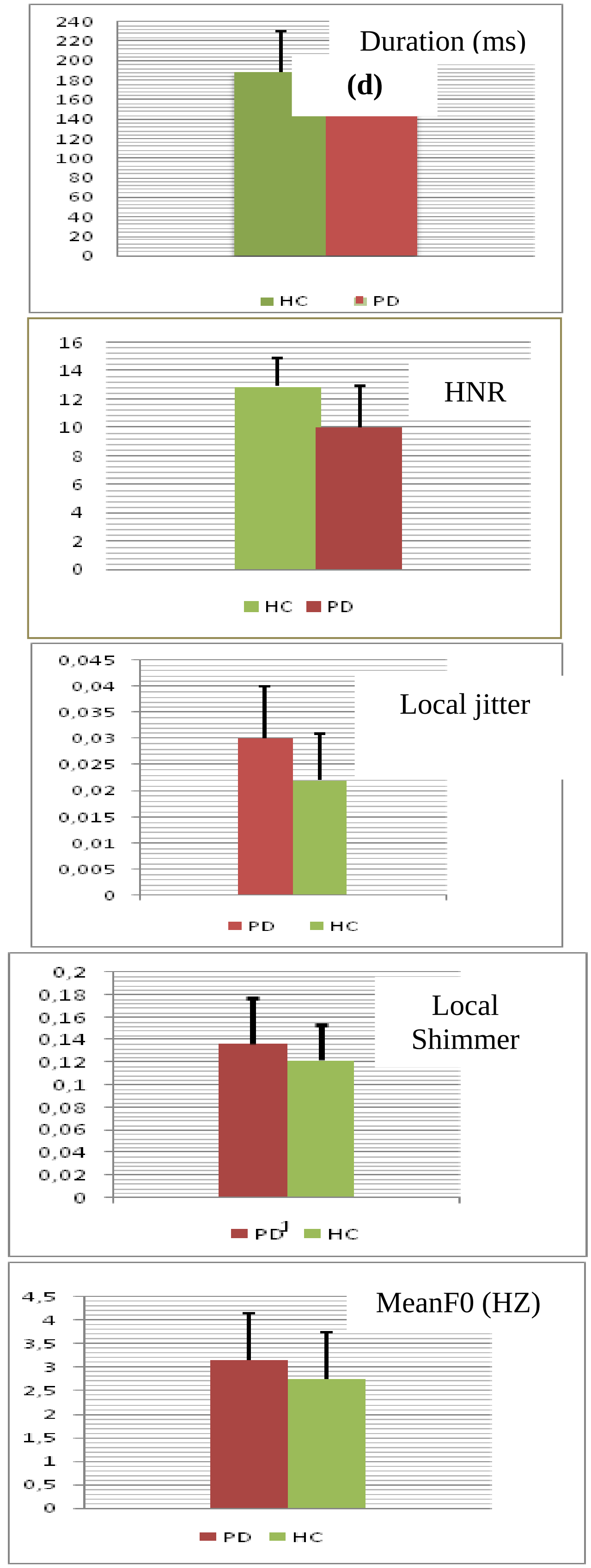


Fig. 4 Means and standard deviations for the extracted features for PD and HC people for the task of DDK evaluation (bars indicate means values and error bars standard deviations). HC= Healthy control, PD= Parkinson Disease.

In order to report the variability of the per-frame values of the extracted features over the length of an utterance, the quartiles of the standard deviations of the per-frame values of these features (in dB) are given in Table 2 for all speakers. We can remark from this table that the standard deviations of the per-frame values of the extracted features of PD and HC speakers are situated in the same interval. Also, the two-tailed t-test has demonstrated that they do not differ significantly between both PD and HC groups.

Table 2: Quartiles and mean of the standard deviations of the per-frame extracted feature values over the length of an utterance in dB for control and PD speakers

Feature	***Category***	***Min***	***Quartille1***	***Median***	***Quartille3***	***Max***
duration	HC	107	156	190	218	304
	PD	94	140	163	179	243
Mean F0	HC	0.76	1.5	2.4	3.7	6.3
	PD	0.8	1.7	2.9	4.5	9.1
HNR	HC	4.5	8.1	9.7	11.08	17.04
	PD	3.085	7.7	9.8	12.1	19.07
Jitter	HC	0.009	0.017	0.022	0.029	0.06
	PD	0.006	0.018	0.024	0.032	0.065
Shimmer	HC	0.60	1.09	1.28	1.40	1.73
	PD	0.72	1.1	1.3	1.45	1.94

5.2 Results

For the results of different performances obtained with the tested classifiers and using all vocal features in the dataset, they are shown in Table 3. Indeed, FC-DNN classifier has achieved the best performances in terms of all metrics (accuracy=100%, precision=0.991, recall=0.99, F-Measure=0.991 and MCC=0.95) using 10-fold cross-validation with comparison to all tested classifiers. For the second best performances, they were recorded using XGBoost classifier (accuracy=91%, precision=0.91, recall =0.90, F-Measure =0.90 and MCC=0.79, respectively). As regards to the third best results of these performances, they were reached by Linear discriminator analysis (LDA) (accuracy=87%, precision=0.89, recall=0.88, F-Measure=0.88 and MCC=0.75). Nevertheless, the worst results of these performances were obtained using Support Vector Clustering (SVC) (accuracy =77%, precision=0.77, recall=0.78, F-measure=0.77 and MCC=0.46). Also, we can remark that the FC-DNN classifier has shown its superiority as a deep learning algorithm with comparison to those of machine learning algorithm (for example Multi-layer perceptron (MLP) classifier). Moreover, we can summarize that the results of the evaluated performances, especially those achieved using FC-DNN classifier for MCC metric (MCC=0.95) have obviously demonstrated the strength of this classifier for the discrimination between healthy people and PD patients. This record was also achieved in virtue of the high efficiency of the selected features which have proved to be so pertinent in this discrimination.

To ensure about the efficiency of the proposed classifier (FC-DNN) for the task of DDK evaluation, we have compared the obtained results to those reached with other methods using the same dataset (as shown in Table 4). From Table IV, we can see clearly that the combination of unvoiced features with SVM classifier have contributed to reach the best performances for DDK evaluation using Spanish, German and Czech languages with comparison to other classic methods like GMM-UBM, Noise+F1F2+MFCC and prosody. Furthermore, the best results using this combination were reached with Spanish language (accuracy=99±3.2%, sensitivity=99±0.0% and specificity=98.6±6.3) with comparison to other tested languages. Comparing to the best method in [52], we can say that the proposed model in this work composed of 14 vocal features and FC-DNN classifier has led to outperform successfully that method for the task of DDK evaluation.

Table 3: The results of different performances obtained with different classifiers using 10-fold Cross-Validation

	Accuracy (%)	***Precision***	***Recall***	***F-measure***	***MCC***
Support Vector Clustering (SVC)	77	0,77	0,78	0,77	0,46
Linear Regression (LR)	78	0,77	0,78	0,78	0,49
K-Nearest Neighbors (KNN)	82	0,84	0,82	0,81	0,63
Decision Tree (DT)	83	0,84	0,83	0,82	0,65
Multi Layer Perceptron (MLP)	85	0,86	0,85	0,85	0,69
Linear	87	0,89	0,88	0,88	0,75

Discriminate Analysis (LDA)					
XGBoost	90	0,91	0,9	0,9	0,79
FC-DNN	**100**	**0.991**	**0.99**	**0.991**	**0.95**

Table 4: The different performances obtained for DDK evaluation using different recordings in Spanish, German and Czech languages [52]

		Accuracy (%)	***Sensitivity (%)***	***Specificity (%)***
Spanish	Noise+F1F2 +MFCC	80.6±9.4	90.61±4.1	70.61±9.4
	Prosody	80.6±6.7	88.61±3.9	72.61±3.9
	GMM-UBM	82.6±9.2	96.6±8.4	68.62±1.5
	Unvoiced	**99±3.2**	**99±0.0**	**98.6±6.3**
German	Unvoiced	**97.8±29**	**98.96±3.5**	**96.56±5.6**
Czech	Unvoiced	**93.66±16**	**99.36±2.7**	**87.96±31.4**

6. Conclusion and Further work

In this paper, we have proposed a model composed of Fully-connected deep neural network (FC-DNN) as a classifier and a set of vocal features for the task of DDK evaluation. The suggested approach has tested on Spanish dataset and it has led to reach good results in terms of the evaluated performances (accuracy=100%, precision=0.991, recall=0.99, F-Measure=0.991 and MCC=0.95). Also, the proposed approach has succeeded to outperform the state-of-the-art methods like that one composed of unvoiced features and SVM classifier.
To ensure about the real-time performances of the proposed model, we suggest as a further work to implement it on embedded architectures, such as raspberry Pi 4 and FPGA boards.

Acknowledgment

The authors would like to express their cordial thanks to Pr. Adnen Cherif for his valuable advice. Also, they want to address their warm salutations to the following authors for their corporation by offering them the opportunity to work on Spanish database. These authors are Orozco-Arroyave, J.R., Arias-Londoño, J.D., Vargas-Bonilla, J.F., González-Rátiva, M.C., and Nöth E.

References

[1] J. G. Goldman, S. K. Holden, I. Litvan, I. McKeith, G. T. Stebbins, & J. P. Taylor, “Evolution of diagnostic criteria and assessments for Parkinson’s disease mild cognitive impairment,”Movement Disorders., vol.33, pp.503–510, 2018.
[2] S. Ray, “A Predictive Diagnosis for Parkinson’s Disease Through Machine Learning”, The Canadian Science Fair Journal, Vol.2, pp.120-128, 2019.
[3] P. Rizek, N. Kumar, & M. S. Jog, “An update on the diagnosis and treatment of Parkinson disease,” Cmaj., vol.188, pp.1157–1165, 2016.
[4] S. Saria, & A. Zhan, measuring medication response using wearables for parkinson’s disease, Google Patents, 2018.
[5] Wong, D., & Yip, S. (2018). Machine learning classifies cancer. Nature, 555 (7697), 446–447.
[6] S. Mallya, M. Overhage, N. Srivastava, T. Arai, & C. Erdman, “Effectiveness of LSTMs in Predicting Congestive Heart Failure Onset,” arXiv.,vol.20, pp.1902.02443, 2019.
[7] K.Sakai, & K. Yamada, “Machine learning studies on major brain diseases: 5-year trends of 2014–2018,” Japanese Journal of Radiology., vol.37, pp.34–72,2019.
[8] J. De Fauw, J. R. Ledsam, B. Romera-Paredes, S. Nikolov, N. Tomasev, S. Blackwell,O. Ronneberger, “Clinically applicable deep learning for diagnosis and referral in retinal disease,” Nature medicine., vol.24, pp.1342– 1350, 2018
[9] S. H. Park, “Regulatory Approval versus Clinical Validation of Artificial Intelligence Diagnostic Tools,” Radiology.,vol.288, pp.910–911, 2018.
[10] M. K. Sugai, S. Nomura, S. Gilmour, G. A. Stevens, & K. Shibuya, “Demographic and clinical factors associated with having ischemic heart disease as a multiple contributing causes of death among diabetes mellitus deaths in the united states and brazil,” Endocrine Abstracts vol.56, pp.84, 2018.
[11] A. Rojas, J. Górriz, J. Ramírez, I. Illán, F. Martínez-Murcia, A. , Ortiz, et al., “Application of empirical mode decomposition (emd) on
[12] datscan spect images to explore Parkinson disease, “Exp. Syst. Appl., Vol.40, pp.2756–2766, 2013.
[13] F. J. Martínez-Murcia, J. M. Górriz, J. Ramírez, I. A. Illán, and A. Ortiz, “Automatic detection of parkinsonism using significance measures and component analysis in datscan imaging,” Neurocomputing.,vol.126, pp.58–70, 2014.
[14] F. Martinez-Murcia, J. Górriz, J. Ramírez, A. Ortiz,“ for the Alzheimer’s Disease Neuroimaging Initiative. A spherical brain mapping of MR images for the detection of Alzheimer’s disease,” Curr. Alzheimer Res., Vol.13, pp.575–588, 2016.
[15] L. Khedher, J. Ramírez, J. Górriz, A. Brahim, and F. Segovia, “Early diagnosis of disease based on partial least squares, principal component analysis and support vector machine using segmented MRI images,” Neurocomputing., vol.151, pp.139–150, 2015.
[16] C. R. Pereira, D. R. Pereira, F. A. d. Silva, C. Hook, S. A. T. Weber, L. A. M. Pereira, et al., “A step towards the automated diagnosis of parkinson’s disease: analyzing handwriting movements,” in 2015 IEEE 28th International Symposium on Computer-Based Medical Systems (Sao Carlos),vol.12, pp.171–176, 2015.
[17] S. Badoud, D. V. D. Ville, N. Nicastro, V. Garibotto, P. R. Burkhard, and S. Haller, “Discriminating among

degenerative parkinsonisms using advanced 123i-ioflupane spect analyses," NeuroImage.,vol.12, pp.234–240, 2016.
[18] H. D. Tagare, C. DeLorenzo, S. Chelikani, L. Saperstein, and R. K. Fulbright, "Voxel-based logistic analysis of PPMI control and Parkinson's disease DaTscans,"NeuroImage,vol. 152, pp.299–311, 2017.
[19] D. J. Towey, P. G. Bain, andK. S. Nijran, "Automatic classification of 123I-FP-CIT (DaTSCAN) SPECT images," Nucl. Med. Commun.,vol. 32, pp.699–707, 2011.
[20] Y. LeCun, Y. Bengio, and G. Hinton, " Deep learning," Nature, vol.521, pp.436–444, 2015.
[21] G. E. Hinton, L. Deng, D. Yu, G. E. Dahl,A. rahman Mohamed, N. Jaitly, et al, "deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups," IEEE Signal Process. Mag., vol.9, pp2–97, 2012.
[22] B. Alipanahi, A. Delong, M. T. Weirauch, and B. J. J Frey, Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning, 2015.
[23] H. Chen, O. Engkvist, Y. Wang, M. Olivecrona, and T. Blaschke, "The rise of deep learning in drug discovery," Drug Discov. Today.,vol.23, pp.1241–1250, 2018.
[24] S. R. Kheradpisheh, M. Ghodrati, M. Ganjtabesh, and T. Masquelier, " Deep networks can resemble human feed-forward vision in invariant object recognition", Sci. Rep.,vol.6, pp.32672, 2016.
[25] K. He, X..Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (Las Vegas, NV)., vol.10, pp.770–778, 2016.
[26] M. A. Little, P. E. McSharry, E. J. Hunter, J. Spielman, and L. O. Ramig, "Suitability of dysphonia measurements for telemonitoring of Parkinson's disease," IEEE Trans. Biomed. Eng., vol. 56,pp. 1015_1022, 2009.
[27] B. E. Sakar, M. E. Isenkul, C. O. Sakar, A. Sertbas, F. Gurgen, S. Delil, H. Apaydin, and O. Kursun, "Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings," IEEE J. Biomed. Health Inform., vol. 17, pp. 828_834, 2013.
[28] A. Tsanas, M. A. Little, P. E. McSharry, J. Spielman, and L. O. Ramig, "Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease," IEEE Trans. Biomed. Eng., vol. 59, pp. 1264_1271, 2012.
[29] C. O. Sakar, G. Serbes, A. Gunduz, H. C. Tunc, H. Nizam, B. E. Sakar, M. Tutuncu, T. Aydin, M. E. Isenkul, and H. Apaydin, "A comparative analysis of speech signal processing algorithms for Parkinson's disease classi_cation and the use of the tunable Q-factor wavelet transform,"Appl. Soft Comput., vol. 74, pp. 255-263, 2019.
[30] H. K. Rouzbahani and M. R. Daliri, "Diagnosis of Parkinson's disease in human using voice signals," Basic Clin. Neurosci., vol. 2, pp. 12-20, 2011.
[31] A. Sharma and R. N. Giri, "Automatic recognition of Parkinson's disease via arti_cial neural network and support vector machine," Int. J. Innov. Technol. Exploring Eng., vol. 4, pp. 2278-3075, 2014.
[32] L. Parisi, N. RaviChandran, and M. L. Manaog, "Feature-driven machine learning to improve early diagnosis of Parkinson's disease,"Expert Syst. Appl., vol. 110, pp. 182-190, 2018.
[33] C. O. Sakar, G. Serbes, A. Gunduz, H. C. Tunc, H. Nizam, B. E. Sakar, M. Tutuncu, T. Aydin, M. E. Isenkul, and H. Apaydin, "A comparative analysis of speech signal processing algorithms for Parkinson's disease classi_cation and the use of the tunable Q-factor wavelet transform,"Appl. Soft Comput., vol. 74, pp. 255-263, 2019.
[34] B. M. Esko_er, S. I. Lee, J.-F. Daneault, F. N. Golabchi, G. Ferreira-Carvalho, G. Vergara-Diaz, S. Sapienza, G. Costante, J. Klucken, and T. Kautz, "Recent machine learning advancements in sensor-based mobility analysis: Deep learning for Parkinson's disease assessment," in Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC), vol.12., pp. 655-658, 2016.
[35] S. L. Oh, Y. Hagiwara, U. Raghavendra, R. Yuvaraj, N. Arunkumar, M. Murugappan, and U. R. Acharya, "A deep learning approach for Parkinson's disease diagnosis from EEG signals," Neural Comput. Appl., pp. 1-7, 2018.
[36] C. R. Pereira, S. A. T. Weber, C. Hook, G. H. Rosa, and J. P. Papa, "Deep learning-aided Parkinson's disease diagnosis from handwritten dynamics,"in Proc. IEEE 29th SIBGRAPI Conf. Graph., Patterns Images (SIBGRAPI)., vol.12, pp. 340-346, 2016.
[37] A. Caliskan, H. Badem, A. Basturk, and M. E. Yuksel, "Diagnosis of the Parkinson disease by using deep neural network classifier," Istanbul Univ.- J. Elect. Electron. Eng., vol. 17, pp. 3311-3318, 2017.
[38] T. J. Wroge, Y. Ö zkanca, C. Demiroglu, D. Si, D. C. Atkins, and R. H. Ghomi, "Parkinson's disease diagnosis using machine learning and voice," in Proc. IEEE Signal Process. Med. Biol. Symp. (SPMB)., vol.3, pp. 107, 2018.
[39] H. GUNDUZ, "Deep Learning-Based Parkinson's Disease Classification Using Vocal Feature Sets," IEEE Access journal., Vol.7, 2019.
[40] I. Luong M-T, L.Sutskever, O. Vinyals, W. Zaremba, "Addressing the rare word problem in neural machine translation," In Association for Computational Linguistics (ACL)., vol.57, pp.11–19, 2015.
[41] Sak H, Senior A, Beaufays F, "Long short-term memory recurrent neural network architectures for large scale acoustic modeling", In International Speech Communication Association,.vol.17, pp.338–342, 2014.
[42] E. Grosicki, H. El Abed, "ICDAR handwriting recognition competition," In International Conference on Document Analysis and Recognition.,vol.79,pp.1398–1402, 2009.
[43] W. Zaremba, I. Sutskever, O. Vinyals, "Recurrent neural network regularization," In arXiv preprint arXiv., vol.140, pp.2329, 2014.
[44] K.Cho, B. Van Merrienboer, C. Gulcehre, et al., "Learning phrase representations using RNN encoder-decoder for statistical machine translation," In Empirical Methods in Natural Language Processing (EMNLP)., vol.7, pp.1724–1734, 2014.
[45] E.Choi, A.Schuetz, W.F Stewart, and J. Sun, "Using recurrent neural network models for early detection of heart failure onset", Journal of the American Medical Informatics Association, vol.24, pp.361–370, 2017.
[46] P. Malhotra, T. V. Vishnu, L. Vig, P. Agarwal, and G. Shroff, "TimeNet: Pre-trained deep recurrent neural network for time series classification," in Proc. 25th European Symp. Artificial Neural Networks, Computational Intelligence and

Machine Learning, Bruges, Belgium., vol.12, pp. 607–612, 2017.
[47] X. J. Shi, Z. R. Chen, H. Wang, D. Y. Yeung, W. K. Wong, and W. C. Woo, “Convolutional LSTM network: a machine learning approach for precipitation nowcasting,” in Proc. 28th Int. Conf. Neural Information Processing Systems, Montreal, Canada.,vol.87, pp. 802–810, 2015.
[48] Y. Jadoul, B. Thompson, and B. De Boer, “Introducing Parselmouth: A Python interface to Praat,” Journal of Phonetics., vol. 71, pp.1–15, 2018.
[49] G.Al-Bdour, R.Al-Qurran, M.Al-Ayyoub, A.Shatnawi, “A detailed comparative study of open source deep learning frameworks”, arXiv:1903.00102v1 [cs.LG].,vol.25, pp.2101-2110, 2019.
[50] A. Grigorevich Ivakhnenko and V. G. Lapa, Cybernetic predicting devices, Technical report, DTIC Document, 1966.
[51] https://heartbeat.fritz.ai/classification-with-tensorflow-and-dense-neural-networks 8299327a818a.
[52] J.R. Orozco-Arroyave, J.D. Arias-Londoño, J.F. Vargas-Bonilla, M.C. González-Rátiva, and E.Nöth., “New Speech Corpus Database for the Analysis of People Suffering From Parkinson’s Disease,” In Proc. Of the International Conference on Language Resources and Evaluation (lrec).,vol103, pp.342-347, 2014.
[53] J. R. Orozco-Arroyave, F. Hönig, J. D. Arias-Londoño and J. F. Vargas-Bonilla, K. Daqrouq, S. Skodda, J. Rusz, E.NöthCYE, “Automatic detection of Parkinson's disease in running speech spoken in three different,” J. Acoust. Soc. Am., vol.39, 2016.