Pattern Recognition Letters 168 (2023) 64–70

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Improving Parkinson's disease recognition through voice analysis using deep learning

Rania Khaskhoussy*, Yassine Ben Ayed

MIRACL: Multimedia InfoRmation System and Advanced Computing Laboratory, University of Sfax, National Engineering School of Sfax (ENIS), BP 1173, Sfax 3038, Tunisia

ARTICLE INFO

Article history:
Received 22 March 2022
Revised 23 February 2023
Accepted 7 March 2023
Available online 9 March 2023

Edited by: Maria De Marsico

Keywords:
Parkinson's disease
SVM
CNN
I-vector features
Speech

ABSTRACT

Parkinson's disease (PD) corresponds to one of the most common neurological diseases in the world, which is mainly manifested by motor, cognitive, and language disorders. The change in the patient's voice is one of the most striking clinical signs and proves to be an element of interest to support the diagnosis and the assessment of PD. In this research paper, a new approach based on speech signal analysis is set forward to automatically detect Parkinson's disease. The approach evaluates two learning techniques, namely Support Vector Machines (SVM) and Convolutional Neural Networks (CNN), to classify data obtained from speech tasks. Two input data,i.e., the raw speech signal's values and the i-vector features of dimensions 100, 200 and 300 are extracted in this study. Eventually, an evaluation step is undertaken through the use of five evaluation metrics which are accuracy, precision, recall/sensitivity, specificity and f-score. The most pertinent obtained results for a test dataset composed of 28 participants are recorded as follows: an accuracy of 100%, precision of 0.99, recall/sensitivity of 0.98, specificity of 0.96 and f-score of 0.98. These findings corroborate that our approach can help in terms of PD detection.

© 2023 Elsevier B.V. All rights reserved.

1. Introduction

Parkinson's disease stands for the second neurodegenerative disease following Alzheimer's disease, as described by the british physician James Parkinson in 1817. It is a major cause of motor disability among elderly people [3]. According to World Health Organization (WHO), this pathology affects approximately 6.3 million people worldwide and its distribution is fairly homogeneous. PD basically appears after the age of 65 and in some cases before reaching the age of 50. Notably, men are slightly more affected than women [12].

PD refers to a global public health issue since its prevalence grows as the population gets older. Along with other neurodegenerative diseases such as Alzheimer's disease, PD is expected to transcend cancer disease in the second rank of mortality by 2040 (WHO).

Idiopathic PD is characterized by the destruction of a specific population of neurons, namely the dopamine neurons of the brain's substantia nigra. This destruction leads to motor, cognitive, and language disorders [7].

The appearance of dysarthria of hypokinetic type (monotonous speech with difficulty initiating articulation, variable rate and a hoarse, breathy voice) is often associated with PD. These speech disorders emerge at the beginning of PD and even during the presymptomatic period [8,17]. Thus, this observation can help in terms of the detection of PD using voice analysis. In this research paper, we set forward a new approach based on speech signal analysis for PD detection. The main novelty of our approach resides in the extraction of new features which have not yet been extracted from the used dataset, called i-vector features and CNN deep features. We extracted from each voice sample three i-vectors of different sizes 100, 200, and 300 using Gaussian Mixture Models (GMM) based on Universal Background Model (UBM) applied on 39-dimensional Mel-Frequency Cepstral Coefficients (MFCC). As a matter of fact, three different models were obtained after using i-vector features and raw signal values as inputs for both learning techniques, namely CNN and SVM. Finally, we calculated the accuracy, precision, recall/sensitivity, specificity and f-score for each model in order to assess its performance in terms of PD detection.

The remainder of this paper is laid out as follows. Section 2 displays the most prominent contributions using the speech signal for PD detection. Section 3 introduces the proposed approach. In Section 4, the experimental setup and the results are exhibited and Section 5 concludes the whole work and provides pertinent perspectives for future works.

* Corresponding author.
E-mail addresses: rania.khaskhoussy@enis.tn (R. Khaskhoussy), yassine.benayed@isims.usf.tn (Y.B. Ayed).

https://doi.org/10.1016/j.patrec.2023.03.011
0167-8655/© 2023 Elsevier B.V. All rights reserved.

2. Related works

Harel et al. [8] and Postuma et al. [17] demonstrated that some voice disorders that characterize PD occur early and even before the motor symptoms. Several studies departed from this idea and elaborated approaches based on speech analysis to identify PD. A multi-model framework based on the use of several Machine Learning (ML) techniques of diverse nature, was developed by Ali et al. [1] for PD detection. Six groups of features were extracted in the study involving frequency, pulse, amplitude, voicing, pitch, and harmonicity to train the different ML techniques. The performed analysis revealed that vowel samples provide additional information for PD. The study entailed also that SVM and logistic regression present the best accuracy amounting to 70% compared to linear discriminant analysis, gaussian naive bayes, decision tree, and k-nearest neighbours.

An optimization of k-nearest neighbours algorithm was introduced by Amato et al. [4], to discriminate people with PD from healthy subjects. The authors combined 126 features extracted from speech database, in order to compare the performance of early and late feature fusion schemes, so as to identify the best pattern configuration. The study corroborated the feasibility of automatic PD assessment using speech recordings and that a selection of the most relevant characteristics allowed the possibility of voice processing using a simple smartphone application. In another approach, Yaman et al. [24] applied a statistical method for PD recognition. The method generated 177 new features from 44 acoustic features. Subsequently, these features were normalized by the ReliefF method to select the most discriminating features. The study concluded that the most weighed features exhibited a better classification result of 91.25% using SVM classifier.

In an investigation conducted by Senturk [20], a features selection method was developed for early PD detection. The method was performed on 22 phonetic features extracted from the voice samples of PD patients and healthy subjects. The investigation disclosed that SVM classifier with feature selection presented the highest classification performance 93.84% compared to pure SVM. In the same context, an optimization of features selection method using Modified Grey Wolf Optimization (MGWO) model was performed by Sharma et al. [21] for PD prediction. The authors assessed the proposed model on various types of datasets using random forest, k-nearest neighbor and decision tree classifier and confirmed the feasibility of this optimization in the prediction of PD through an accuracy of 94.83%.

ML techniques continue to be extensively adopted to assess PD departing from speech signals, in the approach of Johri et al. [10] where two neural network based models were introduced, to diagnose PD at an early stage. Relying on two symptoms walking patterns (gait) and speech impairment, the study yielded that the proposed models were more efficient and fulfilled a better accuracy of about 89.15% compared to other tested models. In another approach, Ali et al. [2] suggested a new hybrid intelligent system using the linear discriminant analysis, genetic algorithm and neural network, to perform an acoustic analysis of voice signals for detecting PD. The authors proved that after eliminating bias in the classification task, the proposed system provided a classification result of 95%.

Within the same ML context, concerning particularly the use of convolutional neural networks for PD diagnosis. Bhattacharjee et al. [5] proposed several CNN models trained on the fusion of baseline MFCC with other features. The study reveals that the fusion of f_0 and MFCC achieves the highest classification accuracy under both clean and matched train-test conditions. In another approach, Mallela et al. [16] developed a cascaded architecture comprising CNN and BLSTM for the classification of PD and Healthy Controls (HC). The study concludes that the learned representations obtained by CNN from raw speech perform better than the baseline with an accuracy of 98.27%. Likewise, the study of Suhas et al. [22] proved that CNN model trained on log mel spectrograms of PD and HC presents good results in terms of PD classification and outperforms the CNN model trained on the baseline MFCC.

For additional works related to the diagnosis of PD, the reader may consult review of Saravanan et al. [19].

3. Proposed approach

The proposed approach architecture for PD detection is portrayed in Fig. 1. Three different models are generated from three systems (S1, S2 and S3) that build up our approach.

In the first system, we propose the use of CNN in which we utilize the raw signal for deep features extraction. In the classification phase, we suggest the use of MultiLayer Perceptron (MLP).

For the second system, we invest the deep features obtained by the CNN to feed the classification phase which will be carried out by an SVM with different kernels, namely linear, sigmoid, polynomial and RBF.

In the third system, we use i-vectors as input characteristics, which are obtained from the MFCC coefficients. These coefficients are used by the GMM-UBM model in order to extract the i-vector features. For the classification phase, we use SVM with different kernels, namely linear, sigmoid, polynomial and RBF.

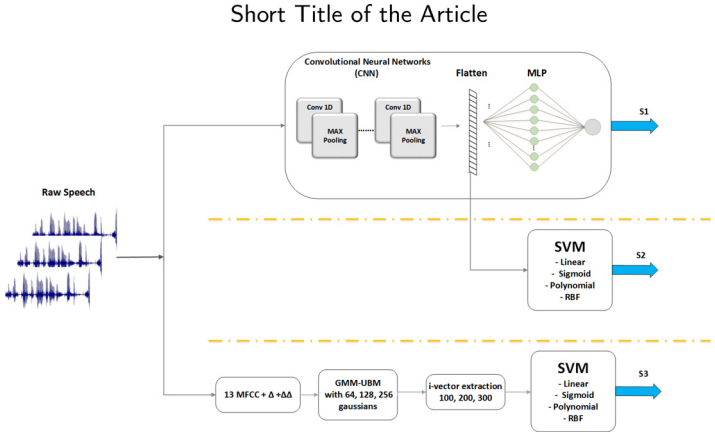


Fig. 1. Parkinson's disease detection with the proposed approach .

3.1. Dataset description

The dataset used in this work was collected by Sakar et al. [18] to provide speech data for phonetic-prosodic studies and to develop and assess automatic recognition applications for PD [11,13,15]. This dataset rests upon 1208 voice recordings in WAV format acquired by a reliable MC-1500 microphone placed at 10 cm far from the subjects with a frequency interval ranging between 50 Hertz (Hz) and 13 kHz.

This dataset is composed of two data subsets. The first is used for the training phase, which contains 1040 voice files recorded from 40 participants where 20 are parkinsonian (6 women and 14 men) and 20 are healthy subjects (10 women and 10 men). For all participants of this subset, only 26 voice samples including sustained vowels (/ a /, / o / and / u /), numbers and words are recorded with a duration of 5 s.

The second subset is used for the testing phase and contains 168 voice files recorded from 28 participants where 14 are parkinsonian (7 women and 7 men) and 14 are healthy subjects (7 women and 7 men). With these participants, two sustained vowels (/ a / and / o /) are recorded three times with a duration of 5 s.

For PD patients, the time since diagnosis varies between 0 and 6 years for the first subset and between 0 and 13 years for the second subset. The ages of the participants vary between 45 and 83 years for healthy subjects and between 39 and 79 years for PD patients.

3.2. Features extraction

3.2.1. Mel-frequency cepstral coefficients extration

MFCC stand for acoustic parameters that were first introduced in 1980 by Davis and Mermelstein [6]. These parameters characterize the spectral envelope of the voice and therefore reflect the shape and volume of the vocal tract. MFCC exploit both the decorrelation properties of the cepstrum and the psychoacoustic principles of the human ear. Calculating the 39 MFCC is carried out according to 6 steps (Pre-emphasis, Windowing, Fast Fourier Transform (FFT) on each frame, MEL band pass filtering, Log of the energy of each filter / Discrete cosine transform, Temporal derivatives).

The MFCC cepstral parameters are obtained from the energies of a filters bank according to the Mel scale. To deduce the first 13 coefficients, the logarithm of the energies leaving the bank of filters M is applied by the discrete cosine transform (DCT).

The cepstral coefficients are extracted over windows of length 25 ms with overlaps of 10 ms. Indeed, the typical duration for the windows of the speech signal varies between 20 ms and 30 ms with an overlap which ranges between 10 ms and 15 ms. This refers to the fact that during this period the signal is considered stationary; however, over a longer period, it may not be stationary. Finally, 13 coefficients (12 coefficients + energy) are considered along with their first and second derivatives ($\Delta MFCC$ and $\Delta\Delta MFCC$), forming in all vectors with 39 components.

3.2.2. I-vectors extraction

I-vectors correspond to a compress representation of supervectors obtained by the GMM-UBM applied on the MFCC vectors [14].

As a first step, a Universal Background Model of size 64, 128 and 256 gaussians was trained on MFCC vectors using the Expectation Maximization (EM) algorithm. Participants-specific models were then obtained using Maximum A Posteriori (MAP) mean adaptation.

After adaptation, the models were represented as high-dimensional supervectors of mean distributions (39×64, 39×128 and 39×256). These supervectors were then represented in the form of low-dimensional vectors, which are the i-vectors (100, 200,300) using factor analysis.

3.3. Support vector machines for PD detection

SVM is a discriminant classifier which separates two classes by a separation hyperplane. It is a widely used technique in terms of classification owing to its powerful generalization [23].

For PD detection, SVM models are trained using different kernel functions with data from PD patients (with label 1) as well as data from healthy subjects (with label 0). The used data are i-vector features and CNN features. The used kernel functions are linear, polynomial with degree equal to 3, sigmoid, and RBF.

The linear kernel formula is identified as:

$$K(x_i, x_j) = x_i^T x_j \tag{1}$$

The sigmoid kernel formula is indicated by:

$$K(x_i, x_j) = \tanh(x_i x_j + c) \tag{2}$$

The polynomial kernel formula is expressed as follows:

$$K(x_i, x_j) = (x_i^T x_j + c)^d, \tag{3}$$

where c and d are respectively a constant and the degree of the kernel.

The RBF kernel formula is determined as:

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right), \tag{4}$$

where σ is a positive real that represents the bandwidth of the kernel.

3.4. Convolutional neural networks

The CNN architecture is inspired from the organization of the visual cortex of animals [9]. It is a multilayer neural network composed of two different types of layers, which are the convolution layers and the pooling layers. Both layers are connected alternately and form the central part of the network.

For PD detection, the CNN is trained on the raw signals of PD patients and healthy participants. The input signal is convolved with trainable filters so as to produce feature matrices in the first convolution layer. A connection weight layer is included in each filter. Moving through an activation function, these values generate new feature matrices in the first pooling layer. This procedure continues to obtain the feature matrices in subsequent convolution and pooling layers. Eventually, the values of these features are flattened into a single vector which will be considered as the input of the MLP network and SVM.

4. Experiments and results

In this part, we handle all experiments that we conducted on the used dataset by considering that all the measurements of each participant are independent. Additionally, we elaborate and display the corresponding obtained results. The central objective of these experiments lies in demonstrating, on the one hand, the significance of CNN features and i-vector features in terms of the detection of PD and on the other hand, the impact of the models obtained by i-vector features and through the hybridization of CNN features with SVM classifier on the performance of our approach.

To ensure the effectiveness of our proposed systems, we conducted all experiments with cross-validation (cv = 5 fold) and without cross-validation scheme. In each fold of cv, we took in the training phase from the set of 40 subjects, 80% from each voice recording for training and 20% from each voice recording for the

Table 1
Classification results using CNN features and SVM.

Experiments	SVM kernels	Acc (%)	P	Se	Sp	*F*-score
With cross-validation (cv = 5)	Linear	68.00 ± 12.0	0.69 ± 0.07	0.60 ± 0.03	0.61 ± 0.11	0.65 ± 0.04
	Sigmoid	71.00 ± 6.1	0.68 ± 0.02	**0.65 ± 0.01**	0.64 ± 0.03	**0.66 ± 0.01**
	Polynomial	70.00 ± 5.0	0.65 ± 0.03	0.65 ± 0.04	0.63 ± 0.01	0.65 ± 0.03
	RBF	**75.00 ± 4.0**	**0.74 ± 0.03**	0.62 ± 0.02	**0.65 ± 0.01**	0.65 ± 0.03
Without cross-validation	Linear	76.00	0.77	0.66	0.64	0.68
	Sigmoid	78.00	0.75	**0.73**	0.70	**0.73**
	Polynomial	75.00	0.71	0.70	0.70	0.71
	RBF	**79.00**	**0.79**	0.70	**0.71**	0.72

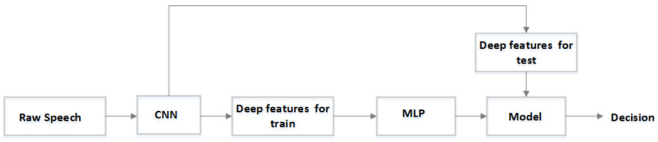


Fig. 2. CNN features and MLP based system architecture.

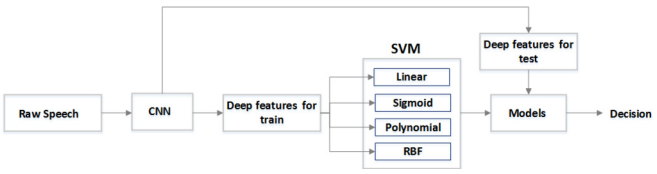


Fig. 3. CNN features and SVM based system architecture.

validation step. As for the test phase, we used the recordings of the 28 subjects since the used dataset is benchmark. Then, Five metrics of evaluations, namely Accuracy (Acc), Precision (P), Recall (R)/Sensitivity (Se), Specificity (Sp) and *F*-score (*F*) are selected to assess the performance of our approach, which are defined as:

$$Acc = \left(\frac{TP + TN}{All\ samples}\right), \qquad P = \left(\frac{TP}{TP + FP}\right)$$

$$R/Se = \left(\frac{TP}{TP + FN}\right), \qquad F\text{-score} = 2\left(\frac{P * R}{P + R}\right)$$

$$Sp = \left(\frac{TN}{TN + FP}\right)$$

where, TP, True Positive (PD patient classified as Parkinsonian); TN, True Negative (Healthy person classified as healthy person);FP, False Positive (Healthy person classified as Parkinsonian);FN, False Negative (Parkinsonian classified as healthy person).

4.1. CNN features and MLP based system: S1

For PD detection, S1 used the raw speech of the participants, in numerical values form, as input to CNN for deep features extraction. Afterwards, these features were used by a multilayer perceptron to build up the appropriate model of the system, in order to determine the class of a given sample as indicated in Fig. 2.

The CNN architecture used in this system relies upon three convolutional layers as well as three max pooling layers. The activation functions are all ReLU and the kernel size is equal to 3 in the convolutional layers. The number of filters in each CNN layers is 6, the size of pooling layer is 2, optimizer algorithm is adam, the loss function is binary_crossentropy, batch size is 32, learning rate is 0.01 and validation split is equal to 0.2.

As for the MLP architecture, it is made up of three layers: input layer, one hidden layer and output layer composed of a single neuron. The activation function is ReLU for the input and hidden layers and sigmoid for the output layer.

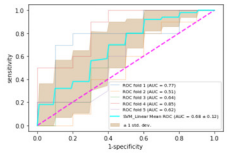
After 100 epochs of training, we got promising classification results with a five-fold cross-validation scheme. These results correspond to an accuracy of 60%, a precision of 0.52, a sensitivity of 0.48, a specificity of 0.50 and an f-score of 0.50.

4.2. CNN features and SVM based system: S2

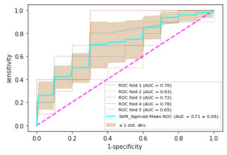
As illustrated in Fig. 3, the raw signals were provided as input to CNN for deep features extraction (CNN features). Then, these features were flattened in a vector of size 9854 and fed to SVM for the training step. Finally, the accuracy, precision, sensitivity, specificity and f-score were computed using a test set.

The CNN architecture used in S2 proved to be the same as the one invested in the previous system S1. Table 1 summarizes the classification results obtained by the SVM models with and without cross-validation. It is noticeable that there is a significant improvement in terms of the results of S2 compared to S1. We can detect that a maximum classification accuracy of 79%, precision of 0.79 and specificity of 0.71 are achieved using RBF kernel of SVM, where the maximum sensitivity and f-score of 0.73 are reached by the sigmoid kernel. Departing from Table 1, we can observe that all evaluation metrics values increased, as the accuracy rose from 61% to 75 to 79%, which was the case also for precision, sensitivity, specificity and f-score which increased by a value between 0.22 and 0.27 compared to S1.

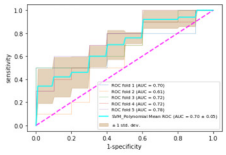
Figures 4 and 5 demonstrate the performance of the S2 system models using the area under the receiver operating characteristic curve (AUC-ROC) with and without cross-validation. It is clear that there is a slight difference between the performance of models with cv and without cv, but it is basically much better than the previous system S1.



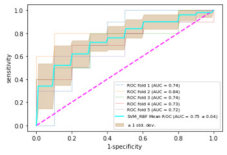
(a) AUC-ROC curves of S2 with linear SVM



(b) AUC-ROC curves of S2 with sigmoid SVM

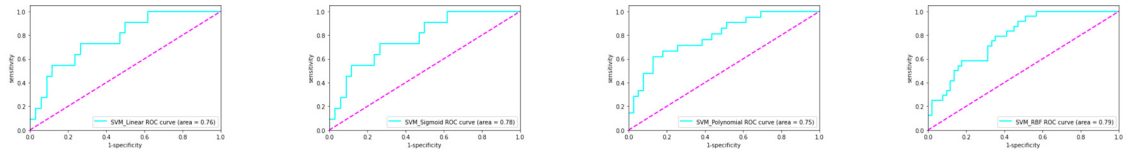


(c) AUC-ROC curves of S2 with polynomial SVM



(d) AUC-ROC curves of S2 with RBF SVM

Fig. 4. AUC-ROC curves of S2 with cv.



(a) AUC-ROC curves of S2 with linear SVM

(b) AUC-ROC curves of S2 with sigmoid SVM

(c) AUC-ROC curves of S2 with polynomial SVM

(d) AUC-ROC curves of S2 with RBF SVM

Fig. 5. AUC-ROC curves of S2 without cv.

Table 2
Classification results using linear kernel.

Experiments	I-vectors	100					200					300				
	UBM size	Acc (%)	P	Se	Sp	*F*-score	Acc (%)	P	Se	Sp	*F*-score	Acc (%)	P	Se	Sp	*F*-score
With cv	64	75.30	0.75	0.76	0.75	0.75	81.00	0.80	0.81	0.82	0.80	74.03	0.73	0.75	0.74	0.74
		±3.2	±0.04	±0.01	±0.02	±0.02	±2.4	±0.05	±0.01	±0.03	±0.04	±3.0	±0.01	±0.03	±0.03	±0.02
	128	88.00	0.89	0.87	0.86	0.88	**96.74**	**0.95**	**0.96**	**0.95**	**0.95**	82.33	0.80	0.79	0.78	0.79
		±1.1	±0.05	±0.04	±0.02	±0.03	**±3.1**	**±0.02**	**±0.03**	**±0.04**	**±0.02**	±2.5	±0.02	±0.03	±0.03	±0.02
	256	66.00	0.64	0.65	0.63	0.64	74.00	0.72	0.74	0.70	0.73	69.02	0.67	0.70	0.68	0.68
		±1.7	±0.07	±0.09	±0.01	±0.09	±0.96	±0.05	±0.05	±0.08	±0.06	±3.5	±0.09	±0.04	±0.02	±0.06
Without cv	64	80.71	0.78	0.80	0.75	0.79	86.12	0.84	0.86	0.83	0.85	76.12	0.76	0.77	0.74	0.77
	128	89.24	0.89	0.88	0.85	0.87	**98.18**	**0.97**	**0.96**	**0.93**	**0.98**	80.00	0.78	0.81	0.80	0.79
	256	68.53	0.68	0.64	0.67	0.66	75.51	0.74	0.73	0.71	0.73	73.15	0.71	0.73	0.72	0.72

Table 3
Classification results using sigmoid kernel.

Experiments	I-vectors	100					200					300				
	UBM size	Acc (%)	P	Se	Sp	*F*-score	Acc (%)	P	Se	Sp	*F*-score	Acc (%)	P	Se	Sp	*F*-score
With cv	64	80.00	0.81	0.82	0.82	0.81	95.26	0.92	0.93	**0.92**	0.92	71.36	0.72	0.70	0.71	0.71
		±2.2	±0.05	±0.04	±0.08	±0.05	±4.8	±0.03	±0.07	**±0.06**	±0.05	±3.0	±0.02	±0.06	±0.06	±0.05
	128	85.55	0.84	0.83	0.84	0.83	**96.00**	**0.95**	**0.92**	0.91	**0.93**	73.58	0.74	0.72	0.73	0.73
		±4.6	±0.04	±0.01	±0.01	±0.03	**±2.2**	**±0.03**	**±0.04**	±0.08	**±0.04**	±1.5	±0.03	±0.04	±0.01	±0.03
	256	66.20	0.67	0.64	0.66	0.65	95.00	0.91	**0.93**	0.89	0.92	70.19	0.69	0.71	0.66	0.69
		±1.3	±0.05	±0.01	±0.01	±0.03	±6.5	±0.02	**±0.05**	±0.04	±0.04	±3.8	±0.09	±0.04	±0.02	±0.06
Without cv	64	80.73	0.82	0.80	0.79	0.81	99.00	0.92	**0.94**	0.91	**0.92**	72.50	0.71	0.72	0.70	0.71
	128	89.27	0.90	0.89	0.88	0.89	**100**	**0.93**	0.91	**0.93**	**0.92**	75.00	0.76	0.72	0.74	0.75
	256	68.54	0.68	0.66	0.64	0.67	98.05	0.92	0.91	0.90	**0.92**	74.00	0.74	0.75	0.72	0.75

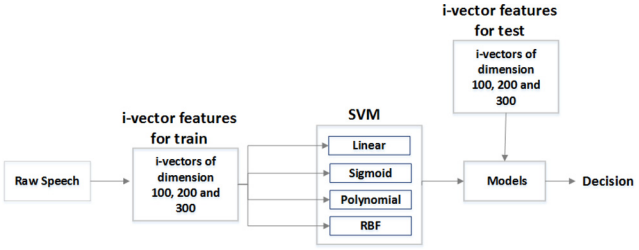


Fig. 6. I-vectors and SVM based system architecture.

Referring to these results, we conclude that the hybridization of CNN features and SVM displays a good performance in terms of PD detection, which proves that the idea of hybridizing CNN features with SVM improves the results of our approach compared to a neural model.

4.3. I-vectors and SVM based system: S3

Another system S3 completes the construction of our approach to optimize forward the detection task of PD. The architecture of this system is highlighted in Fig. 6. Three i-vectors of dimensions 100, 200 and 300 are used as input features to S3. For the classification step, we used SVM as a binary classifier, since it is a binary classification, in order to train the model; then, we calculated the accuracy, precision, sensitivity, specificity and f-score using a test set.

The obtained results are outlined inTables 2–5. Each table records the accuracies, precisions, sensitivity, specificity and f-scores achieved by each i-vector dimensionality with a specific UBM size.

By analyzing the results presented in these tables, we infer that through the models obtained by SVM kernels, the i-vectors of dimensions 200 exhibit the best accuracies and f-scores in the discrimination between PD patients and healthy people, with and without cv compared to the i-vectors of dimensions 100 and 300.

Based on the research results, we deduce from Tables 2 to 4 that when we perform experiments with cross-validation scheme, we obtain the highest accuracy of 97.68% with polynomial SVM, while the rest best evaluation metrics values sensitivity of 0.96, precision, specificity and f-score of 0.95 were reached by linear SVM. Meanwhile, we infer that without cv, the best classification accuracy is 100%, precision is 0.99, specificity is 0.96, sensitivity and f-score is 0.98. These values are achieved using the polynomial kernel of SVM with i-vectors of dimensions 200 and UBM of size 128 gaussians.

According to Tables 2–5, we find that the majority of evaluation metrics values obtained by SVM models improve between the i-vectors of dimensions 100 and 200, then decrease between the i-vectors of dimensions 200 and 300, which implies that increasing and decreasing i-vectors dimensions influences the classification results of the used data.

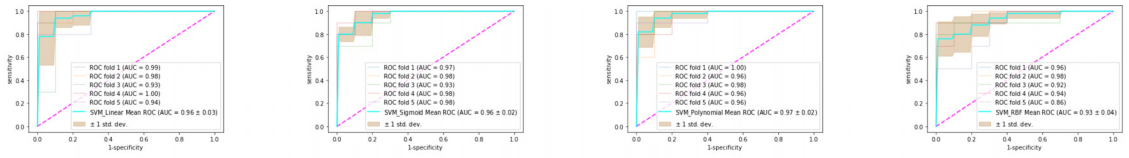
Another outstanding finding drawn from the S3 AUC-ROC curves in Figs. 7 and 8, relates to the fact that most evaluation metrics values obtained by S3 are much better than those of S1

Table 4
Classification results using polynomial kernel.

Experiments	I-vectors	100					200					300				
	UBM size	Acc (%)	P	Se	Sp	*F*-score	Acc (%)	P	Se	Sp	*F*-score	Acc (%)	P	Se	Sp	*F*-score
With cv	64	77.50	0.76	0.75	0.75	0.75	90.56	0.88	0.90	0.89	0.88	81.00	0.79	0.80	0.75	0.80
		±1.8	±0.01	±0.06	±0.02	±0.04	±7.3	±0.08	±0.05	±0.06	±0.06	±2.5	±0.07	±0.07	±0.03	±0.06
	128	91.00	0.91	0.85	0.82	0.88	**97.68**	**0.94**	**0.96**	**0.93**	**0.94**	79.02	0.81	0.78	0.80	0.84
		±2.3	±0.03	±0.02	±0.06	±0.03	**±2.1**	**±0.02**	**±0.04**	**±0.03**	**±0.03**	±2.9	±0.06	±0.04	±0.01	±0.06
	256	82.11	0.80	0.81	0.78	0.81	94.16	0.92	0.90	0.89	0.91	76.50	0.77	0.72	0.74	0.75
		±2.0	±0.04	±0.07	±0.03	±0.06	±4.5	±0.03	±0.02	±0.05	±0.03	±1.8	±0.01	±0.03	±0.01	±0.02
Without cv	64	81.74	0.79	0.81	0.82	0.80	99.01	0.97	0.91	0.93	0.95	84.44	0.87	0.83	0.84	0.83
	128	93.74	0.98	0.94	0.95	0.97	**100**	**0.99**	**0.98**	**0.96**	**0.98**	87.65	0.88	0.86	0.85	0.86
	256	87.48	0.84	0.87	0.84	0.85	99.00	0.97	0.95	0.94	0.96	78.00	0.84	0.78	0.81	0.77

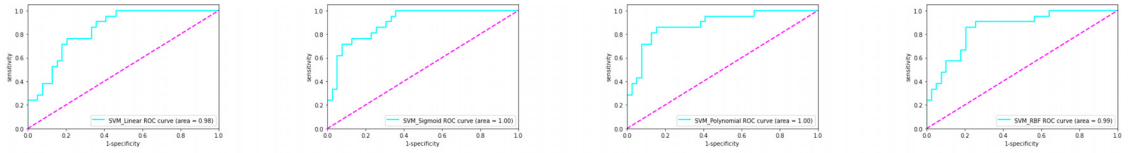
Table 5
Classification results using RBF kernel.

Experiments	I-vectors	100					200					300				
	UBM size	Acc (%)	P	Se	Sp	*F*-score	Acc (%)	P	Se	Sp	*F*-score	Acc (%)	P	Se	Sp	*F*-score
With cv	64	57.23	0.55	0.58	0.54	0.56	88.50	0.86	0.83	0.79	0.85	78.00	0.74	0.71	0.76	0.73
		±1.6	±0.02	±0.04	±0.01	±0.02	±2.1	±0.03	±0.01	±0.01	±0.02	±3.0	±0.02	±0.03	±0.04	±0.03
	128	66.00	0.62	0.65	0.59	0.63	**93.68**	**0.95**	**0.91**	**0.89**	**0.93**	78.11	0.80	0.76	0.71	0.72
		±1.4	±0.03	±0.01	±0.02	±0.02	**±4.0**	**±0.06**	**±0.05**	**±0.05**	**±0.04**	±1.9	±0.05	±0.04	±0.01	±0.05
	256	62.33	0.59	0.61	0.57	0.60	76.05	0.78	0.74	0.71	0.75	75.00	0.70	0.74	0.61	0.71
		±3.5	±0.05	±0.06	±0.03	±0.05	±3.0	±0.04	±0.02	±0.03	±0.04	±4.6	±0.04	±0.01	±0.02	±0.02
Without cv	64	61.00	0.63	0.59	0.55	0.61	91.86	0.89	0.92	0.91	0.90	76.00	0.76	0.77	0.74	0.77
	128	62.62	0.61	0.62	0.60	0.61	**99.39**	**0.98**	**0.97**	**0.94**	**0.97**	80.6	0.82	0.80	0.77	0.80
	256	60.57	0.61	0.59	0.54	0.59	81.19	0.80	0.81	0.78	0.80	76.50	0.80	0.78	0.69	0.74



(a) AUC-ROC curves of S3 with linear SVM (b) AUC-ROC curves of S3 with sigmoid SVM (c) AUC-ROC curves of S3 with polynomial SVM (d) AUC-ROC curves of S3 with RBF SVM

Fig. 7. AUC-ROC curves of S3 (i-vectors of 200 and UBM of 128 gaussians) with cv.



(a) AUC-ROC curves of S3 with linear SVM (b) AUC-ROC curves of S3 with sigmoid SVM (c) AUC-ROC curves of S3 with polynomial SVM (d) AUC-ROC curves of S3 with RBF SVM

Fig. 8. AUC-ROC curves of S3 (i-vectors of 200 and UBM of 128 gaussians) without cv.

Table 6
Performance comparison with related works.

Methods	Acc (%)	Classifier
Bhattacharjee et al. [5]	87.95	CNN-LSTM
Johri et al. [10]	89.15	Deep Neural Networks
Suhas et al. [22]	93	CNN
Sharma et al. [21]	94.83	*K*-Nearest Neighbor
Ali et al. [2]	95	Neural Networks
Mallela et al. [16]	97.28	CNN-BLSTM
Proposed	**100**	**SVM-poly**

and S2. This is indicative that S3 presents a good performance in terms of PD detection compared to S2 and S1.

For a deeper assessment, we referred to other approaches that used the same dataset. The comparative results are displayed in Table 6. The accuracy values obtained with our approach is better than those reported in related works approaches. It can be detected that compared to Johri et al. [10], the accuracy rose from 89.15% to achieve 100%, whereas in Sharma et al. [21] and Ali et al. [2]'s works the accuracy improved with a value of 5% to reach 100%.

Similarly, when we compare our approach to other CNN-based approaches [5,16,22] which uses other data sets; we realize that our approach performs better with S3 system compared to other ones. This high accuracy further confirms the effectiveness, feasibility and reliability of i-vectors features in terms of PD detection.

5. Conclusion

In this paper, a new approach in which we address the issue of Parkinson's disease detection was elaborated. The approach assesses the use of i-vectors as well as CNN features and their effects at the level of the discrimination between PD patients and

healthy people. We explored how the i-vectors extractor parameters such as dimension influence the classification results. Relying on the obtained results, we inferred the significance and potential of SVM models with i-vectors features and CNN features in Parkinson's disease detection.

To sum up, we would simply assert that our research is a step that can be extented and built upon. Indeed, it serves as an enlightening guideline for future works in terms of using other deep learning techniques for the classification step. It also provides a theoretical basis for future researchers in terms of a deeper and more extensive experimentation of this research with other large data sets and additional types of decisive features that will be conducive to more accurate and earlier PD detection. We equally plan to enhance further our approach through the incorporation of a new system based on wav2vec model as well as the assessment of the obtained results compared to those of our proposed systems.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

References

[1] L. Ali, S.U. Khan, M. Arshad, S. Ali, M. Anwar, A multi-model framework for evaluating type of speech samples having complementary information about Parkinson's disease, in: International Conference on Electrical, Communication, and Computer Engineering, 2019, pp. 1–5.
[2] L. Ali, C. Zhu, Z. Zhang, Y. Liu, Automated detection of Parkinson's disease based on multiple types of sustained phonations using linear discriminant analysis and genetically optimized neural network, IEEE J. Transl. Eng. Health Med. 7 (2019) 1–10.
[3] J.S. Almeida, P.P. Rebouças Filho, T. Carneiro, W. Wei, R. Damaševičius, R. Maskeliūnas, V.H.C. de Albuquerque, Detecting Parkinson's disease with sustained phonation and speech signals using machine learning techniques, Pattern Recognit. Lett. 125 (2019) 55–62.
[4] F. Amato, L. Borzì, G. Olmo, J.R. Orozco-Arroyave, An algorithm for Parkinson's disease speech classification based on isolated words analysis, Health Inf. Sci. Syst. 9 (1) (2021) 1–15.
[5] T. Bhattacharjee, J. Mallela, Y. Belur, A. Nalini, R. Yadav, P. Reddy, D. Gope, P.K. Ghosh, Source and vocal tract cues for speech-based classification of patients with Parkinson's disease and healthy subjects, in: INTERSPEECH, 2021, pp. 2961–2965.
[6] S. Davis, P. Mermelstein, Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences, IEEE Trans. Acoust., Speech, Signal Process. 28 (4) (1980) 357–366.
[7] C.G. Goetz, The history of Parkinson's disease: early clinical descriptions and neurological therapies, Cold Spring Harb. Perspect. Med. 1 (1) (2011) 62–88.
[8] B. Harel, M. Cannizzaro, P.J. Snyder, Variability in fundamental frequency during speech in prodromal and incipient Parkinson's disease: a longitudinal case study, Brain Cogn. 56 (1) (2004) 24–29.
[9] D.H. Hubel, T.N. Wiesel, Receptive fields, binocular interaction and functional architecture in the cat's visual cortex, J. Physiol. 160 (1) (1962) 106–154.
[10] A. Johri, A. Tripathi, et al., Parkinson disease detection using deep neural networks, in: Twelfth International Conference on Contemporary Computing, 2019, pp. 1–4.
[11] R. Khaskhoussy, Y. BenAyed, Automatic detection of Parkinson's disease from speech using acoustic, prosodic and phonetic features, in: International Conference on Intelligent Systems Design and Applications, 2019, pp. 80–89.
[12] R. Khaskhoussy, Y. BenAyed, Detecting Parkinson's disease according to gender using speech signals, in: International Conference on Knowledge Science, Engineering and Management, 2021, pp. 414–425.
[13] R. Khaskhoussy, Y. BenAyed, A deep convolutional autoencoder-based approach for Parkinson's disease diagnosis through speech signals, in: International Conference on Advanced Data Mining and Applications, Springer, 2022, pp. 15–26.
[14] R. Khaskhoussy, Y. BenAyed, An i-vector-based approach for discriminating between patients with Parkinson's disease and healthy people, in: Fourteenth International Conference on Machine Vision, vol. 12084, 2022, pp. 69–77.
[15] R. Khaskhoussy, Y. BenAyed, Speech processing for early Parkinson's disease diagnosis: machine learning and deep learning-based approach, Soc. Netw. Anal. Min. 12 (1) (2022) 1–12.
[16] J. Mallela, A. Illa, Y. Belur, A. Nalini, R. Yadav, P. Reddy, D. Gope, P.K. Ghosh, Raw speech waveform based classification of patients with ALS, Parkinson's disease and healthy controls using CNN-BLSTM, in: INTERSPEECH, 2020, pp. 4586–4590.
[17] R. Postuma, A. Lang, J. Gagnon, A. Pelletier, J. Montplaisir, How does parkinsonism start? Prodromal parkinsonism motor changes in idiopathic REM sleep behaviour disorder, Brain 135 (6) (2012) 1860–1870.
[18] B.E. Sakar, M.E. Isenkul, C.O. Sakar, A. Sertbas, F. Gurgen, S. Delil, H. Apaydin, O. Kursun, Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings, IEEE J. Biomed. Health Inform. 17 (4) (2013) 828–834.
[19] S. Saravanan, K. Ramkumar, K. Adalarasu, V. Sivanandam, S.R. Kumar, S. Stalin, R. Amirtharajan, A systematic review of artificial intelligence (AI) based approaches for the diagnosis of Parkinson's disease, Arch. Comput. Methods Eng. 29 (6) (2022) 3639–3653.
[20] Z.K. Senturk, Early diagnosis of Parkinson's disease using machine learning algorithms, Med. Hypotheses 138 (2020) 109603.
[21] P. Sharma, S. Sundaram, M. Sharma, A. Sharma, D. Gupta, Diagnosis of Parkinson's disease using modified grey wolf optimization, Cogn. Syst. Res. 54 (2019) 100–115.
[22] B. Suhas, J. Mallela, A. Illa, B. Yamini, N. Atchayaram, R. Yadav, D. Gope, P.K. Ghosh, Speech task based automatic classification of ALS and Parkinson's disease and their severity using log Mel spectrograms, in: International Conference on Signal Processing and Communications, 2020, pp. 1–5.
[23] V. Vapnik, The support vector method of function estimation, in: Nonlinear Modeling, 1998, pp. 55–85.
[24] O. Yaman, F. Ertam, T. Tuncer, Automated Parkinson's disease recognition based on statistical pooling method using acoustic features, Med. Hypotheses 135 (2020) 109483.