International Journal of Electrical and Computer Engineering (IJECE)
Vol. 13, No. 3, June 2023, pp. 3204~3216
ISSN: 2088-8708, DOI: 10.11591/ijece.v13i3.pp3204-3216

Parkinson's diagnosis hybrid system based on deep learning classification with imbalanced dataset

Asmae Ouhmida[1], Abdelhadi Raihani[1], Bouchaib Cherradi[1,2], Sara Sandabad[1]
[1]EEIS Laboratory, ENSET Mohammedia, Hassan II University of Casablanca, Mohammedia, Morocco
[2]STIE Team, CRMEF Casablanca-Settat, Provincial section of El Jadida, El Jadida, Morocco

Article Info

Article history:

Received Apr 27, 2022
Revised Sep 16, 2022
Accepted Oct 1, 2022

Keywords:

Acoustic dataset
Classification
Deep learning
Feature selection
Parkinson's disease
Synthetic minority oversampling technique

ABSTRACT

Brain degeneration involves several neurological troubles such as Parkinson's disease (PD). Since this neurodegenerative disorder has no known cure, early detection has a paramount role in improving the patient's life. Research has shown that voice disorder is one of the first symptoms detected. The application of deep learning techniques to data extracted from voice allows the production of a diagnostic support system for the Parkinson's disease detection. In this work, we adopted the synthetic minority oversampling technique (SMOTE) technique to solve the imbalanced class problems. We performed feature selection, relying on the Chi-square feature technique to choose the most significant attributes. We opted for three deep learning classifiers, which are long-short term memory (LSTM), bidirectional LSTM (Bi-LSTM), and deep-LSTM (D-LSTM). After tuning the parameters by selecting different options, the experiment results show that the D-LSTM technique outperformed the LSTM and Bi-LSTM ones. It yielded the best score for both the imbalanced original dataset and for the balanced dataset with accuracy scores of 94.87% and 97.44%, respectively.

Corresponding Author:

Asmae Ouhmida
Electrical Engineering and Intelligent Systems Laboratory, ENSET Mohammedia, Hassan II University of Casablanca
159 Hassan II Avenue, Mohammedia, Morocco
Email: asmaeouhmida1995@gmail.com

1. INTRODUCTION

Parkinson's disease is the second slowest neurodegenerative disease in terms of progression, after Alzheimer's disease [1]. This neurological disorder affects about seven to ten million people worldwide [2]. The loss of nerve cell function in the brain, specifically in the substantia nigra, over time is the principal cause of Parkinson's. Neurons produce a brain chemical called neurotransmitter. This substance is less produced when neurons begin to be damaged or die [3]. Research has shown that although Parkinson's disease (PD) cannot be cured, early diagnosis can prevent a severe condition and allow for appropriate treatment [4], [5]. A variety of motor and non-motor symptoms characterize PD; they develop slowly over time [6]. Resting tremors, bradykinesia and rigidity are some of the common PD symptoms, which affect balance, hand movement, gait, and speech [7].

It is worth mentioning that the clinical assessment represents the professional health care of PD; it is the basis of the identification of PD, and it is important for monitoring the patient's condition. In order to facilitate the diagnosis of PD and other diseases, the application of Artificial Intelligence in the diagnosis and telemedicine field has made tremendous progress [8]. Various tools have been used to simplify the detection

of PD, such as electromyography (EMG) signals, freezing of gait (FoG), handwritten images recognition [9], [10], electroencephalography (EEG) signals, and magnetic resonance imaging (MRI) images.

Around 90% of patients with Parkinson's disease suffer from vocal trouble problems, and they are detected earlier than other symptoms. These disorders in voice can be difficulty in articulation (dysarthria), reduced pitch (monotone), reduced volume (hypophonia), or defective volume (dysphonia) [11]. Whereof, for our study, we chose changes that affect speech to diagnose Parkinson's disease.

In the literature, several classification techniques have been used to identify people with Parkinson's disease [12]. In this context, deep learning models can make decisions much faster and medical datasets can analyze the disease readily [13]. However, for better accuracy, these classifiers require a balanced class distribution during the training phase to avoid biased learning towards the majority classes at the expense of the minority classes. Thus, data normalization and balancing techniques are the first steps in the preprocessing phase.

In general, class imbalance is a global data problem suffered by several real-world fields, such as remote sensing [14], face recognition [15], and text mining [16]. Regarding binary dataset classification, the imbalanced class problem takes place when the number of negative samples (majority class) exceeds the number of positive class samples (minority class) [17], [18]. This underrepresentation of the minority class in the dataset reduces the classification accuracy. Therefore, one of the challenges is how to deal with imbalanced data in the classification domain. Moreover, dealing with all the features in dataset makes the classification problem cumbersome and costly, especially since the large size of the features can hamper the progress of the prediction model. That is why it is often useful and essential to perform a selection of the most relevant features in the preprocessing phase. In this paper, we propose an approach to improve the performance of PD classification with an imbalanced dataset. It attempts to address resampling methods to deal with imbalanced classes and obtain a good classification score. Three deep learning algorithms are studied to choose the most efficient one. Overall, the main contributions of the proposed work are summarized in 3 main objectives:

- Objective 1: To highlight the main impact of applying the synthetic minority oversampling technique (SMOTE) technique to imbalanced data to generate a balanced distribution of classes.
- Objective 2: To prove the role of the Chi-square feature selection method in the pre-processing step, and its effect on improving the classification models by considering only the most relevant and informative features.
- Objective 3: To show the effectiveness of deep learning models and their comparison in order to highlight their efficiency and select the most powerful classification algorithm.

In the next section, the literature review related to the Parkinson's disease detection is presented. In section 3, the dataset used and its visualization are presented as well as the proposed methodology. Section 4 introduces a description of the architecture of the proposed system. The experiment setting and results are discussed in section 5. The conclusion and future prospects are outlined in section 6.

2. LITTERATURE REVIEW

This section summarizes the most related works in the field of Parkinson's disease prediction. It also reviews some works dealing with the SMOTE method to increase the minority instances to reach the majority ones. In this context, various oversampling and undersampling methods and their combinations are analyzed in [19]. The authors applied those methods to the detection of Alzheimer's disease, including K-medoids and random undersampling, SMOTE, and random oversampling techniques. Their experiments proved that K-medoids and SMOTE sampling methods are more technical and perform greater than the random oversampling and undersampling techniques.

Sharma *et al.* [20] illustrated the weak performance of the SMOTE technique of oversampling in case of insufficient minority class instances. The information provided by the minority patterns is limited and not enough to produce representative instances. Thereby, they established a new method named sampling with the majority (SWIM) for oversampling. Based on abundant majority instances, the SWIM technique determines the approximative distribution of the majority class. And using the same Mahalanobis distance as the primary minority patterns, minority patterns are produced. In the PD classification area, many works have studied the SMOTE technique to handle imbalanced datasets.

Zeng *et al.* [21] presented their preprocessing method based on the combination of the Tomek links technique and SMOTE algorithm. They applied this combination to three imbalanced datasets, which are vertebral column pathologies, Parkinson's disease, and diabetes datasets. Their model consisted of 8 classifiers and to evaluate its efficiency, they compared the use of the SMOTE algorithm alone to the experimental results. It appeared that the application of the combination of algorithms yielded better performances for the three diseases compared to the application of the SMOTE algorithm alone. Alqahtani

et al. [22] presented their experiment to diagnose Parkinson's disease. They detected the speech disorder and analyzed voice signals based on NNge classification algorithms. Furthermore, the dataset used was balanced by applying the SMOTE technique. Eventually, the NNge algorithm with the optimized AdaBoost gives an accuracy of 96.30%. Different machine learning techniques are used to introduce a framework for automatic prediction of Parkinson's disease, and to build decision support systems [23]. The authors used SMOTE, gradient boosting and random forest algorithms to widen the training dataset. The oversampling evinces some significant improvements, with a prediction accuracy score of 91.5 %.

Grover *et al.* [24] used a deep neural network (DNN) to predict PD severity and studied the Unified Parkinson's disease rating scale (UPDRS) score. They applied the Total UPDRS and Motor-UPDRS scores to perform two experiments. As a result, the classification accuracy obtained using the Motor-UPDRS score was greater than the Total-UPDRS score accuracy, with a rate of 81.66%. The voice signals were used to present a PD detection system [25]. Recursive feature elimination (RFE) and minimum redundancy maximum relevancy (mRMR) techniques were applied in the feature selection phase and eight classifiers performed the classification. Thus, the best accuracy was achieved by combining XGBoost with RFE, with a rate of 95.39%. Gunduz [26] aimed to classify PD, by adopting two CNN-based frameworks. He used the UCI repository acoustic dataset. The proposed frameworks are model-level combinations and feature-level combinations. The highest score of accuracy, 86.9%, was obtained using the model level framework. In [27], the author used speech signals to create a hybrid model for the diagnosis of Parkinson's disease. The SMOTE technique was adopted in the preprocessing stage. The random forest classifier achieved a maximum accuracy of 94.8% during classification. Others have used an optimized version of the BAT classifier to present a Parkinson's disease diagnostic system [28]. They worked on the UCI Parkinson's disease classification dataset and they selected only 23 features. To create their model, the selected features feed 23 neurons in the input layer. The proposed model achieved an accuracy of 96.74% with a loss of 3.27%.

A nonlinear support vector machine (SVM) algorithm for classification and eight techniques of pattern ranking of feature selection was adopted in [29] in order to determine whether a patient has PD or not. The Wilcox rank method achieved the highest accuracy score of 92.91%. Wroge *et al.*[30] collected speech recordings and bio-markers from healthy and PD patients using a mobile application. They applied the mRMR feature selection technique to the features extracted from the speech signals, for the first set of features. Mel frequency cepstral coefficients (MFCC) were used as the second feature set. DNN, among other deep learning classifiers, was applied to both feature sets. In terms of accuracy, the DNN model outperformed the different models and achieved the highest rate of 85%. To sum up, Table 1 summarizes some of the research contributions to Parkinson's disease classification.

Table 1. State of art of different approaches in Parkinson's detection

Ref and year	N° of features	FS method	Method	SMOTE	Accuracy (%)
[22] 2018	22	--	NNge + AdaBoost	Yes	96.30
[23] 2021	754	--	Random Forest + Gradient Boosting	Yes	91.5
[24] 2018	16	--	DNN	No	81.66
[25] 2019	754	RFE/mRMR	XGboost	No	95.39
[26] 2019	Combination of three feature sets	Chi-square	CNN	No	86.9
[27] 2019	753	--	Random Forest	Yes	94.8
[28] 2020	23	--	BAT	No	96.74
[29] 2019	14	Wilcoxon-based pattern ranking technique	SVM	No	92.91
[30] 2018	1200	mRMR	DNN	No	85

3. MATERIALS AND METHODS

3.1. Dataset

In this research work, we use the dataset originally generated by Little *et al.* [31], for the diagnosis of Parkinson's disease, based on patients' speech signals. The study is based on voice recordings of 31 patients, which include 23 with Parkinson's disease and 8 healthy patients. The dataset contains 195 vocal recordings with 23 attributes. To assess whether a voice recording belongs to a PD person or a healthy person, measured 22 voice features is measured for 195 biomedical voice measurements. The discrimination is done according to the "status" column, which takes the value of 1 for PD and 0 for healthy patients. The 22 acoustic features are depicted in Table 2.

Figure 1 visualizes a histogram representing the acoustic features of the studied dataset. Each plot shows the frequency (y-axis) of unique values (x-axis) of a given feature. The heatmap, shown in Figure 2,

makes data visualization easier, and allows us to easily compare different features of the complex dataset. Thereby, the most relevant ones can be readily located through visual marks based on color intensity. The figure provides a correlation matrix of 22 attributes. A perfect correlation between features is specified by black color whilst no correlation is indicated between vocal attributes by white color.

Table 2. Acoustic features of Parkinson's dataset

Feature category	Vocal Feature	Description
Variation in fundamental frequency	MDVP: Jitter (abs)	Absolute jitter in microseconds
	MDVP: Jitter (%)	Fundamental frequency perturbation (%)
	MDVP: PPQ	Five-point period perturbation quotient
	MDVP: DDP	Average absolute difference of differences between cycles, divided by the average period
	MDVP: RAP	Relative amplitude perturbation
Variation in amplitude	MDVP: Shimmer	Shimmer local amplitude perturbation
	MDVP: Shimmer (dB)	Local amplitude perturbation (decibels)
	MDVP: APQ3	3-point amplitude perturbation quotient
	MDVP: APQ5	5-point amplitude perturbation quotient
	MDVP: APQ	11-point amplitude perturbation quotient
	MDVP: DDA	Average absolute difference between the amplitudes of consecutive periods
Vocal fundamental frequency	MDVP: Fo (Hz)	Average vocal fundamental frequency
	MDVP: Fhi (Hz)	Maximum vocal fundamental frequency
	MDVP: Flo (Hz)	Minimum vocal fundamental frequency
Ratio of noise to tonal components in the voice	NHR	Noise-to-harmonics ratio
	HNR	Harmonics-to-noise ratio
Non-linear measures of fundamental frequency variation	Spread 1	Fundamental frequency variation
	Spread 2	Fundamental frequency variation
	PPE	Pitch period entropy
Non-linear dynamical complexity measures	RPDE	Recurrence period density entropy
	D2	Correlation dimension
Signal fractal scaling exponent	DFA	Detrended fluctuation analysis

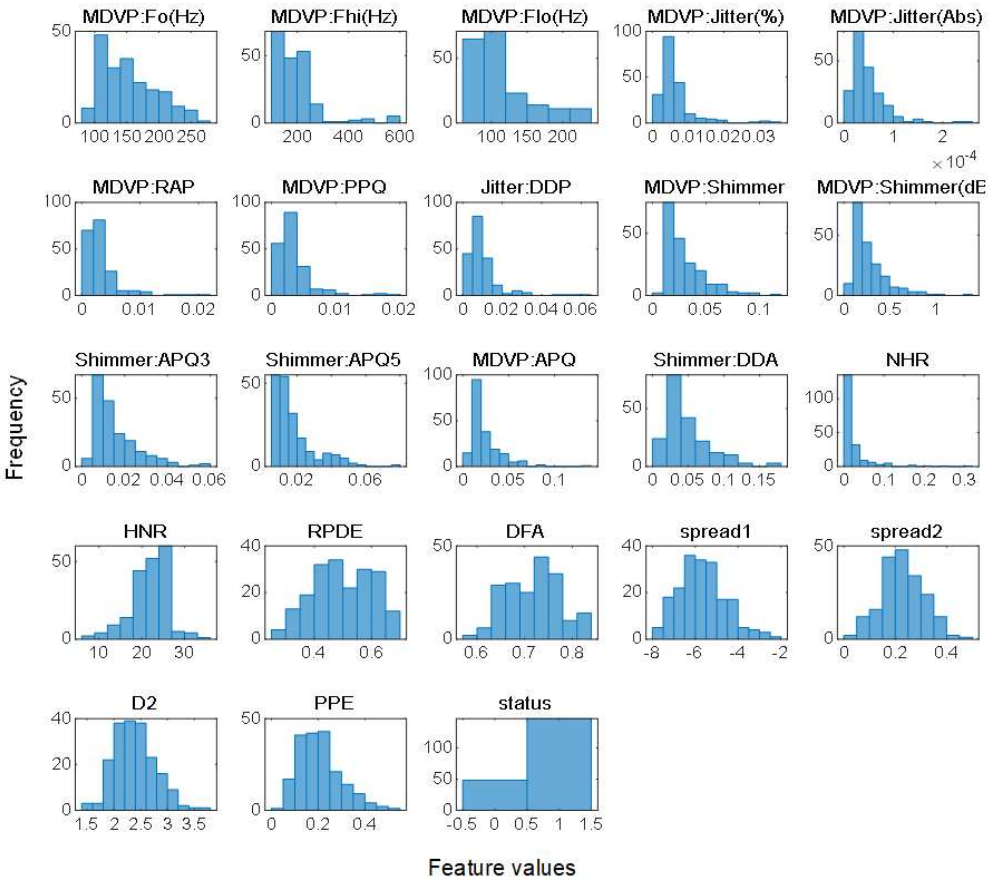

Figure 1. Histogram representation of the acoustic features

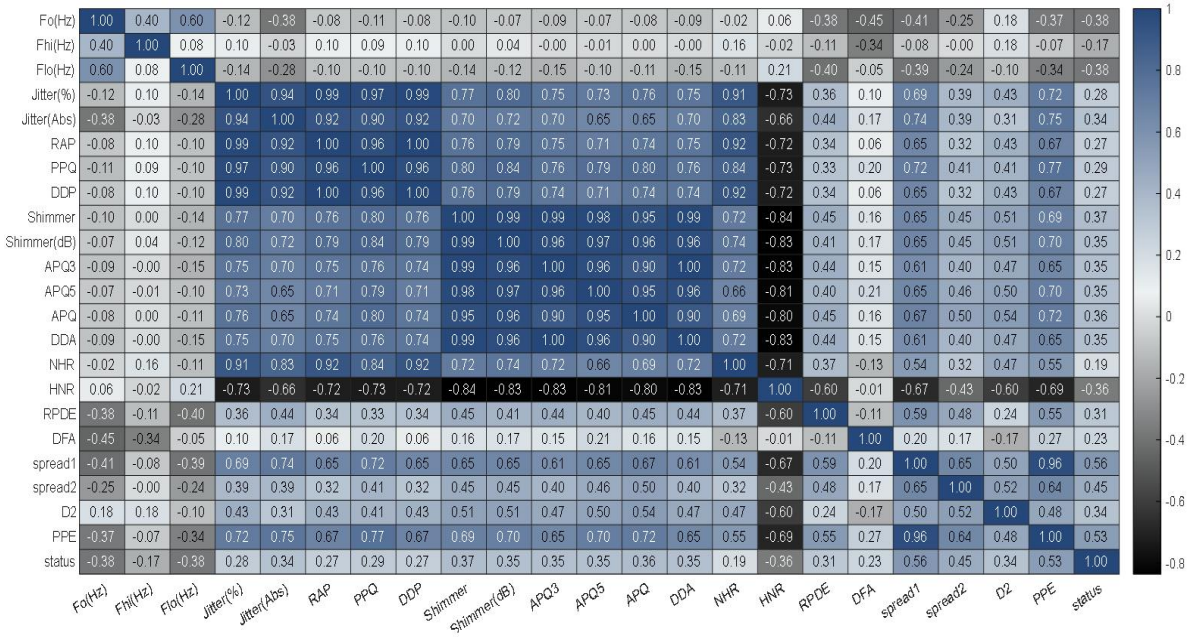

Figure 2. Heatmap of correlation between features

The coefficients indicate the correlation level. It is noticeable that HNR is very discriminating, it shows the lowest correlation with the majority of the other features. We can also notice that the vocal fundamental frequency attributes, namely *Flo*, *Fhi*, and *Fo* are reciprocally correlated, but denote a low correlation coefficient with the remaining features.

The proposed methodology aims to exploit a deep learning framework to identify individuals with and without Parkinson's disease. The key steps required to achieve this are: i) resampling techniques to handle imbalanced datasets in real-world classification problems; ii) implementing feature selection to find relevant features to build the best performing classifier; iii) obtaining an efficient and accurate deep learning classifier among LSTM, Bi-LSTM, and D-LSTM algorithms; and iv) presenting the evaluation criteria. These points are discussed extensively in the rest of this section.

3.2. Sampling methods for handling imbalanced data

Dealing with imbalanced classes of datasets in machine learning and/or deep learning in real classification applications is a significant challenge for scientists, as the learning model will be biased towards the majority class, which can significantly affect accuracy. There are several methods to learn from the imbalanced dataset. The most common, based on data augmentation technique, is known as SMOTE. Many authors approve that even random oversampling aims to create a balanced distribution randomly, it can make copies of already existing examples. Thus, the overfitting occurring probability can be increased [17].

The SMOTE technique, proposed by Chawla *et al.* [32], is the most popular oversampling method. It is used in several applications and has known large success [33]. It avoids the problem of overfitting. Its fundamental idea is to interpolate diverse minority class instances that are together, and then produce new minority class instances [34]. This process is based on finding similarities in the minority class feature space between existing samples [32]. It is important to note that in classification problems based on deep learning models, the SMOTE technique is only applied to the training subset in order to correctly train the classifier. The test subset remains unchanged so that it correctly represents the original data as illustrated by the diagram in Figure 3.

3.3. Feature selection

The feature selection process has a crucial role in the preprocessing phase since a large number of features in a dataset can produce a poor decision model, which will take more time for the training stage. Thus, it is essential to select the most relevant features from the dataset used. And most of the time, as consequence, this technique increases the performance of the decision model.

Chi-square is a statistical algorithm for feature selection. Its main role is to exclude some irrelevant features and leave the features that are most needed to correctly represent the data. These attributes improve

the performance of the classification algorithms and reduce the learning duration [35]. Thus, it is used to evaluate whether the class label and the feature are independent [36]. The formula to calculate the chi-square score with r values and C class is defined as:

$$x^2 = -\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(n_{ij}-\mu_{ij})}{\mu_{ij}} \quad (1)$$

$$\mu_{ij} = \frac{(n_{*j}n_{i*})}{n} \quad (2)$$

where, n_{ij} represents the number of samples value with the i^{th} value of the feature, n_{*j} represents the number of samples in class j, n_{i*} represents the number of samples with the ith feature value, and n represents the number of samples. Figure 4 visualizes the feature ranking calculated using the Chi-square technique.

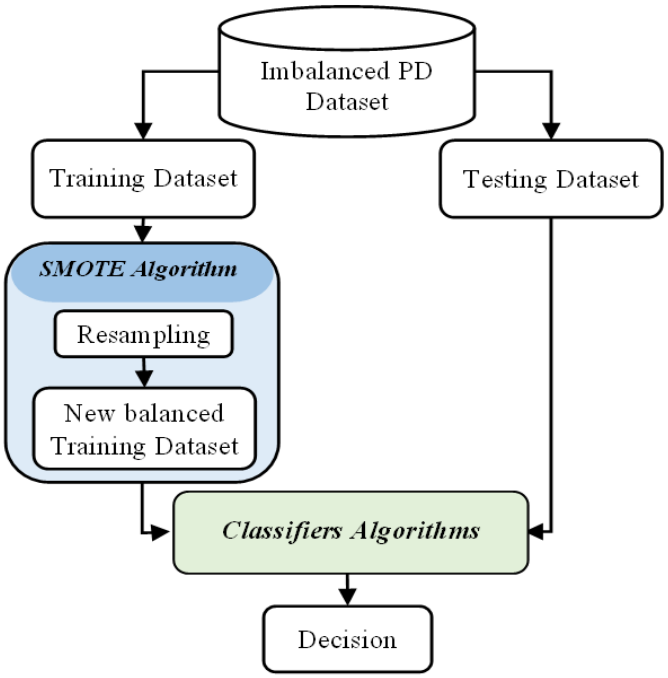

Figure 3. Diagnosis system architecture based on SMOTE technique

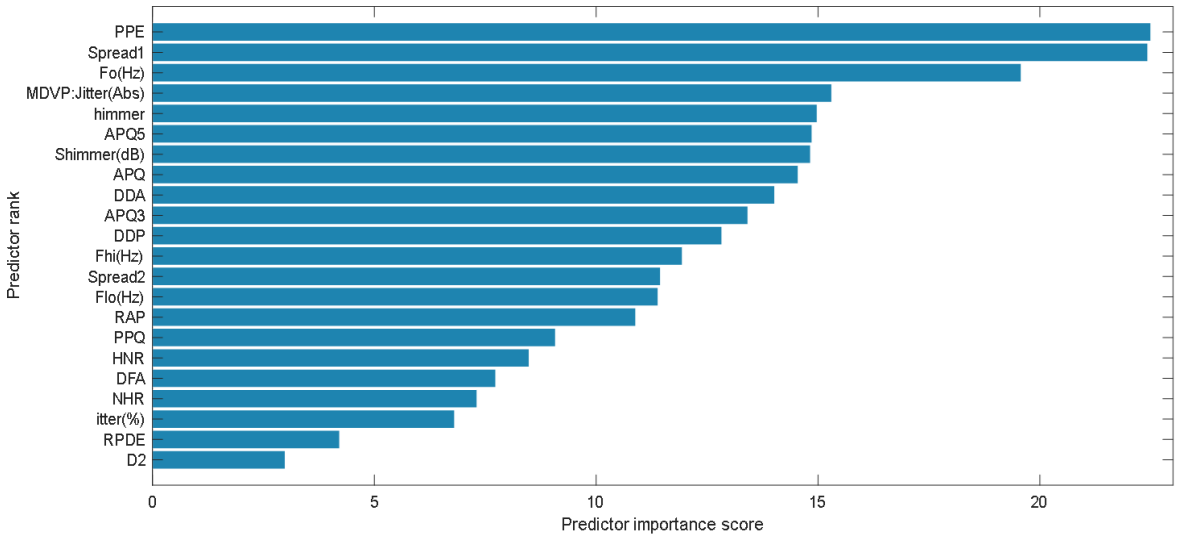

Figure 4. Acoustic feature ranking based on Chi-square technique

3.4. Network essentials

3.4.1. Long-short term memory

The Long-short term memory (LSTM) is suggested by Hochreiter and Schmidhuber for sequence learning. It represents the main building block of the recurrent neural network (RNN). The standard form of an LSTM cell contains three primary cell-memory networks and three separate gates [37]. The architecture of a typical LSTM cell is shown in Figure 5 and the proposed LSTM model is represented in Figure 6. The equations of the cell-memory networks and the gates are represented below, respectively.

– Input cell equation:

$$g_t = \tanh(W_{xg}x_t + W_{hg}h_{t-1} + b_g) \tag{3}$$

– Memory cell equation:

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t \tag{4}$$

– Output cell equation:

$$h_t = o_t \odot tanh(c_t) \tag{5}$$

– Input gate equation:

$$i_t = sigmoid(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \tag{6}$$

– Forget gate equation:

$$f_t = sigmoid(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \tag{7}$$

– Output gate equation:

$$o_t = sigmoid(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_{t-1} + b_o). \tag{8}$$

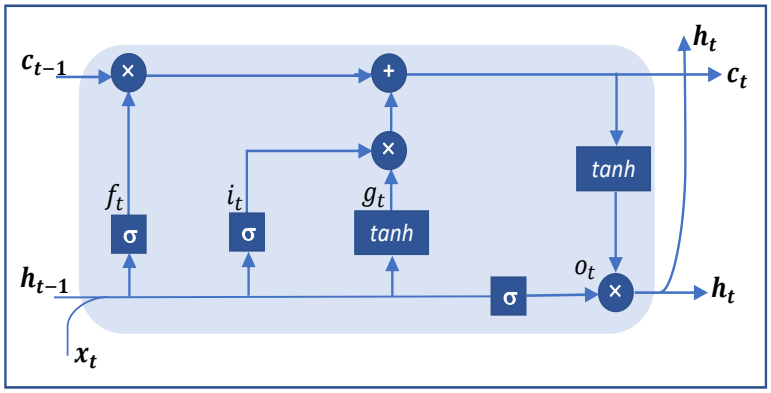

Figure 5. Structure of LSTM cell

3.4.2. Bidirectional LSTM

The bi-directional LSTM (Bi-LSTM) technique was developed by Schuster and Paliwal [38] to overcome the limitation of the standard LSTM since it only includes past information [37]. Thereby, this approach investigates efficiently both the past and future context. Figure 7 illustrates the Bi-LSTM architecture. The (9) and (10) define the Bi-LSTM.

$$\vec{h} = \sigma(x_t U + \vec{h}_{t-1} W + b_t) \tag{9}$$

$$\overleftarrow{h} = \sigma(x_t U + \overleftarrow{h}_{t-1} W + b_t) \tag{10}$$

where $\overleftarrow{h}$ represents the backward hidden state and $\overrightarrow{h}$ represents the forward hidden state. The hidden state is obtained by concatenating the forward and backward hidden states, at time t.

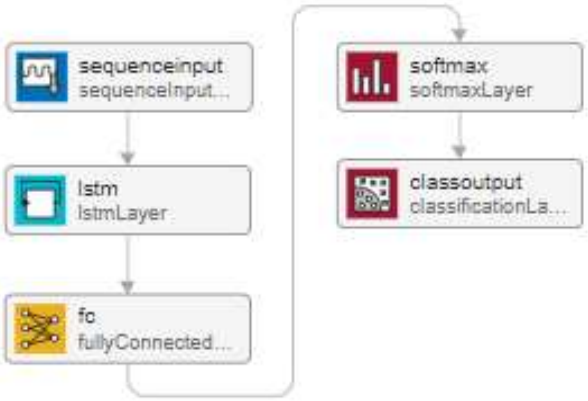

Figure 6. Architecture of LSTM network

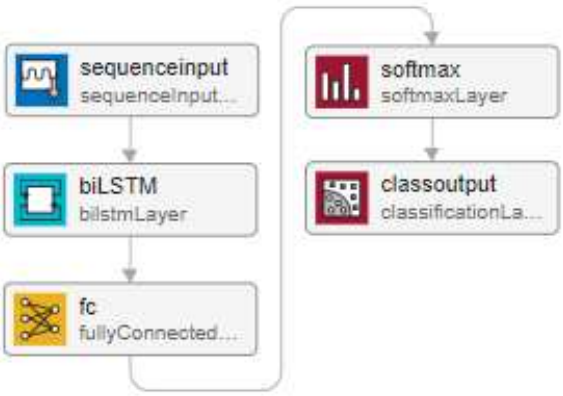

Figure 7. Architecture of Bi-LSTM network

3.4.3. Deep LSTM

Researches demonstrate that enhancing the depth of the neural network can be a good way to upgrade the overall performance [39]. Thus, by inserting additional LSTM layers, we can develop a deeper LSTM (D-LSTM) network. Figure 8 shows the proposed architecture of the D-LSTM network.

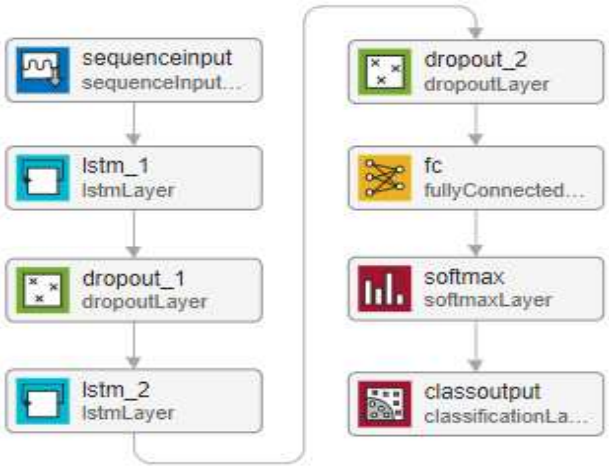

Figure 8. Architecture of D-LSTM network

3.5. Performance metrics

To create the classification model, we divided the database into the training part which takes 70% of the acoustic dataset, the validation part which takes 10%, and the testing part with the remaining 20%. To assess the performance, we used a ten-fold cross-validation technique for the generalization of the results. To evaluate the performance of our model during the test stage, we opted for multiple evaluation metrics such as accuracy, sensitivity, specificity, F1-score, precision, and AUC, as in (11). Accuracy measures the correct results among all predictions. Sensitivity measures true positives proportion, it represents patients that, truly have PD disease. Specificity measures true negatives proportion, it represents patients that, truly are healthy persons. Precision measures the rate that PD detection results are true. F1-score value is related to precision and sensitivity, it returns values between 0 and 1. The false alarm rate (FAR) measures the false results proportion of detecting PD. Area under the ROC curve (AUC) score measured the separability degree, it takes values between 0 and 1. The closer its value is to 1, the model better performs for PD diagnosis.

$$\begin{cases} Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \\ Sensitivity = \frac{TP}{TP+FN} \times 100 \\ Specificity = \frac{TN}{TN+FP} \times 100 \\ Precision = \frac{TP}{TP+FP} \times 100 \\ F1-score = 2 \times \frac{Precision \times Sensitivity}{Precision+Sensitivity} \\ FAR = \frac{FP}{TP+FP} \times 100 \end{cases} \quad (11)$$

4. FRAMEWORK ARCHITECTURE

As shown in Figure 9, the proposed system is built on the basis of 4 stages, starting with preprocessing, classification, evaluation, and finally, decision making. The pre-processing step consists of SMOTE algorithm, which is applied to the training data to balance the class distribution, and Chi-square feature selection to sort the relevant features according to their importance score and extract the most relevant features. Afterward, using the original balanced training data and the result of feature selection in the classification process, we make the LSTM, Bi-LSTM, and D-LSTM models, based on a training operation. The three developed classifiers were tested on the original and reduced testing data. In the last stage, the performance of each model is checked using accuracy, sensitivity, specificity, F1-score, precision, FAR, and AUC. The result of this block allows us to decide if a patient has Parkinson's disease or not.

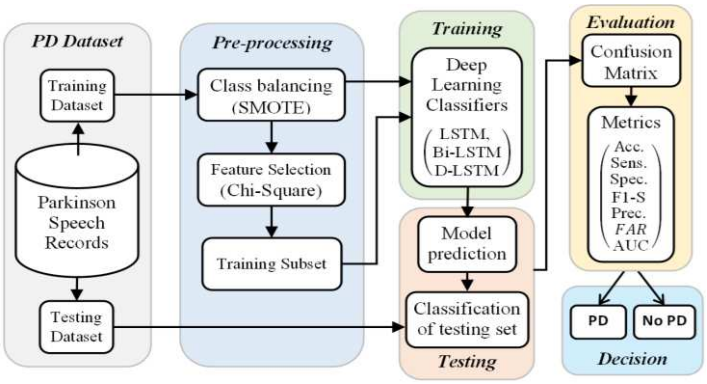

Figure 9. The proposed Parkinson's disease diagnosis architecture

5. RESULTS AND DISCUSSION

This section presents the implementation setup and the results of our experiments with reference to the literature. In our experiment, the SMOTE technique represents a primordial stage in the PD classification process. The original dataset included 195 samples, of which 75% represented PD patients and 25% represented healthy patients. Afterward, the application of the SMOTE technique generated a synthetic balanced training subset, which contained 312 instances, 50% of them represented healthy patients and 50% representing PD patients. Before applying this oversampling method, the results obtained were found to be poor compared to the measurement acquired after oversampling the minority class instances, as shown in Table 3. This illustrates the main role of using the SMOTE technique to handle imbalanced data in a classification application.

Table 3. Comparison between original and synthetic dataset results

	Acc (%)	Sens (%)	Spec (%)	Precision (%)	F1 (%)	FAR (%)	AUC (%)
Original Dataset	84.62	100	0	84.62	91.67	15.38	88.89
Synthetic Dataset	87.18	100	16.67	86.84	92.96	13.16	90.91

5.1. Implementation setup

To validate our proposed framework, several scenarios were investigated and evaluated to choose the best model for each of the LSTM, Bi-LSTM, and D-LSTM classifiers. To achieve efficient results, we

tuned the neural network parameters using an adaptive moment estimation (Adam) optimizer. After multiple experimental tests, we deduced that the configuration according to model 5 is the most appropriate showing high performance of the classifiers as depicted in Table 4. For this model, we have a maximum number of epochs of 200, a mini-batch size of 80, and 20 hidden units per layer. Except for the D-LSTM classifier, we used two LSTM layers with the same number of hidden units as the other classifiers, followed by two dropout layers, which set the inputs randomly to zero with a specific probability of 0.3 and 0.2, for the first and second layers, respectively. This model showed good results for all experiments conducted in the following.

5.2. Performance measures without feature selection

The experimental results, presented in Table 4 and depicted in Figure 10 show that the D-LSTM algorithm performed greatly compared to the other algorithms for all evaluation criteria. D-LSTM realize 94.87% accuracy, 100% sensitivity, 66.67% specificity, 94.29% precision, 97.06% F1-score, 5.71% FAR and 95.45% AUC.

Table 4. Classifiers performance for full features dataset

Methods	Parameters	Acc (%)	Sens (%)	Spec (%)	Precision (%)	F1 (%)	FAR (%)	AUC (%)
LSTM	Model 1	87.18	100	16.67	86.84	92.96	13.16	84.85
	Model 2	84.62	100	0	84.62	91.67	15.38	72.24
	Model 3	84.62	100	0	84.62	91.67	15.38	81.82
	Model 4	87.18	100	16.67	86.84	92.96	13.16	90.40
	Model 5	89.74	100	33.33	89.19	94.29	10.81	90.91
Bi-LSTM	Model 1	87.18	100	16.67	86.84	92.96	13.16	89.39
	Model 2	92.31	100	50.00	91.67	95.65	8.33	91.41
	Model 3	87.18	100	16.67	86.84	92.96	13.16	87.88
	Model 4	84.62	100	0	84.62	91.67	15.38	86.87
	Model 5	9231	100	50.00	91.67	95.65	8.33	93.94
D-LSTM	Model 1	84.62	100	0	84.62	91.67	15.38	75.76
	Model 2	89.74	100	33.34	89.19	94.29	10.81	93.94
	Model 3	87.18	100	16.67	86.84	92.96	13.16	88.38
	Model 4	84.92	100	0	84.62	91.67	15.38	80.81
	Model 5	94.87	100	66.67	94.29	97.06	5.71	95.45

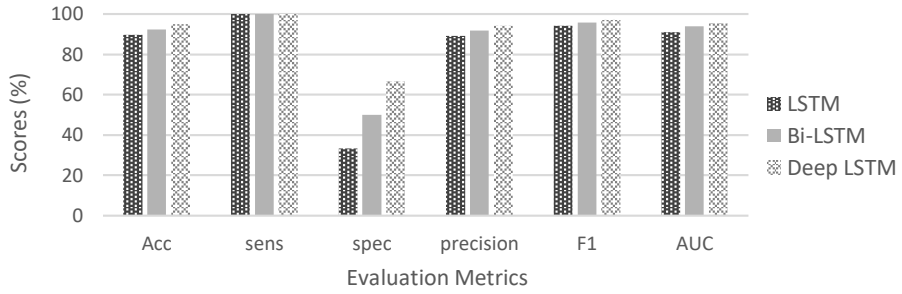

Figure 10. Performance comparison of deep learning techniques without feature selection

5.3. Performance measures on selected features by Chi-square

After sorting the features according to their importance and selecting the most relevant ones, we applied the reduced dataset to the three deep learning algorithms. In fact, multiple tests were performed, showing that the first 15 features guaranteed stable results and the remaining features had no effect on performance improvement. We can notice, as shown in Table 5, that the highest rates in all performance measures are associated with the D-LSTM algorithm, using the reduced features set. It scored 97.44%, 100%, 83.33%, 97.06%, 98.51%, 2.94% and, 99.49% in terms of accuracy, sensitivity, specificity, precision, F1-score, FAR and, AUC, respectively as depicted in Figure 11.

Table 5. Classifiers performance using Chi-square algorithm

Methods	Acc (%)	Sens (%)	Spec (%)	Precision (%)	F1 (%)	FAR (%)	AUC (%)
LSTM	94.87	100	66.67	94.29	97.06	5.71	96.97
Bi-LSTM	97.44	100	83.33	97.06	98.51	2.94	98.99
D-LSTM	98.72	100	83.33	97.06	98.51	2.94	99.49

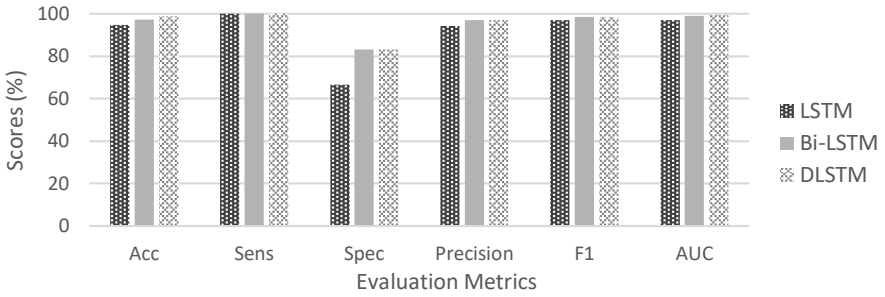

Figure 11. Performance comparison of deep learning techniques after using feature selection

5.4. Discussion

Experiments conducted throughout this work, have shown that dealing with a balanced dataset and a feature selection technique allows us to achieve good performance to identify PD patients. Indeed, as shown in Table 3, the SMOTE technique achieved an accuracy of 87.18% instead of 84.62% while improving all other performance indicators. Likewise, the studied classifiers performed best with the Chi-square feature selection technique when comparing the results shown in Table 3 and Table 4. Thereby, the reduced features composed of 15 features perfectly describe the attributes of the PD dataset. In Table 4, it is noted that both the Bi-LSTM and D-LSTM algorithms achieved the highest accuracy of 97.44%. Nevertheless, the D-LSTM classifier remained the best neural architecture due to its higher AUC score than the Bi-LSTM. Overall, the D-LSTM network outperformed both the LSTM and Bi-LSTM, as it represented the deepest learning model compared to the others. The success of our prediction approach lies in the depth of the neural networks. Even though it is not the best, the Bi-LSTM algorithm remains better than the LSTM algorithm. An additional LSTM layer is added to the Bi-LSTM architecture. Its main role is to reverse the orientation of the information flow. Thus, it facilitates the use of information in both directions. Therefore, using this architecture model can generate more meaningful classification results.

6. CONCLUSION

In this work, we have proposed a hybrid solution based on the use of three deep learning Algorithms combined with data preprocessing for PD classification with the imbalanced dataset. Thus, the hybrid model, combining the SMOTE technique to balance the dataset, the Chi-square feature selection technique to extract the most informative attributes, and the D-LSTM deep learning classifier, enhanced the detection process of Parkinson's disease. The classification phase required training three different classifiers namely LSTM, Bi-LSTM, and D-LSTM. The testing process evinced that D-LSTM deep learning algorithm is the most satisfactory classifier in terms of detecting people with PD among healthy people. The D-LSTM algorithm, which is the best classifier, applied to the imbalanced PD dataset, achieved an accuracy and AUC of 94.87% and 95.45% respectively. The model shows a great performance rate compared to the literature results once we applied the Chi-square feature selection algorithm, with an accuracy and AUC score of 97.44% and 99.49% respectively. We can affirm that our contribution can be a second solution for the diagnosis of Parkinson's disease, but it certainly cannot replace medical professionals. In future works, we will attempt to adopt large acoustic databases and apply more efficient algorithms for both pre-processing and classification stages in order to assess Parkinson's disease. We also expect to study the disease progression, by identifying the degree of illness for a PD patient, based on UPDRS scores in the given dataset.

REFERENCES

[1] C. O. Sakar *et al.*, "A comparative analysis of speech signal processing algorithms for Parkinson's disease classification and the use of the tunable Q-factor wavelet transform," *Applied Soft Computing*, vol. 74, pp. 255–263, Jan. 2019, doi: 10.1016/j.asoc.2018.10.022.

[2] T. Tuncer, S. Dogan, and U. R. Acharya, "Automated detection of Parkinson's disease using minimum average maximum tree and singular value decomposition method with vowels," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 1, pp. 211–220, Jan. 2020, doi: 10.1016/j.bbe.2019.05.006.

[3] R. Lamba, T. Gulati, H. F. Alharbi, and A. Jain, "A hybrid system for Parkinson's disease diagnosis using machine learning techniques," *International Journal of Speech Technology*, vol. 25, no. 3, pp. 583–593, 2022, doi: 10.1007/s10772-021-09837-9.

[4] R. Lamba, T. Gulati, and A. Jain, "Comparative analysis of Parkinson's disease diagnosis system: a review," *Advances in Mathematics: Scientific Journal*, vol. 9, no. 6, pp. 3401–3408, Jul. 2020, doi: 10.37418/amsj.9.6.20.

[5] F. N. Emamzadeh and A. Surguchov, "Parkinson's disease: biomarkers, treatment, and risk factors," *Frontiers in Neuroscience*, vol. 12, Aug. 2018, doi: 10.3389/fnins.2018.00612.

[6] A. Ouhmida, A. Raihani, B. Cherradi, and O. Terrada, "A novel approach for Parkinson's disease detection based on voice classification and features selection techniques," *International Journal of Online and Biomedical Engineering (IJOE)*, vol. 17, no. 10, Oct. 2021, doi: 10.3991/ijoe.v17i10.24499.

[7] X. S. Zhang, L. He, K. Chen, Y. Luo, J. Zhou, and F. Wang, "Multi-view graph convolutional network and its applications on neuroimage analysis for Parkinson's disease," *arXiv:1805.08801,* May 2018.

[8] B. Cherradi, O. Terrada, A. Ouhmida, S. Hamida, A. Raihani, and O. Bouattane, "Computer-aided diagnosis system for early prediction of atherosclerosis using machine learning and k-fold cross-validation," in *2021 International Congress of Advanced Technology and Engineering (ICOTEN)*, Jul. 2021, pp. 1–9, doi: 10.1109/ICOTEN52080.2021.9493524.

[9] S. Hamida, B. Cherradi, A. Raihani, and H. Ouajji, "Performance evaluation of machine learning algorithms in handwritten digits recognition," in *2019 1st International Conference on Smart Systems and Data Science (ICSSD)*, Oct. 2019, pp. 1–6, doi: 10.1109/ICSSD47982.2019.9003052.

[10] O. Asmae, R. Abdelhadi, C. Bouchaib, S. Sara, and K. Tajeddine, "Parkinson's disease identification using KNN and ANN algorithms based on voice disorder," in *2020 1st International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)*, Apr. 2020, pp. 1–6, doi: 10.1109/IRASET48871.2020.9092228.

[11] T. A. Zesiewicz, Y. Bezchlibnyk, N. Dohse, and S. D. Ghanekar, "Management of early Parkinson disease," *Clinics in Geriatric Medicine*, vol. 36, no. 1, pp. 35–41, Feb. 2020, doi: 10.1016/j.cger.2019.09.001.

[12] A. Ouhmida, A. Raihani, B. Cherradi, and Y. Lamalem, "Parkinson's disease classification using machine learning algorithms: performance analysis and comparison," in *2022 2nd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)*, Mar. 2022, pp. 1–6, doi: 10.1109/IRASET52964.2022.9738264.

[13] A. Ouhmida, O. Terrada, A. Raihani, B. Cherradi, and S. Hamida, "Voice-based deep learning medical diagnosis system for Parkinson's disease prediction," in *2021 International Congress of Advanced Technology and Engineering (ICOTEN)*, Jul. 2021, pp. 1–5, doi: 10.1109/ICOTEN52080.2021.9493456.

[14] T. Sun, L. Jiao, J. Feng, F. Liu, and X. Zhang, "Imbalanced hyperspectral image classification based on maximum margin," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 3, pp. 522–526, Mar. 2015, doi: 10.1109/LGRS.2014.2349272.

[15] Y. Liu and Y. Chen, "Total margin based adaptive fuzzy support vector machines for multiview face recognition," in *2005 IEEE International Conference on Systems, Man and Cybernetics*, 2005, vol. 2, pp. 1704–1711, doi: 10.1109/ICSMC.2005.1571394.

[16] T. Munkhdalai, O.-E. Namsrai, and K. H. Ryu, "Self-training in significance space of support vectors for imbalanced biomedical event data," *BMC Bioinformatics*, vol. 16, Dec. 2015, doi: 10.1186/1471-2105-16-S7-S6.

[17] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera, "A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 4, pp. 463–484, Jul. 2012, doi: 10.1109/TSMCC.2011.2161285.

[18] N. Japkowicz and S. Stephen, "The class imbalance problem: A systematic study1," *Intelligent Data Analysis*, vol. 6, no. 5, pp. 429–449, Nov. 2002, doi: 10.3233/IDA-2002-6504.

[19] R. Dubey, J. Zhou, Y. Wang, P. M. Thompson, and J. Ye, "Analysis of sampling techniques for imbalanced data: An n=648 ADNI study," *NeuroImage*, vol. 87, pp. 220–241, Feb. 2014, doi: 10.1016/j.neuroimage.2013.10.005.

[20] S. Sharma, C. Bellinger, B. Krawczyk, O. Zaiane, and N. Japkowicz, "Synthetic oversampling with the majority class: a new perspective on handling extreme imbalance," in *2018 IEEE International Conference on Data Mining (ICDM)*, Nov. 2018, pp. 447–456, doi: 10.1109/ICDM.2018.00060.

[21] M. Zeng, B. Zou, F. Wei, X. Liu, and L. Wang, "Effective prediction of three common diseases by combining SMOTE with Tomek links technique for imbalanced medical data," in *2016 IEEE International Conference of Online Analysis and Computing Science (ICOACS)*, May 2016, pp. 225–228, doi: 10.1109/ICOACS.2016.7563084.

[22] E. J. Alqahtani, F. H. Alshamrani, H. F. Syed, and S. O. Olatunji, "Classification of Parkinson's disease using NNge classification algorithm," in *2018 21st Saudi Computer Society National Computer Conference (NCC)*, Apr. 2018, pp. 1–7, doi: 10.1109/NCG.2018.8592989.

[23] D. Jain, A. K. Mishra, and S. K. Das, "Machine learning based automatic prediction of Parkinson's disease using speech features," in *Proceedings of International Conference on Artificial Intelligence and Applications*, Springer Singapore, 2021, pp. 351–362.

[24] S. Grover, S. Bhartia, Akshama, A. Yadav, and S. K.R., "Predicting severity of Parkinson's disease using deep learning," *Procedia Computer Science*, vol. 132, pp. 1788–1794, 2018, doi: 10.1016/j.procs.2018.05.154.

[25] I. Nissar, D. Rizvi, S. Masood, and A. Mir, "Voice-based detection of Parkinson's disease through ensemble machine learning approach: a performance study," *EAI Endorsed Transactions on Pervasive Health and Technology*, vol. 5, no. 19, Aug. 2019, doi: 10.4108/eai.13-7-2018.162806.

[26] H. Gunduz, "Deep learning-based Parkinson's disease classification using vocal feature sets," *IEEE Access*, vol. 7, pp. 115540–115551, 2019, doi: 10.1109/ACCESS.2019.2936564.

[27] K. Polat, "A hybrid approach to parkinson disease classification using speech signal: the combination of SMOTE and random forests," in *2019 Scientific Meeting on Electrical-Electronics and Biomedical Engineering and Computer Science (EBBT)*, Apr. 2019, pp. 1–3, doi: 10.1109/EBBT.2019.8741725.

[28] R. Olivares *et al.*, "An optimized brain-based algorithm for classifying Parkinson's disease," *Applied Sciences*, vol. 10, no. 5, Mar. 2020, doi: 10.3390/app10051827.

[29] S. Lahmiri and A. Shmuel, "Detection of Parkinson's disease based on voice patterns ranking and optimized support vector machine," *Biomedical Signal Processing and Control*, vol. 49, pp. 427–433, Mar. 2019, doi: 10.1016/j.bspc.2018.08.029.

[30] T. J. Wroge, Y. Ozkanca, C. Demiroglu, D. Si, D. C. Atkins, and R. H. Ghomi, "Parkinson's disease diagnosis using machine learning and voice," in *2018 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, Dec. 2018, pp. 1–7, doi: 10.1109/SPMB.2018.8615607.

[31] 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 Transactions on Biomedical Engineering*, vol. 56, no. 4, pp. 1015–1022, Apr. 2009, doi: 10.1109/TBME.2008.2005954.

[32] N. V Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, Jun. 2002, doi: 10.1613/jair.953.

[33] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009, doi: 10.1109/TKDE.2008.239.

[34] G. E. A. P. A. Batista, R. C. Prati, and M. C. Monard, "A study of the behavior of several methods for balancing machine learning training data," *ACM SIGKDD Explorations Newsletter*, vol. 6, no. 1, pp. 20–29, Jun. 2004, doi: 10.1145/1007730.1007735.

[35] H. Liu and R. Setiono, "Chi2: feature selection and discretization of numeric attributes," in *Proceedings of 7th IEEE International Conference on Tools with Artificial Intelligence*, 1995, pp. 388–391, doi: 10.1109/TAI.1995.479783.

[36] N. Rachburee and W. Punlumjeak, "A comparison of feature selection approach between greedy, IG-ratio, Chi-square, and mRMR in educational mining," in *2015 7th International Conference on Information Technology and Electrical Engineering (ICITEE)*, Oct. 2015, pp. 420–424, doi: 10.1109/ICITEED.2015.7408983.
[37] S. B. Shuvo, S. N. Ali, S. I. Swapnil, M. S. Al-Rakhami, and A. Gumaei, "CardioXNet: a novel lightweight deep learning framework for cardiovascular disease classification using heart sound recordings," *IEEE Access*, vol. 9, pp. 36955–36967, 2021, doi: 10.1109/ACCESS.2021.3063129.
[38] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997, doi: 10.1109/78.650093.
[39] A. Sagheer and M. Kotb, "Time series forecasting of petroleum production using deep LSTM recurrent networks," *Neurocomputing*, vol. 323, pp. 203–213, Jan. 2019, doi: 10.1016/j.neucom.2018.09.082.

BIOGRAPHIES OF AUTHORS

Asmae Ouhmida received her Master's degree in 2018 in distributed information systems at ENSET Institute, Mohammedia, Hassan II University of Casablanca, Morocco. She is currently pursuing the Ph.D. degree at the Electrical Engineering and Intelligent Systems (EEIS) Laboratory at ENSET Mohammedia of the same University. Her research interest are biomedical engineering, signal processing, and artificial intelligence area. She can be contacted at email: asmaeouhmida1995@gmail.com.

Abdelhadi Raihani was appointed as a professor in Electronics Engineering at Hassan II University of Casablanca, ENSET Institute, Mohammedia Morocco since 1991. He received the B.S. degree in Applied Electronics in 1991 from the ENSET Institute. He has his DEA diploma in Information Processing from Ben M'sik University of Casablanca in 1994. He received the Ph.D. in Parallel Architectures Application and Image Processing from the Ain Chock University of Casablanca in 1998. His current research interests are in the medical image processing areas, electrical engineering fields, particularly in renewable energy, energy management systems, power and energy systems control. He supervised several Ph.D. and Engineers students in these topics. He can be contacted at email: abraihani@yahoo.fr.

Bouchaib Cherradi was born in 1970 at El Jadida, Morocco. He received the B.S. degree in Electronics in 1990 and the M.S. degree in Applied Electronics in 1994 from the ENSET Institute of Mohammedia, Morocco. He received the DESA diploma in instrumentation of measure and control (IMC) from Chouaib Doukkali University at El Jadida in 2004. He received his Ph.D. in Electronics and Image processing from the Faculty of Science and Technology, Mohammedia. His research interests include applications of Massively Parallel Architectures, Cluster Analysis, Pattern Recognition, Image Processing, Fuzzy Logic systems, Artificial Intelligence, Machine Learning and Deep Learning in Medical and educational data analysis. Dr. Cherradi works actually as an Associate professor in CRMEF Casablanca-Settat (Provincial section of El Jadida). In addition, he is associate researcher member of Electrical Engineering and Intelligent Systems (EEIS) Laboratory at ENSET of Mohammedia, Hassan II University of Casablanca (UH2C), and LaROSERI Laboratory on leave from the Faculty of Science El Jadida (Chouaib Doukali University), Morocco. He is a supervisor of several Ph.D. students. He can be contacted at email: bouchaib.cherradi@gmail.com.

Sara Sandabad was born in Casablanca, on May 11, 1988. She received the B.Sc. and the M.Sc. degrees in Electrical Engineering from ENSET, Mohamed 5 University of Rabat (Morocco) and the Ph.D. degree from ENSIAS from the Mohamed 5 University of Rabat, all in Electrical Engineering, in 2010, 2012, and 2016, respectively. Currently, he is a Professor of Electrical Engineering at ISEM, Hassan 2 University of Casablanca. She is also an academic member of Laboratory of Control and Mechanical Characterization of Materials and Structures, ENSEM, Hassan II University of Casablanca, Morocco. Her research interests are in the area artificial intelligence, electrical engineering, signal processing, information technologies, and biomedical engineering. She can be contacted at email: sandabad.sandabad@gmail.com.