Use of ANN, LSTM and CNN Classifiers for the New MSCC and BSCC Methods in the Detection of Parkinson's Disease by Voice Analysis

Miyara Mounia[1], Boualoulou Nouhaila[2], Nsiri Benayad[3], Belhoussine Drissi Taoufiq[4]
Computer Science and Systems Laboratory (LIS)-Faculty of Science Ain Chock, University Hassan II, Casablanca, Morocco[1]
Laboratory Electrical and Industrial Engineering-Information Processing-Informatics and Logistics (GEITIIL)-Faculty of Science Ain Chock, University Hassan II, Casablanca, Morocco[2, 4]
Research Center STIS, M2CS-National Higher School of Arts and Craft Rabat (ENSAM), Mohammed V University in Rabat, Rabat, Morocco[2, 3]

Abstract—**Parkinson's disease (PD) is a neurodegenerative condition that impacts a significant global population. The timely and precise identification of PD plays a pivotal role in facilitating early intervention and the efficient management of the condition. Recently, speech analysis has emerged as a promising non-invasive technique for the detection of PD due to its accessibility and ability to reveal subtle vocal biomarkers associated with the disease. This research introduces an innovative approach utilizing Short-Time Fourier Transform (STFT) to generate spectrograms, specifically Bark Spectrogram Cepstral Coefficients (BSCC) and Mel Spectrogram Cepstral Coefficients (MSCC). These coefficients are compared with traditional and well-known coefficients, namely Mel-Frequency Cepstral Coefficients (MFCC) and Bark Frequency Cepstral Coefficients (BFCC). To extract the most effective coefficients for Parkinson's disease detection, three robust classification techniques—Long Short-Term Memory neural networks (LSTM), Convolutional Neural Networks (CNN), and Artificial Neural Networks (ANN)—are employed. As a result, the BSCC and MSCC algorithms achieve a maximum accuracy rate of 90%, surpassing the accuracy of the traditional MFCC and BFCC coefficients. Therefore, these newly proposed coefficients prove to be more precise in diagnosing Parkinson's disease compared to the conventional MFCC and BFCC coefficients.**

Keywords—Parkinson's Disease (PD); Bark Spectrogram Cepstral Coefficients (BSCC); Mel Spectrogram Cepstral Coefficients (MSCC); Long-Term Memory Neural Networks (LSTM); Convolutional Neural Networks (CNN); Artificial Neural Networks (ANN)

I. Introduction

Parkinson's disease (PD) ranks among the most widespread and incapacitating neurodegenerative conditions, impacting millions of individuals across the globe. Named after the pioneering work of British physician James Parkinson, who first documented its clinical manifestations in 1817, Parkinson's disease has emerged as a profound challenge in modern medicine. Characterized by its progressive deterioration of motor function, PD also presents an intricate array of non-motor symptoms, encompassing sleep disturbances, cognitive impairments, emotional alterations, and autonomic dysfunctions.

This neurological condition is mainly characterized by the deterioration of dopaminergic neurons located in the substantia nigra part of the brain, resulting in a significant reduction in dopamine production. The repercussions of this disruption within the central nervous system manifest through a range of distinctive clinical symptoms, including muscle rigidity, resting tremors, bradykinesia (slowness of voluntary movements), and postural instability.

The timely and precise detection of PD is of paramount importance in providing timely and effective medical interventions. While current diagnostic methods primarily rely on clinical evaluation and specialized imaging techniques, an expanding body of research suggests that significant insights into the early detection of Parkinson's disease may be gleaned from analyzing the human voice. This hypothesis is rooted in the notion that subtle vocal changes, often imperceptible to the human ear, may serve as early indicators of the disease. Leveraging the advancements in machine learning and voice analysis technologies, researchers are increasingly exploring the potential of voice-based biomarkers for PD diagnosis.

In the academic literature, there is increasing attention to the application of speech-based techniques using both machine learning and deep learning for the detection of Parkinson's disease. Numerous research papers have investigated the application of machine learning methodologies, including SVM [1] – [4], KNN [5], [6] , DT [7] , and genetic algorithms [26] in this context. Simultaneously, the PD identification has been addressed through the employment of established convolutional neural network (CNN) architectures such as AlexNet, DenseNet, LSTM, SqueezeNet, VGG19, and others, as well as custom-designed CNN architectures developed by researchers for deep learning investigations [8], [9]. Notably, CNN architectures have demonstrated enhanced performance in the domain of feature extraction [8].

Moreover, CNN techniques have exhibited success in various research domains, encompassing tasks like ocean noise detection [10], COVID-19 detection via X-ray images [11]–[13], mammography image segmentation and classification [14], classification of environmental sounds [15], Alzheimer's disease detection [16], skin cancer diagnosis [17], identification of cartilage lesions [18], fatigue diagnosis based on heart sounds [19], diagnosis of joint disorders [20], premature retinopathy assessment [21], and even the diagnosis of idiopathic Parkinson's disease [22].

In this research, we explore the use of spectrogram-based techniques, specifically BSCC (Bark Spectral Cepstral Coefficients) and MSCC (Mel Spectral Cepstral Coefficients), to extract pertinent vocal features for disease classification. Our dataset, sourced from the SAKAR database, comprises 38 audio recordings, encompassing 18 patients diagnosed with PD and 20 healthy individuals. The stratification of the dataset allows for a comprehensive examination of the proposed methodologies in distinguishing between healthy and affected individuals.

This paper is structured as follows: Section I presents an extensive review of the background, contextualizing the significance of voice analysis in Parkinson's disease diagnosis. Section II delves into related work. Section III is about database. Section IV outlines the proposed methodology, elucidating the intricacies of the BSCC and MSCC techniques and their application in our study. Section V and Section VI encompasses the discussion of results, presenting findings, and insights derived from our experiments. Lastly, we summarize our main findings and their implications for further research and clinical applications in Section VII.

II. Related Work

The study of Mehmet Bilal et al. outlines a new approach for detecting PD based on voice signals using LSTM and pre-trained deep networks. The process involves four steps: noise reduction, mel-spectrogram extraction, deep feature extraction using pre-trained ResNet models, and classification with an LSTM model. Experiments conducted with the PC-GITA dataset, a widely used dataset, demonstrate superior classification performance compared to existing methods, emphasizing the importance of early diagnosis for speech-related Parkinson's symptoms, an accuracy of 98.61% was attained [23], for Gaffari selik et al. In their study, the research leverages advanced deep learning and machine learning techniques to diagnose PD using voice signal datasets from both PD patients and healthy individuals (PDO_Dataset and PD_Dataset). The study examines existing machine learning and CNN algorithms for PD diagnosis, conducting a comparative performance analysis. Furthermore, a novel approach named SkipConNet + RF, combining RF and CNN, is introduced for PD detection. SkipConNet extracts crucial features from voice signals and then employs the RF algorithm for estimation. This approach significantly enhances RF algorithm performance, achieving an improvement ranging from 3% to 17.19%. Remarkably, the SkipConNet + RF method achieves remarkable accuracy, with a 98.30% on the PDO_Dataset dataset and 99.11% success rate on the PD_Dataset dataset, showcasing its potential as a highly effective tool for PD diagnosis [24] .

The article of Quan et al. presents a novel deep learning approach for detecting PD through voice signals. The approach utilizes time-distributed 2D-CNNs to extract dynamic time series features and employs a 1D-CNN to capture dependencies between these features. The model's performance was evaluated on two databases. On Database-1, it outperformed traditional machine learning models, achieving accuracies of 81.6% for sustained vowel /a/ and 75.3% for a short Chinese sentence. On Database-2, the model attained up to 92% accuracy across various speech tasks, including reading sentences in Spanish. The model's learned time series features effectively captured variability and the reduced frequency range in PD sounds, crucial for diagnosis. Furthermore, the study highlights the significance of the low-frequency region in Mel-spectrograms for PD recognition from voice, surpassing the influence of the high-frequency region [25], For Karan et al. Their study addresses the use of speech as an early marker for PD detection, given its impact on several speech components. To overcome challenges related to non-stationarity and discontinuity in speech signals, the researchers introduce a novel feature called IMFCC based on empirical mode decomposition. The performance of these proposed IMFCC features is evaluated using two datasets, each comprising 25 Parkinson-affected individuals and 20 normal. The findings demonstrate that IMFCC features offer significantly improved classification accuracy in both datasets, with an impressive increase of 10–20% compared to MFCC features. This suggests the potential of IMFCC as a highly effective tool for PD identification through voice analysis [26]. Chen et al. employed an architecture that utilized the HHT and KNN algorithms. They extracted a total of 21 characteristics, consisting of 12 from each sound sample using the HHT algorithm and nine using the LPCC algorithm. Subsequently, these extracted characteristics were individually classified using KNN, DT and RF algorithms. The authors reported the best performance using the KNN, achieving an accuracy of 93.3% [27].

The study of Tasnas et al. Investigate the connection between speech dysfunction and PD, particularly focusing on novel dysphonia measures aimed at predicting PD symptom severity through speech signals. A sum of 132 dysphonia metrics was calculated based on sustained vowel sounds. These measures were then reduced to four parsimonious subsets using feature selection algorithms. These subsets were utilized for binary classification, employing both RF and SVM as statistical classifiers. The research leveraged a database with 263 samples from 43 subjects and demonstrated that these novel dysphonia measures can achieve remarkable results, with an overall classification accuracy of nearly 99% using only ten dysphonia features. The study highlights the complementarity of these newly proposed measures with existing algorithms, enhancing the classifiers' ability to distinguish PD subjects from healthy controls. These findings represent a significant advancement towards non-invasive diagnostic decision support for PD [28]. Yaman et al. utilized a freely accessible dataset from the UCI dataset platform, comprising 240 speech samples, with 40 from PD patients and 40 from healthy ones. Their approach involved augmenting the dataset's attributes through the application of a statistical pooling method. Subsequently, they derived weighted features employing the ReliefF technique. These weighted feature vectors were then subjected to classification using both SVM and KNN techniques, yielding a commendable 91.25% success rate with the SVM algorithm and 91.23% with the KNN method [29]. The paper of Kavita Bhatt et al. Highlights the significance of early PD detection in the elderly, emphasizing common symptoms such as dysarthria, tremors, cognitive changes, and muscle stiffness. The study proposes a Deep Neural Network algorithm using spectrograms generated by the Superlet Transform for PD detection from speech signals. The method achieved an impressive 96% on the ItalianPVS dataset and 92% overall accuracy on the PC-GITA dataset, outperforming state-of-the-art methods in PD detection from diverse speech data sources [30].

The research conducted by Mahboobeh and colleagues focuses on the demanding task of diagnosing PD earlier, particularly focusing on gender-specific differences in speech characteristics. A hybrid method is proposed, leveraging features' scores based on a two-dimensional projection. After removing gender-specific features, the approach employs Classification-Based Feature Score (CBFS) and Statistical-Based Feature Score (SBFS) to rank the remaining features. Resampling enhances feature selection stability. Various classifiers (KNN, NSVM, LSVM, RF, and NB) are applied, achieving 84% and 86% accuracy rates for women and men, respectively, with fewer selected features compared to prior studies. The results highlight feature commonalities despite gender differences and validate using an independent dataset for added robustness [31]. The proposed model comprises three key stages. Firstly, noise is eliminated from the signals using the DWT-EMD and EMD-DWT methods. Secondly, MFCC and GTCC features are extracted from the enhanced audio signals. The final step involves classification, where these features are input into CNN and LSTM models designed to capture sequential information from the extracted features. In the experimental phase, the study employs the PC-GITA and Sakar datasets, applying a ten-fold cross-validation method. Impressively, the highest classification accuracy for the PC-GITA dataset, the accuracy reaches 100% for EMD-DWT-GTCC-CNN and 96.55% for DWT-EMD-GTCC-CNN. For the Sakar dataset reaches 100% for both the EMD-DWT-GTCC-CNN and DWT-EMD-GTCC-CNN combinations. These findings underscore the effectiveness of GTCC features over MFCC in Parkinson's disease assessment. This research showcases a promising avenue for accurate and timely detection of PD through speech analysis, potentially improving the effectiveness of telemedicine-based diagnosis and monitoring systems [32].

III. Database

The study utilizes a dataset from a prior investigation, comprising 38 voice recordings. This dataset includes 20 individuals diagnosed with Parkinson's disease (PD) and 18 individuals classified as healthy controls. During these recording sessions, participants were specifically instructed to articulate the vowel 'a' using a standard microphone with a sampling frequency of 44,100 Hz, and the recordings were conducted on a desktop computer equipped with a 16-bit sound card. These voice recordings form the foundation of our analysis [33].

The proposed methodology involves the utilization of advanced feature extraction techniques as shown Fig. 1, including MFCC, MSCC, and BSCC, BFCC which will be detailed subsequently. These extracted acoustic features are then fed into powerful classification models, such as ANN, CNN and LSTM. The integration of these cutting-edge feature extraction methods and state-of-the-art classification algorithms forms the core of our approach, ultimately facilitating the accurate differentiation between individuals affected by PD and those in a healthy control group. The subsequent sections will provide a comprehensive explanation of each method and its role in our classification framework.

IV. Proposed Method

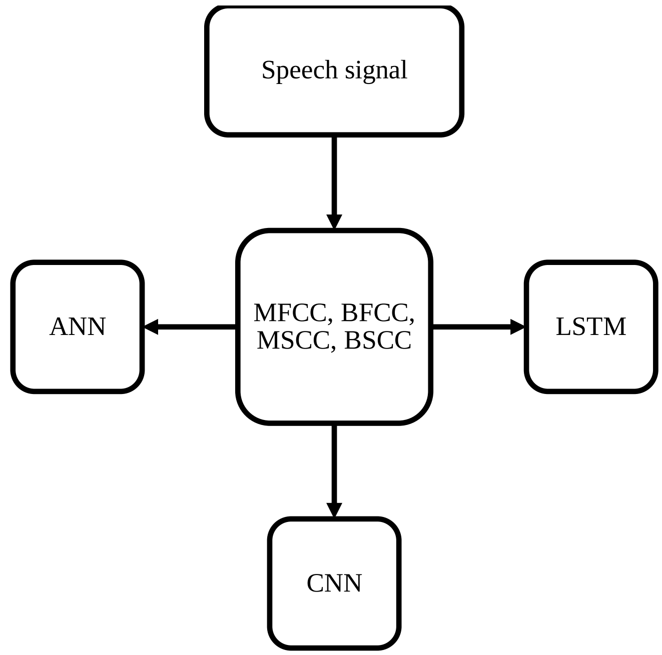

Fig. 1. The schematic for the proposed method.

A. MFCC

MFCC stands as a commonly employed method for extracting features in the realm of speech and audio signal processing, Fig. 2 depicts the different steps to follow in order to obtain the MFCC coefficients.

Pre-emphasis: The audio signal is pre-emphasized by applying a high-pass filter to amplify high-frequency components, as described this equation with k = 0.97.

$$H(z) = 1 - kz^{-1} \quad (1)$$

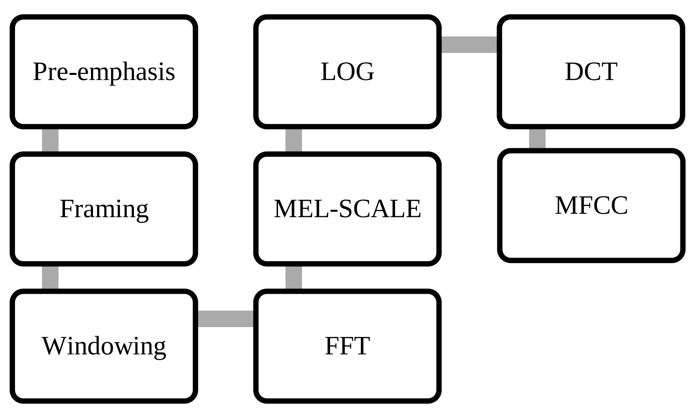

Fig. 2. The steps for calculating MFCC coefficients.

Framing: The signal with pre-emphasis undergoes segmentation into short, overlapping frames to capture temporal attributes.

Windowing: Each frame is windowed using the Hamming window function to prepare it for Fourier analysis.

$$w(n) = 0{,}54 - 0{,}46\cos\left(\frac{2\pi n}{N-1}\right) \quad (2)$$

FFT (Fast Fourier Transform): The FFT is applied to each frame to convert it into the frequency domain.

$$S_n = \sum_{k=0}^{N-1} S_k e^{-j2\pi\frac{kn}{N}} \quad (3)$$

Mel Scale: The resulting spectrum is transformed onto the Mel scale, which simulates human auditory perception.

$$\text{Mel} = 2595\log_{10}\left(1 + \frac{f}{700}\right) \quad (4)$$

Discrete Cosine Transform (DCT): The DCT is used to decorrelate the Mel-scaled spectrum, reducing redundancy.

$$c_i = \sqrt{\frac{2}{N}} \sum_{j=1}^{M} m_j \cos\left(\frac{\pi i}{N}(j - 0{,}5)\right) \quad (5)$$

MFCC Calculation: Finally, a subset of the DCT coefficients is selected as Mel-Frequency Cepstral Coefficients, which indicate the spectral properties of the audio signal.

B. BFCC

BFCC is similar to MFCC, but it uses the Bark scale instead of the Mel scale. In Fig. 3, the various stages required to acquire the BFCC coefficients are illustrated.

Pre-emphasis, Framing, Windowing, FFT: These initial steps are identical to those in MFCC.

Bark Scale: Instead of the Mel scale, the FFT results are transformed onto the Bark scale, which is another representation of auditory frequency perception.

$$\text{Bark}(f) = \frac{26{,}81f}{1960+f} + 0{,}53 \quad (6)$$

DCT: The DCT is applied to the Bark-scaled spectrum to reduce redundancy and extract cepstral coefficients.

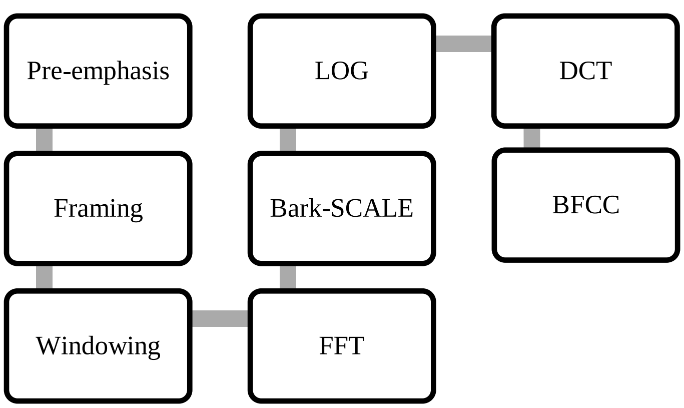

Fig. 3. The steps for calculating BFCC coefficients.

C. MSCC

MSCC is a variant of MFCC that directly uses the STFT instead of the FFT. The different procedures for obtaining the MSCC coefficients are presented in Fig. 4.

Pre-emphasis, Framing, Windowing: These steps remain the same as in MFCC.

STFT: Instead of the FFT, the STFT is used to attain a time-varying spectral representation for each frame.

Mel Scale and DCT: The Mel scale is implemented to the STFT results, followed by DCT, to calculate Mel Spectral Cepstral Coefficients.

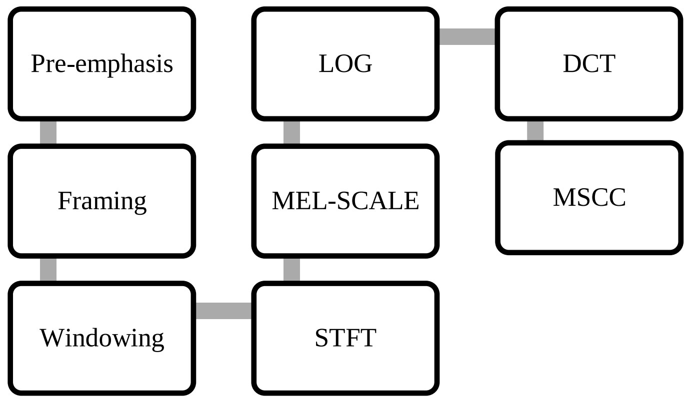

Fig. 4. The steps for calculating MSCC coefficients.

D. BSCC

BSCC is similar to MSCC, but it uses the Bark scale instead of the Mel scale in the final step. Fig. 5 outlines the sequential steps for deriving the BSCC coefficients.

Pre-emphasis, Framing, Windowing and STFT: These initial steps are identical to those in MSCC.

Bark Scale: Instead of the Mel scale, the STFT results are transformed onto the Bark scale, and then DCT is applied to extract Bark Spectral Cepstral Coefficients.

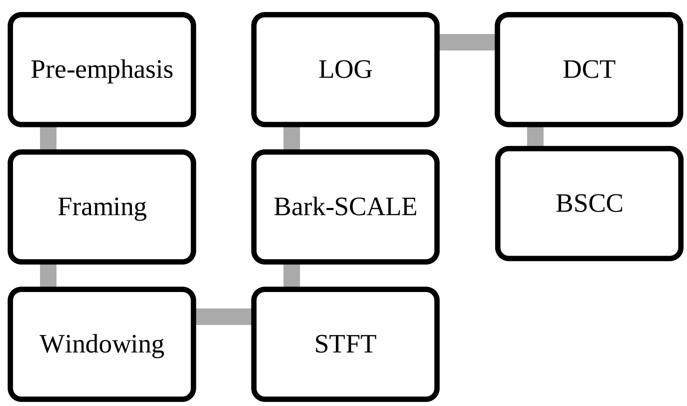

Fig. 5. The steps for calculating BSCC coefficients.

E. ANN

An Artificial Neural Network (ANN), modeled on the structure of biological neural networks, is a computational model composed of interconnected processing units, or neurons, organized into layers. These neurons process and transmit information through weighted connections. ANNs are designed for supervised learning and have the capacity to memorize complicated patterns and relationships within data. They include an input layer, one or more hidden layers, and an output layer. ANNs are widely employed for diversity of applications, including classification, regression, and function approximation.

F. LSTM

An LSTM is a form of RNN specifically created to tackle the challenge of the vanishing gradient issue encountered in conventional RNNs. LSTMs are notably effective for tasks that encompass sequential data. Time series data, or natural language processing. They incorporate specialized memory cells that can capture long-range dependencies in data, making them capable of retaining information over extended time intervals. LSTMs are crucial in applications such as speech recognition, and sentiment analysis, where understanding and modeling sequential patterns are essential.

G. CNN

A CNN is a deep learning structure mainly utilized for the analysis of images and spatial data. CNNs excel at automatically extracting hierarchical and spatial features from input data using convolutional layers, pooling layers, and fully connected layers. Convolutional layers apply convolutional operations to scan and detect local patterns within the input, making CNNs highly effective in image classification, object detection, and image generation tasks. They have also found applications beyond image processing, including in natural language processing and reinforcement learning.

V. RESULT

In this section, the experimental outcomes are presented and the performance of the new approaches is assessed, MSCC and BSCC, alongside the BFCC and MFCC, in the context of PD detection through speech analysis. We also discuss the outcomes obtained using ANN, LSTM, and CNN for classification. Fig. 6 to Fig. 9 sequentially display the MFCC, MSCC, BFCC, and BSCC coefficients obtained for an individual diagnosed with Parkinson's disease.

The Fig. 6, 7, 8, and 9 displays the initial twelve coefficient values of MFCC, MSCC, BFCC, and BSCC, respectively. These coefficients encompass numerous frames that demand significant processing time for classification, hindering the accurate diagnostic decision-making process. To address this issue, the average value of these frames is computed to obtain the voiceprint and mitigate the processing burden.

The results of our experiments reveal varying levels of accuracy across different feature extraction methods and classification algorithms. For Mel Spectrogram-based Cepstral Coefficients (MSCC), we observed high accuracy rates, with ANN achieving 90% accuracy and CNN also reaching 90%, underscoring their effectiveness. LSTM, while slightly lower at 81%, still performed well with MSCC. Similarly, Bark Spectrogram-based Cepstral Coefficients (BSCC) exhibited strong accuracy, with ANN achieving 81% accuracy and LSTM and CNN both reaching 90%. In contrast, conventional BFCC and MFCC yielded comparatively lower accuracy rates, with ANN achieving 54% and 64%, respectively, and LSTM and CNN hovering around 80%. These findings suggest that MSCC and BSCC may offer superior feature extraction capabilities in the field of diagnosing PD, potentially revolutionizing the field with their higher accuracy rates when coupled with advanced classification techniques like CNN and LSTM.

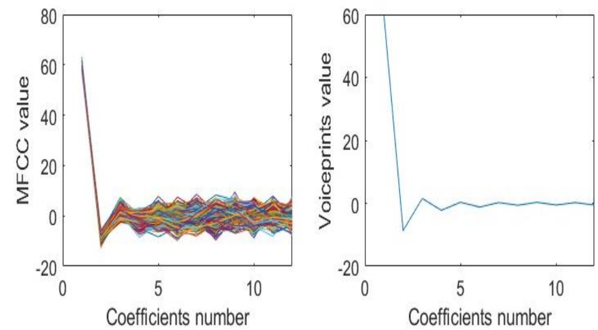

Fig. 6. The 12 MFCC coefficients for an individual with Parkinson's disease.

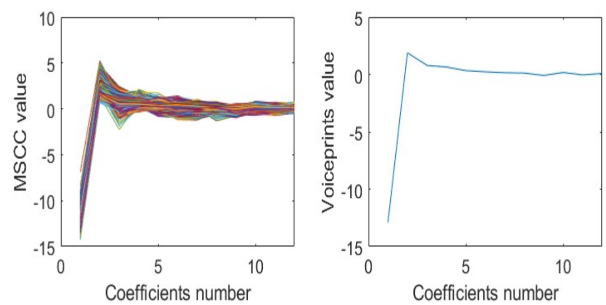

Fig. 7. The 12 MSCC coefficients for an individual with Parkinson's disease.

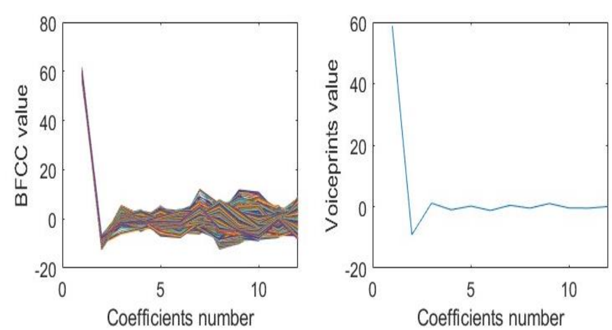

Fig. 8. The 12 BFCC coefficients for an individual with Parkinson's disease.

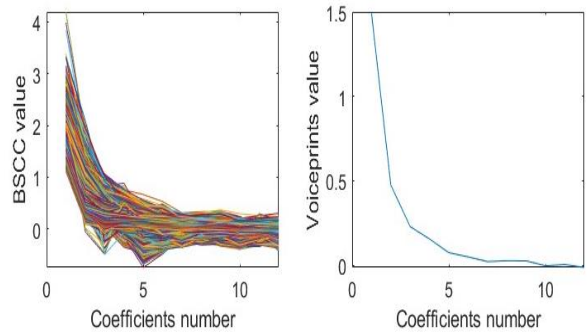

Fig. 9. The 12 BSCC coefficients for an individual with Parkinson's disease.

The achieved accuracy rates with MSCC and BSCC in combination with ANN, LSTM, and CNN are notable and indicate their promise as powerful features for PD detection. The consistent high accuracy rates of 90% with MSCC and BSCC when coupled with CNN highlight their efficacy in feature extraction. MSCC and BSCC capture spectral characteristics more effectively than traditional MFCC and BFCC, which may account for their superior performance. These findings suggest that MSCC and BSCC hold potential as valuable additions to the arsenal of voice-based PD detection methodologies.

In contrast, the results obtained using MFCC and BFCC, particularly with ANN, show comparatively lower accuracy rates of 54% and 64%, respectively. While these traditional coefficients have been widely used in voice analysis, our findings suggest that MSCC and BSCC surpass them in the context of PD detection. The improved accuracy achieved with MSCC and BSCC underscores their ability to capture nuanced vocal characteristics associated with PD more effectively.

TABLE I. RESULTS OBTAINED WITH THESE METHODS

	ANN			**LSTM**			**CNN**		
	Accuracy (%)	***Sensitivity (%)***	***Specificity (%)***	**Accuracy (%)**	**Sensitivity (%)**	**Specificity (%)**	**Accuracy (%)**	**Sensitivity (%)**	**Specificity (%)**
MSCC	90	60	87	81	50	50	90	50	100
BSCC	81	55	100	90	55	100	90	50	100
MFCC	54	100	100	81	66	100	81	71	100
BFCC	64	100	100	81	50	55	81	80	100

The accuracy rates obtained in this study are promising for the development of reliable voice-based PD detection systems. The utilization of MSCC and BSCC, combined with advanced deep learning techniques like CNN, offers a potential breakthrough in early PD diagnosis. These discoveries could hold noteworthy consequences for the development of non-invasive, cost-effective, and accessible PD screening tools, ultimately aiding in the timely intervention and management of this debilitating disease. Table I presents the results obtained for each of the MFCC, BFCC, MSCC, and BSCC coefficients using ANN, LSTM, and CNN.

In conclusion, our results suggest that MSCC and BSCC, in conjunction with CNN, represent a promising avenue for enhancing the accuracy of voice-based PD detection systems, potentially revolutionizing the way we diagnose and manage Parkinson's disease. Additional investigation and validation on larger datasets are warranted to confirm these findings and pave the way for practical clinical applications.

Three measures of performance were used in this study to evaluate the efficiency of classifiers on data sets: accuracy (see Eq. 7), sensitivity (see Eq. 8) and specificity (see Eq. 9). Accuracy is considered to be the percentage of precise results. Their definitions are as below:

$$\text{Accuracy} = \frac{\text{TN+TP}}{\text{TN+TP+FP+FN}} \quad (7)$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP+FN}} \quad (8)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN+FP}} \quad (9)$$

With:

Subjects without Parkinson's disease correctly categorized are True Negatives (TP).

Subjects with Parkinson's disease correctly categorized are True Positives (TN).

Subjects with Parkinson's disease incorrectly categorized are False Positives (FP).

Subjects without Parkinson's disease incorrectly categorized are False Negatives (FN).

VI. DISCUSSION

In the domain of diagnosing PD through vocal analysis, several methods have been proposed in the literature. This paper stands out, utilizing a combination of BFCC, MFCC, MSCC, and BSCC, combined with LSTM, ANN and CNN classifiers. Notably, the CNN-based method with MSCC ET BSCC achieved the highest accuracy of 90%, suggesting promising results for PD detection. Table II provides a comparison of this study with recently published articles.

TABLE II. COMPARISON WITH RECENT RESEARCH

Study	**Dataset**	**Method**	**Accuracy**
El-Hasnony et al.[34]	Sakar dataset	ANFIS + PSOGWO	87,5 %
Vasquez-Correa et al. [35]	Spanish datasets	CNN	89 %
Gunduz [36]	Sakar dataset	CNN	86,9 %
Sakar et al. [37]	Sakar dataset	SVM	86 %
Belhoussine et al. [38]	Sakar dataset	DWT-MFCC	86,84 %
Zayrit et al. [39]	Sakar dataset	SVM rbf SVM lin	81 % 79 %

In the domain of diagnosing PD through vocal analysis, several methods have been proposed in the literature. This paper stands out, utilizing a combination of BFCC, MFCC, MSCC, and BSCC, combined with LSTM, ANN and CNN classifiers. Notably, the CNN-based method with MSCC et BSCC achieved the highest accuracy of 90%, suggesting promising results for PD detection. Table II provides a comparison of this study with recently published articles.

Comparatively, previous research has explored various techniques for PD diagnosis through vocal analysis. El-Hasnony et al. [34] introduced a fog-based ANFIS+PSOGWO model, achieving an accuracy of 87.5% and outperforming other optimization methods Vasquez-Correa et al. [35] employed a CNN approach using STFT and continuous wavelet transform, achieving an accuracy of up to 89% in distinguishing PD patients from healthy speakers. Gunduz [36] proposed two CNN frameworks with deep feature extraction and achieved an accuracy of up to 86.90%. Sakar et al. [37] utilized tunable Q-factor wavelet transform for feature extraction and obtained an accuracy of up to 86%. Belhoussine et al. [38] focused on optimized wavelet selection in combination with MFCC features and SVM classification, achieving an accuracy of 86.84%. Finally, Zayrit et al. [39] used SVM classification with RBF kernel with Daubechies wavelet transform and MFCC features, obtaining an accuracy of 81%, but an accuracy of 79% by using SVM with linear Kernel.

Comparing these methods to the novel approach, which leverages MSCC, and BSCC with CNN classification to reach a 90% accuracy rate, it is evident that these methods outperform most of the previously mentioned techniques. This suggests that the combination of these advanced cepstral coefficients and CNN classification represent a significant advancement in the field of PD diagnosis through vocal analysis, potentially offering more accurate and reliable results. Further research and validation are necessary to confirm these findings and assess the practicality of implementing the proposed method in clinical settings.

VII. CONCLUSION

In summary, this document has delved into the exploration of Parkinson's disease through the lens of voice analysis, employing various feature extraction methods MFCC, MSCC, BFCC, and BSCC, coupled with a classification approach employing ANN, CNN, and LSTM. Notably, our findings consistently demonstrate that MFCC and BFCC methods consistently outperform others, achieving an impressive accuracy rate of 90%. These results underscore the potential of voice-based diagnostic tools in advancing our understanding and early identification of PD, highlighting the promising avenues for future research and clinical applications in this critical domaine. Subsequent investigations into the diagnosis of PD ailment aim to employ diverse neural network architectures and algorithms for feature selection on an expanded dataset.

REFERENCES

[1] T. Zhang, Y. Zhang, H. Sun, and H. Shan, "Parkinson disease detection using energy direction features based on EMD from voice signal," Biocybern Biomed Eng, vol. 41, no. 1, pp. 127–141, Jan. 2021, doi: 10.1016/j.bbe.2020.12.009.

[2] A. M. García et al., "Cognitive Determinants of Dysarthria in Parkinson's Disease : An Automated Machine Learning Approach," Movement Disorders, vol. 36, no. 12, pp. 2862–2873, Dec. 2021, doi: 10.1002/mds.28751.

[3] N. P. Narendra, B. Schuller, and P. Alku, "The Detection of Parkinson's Disease From Speech Using Voice Source Information," IEEE/ACM Trans Audio Speech Lang Process, vol. 29, pp. 1925–1936, 2021, doi: 10.1109/TASLP.2021.3078364.

[4] H. Gunduz, "An efficient dimensionality reduction method using filter-based feature selection and variational autoencoders on Parkinson's disease classification," Biomed Signal Process Control, vol. 66, p. 102452, Apr. 2021, doi: 10.1016/j.bspc.2021.102452.

[5] Z. Soumaya, B. Taoufiq, N. Benayad, B. Achraf, and A. Ammoumou, "A hybrid method for the diagnosis and classifying parkinson's patients based on time–frequency domain properties and K-nearest neighbor," J Med Signals Sens, vol. 10, no. 1, p. 60, 2020, doi: 10.4103/jmss.JMSS_61_18.

[6] 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," Biocybern Biomed Eng, vol. 40, no. 1, pp. 211–220, Jan. 2020, doi: 10.1016/j.bbe.2019.05.006.

[7] B. Nouhaila, B. D. Taoufiq, and N. Benayad, "An Intelligent Approach based on the Combination of the Discrete Wavelet Transform, Delta Delta MFCC for Parkinson's Disease Diagnosis," International Journal of Advanced Computer Science and Applications, vol. 13, no. 4, pp. 562–571, 2022, doi: 10.14569/IJACSA.2022.0130466.

[8] M. B. Er, E. Isik, and I. Isik, "Parkinson's detection based on combined CNN and LSTM using enhanced speech signals with Variational mode decomposition," Biomed Signal Process Control, vol. 70, p. 103006, Sep. 2021, doi: 10.1016/j.bspc.2021.103006.

[9] L. Zahid et al., "A Spectrogram-Based Deep Feature Assisted Computer-Aided Diagnostic System for Parkinson's Disease," IEEE Access, vol. 8, pp. 35482–35495, 2020, doi: 10.1109/ACCESS.2020.2974008.

[10] B. Mishachandar and S. Vairamuthu, "Diverse ocean noise classification using deep learning," Applied Acoustics, vol. 181, p. 108141, Oct. 2021, doi: 10.1016/j.apacoust.2021.108141.

[11] S. Sheykhivand et al., "Developing an efficient deep neural network for automatic detection of COVID-19 using chest X-ray images," Alexandria Engineering Journal, vol. 60, no. 3, pp. 2885–2903, Jun. 2021, doi: 10.1016/j.aej.2021.01.011.

[12] M. Aminu, N. A. Ahmad, and M. H. Mohd Noor, "Covid-19 detection via deep neural network and occlusion sensitivity maps," Alexandria Engineering Journal, vol. 60, no. 5, pp. 4829–4855, Oct. 2021, doi: 10.1016/j.aej.2021.03.052.

[13] K. Shankar et al., "An optimal cascaded recurrent neural network for intelligent COVID-19 detection using Chest X-ray images," Appl Soft Comput, vol. 113, p. 107878, Dec. 2021, doi: 10.1016/j.asoc.2021.107878.

[14] W. M. Salama and M. H. Aly, "Deep learning in mammography images segmentation and classification: Automated CNN approach," Alexandria Engineering Journal, vol. 60, no. 5, pp. 4701–4709, Oct. 2021, doi: 10.1016/j.aej.2021.03.048.

[15] A. M. Tripathi and A. Mishra, "Self-supervised learning for Environmental Sound Classification," Applied Acoustics, vol. 182, p. 108183, Nov. 2021, doi: 10.1016/j.apacoust.2021.108183.

[16] M. EL-Geneedy, H. E.-D. Moustafa, F. Khalifa, H. Khater, and E. AbdElhalim, "An MRI-based deep learning approach for accurate detection of Alzheimer's disease," Alexandria Engineering Journal, vol. 63, pp. 211–221, Jan. 2023, doi: 10.1016/j.aej.2022.07.062.

[17] A. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, no. 7639, pp. 115–118, Feb. 2017, doi: 10.1038/nature21056.

[18] F. Liu et al., "Deep Learning Approach for Evaluating Knee MR Images: Achieving High Diagnostic Performance for Cartilage Lesion Detection," Radiology, vol. 289, no. 1, pp. 160–169, Oct. 2018, doi: 10.1148/radiol.2018172986.

[19] C. Yin et al., "Diagnosis of exercise-induced cardiac fatigue based on deep learning and heart sounds," Applied Acoustics, vol. 197, p. 108900, Aug. 2022, doi: 10.1016/j.apacoust.2022.108900.

[20] U. Taşkıran and M. Çunkaş, "A deep learning based decision support system for diagnosis of Temporomandibular joint disorder," Applied Acoustics, vol. 182, p. 108292, Nov. 2021, doi: 10.1016/j.apacoust.2021.108292.

[21] J. M. Brown et al., "Automated Diagnosis of Plus Disease in Retinopathy of Prematurity Using Deep Convolutional Neural Networks," JAMA Ophthalmol, vol. 136, no. 7, p. 803, Jul. 2018, doi: 10.1001/jamaophthalmol.2018.1934.

[22] D. H. Shin et al., "Automated assessment of the substantia nigra on susceptibility map-weighted imaging using deep convolutional neural networks for diagnosis of Idiopathic Parkinson's disease," Parkinsonism Relat Disord, vol. 85, pp. 84–90, Apr. 2021, doi: 10.1016/j.parkreldis.2021.03.004.

[23] M. B. Er, E. Isik, and I. Isik, "Parkinson's detection based on combined CNN and LSTM using enhanced speech signals with Variational mode decomposition," Biomed Signal Process Control, vol. 70, p. 103006, Sep. 2021, doi: 10.1016/j.bspc.2021.103006.

[24] G. Celik and E. Başaran, "Proposing a new approach based on convolutional neural networks and random forest for the diagnosis of Parkinson's disease from speech signals," Applied Acoustics, vol. 211, p. 109476, Aug. 2023, doi: 10.1016/j.apacoust.2023.109476.

[25] C. Quan, K. Ren, Z. Luo, Z. Chen, and Y. Ling, "End-to-end deep learning approach for Parkinson's disease detection from speech signals," Biocybern Biomed Eng, vol. 42, no. 2, pp. 556–574, Apr. 2022, doi: 10.1016/j.bbe.2022.04.002.

[26] B. Karan, S. S. Sahu, and K. Mahto, "Parkinson disease prediction using intrinsic mode function based features from speech signal," Biocybern Biomed Eng, vol. 40, no. 1, pp. 249–264, Jan. 2020, doi: 10.1016/j.bbe.2019.05.005.

[27] L. Chen, C. Wang, J. Chen, Z. Xiang, and X. Hu, "Voice Disorder Identification by using Hilbert-Huang Transform (HHT) and K Nearest Neighbor (KNN)," Journal of Voice, vol. 35, no. 6, pp. 932.e1-932.e11, Nov. 2021, doi: 10.1016/j.jvoice.2020.03.009.

[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, no. 5, pp. 1264–1271, May 2012, doi: 10.1109/TBME.2012.2183367.

[29] O. Yaman, F. Ertam, and T. Tuncer, "Automated Parkinson's disease recognition based on statistical pooling method using acoustic features," Med Hypotheses, vol. 135, p. 109483, Feb. 2020, doi: 10.1016/j.mehy.2019.109483.

[30] K. Bhatt, N. Jayanthi, and M. Kumar, "High-resolution superlet transform based techniques for Parkinson's disease detection using speech signal," Applied Acoustics, vol. 214, p. 109657, Nov. 2023, doi: 10.1016/j.apacoust.2023.109657.

[31] M. Hasanzadeh and H. Mahmoodian, "A novel hybrid method for feature selection based on gender analysis for early Parkinson's disease diagnosis using speech analysis," Applied Acoustics, vol. 211, p. 109561, Aug. 2023, doi: 10.1016/j.apacoust.2023.109561.

[32] N. Boualoulou, T. Belhoussine Drissi, and B. Nsiri, "Cnn and Lstm for the Classification of Parkinson's Disease Based on the GTCC and MFCC," Applied Computer Science, vol. 19, no. 2, pp. 1–24, Jun. 2023, doi: 10.35784/acs-2023-11.

[33] B. E. Sakar et al., "Collection and Analysis of a Parkinson Speech Dataset With Multiple Types of Sound Recordings," IEEE J Biomed Health Inform, vol. 17, no. 4, pp. 828–834, Jul. 2013, doi: 10.1109/JBHI.2013.2245674.

[34] I. M. El-Hasnony, S. I. Barakat, and R. R. Mostafa, "Optimized ANFIS Model Using Hybrid Metaheuristic Algorithms for Parkinson's Disease Prediction in IoT Environment," IEEE Access, vol. 8, pp. 119252–119270, 2020, doi: 10.1109/ACCESS.2020.3005614.

[35] J. C. Vásquez-Correa, J. R. Orozco-Arroyave, and E. Nöth, "Convolutional neural network to model articulation impairments in patients with Parkinson's disease," in Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, International Speech Communication Association, 2017, pp. 314–318. doi: 10.21437/Interspeech.2017-1078.

[36] 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.

[37] 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," Appl Soft Comput, vol. 74, pp. 255–263, Jan. 2019, doi: 10.1016/j.asoc.2018.10.022.

[38] T. B. DRISSI, S. ZAYRIT, B. NSIRI, and A. AMMOUMMOU, "Diagnosis of Parkinson's Disease based on Wavelet Transform and Mel Frequency Cepstral Coefficients," International Journal of Advanced Computer Science and Applications, vol. 10, no. 3, 2019, doi: 10.14569/IJACSA.2019.0100315.

[39] Z. Soumaya, B. D. Taoufiq, N. Benayad, B Achraf, & A. Ammoumou, (2020). A hybrid method for the diagnosis and classifying parkinson's patients based on time–frequency domain properties and K-nearest neighbor. Journal of medical signals and sensors, 10(1), 60.