Applied Computer Science, vol. 19, no. 2, pp. 1–24
doi: 10.35784/acs-2023-11

Submitted: 2023-04-26 | Revised: 2023-05-17 | Accepted: 2023-05-23

Keywords: Parkinson's disease, voice signal, GTCC, MFCC, DWT, EMD, CNN and LSTM

Nouhaila BOUALOULOU [0000-0001-8565-8239]*,
Taoufiq BELHOUSSINE DRISSI [0000-0003-2958-070X]*,
Benayad NSIRI [0000-0003-3885-9534]**

CNN AND LSTM FOR THE CLASSIFICATION OF PARKINSON'S DISEASE BASED ON THE GTCC AND MFCC

Abstract

Parkinson's disease is a recognizable clinical syndrome with a variety of causes and clinical presentations; it represents a rapidly growing neurodegenerative disorder. Since about 90 percent of Parkinson's disease sufferers have some form of early speech impairment, recent studies on tele diagnosis of Parkinson's disease have focused on the recognition of voice impairments from vowel phonations or the subjects' discourse. This paper presents a new approach for Parkinson's disease detection from speech sounds that are based on CNN and LSTM and uses two categories of characteristics. These are Mel Frequency Cepstral Coefficients (MFCC) and Gammatone Cepstral Coefficients (GTCC) obtained from noise-removed speech signals with comparative EMD-DWT and DWT-EMD analysis. The proposed model is divided into three stages. In the first step, noise is removed from the signals using the EMD-DWT and DWT-EMD methods. In the second step, the GTCC and MFCC are extracted from the enhanced audio signals. The classification process is carried out in the third step by feeding these features into the LSTM and CNN models, which are designed to define sequential information from the extracted features. The experiments are performed using PC-GITA and Sakar datasets and 10-fold cross validation method, the highest classification accuracy for the Sakar dataset reached 100% for both EMD-DWT-GTCC-CNN and DWT-EMD-GTCC-CNN, and for the PC-GITA dataset, the accuracy is reached 100% for EMD-DWT-GTCC-CNN and 96.55% for DWT-EMD-GTCC-CNN. The results of this study indicate that the characteristics of GTCC are more appropriate and accurate for the assessment of PD than MFCC.

* Laboratory Electrical and Industrial Engineering, Information Processing, Informatics, and Logistics (GEITIIL), Faculty of Science Ain Chock, University Hassan II, Casablanca, Morocco, boualoulounouha@gmail.com, taoufiq_belhoussine_drissi@yahoo.fr
** Research Center STIS, M2CS, National Higher School of Arts and Craft, Rabat (ENSAM), Mohammed V University in Rabat, Morocco, nsiri2000@yahoo.fr

1. INTRODUCTION

In 1817, James Parkinson identified Parkinson's disease (PD), which is the second most frequent neurodegenerative illness after Alzheimer's disease. It is marked by a depletion of dopaminergic neurons in the compact substantia nigra (in the midbrain) (Oyinloye et al., 2021). Currently, PD is primarily identified by clinical examination by a neurologist. The diagnosis is usually made if at least two of the following three signs are present: akinesia (inability to initiate movement), rigidity, and resting tremor. However, these motor signs show up only after 50-60% of the dopamine neurons in the substantia nigra and 60-80% of their striatal endings have been lost. Finding methods to diagnose the disease earlier, and possibly to delay its advancement at an early stage, is therefore a key issue for research. Among the numerous clinical symptoms of this illness, the modification of the patient's speech seems to be an element of interest in several aspects.

In this context, the speech signal is presented as a convolution product between the vocal cord vibration as the source signal and the vocal tract as a filter. Moving to the spectral field, this convolution product turns into a scalar product in which it is challenging to distinguish the source and vocal tract contribution, but the relevant information comes from the vocal tract. Therefore, an additional signal conversion is performed to isolate the characteristics of the vocal tract, called the cepstrum. It allows decorrelating the source of the filter by transforming the scalar product into a sum. Several types of parameters can be calculated to describe the cepstrum, such as linear frequency domain prediction cepstral coefficients (FDLPCCs) which have been used for dialect classification (Kethireddy et al., 2022), as well as MFCC coefficients for the early detection of COVID-19 (Dash et al., 2021), we also find the MFCC and HFCC coefficients that are extracted from speech signals to recognize speech emotions (Nagarajan et al., 2020), moreover, articulatory cepstral coefficients (ACC) features are used for Speech recognition (Najnin & Banerjee, 2019). Perceptual Linear Prediction Cepstral Coefficients (PLPCCs) and Mel-frequency cepstral coefficients (MFCCs) used in automatic detection of people with Neurogenic Voice Disorders (NVDs) (Yagnavajjula et al., 2022). Furthermore, in the literature, several studies have employed characteristics derived from the spectral and cepstral analysis of the speech signal. Ali et al. employed two groups of characteristics: the auditory processed spectrum and all-pole model-based cepstral coefficients (APCC) and a Gaussian mixture model for automatic decision (Ali et al., 2016). Moro-Velázquez et al. employed a set of novel and previously suggested measures by application of the modulation spectrum approach, these metrics were optimized taking into account diverse time and frequency bands, to maximize the effectiveness of pathological voice recognition (Moro-Velázquez et al., 2016). In addition, the pathological voice detection due to the EMD-DWT analysis and the characteristics of Higher Order Statistics (HOS) as well as the characteristics of DWT coefficients, the implemented system gives promising results for the SVD database (Hammami et al., 2020).

Different researchers have used various approaches and methods to diagnose Parkinson's disease from the speech signal. A powerful new technique based on multilevel characteristic extraction has been suggested to predict Parkinson's disease from the characteristics of speech recordings of previously identified cases. Characteristic extraction was carried out using the chi-square and L1 norm (CLS) SVM algorithms. In the classification phase, a variety of popular classifiers including KNN, SVM, and DT were employed for machine learning, and the highest efficiency was achieved with the KNN classifier (Demir et al., 2022). The paper

of Terriza et al. introduces a feasibility analysis for a clinical decision aid system for Parkinson's diagnosis and treatment based on acoustic functionalities of laughter using voice recognition processes and automatic classification techniques by assessing various cepstral coefficients to distinguish laughter features of diseased and healthy subjects conjunction with machine learning classification systems, In a database of 20,000 randomly chosen samples from a set of 120 laughs from good health and Parkinson's disease persons, the decision aid system achieved 83% accuracy with an AUC value of 0.86 for the classification of laughter from healthy and Parkinson's disease participants (Terriza et al., 2022). Another paper by Qin et al. proposes an enhanced genetic algorithm coupled with a data optimization tool for Parkinson's disease voice signal identification. These tools, in particular, extract representative features from a speech signal using L1 regularization SVM and then improve the data of the representative features using the SMOTE algorithm. The improved and original features are then employed to train an SVM classifier for voice signal identification. The optimal SVM parameters were calculated utilizing an enhanced genetic algorithm (Qin et al., 2022).

Recent studies have focused more specifically on the early detection of PD by voice as the case of Quan's model extracts the dynamic features of the time series using two-dimensional convolutional neural networks (2D-CNN) distributed in time and then captures the dependencies between these time series using a one-dimensional CNN (1D-CNN). The suggested approach yielded good results, with an accuracy of 81.6% for the vowel /a/ (Quan et al., 2022). Belhoussine Drissi et al. used an advanced signal processing method DWT combined with cepstral coefficients MFCC and SVM classifier; they obtained an accuracy up to 86.84% (Drissi et al., 2019). For Zayrit et al. they applied DWT precisely wavelet of debauchies at level 2 in the third scale from which they extracted the MFCC cepstral coefficients and they employed the KNN classifier, they found an accuracy up to 89% and 98.68%, when applied to 52% and 73% of the database as training data, respectively (Soumaya et al., 2020). For the first time, Sakar et al. used the tunable Q-factor wavelet transform (TQWT) to extract characteristics from the voice signals of Parkinson's disease subjects. The greatest accuracy, up to 85%, is achieved by inputting the extracted characteristics from the normal Q-factor wavelet transform into the multilayer perceptron classifier using the relatively high Q-factor analysis (chosen as 2) (C. O. Sakar et al., 2019). To detect PD from speech sounds, a novel approach based on long-term memory (LSTM) and pre-trained deep networks using mel-spectrograms obtained from noise-free speech signals with variational mode decomposition (VMD). When the learning rate is 0.001 and the batch size is 128, ResNet-101 + LSTM achieves a significant accuracy rate of 98.61% (Er et al., 2021).

The present study proposes less costly computational methods for identifying and categorizing PD in the Sakar Speech Database and the PC-GITA database. For this purpose, the speech signal id processed using two methods. The first is EMD-DWT applied in two stages. In the first stage, the decomposition process is carried out in the empirical mode, while in the second stage the obtained IMFs rre decomposed again using a discrete wavelet transform, in particular Debauchies wavelets at level 2 on the third scale. For the second method DWT-EMD, the voice signals are first decomposed by utilizing the discrete wavelet transform. After that, an empirical modal decomposition is applied to the obtained a3 approximation coefficients, from which the corresponding IMF is obtained. Then several MFCC and GTCC coefficients features are calculated for PD and non-PD voice signals. These features extracted from both methods are fed as an input for the CNN and LSTM classifiers for the classification.

The remainder of the document is arranged as follows. In Section 2, the authors present speech signal datasets, as well as feature extraction (DWT, EMD, GTCC and MFCC), and for classification, they define CNN and LSTM processes. Section 3 presents, analyzes and discusses the results based on overall accuracy, sensitivity and specificity. Section 4 contains conclusions.

2. MATERIALS AND METHODS

In this paper, the proposal for the categorization and detection of PD through voice is divided into two main phases. The first phase focuses on feature extraction, in which the voice signal is conducted into a collection of GTCC and MFCC coefficients features using EMD-DWT analysis and DWT-EMD. In the second phase, these characteristics, which are organized into adequate vectors, as an input to the CNN and LSTM classifiers are used. Figure 1 shows the suggested model. The operation of each stage of the suggested learning model is explained briefly below.

2.1. Data acquisition

Database 1: This paper uses a speech database collected from two populations, namely people with and without PD. The database was collected from 38 participants, 20 PD subjects (6 females and 14 males) and 18 healthy ones (10 females and 8 males) at the Department of Neurology in Cerrahpasa Faculty of Medicine, Istanbul University. The age of the patients in the PD group varied from 43 to 77 years with an average of 64.86 and a standard deviation of 8.97, while the age of the participants in the normal group varied from 45 to 83 years with an average of 62.55 and a standard deviation of 10.79. Data were collected using a Trust MC-1500 microphone with a frequency response of 50 Hz to 13 kHz. During the recording of vocal phonations, the microphone was held 10 cm away from each subject. Each of the participants was requested to utter the vowel /a/ (B. E. Sakar et al., 2013).

Database 2: A PC-GITA dataset was used, which was collected with a professional device from 50 healthy subjects and 50 subjects with Parkinson's disease at 16-bit resolution and 44.1 kHz sampling rate. Women aged 43-76 years (mean age 61.4 years, SD 9.5 years) and men aged 31-86 years (mean age 60.5 years and SD 9.4 years) were considered healthy. While Parkinson's disease affected women aged 49 to 75 and men aged 33 to 81 (Orozco-Arroyave et al., n.d.).

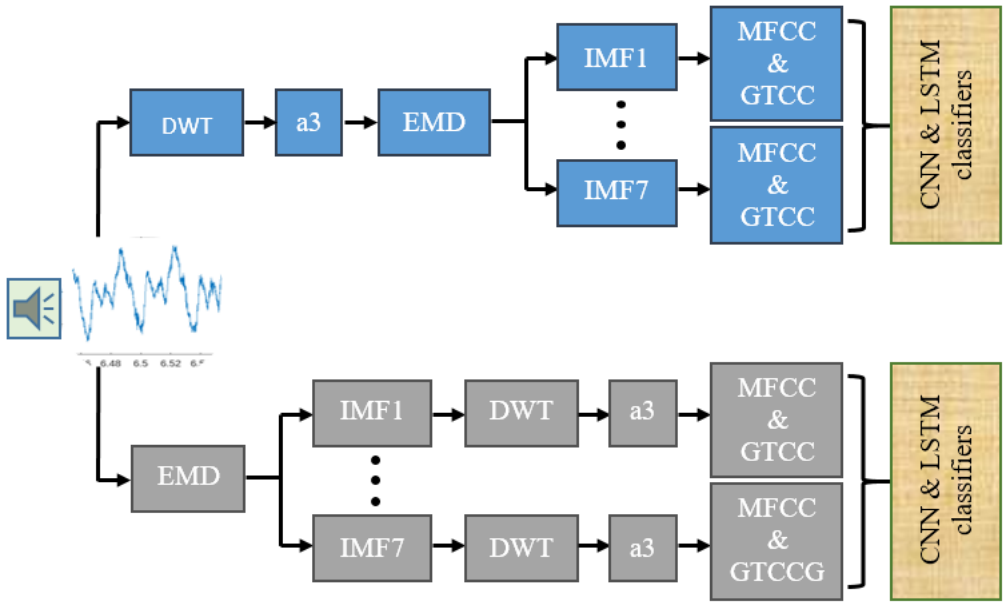

Fig. 1. The suggested method scheme

2.2. Feature extraction

A novel set of high-order statistical features has been captured to recognize speech disorders. As previously stated, the majority of current techniques are focused on traditional parameters such as jitter, formant, pitch, etc., which are challenging to be captured due to the aperiodic nature of speech signals. Furthermore, these characteristics concerning its instability may have a shared value for both PD and healthy voices, making recognition more complicated. In this area, we attempt to identify and categorize speech disorders by extracting GTCC and MFCC parameters from a DWT and EMD combination and its permutation. Wavelet transform analysis is largely used in the field of voice detection. Its coefficients have yielded several characteristics. In this paper, we use EMD and DWT, which are accurate for speech signal processing, in a two-step analysis to design a reliable system capable of classifying voice disorders to diagnose Parkinson's disease utilizing a deep learning techniquess the long short-term memory (LSTM) classifiers and convolutional neural network (CNN).

2.2.1. Mel frequency cepstral coefficients

MFCC: are acoustic parameters first introduced in 1980 by Davis and Mermelstein for automatic speech recognition (Davis & Mermelstein, 1980). These parameters form the spectral envelope of the voice and thus represent the volume and shape of the vocal tract. The MFCCs exploit both the decorrelation properties of the cepstrum and the psychoacoustic principles of the human ear. The calculation of these coefficients is done in seven steps:

– Pre-emphasis: The human ear hears high frequencies better than low frequencies. To reflect this perception, the first step of pre-emphasis is performed by emphasizing the high frequencies by a finite impulse response filter of the first order as shown in the following equation:

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

With k as the pre-emphasis coefficient, it is between $0.9 \leq k \leq 1$.
In our experiment, we pre-emphasized the speech signal samples $\{s_n, n = 1, ..., N\}$ by the following equation, to which we assigned 0.97 to k:

$$s_n' = s_n - ks_{n-1} \tag{2}$$

– Segmentation and windowing: The temporal sound signal is divided into frames of 20 to 40 ms. In our work, we use a frame size of 25 ms and consider the voice signal to be stationary during this time. To avoid abrupt transitions, two consecutive frames are overlapped. According to the following equation, signal multiplication by a Hamming window function allows to avoid of discontinuities between two consecutive frames:

$$s_n'' = \left\{0{,}54 - 0{,}46\cos\left(\frac{2\pi n}{N-1}\right)\right\} s_n' \tag{3}$$

– Fast Fourier Transform (FFT): This stage allows the signal to be transformed from the time to the frequency domain using the following formula:

$$s_n' = \sum_{k=0}^{N-1} s_k e^{-j2\pi\frac{kn}{N}} \tag{4}$$

– Mel filter bank: The human ear can be described as a set of band pass filters more spaced for high frequencies than for low frequencies, implying for humans a better discrimination of two close frequencies in the low frequencies than in the high frequencies. The Mel scale (whose unit is the Mel) was created to reflect this aspect. The Mel-Hertz translation is linear for low frequencies and logarithmic for high frequencies, this Mel scale is obtained according to the next formula:

$$\text{Mel}(f) = 2595\log_{10}\left(1 + \frac{f}{100}\right) \tag{5}$$

– The logarithm of the energies: The logarithm of the energies associated with each filter is then computed to correspond to the human logarithmic perception of sound energy. The logarithm also permits the decorrelation of the filter and the source signal,

it converts the product of the spectral density of the source and the filter into an addition.

– Discrete Cosine Transform (DCT): A DCT is performed on the log energies from the Mel filters. This has the effect of grouping the information of the spectral envelope in its first coefficients, these coefficients are the MFCC, and in our work, we have extracted only 12 MFCC coefficients.

$$c_i = \sqrt{\frac{2}{N}} \sum_{j=1}^{M} m_j \cos\left(\frac{\pi i}{N}(j-0{,}5)\right) \tag{6}$$

In this study, we defined M (number of filters) as 20

– Liftering: allows raising the cepstral coefficients to obtain nearly identical amplitudes because the higher order of these coefficients is too minimal. This step is realized by using the formula below

$$c_n' = \left(1+\frac{L}{2}\sin\left(\frac{\pi n}{L}\right)\right)c_n \tag{7}$$

2.2.2. Gammatone cepstral coefficients (GTCC)

Gammatone cepstral coefficients (GTCCs) are a class of feature representation used in speech processing and analysis. They are derived from the gammatone filterbank based on an Equivalent Rectangular Bandwidth (ERB) scale, a type of filterbank that mimics the way the human auditory system processes sound. They are used to capture the spectral characteristics of speech signals.

The cepstral coefficients of the proposed gammatones are calculated in a similar way to the MFCC extraction in previous steps. They are calculated as the cepstral coefficients of the log magnitude response of a gammatone filterbank with the expression (8) below. The cepstral coefficients are calculated using the discrete cosine transform (DCT) of the logarithm of the frequency response magnitude of a signal. This transform facilitates the analysis of the spectral content of speech signals by decorrelating spectral characteristics and modeling human perception of loudness. The GTCC is determined as follows:

$$g(t) = kt^{(n-1)}e^{-2\pi Bt}\cos(2\pi f_c t+\varphi) \qquad t \succ 0 \tag{8}$$

where: g(t) – is the Impulse Response of Gammatone Filter,
n – is the filter order,
K – is the amplitude factor,
fc – is the central frequency in Hertz,
φ – is the phase shift,
B – is the duration of the impulse response.

$$GTCC_m = \sqrt{\frac{2}{N}} \sum_{n=1}^{N} \log(X_n) \cos\left[\frac{\pi n}{N}(m-\frac{1}{2})\right] \quad 1 \leq m \leq M \tag{9}$$

where: X_n – is the energy of the signal in the n th spectral band,
N – is the number of Gammatone filters (in our study 20),
M – is the number of GTCC (M = 12 in this article).

Previous studies show that the GTCC coefficients, which are based on an auditory periphery model, perform significantly better than the MFCC features under noise conditions (Valero & Alias, 2012). In this study, we extracted these two coefficients using two combined EMD-DWT and DWT-EMD signal processing methods in order to select the appropriate coefficient to detect Parkinson's disease patients.

2.2.3. Empirical mode decomposition

Empirical Mode Decomposition (EMD) is an approach introduced by Huang et al. for the processing of non-stationary signals (Huang et al., 1998). It provides a self-adaptive decomposition because it does not involve external basis functions, but they are obtained from the signal itself. EMD allows the decomposition of an oscillating signal into fundamental contributions known as Intrinsic Mode Functions (IMFs), which are amplitude and frequency-modulated components (AM-FM) and adhere to these two conditions. i) In the entire dataset, the number of extrema and zero crossings must be equal or, at most, equal to one. ii) At any point, the mean of the envelopes defined by the local maxima and local minima is equal to zero. An iterative process called sifting does the extraction of these functions.

The EMD concept can be described by an elementary decomposition operation applied recursively to an oscillating signal s(t). This operation leads to the extraction of the part that oscillates according to a local "high frequency" contribution c(t). Thus, the difference s(t) - c(t) represents the component that oscillates according to a "low frequency" contribution. Then, the procedure is applied again to the difference s(t)-c(t) which is considered a new signal for the extraction of the component which oscillates locally the fastest. The decomposition process repeats and stops if s(t)-c(t) is no longer an oscillating signal. As a consequence of the previous principle, the signal s(t) can be expressed as the sum of the oscillating functions $c_i(t)$ and a residue r(t):

$$s(t) = \sum_{i=1}^{l} c_i(t) + r(t) \tag{10}$$

Where the r(t) is the residue and $c_i(t)$ is the functions are called the Intrinsic Mode Functions (IMFs). Figure 2 shows the decomposition of the speech signal of a healthy patient as a function of intrinsic mode (c_i) and residue (r).

In our experiment, we applied EMD to extract the full oscillatory mode of a speech signal. The resulting IMFs from this method are all employed to build a feature vector that can distinguish between normal and Parkinson's disease participants.

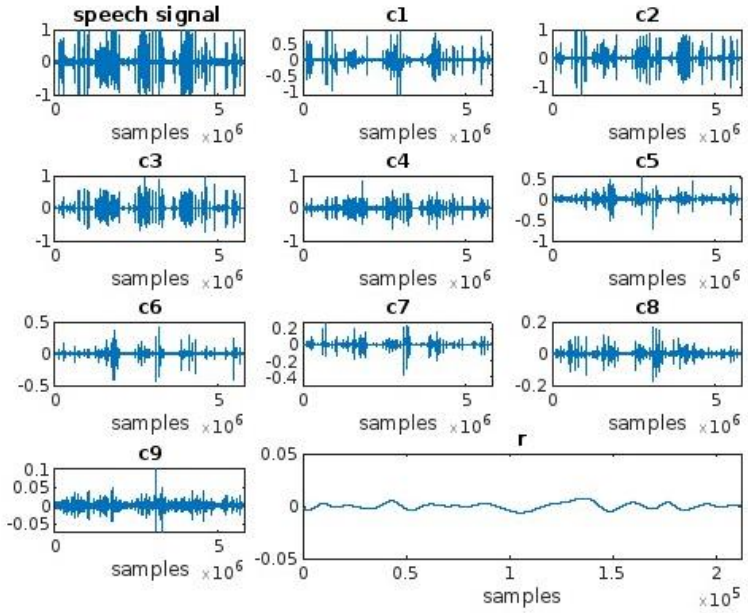

Fig 2. The extracted IMFs from the voice of a healthy participant during a sustained vowel /a/

2.2.4. Wavelets

To surmount the Gabor Transform's limitations (TG), the emergence of wavelets aims to build a tool that can overcome the problem of choosing the window in the TG, using analyzing functions for which the size of the window varies with the frequency.

Wavelet analysis appeared in the early 1980s, in a remarkable article by (Grossmann et al., 1985). This approach appeared initially in geophysics for the analysis of seismic signals. The results obtained by Morlet and formalized by the physicist Alex Grossmann bring out the notion of orthogonal basis, multi-resolution analysis, and compact support wavelets.

Wavelet decomposition is a tool for studying non-stationary signals, the wavelets are formed by translations and dilatations of the mother wavelet Ψ according to the formula below:

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{a}}\Psi\left(\frac{t-b}{a}\right) \tag{11}$$

Where a is the scale factor and b is the translation parameter, the condition for Ψ to be admissible as a wavelet is:

$$\int_{-\infty}^{+\infty} \Psi(t)dt = 0 \tag{12}$$

The Continuous Wavelet Transform (CWT) of an s(t) signal consists in measuring its similarity with the bases of wavelets Ψa,b and is given by the following equation:

$$CWT_s(a,b) = \int_{-\infty}^{+\infty} s(t)\Psi^*_{a,b}(t)dt \quad (13)$$

This transformation is in theory infinitely redundant since the wavelet is continuously translated; however, there is the Discrete Wavelet Transform (DWT) to decrease this redundancy by expanding and translating the wavelet according to discrete values. For the scale factor $a = a_0^j$ and the translation parameter $b = ka_0^j$ discretized, the following formula is used to calculate the DWT:

$$DWT_s(a,b) = \frac{1}{\sqrt{a_0^j}} \int_{-\infty}^{+\infty} s(t)\Psi^*(a_0^{-j}t - k)dt \quad (14)$$

The decomposition of the DWT procedure based on Mallat's multiresolution analysis consists in decomposing the signal into approximations and details. For the level j = 1 means the initial decomposition, the discrete-time voice signal s[k] passes through two complementary filters, the High-Pass filter (HP) and Low-Pass filter (LP), pursued by a sub-sampling operation by a factor equal to 2, and appears as two signals: respectively, the signal of details d_j and the signal of approximations a_j this principle is represented in Figure 3.

$$d_j[i] = w_{high}[i] = \sum_k HP[2\times i-k]\cdot s[k] \quad (15)$$

$$a_j[i] = w_{low}[i] = \sum_k LP[2\times i-k]\cdot s[k] \quad (16)$$

Following the acquisition of d_j and a_j, the approximation a_j is set to s[k], and j is set to 2 (raised by 1), and the preceding operation is reiterated till the j reaches the wavelet decomposition's levels chosen (in our work, j = 3).

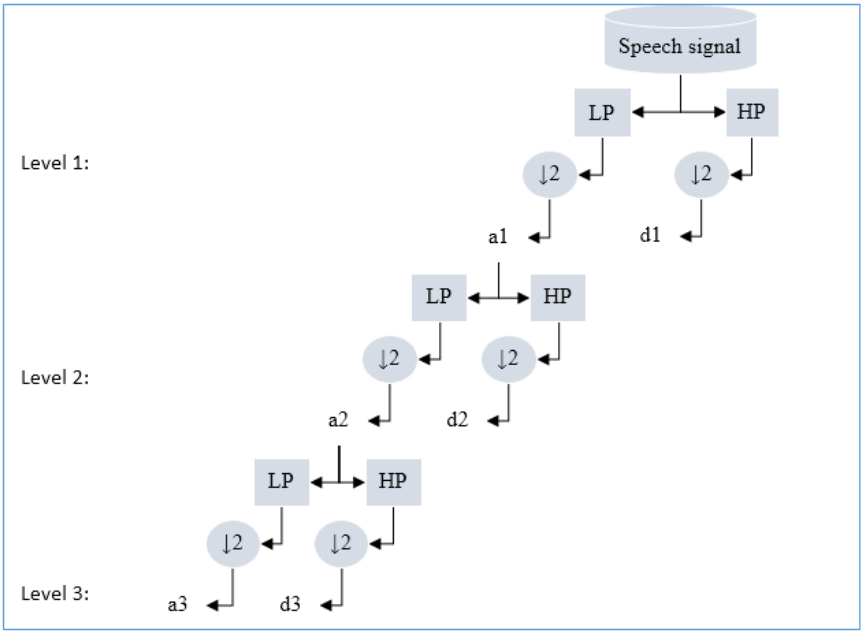

Fig. 3. The DWT's tree structure

2.3. Classification

2.3.1. Convolutional Neural Network (CNN)

Convolutional neural network (CNN): is a type of artificial neural network used in video and image recognition, speech processing, and other applications on multidimensional networks. It is intended for processing data that has a grid-like topology, such as images.

Convolutional layers, activation layers, pooling layers and fully connected layers are all part of a CNN. The model applies a set of filters to the input data in the convolutional layer to identify local patterns or features. To introduce nonlinearity into the model, the activation layer uses a nonlinear function such as ReLU. To reduce spatial dimensions and improve computational efficiency, the pooling layer subsamples the data. Each neuron in the previous layer is connected to each neuron in the next layer in the fully connected layer, resulting in a high-level representation of the input data.

Backpropagation and stochastic gradient descent optimization are commonly used to learn the parameters of a CNN network. The network is trained on a large dataset of labeled sequences, and the optimization process adjusts the parameters to minimize the classification error between the actual labels and predicted ones.

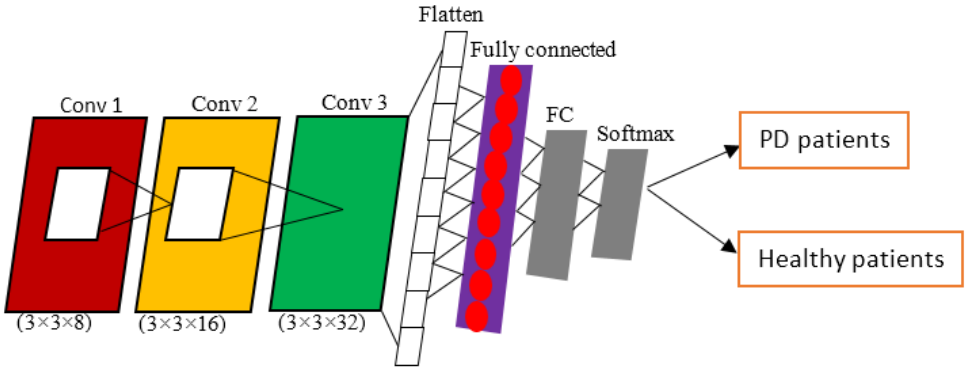

Fig. 4. The suggested deep convolutional neural network architecture for the identification of Parkinson's disease

Figure 4 presents the architecture of CNN network used in this study. The first convolutional layer (Conv 1) in the proposed architecture has 8 square-shaped kernels of size (3×3) that are applied to the input sequences of both GTCC and MFCC feature categories with the same padding and stride setting (1×1). Similarly, the second convolutional layer (Conv 2) has 16 filters with sizes of (3×3) and strides of (1×1). The Conv 3 layer employs the same filter size, stride, and padding as Conv1 and Conv 2, but has 32 filters. This final convolutional layer is followed by a flattening layer, which converts the data's shape to a vector before passing the features to the fully connected layer. The rectified linear unit (RLU) activation function is employed in this suggested CNN, followed by batch normalization to regularize the model after each convolutional layer. To avoid model overfitting, the first Fully Connected (FC) layer is followed by a 25% dropout rate (Srivastava et al., 2014). The last FC layer is powered by the SoftMax classifier to compute the probability of each category.

2.3.2. Long Short-Term Memory (LSTM) networks

Long Short-Term Memory (LSTM) networks: are a kind of Recurrent Neural Network (RNN) commonly used for sequential data treatment. The primary goal of LSTM is to overcome the evanescent gradient problem that exists in traditional RNN. The steps of an LSTM network illustrated in Figure 5 will be explained later:

- Input layer: As a time series, the input characteristics (MFCC and GTCC) are fed into the network.
- LSTM Layer: The core layer of the LSTM network, which includes the memory cells that allow the network to maintain a sense of context over time. The LSTM layer transforms input data into output and hidden states.
- Gate Mechanism: To control the flow of information, the LSTM layer uses three gates: the input gate, the forget gate, and the output gate. These gates control the amount of information that can enter, exit and pass through the memory cells.
- The memory cell: is responsible for retaining contextual information from previous time steps. The forget gate and the input gate update the memory cell, which is used to generate the hidden state.

- The LSTM layer generates the hidden state, which is used to capture contextual information from the preceding steps.
- The output layer, based on the input data and the hidden state, generates an output. The output can be used to predict or categorize the input data.
- Backpropagation in Time (BPTT): The difference between the predicted and actual output is used to update the LSTM layer weight via a process known as backpropagation in time (BPTT).
- Repeat the process of passing the input data through the LSTM layer and updating the weights until the network reaches a satisfactory level of accuracy.

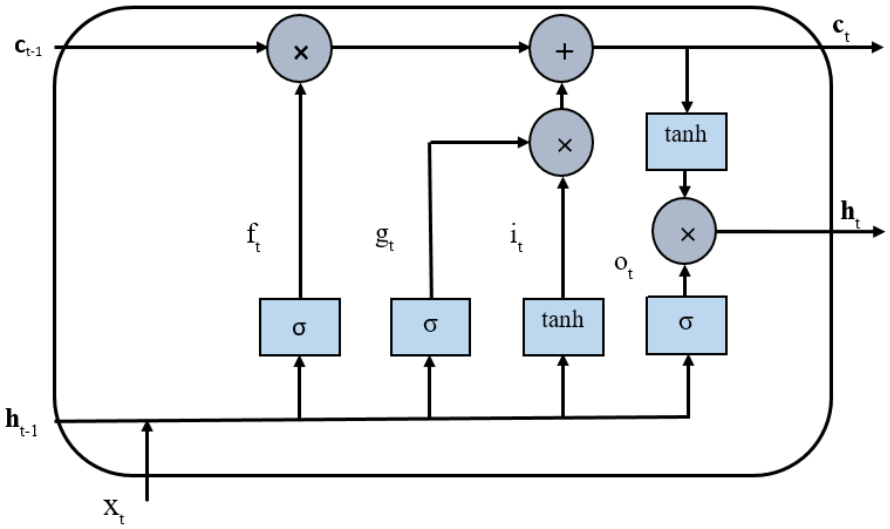

Fig. 5. LSTM structure

This diagram depicts the data flow at time step t, demonstrating how the gates forget, update, and output the cell and hidden states. The symbols c_t and h_t in the diagram represent the cell state at time step t and the output (also known as the hidden state), respectively. At time step t, the block computes the output and updated cell state c_t using the current network state (ct-1, ht-1) and the next time step in the sequence. The layer state consists of hidden state and cell state. The output of the LSTM layer for time step t is stored in the hidden state at time step t. The previous steps' information is stored in the cell state. The layer adds or removes information from the cell state at each time step. These updates are controlled by the layer via gates; the formulas in Table 1 describe the components of these gates at each time step.

Tab. 1. The equations of the gates

Component	Equation
Cell candidate	$g_t = \tanh (W_g x_t + R_g h_t + b_g)$
Forget gate	$f_t = \tanh (W_f x_t + R_f h_t + b_f)$
Input gate	$i_t = \tanh (W_i x_t + R_i h_t + b_i)$
Output gate	$o_t = \tanh (W_o x_t + R_o h_t + b_o)$

With R (Recurrent Weights) W (Input Weights), and b (Bias) are expressed as follow:

$$R = \begin{bmatrix} Rg \\ Rf \\ Ri \\ Ro \end{bmatrix} \quad W = \begin{bmatrix} Wg \\ Wf \\ Wi \\ Wo \end{bmatrix} \quad b = \begin{bmatrix} bg \\ bf \\ bi \\ bo \end{bmatrix}$$

where g_t, f_t, i_t, o_t, tanh and σ denote Cell candidate, Forget gate, Input gate, Output gate, the hyperbolic tangent function and the sigmoid function respectively.

The hidden state and cell state at time step t is obtained by:

$$h_t = o_t \odot \tanh(c_t) \tag{17}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t \tag{18}$$

3. RESULTS AND DISCUSSION

3.2. Performance evaluation

The cost of misclassification, which is the cost of misclassifying a negative instance as a positive instance or vice versa, can be utilized to compare the capabilities of machine learning systems. In this study, the 10-fold crossvalidation approach, which is the preferred method in most real-life circumstances of data, is employed to evaluate the error rate of classification algorithms. The 10-fold crossvalidation approach, in general, divides the data into 10 parts randomly; the nine parts are used for training, while the one tenth is kept for testing, we repeat this process ten times, each time reserving a different tenth for testing. The training data is utilized to train the model, whereas the test data is employed to assess the model's performance.

The accuracy (Acc), sensitivity (Sen), and specificity (Spe), criteria were used to evaluate the effectiveness of this Parkinson's disease identification. The precision with which the suggested technique can categorize all subjects is defined as accuracy (including patients and healthy subjects). The sensitivity indicates how well the proposed system can classify sick people (PD suffers) whereas, specificity, on the other hand, indicates how well the system can categorize normal participants, these parameters can be calculated using the simple equations (19), (20), and (21), respectively.

If TP indicates the number of true positives, TN means the number of true negatives, FP represents the number of false positives, and FN represents the number of false negatives, then the following equations provide the formulation of these assessment metrics:

$$Acc = \frac{TN + TP}{TN + FN + TP + FP} \tag{19}$$

$$Sen = \frac{TP}{TP + FN} \tag{20}$$

$$Spe = \frac{TN}{TN + FP} \tag{21}$$

3.3. Experiment results

This part of the paper discusses the implementation details and the performance attained by GTCC and MFCC for both DWT-EMD and EMD-DWT techniques for the problem of detecting Parkinson's disease by employing two different database. All the experiments were done using Intel (R) Core(TM) i3-3110M CPU @ 2.40GHz 2.40 GHz with 4GB memory and 64-bit Windows operating system, x64 processor. They were also executed using MATLAB R2021b software.

For both EMD-DWT and DWT-EMD methods, the speech signal was split into two permuted methods: empirical mode decomposition followed by discrete wavelet transform Debauchie 2 at level 3 (EMD-DWT) as showed Figure 6, and discrete wavelet transform Debauchie 2 at level 3 followed by empirical mode decomposition (DWT-EMD) as illustrated Figure 7, Debauchie 2 Level 3 was chosen based on the results of (Drissi et al., 2019), who used the same data as Sakar to select the type of wavelet performance. Figure 2 shows a set of signals generated by the EMD from the pronunciation of the letter /a/ by patients with Parkinson's disease. It shows how, after the EMD operation, the high-frequency component of the original voice signal is obtained first, and the frequency of the following IMFs gradually decreases. This is in accordance with the findings of the literature of references (Altuve et al., 2020; Mondal et al., 2018; Zhang et al., 2021), they notice that the IMFs are placed in the decreasing order of frequency components; therefore, we only worked with the first seven IMFs.

Figure 8 presents the first seven IMFs obtained from the approximation a3 by the used of DWT-EMD method, as well as Figure 9 illustrates the first seven approximations a3 extracted from the first seven IMFs for the EMD-DWT method, for a healthy participant and a PD patient, respectively. Therefore, we extracted twelve GTCC and MFCC coefficients features from each function, forming two vectors of seven features, F1 = {$12GTCC_{a3\text{-}IMF1}$; $12GTCC_{a3\text{-}IMF2}$; $12GTCC_{a3\text{-}IMF3}$; $12GTCC_{a3\text{-}IMF4}$; $12GTCC_{a3\text{-}IMF5}$; $12GTCC_{a3\text{-}IMF6}$; $12GTCC_{a3\text{-}IMF7}$} and F2 = {$12MFCC_{a3\text{-}IMF1}$; $12MFCC_{a3\text{-}IMF2}$; $12MFCC_{a3\text{-}IMF3}$; $12MFCC_{a3\text{-}IMF4}$; $12MFCC_{a3\text{-}IMF5}$; $12MFCC_{a3\text{-}IMF6}$; 12MFCCa3-IMF7}, while the indices a3-IMF2, a3-IMF2... represent respectively IMF1 extracted from approximation a3, IMF2 extracted from approximation a3 (for DWT-EMD method)... Similarly, F3 = {$12GTCC_{IMF1\text{-}a3}$; $12GTCC_{IMF2\text{-}a3}$; $12GTCC_{IMF3\text{-}a3}$; $12GTCC_{IMF4\text{-}a3}$; $12GTCC_{IMF5\text{-}a3}$; $12GTCC_{IMF6\text{-}a3}$; $12GTCC_{IMF7\text{-}a3}$} and F4 = {$12MFCC_{IMF1\text{-}a3}$; $12MFCC_{IMF2\text{-}a3}$; $12MFCC_{IMF3\text{-}a3}$; $12MFCC_{IMF4\text{-}a3}$; $12MFCC_{IMF5\text{-}a3}$; $12MFCC_{IMF6\text{-}a3}$; $12MFCC_{IMF7\text{-}a3}$}, while the indices IMF1-a3, IMF2-a3... respectively represent the approximation a3 extracted from IMF1 and the approximation a3 extracted from IMF2 (for EMD-DWT method).

To use the GTCC and MFCC directly as an input to the CNN and LSTM classifier requires a fairly large processing time, to surmount this issue, we averaged the GTCC and MFCC of each voice recording to obtain the voiceprint.

Due to their effectiveness and efficiency, the deep learning's classifiers are used in this study CNN and LSTM, so the extracted GTCC and MFCC are used as a sequences input, The tests are carried out with, the learning rates set to 0.001 and the number of epochs sets

to 20, accordingly for LSTM the number of Hidden sets to 100, these choices are depending to results of reference (Er et al., 2021). Tables 2-5 shows the results achieved with this proposed method parameters.

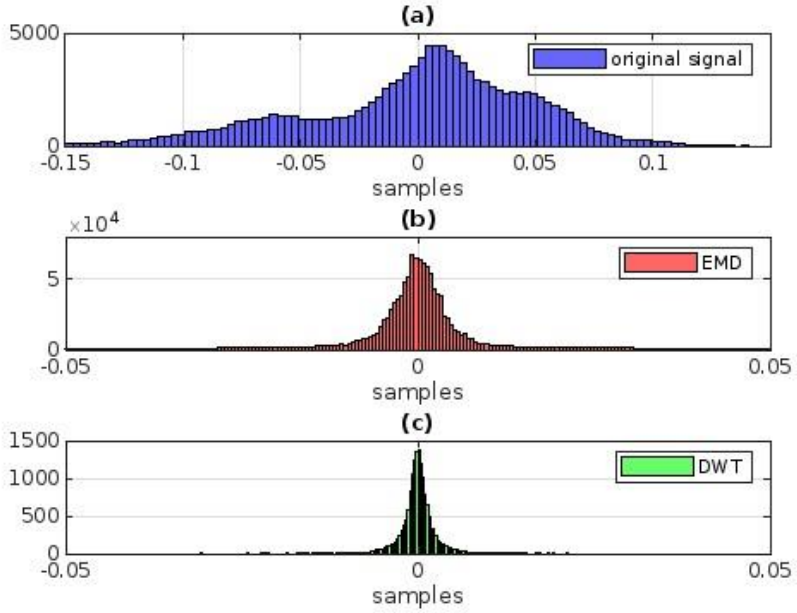

Fig. 6. EMD-DWT decomposition represented by Histogram plot for a PD patient – (a): The original signal, (b): The application of the EMD, (c): The used of DWT after EMD

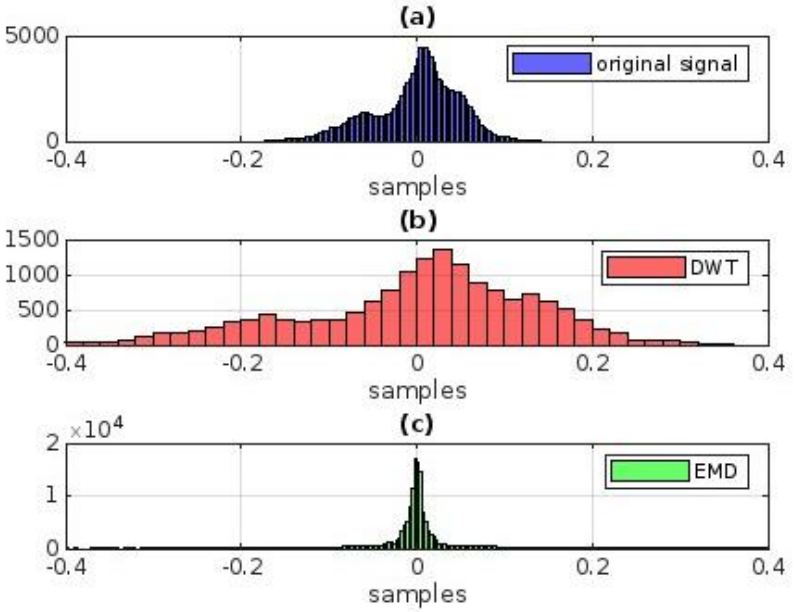

Fig. 7. DWT- EMD decomposition represented by Histogram plot for a PD patient – (a): The original signal, (b): The application of the DWT, (c): The used of EMD after DWT

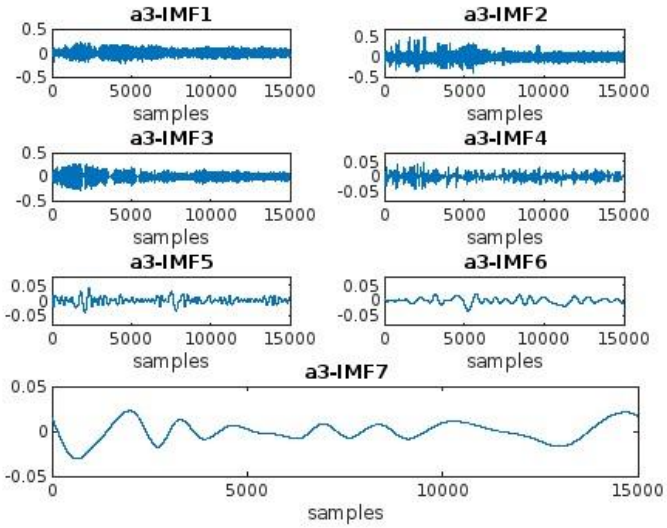

Fig. 8. The first 7 IMFs obtained from the approximation a3 for the vocal signal of the sustained vowel /a/ of a PD person by the DWT-EMD method

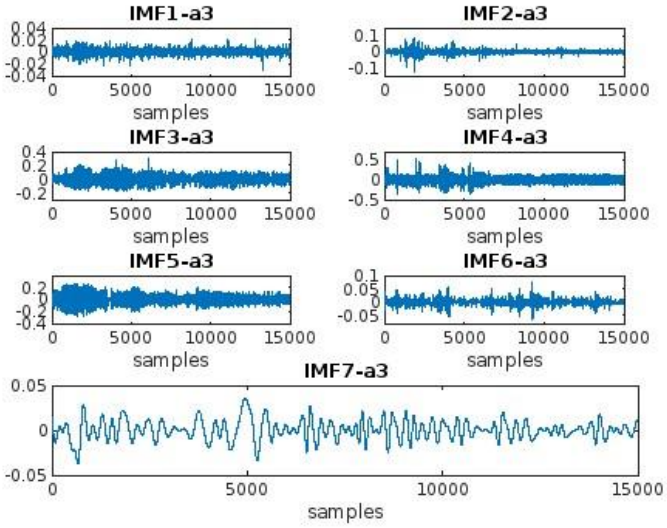

Fig. 9. Approximation a3 for the first 7 IMFs obtained from the vocal signal of the vowel /a/ of a PD person by the EMD-DWT analysis

Tab. 2. The classification results obtained by GTCC for the database 1

	EMD-DWT		DWT-EMD

IMF-a3	GTCC						a3-IMF	GTCC					
	CNN			LSTM				CNN			LSTM		
	Acc %	Spe %	Sen %	Acc %	Spe %	Sen %		Acc %	Spe %	Sen %	Acc %	Spe %	Sen %
IMF1-a3	86.96	78.57	100	86.96	78.57	100	a3-IMF1	65.22	71.42	55.55	95.65	92.85	100
IMF2-a3	69.57	92.85	100	95.65	92.85	100	a3-IMF2	86.96	92.85	77.77	91.30	92.85	88.88
IMF3-a3	86.96	85.71	88.88	91.30	85.71	100	a3-IMF3	86.96	85.71	88.88	95.65	92.85	100
IMF4-a3	100	100	100	95.65	92.85	100	a3-IMF4	100	100	100	82.61	71.42	100
IMF5-a3	100	100	100	100	100	100	a3-IMF5	86.96	78.57	100	82.61	71.42	100
IMF6-a3	100	100	100	100	100	100	a3-IMF6	69.57	64.28	77.77	86.96	85.71	88.88
IMF7-a3	91.30	85.71	100	78.26	78.57	77.77	a3-IMF7	78.26	78.57	77.77	73.91	71.42	77.77

Tab. 3. The classification results obtained by MFCC for the database 1

IMF-a3	EMD-DWT						a3-IMF	DWT-EMD					
	MFCC							MFCC					
	CNN			LSTM				CNN			LSTM		
	Acc %	Spe %	Sen %	Acc %	Spe %	Sen %		Acc %	Spe %	Sen %	Acc %	Spe %	Sen %
IMF1-a3	69.57	57.14	88.88	47.83	28.57	77.77	a3-IMF1	82.61	71.42	100	47.83	50	44.44
IMF2-a3	95.65	71.42	66.66	91.30	85.71	100	a3-IMF2	91.30	92.85	88.88	86.96	78.57	100
IMF3-a3	82.61	92.85	66.66	69.57	64.28	77.77	a3-IMF3	95.65	92.85	100	82.61	71.42	100
IMF4-a3	95.65	92.85	100	82.61	78.57	88.88	a3-IMF4	91.30	85.71	100	73.91	71.42	77.77
IMF5-a3	91.30	92.85	88.88	73.91	64.28	88.88	a3-IMF5	82.61	71.42	100	82.61	85.71	50
IMF6-a3	86.96	92.85	77.77	86.96	100	66.66	a3-IMF6	86.96	85.71	88.88	78.26	78.57	77.77
IMF7-a3	95.65	100	88.88	60.87	64.28	55.55	a3-IMF7	78.26	71.42	88.88	52.17	42.85	66.66

Tab. 4. The classification results obtained by GTCC for the database 2

IMF-a3	EMD-DWT						a3-IMF	DWT-EMD					
	GTCC							GTCC					
	CNN			LSTM				CNN			LSTM		
	Acc %	Spe %	Sen %	Acc %	Spe %	Sen %		Acc %	Spe %	Sen %	Acc %	Spe %	Sen %
IMF1-a3	96.55	92.85	100	62.07	64.28	60	a3-IMF1	79.31	85.71	73.33	72.41	78.57	66.66
IMF2-a3	82.76	85.71	80	75.85	50	100	a3-IMF2	89.66	85.71	93.33	79.31	85.71	73.33
IMF3-a3	89.66	78.57	100	89.66	85.71	93.33	a3-IMF3	93.10	85.71	100	86.21	85.71	86.66
IMF4-a3	93.10	92.85	73.33	93.10	78.57	100	a3-IMF4	96.55	100	93.33	89.86	85.71	93.33
IMF5-a3	89.66	85.71	100	93.10	85.71	100	a3-IMF5	86.21	78.57	93.33	89.66	78.57	100
IMF6-a3	100	100	100	89.66	85.71	93.33	a3-IMF6	82.76	71.42	93.33	82.76	71.42	93.33
IMF7-a3	93.10	85.71	100	89.66	92.85	86.66	a3-IMF7	86.21	78.57	93.33	79.31	64.28	93.33

Tab. 5. The classification results obtained by MFCC for the database 2

IMF-a3	EMD-DWT						a3-IMF	DWT-EMD					
	MFCC							MFCC					
	CNN			LSTM				CNN			LSTM		
	Acc %	Spe %	Sen %	Acc %	Spe %	Sen %		Acc %	Spe %	Sen %	Acc %	Spe %	Sen %
IMF1-a3	75.86	57.14	93.33	86.21	85.71	86.66	a3-IMF1	72.41	64.28	80	48.28	57.14	40
IMF2-a3	79.31	64.28	100	75.85	78.57	73.33	a3-IMF2	93.10	85.71	100	72.41	64.28	80
IMF3-a3	96.55	92.85	100	89.66	78.57	100	a3-IMF3	93.10	85.71	100	75.86	57.14	93.33
IMF4-a3	79.31	71.42	86.66	82.76	71.42	93.33	a3-IMF4	82.76	64.28	100	79.31	64.28	93.33
IMF5-a3	82.76	57.14	100	79.31	85.71	73.33	a3-IMF5	79.31	57.14	100	75.86	57.14	93.33
IMF6-a3	89.66	78.57	100	82.76	71.42	93.33	a3-IMF6	82.76	85.71	80	72.41	50	93.33
IMF7-a3	89.66	85.71	93.33	86.21	85.71	86.66	a3-IMF7	86.21	85.71	86.66	82.76	71.42	93.33

For the Sakar database and for the EMD-DWT method, the extracted GTCC coefficients show a result of 100% for IMF4-a3 using CNN, and for IMF5-a3 and IMF6-a3 using both

CNN and LSTM, while the MFCC coefficients give 95.65% by CNN for IMF2-a3 and IMF4-a3, by the user of the second PC-GITA data with 100% for IMF6-a3 using CNN, whereas the MFCC gives an accuracy up to 96.55% for IMF3-a3 by CNN.

On the other hand, for the DWT-EMD, and for the Sakar database the GTCC yields an accuracy of 100% for a3-IMF4, and 95.65% for the MFCC coefficients a3-IMF3 by CNN. In the PC-GITA database, the GTCC gives, an accuracy of 96.55% by CNN for a3-IMF4, and by CNN 93.10% is achieved using MFCC coefficients for a3-IMF2 and a3-IMF3. We can thus notice that both EMD-DWT methods and vice versa perform better with GTCC coefficients than with MFCC coefficients.

3.4. Discussion

The present study compares GTCC with the well-known MFCC-based features to detect voice tremors in patients with Parkinson's disease. The voice signal features are extracted from various modes of the IMFs-a3 and a3-IMFs of the first seven modes acquired by EMD-DWT and vice versa. Classification experiments with GTCC and MFCC features are carried out using two efficient deep-learning classifiers, CNN and LSTM. The GTCC coefficients achieve the highest classification accuracy of 100% for both EMD-DWT and vice versa with CNN in the Sakar database and an accuracy of 100% and 96.55% in the PC-GITA respectively. The experiments show that CNN clearly outperforms the other classifier LSTM. Since several authors employ the same PC-GITA database in different ways, the outcomes obtained are compared to the outcomes found with the PC-GITA database as well as other database. To determine the best performance of the suggested process, the findings of this study are compared with other approaches used in the literature. Table 6 sums up the main studies on the recognition of Parkinson's disease from the voice signal.

Tab. 6. Performance comparison with other methods

Authors	Methods	Database	Accuracy
(Er et al., 2021)	ResNet-101 & LSTM with VMD	PC Gita (vowel /a/)	98.61%
(Karan & Sekhar Sahu, 2021)	VMD-HAS-HCC-MLP	PC Gita (vowel /a/)	91%
(Karan, Sahu, Orozco-Arroyave, et al., 2020)	HS-IEDCC-SVM	PC Gita (vowel /a/)	82%
(Karan, Sahu, & Mahto, 2020)	EMD-IMFCC-RF	PC Gita (vowel /a/)	96%
(Zahid et al., 2020)	Deep Features with multilayer perceptron	PC Gita (vowel /a/)	99.7%
(López-Pabón et al., 2020)	THH-MFCC-SVM	PC Gita (vowel /a/)	70%
(Nouhaila et al., 2022)	DWT-Delta delta MFCC-DT	Italian database of 50 records (vowel /a/)	97.5%
(Soumaya et al., 2021)	DWT-MFCC and genetic algorithm with SVM	Sakar database (vowel /a/)	91.18%

(Taoufiq et al., 2022)	DWT instead filter bank for MFCC and SVM	Sakar database (vowel /a/)	81.75%
Proposed method	EMD-DWT-GTCC-CNN	PC-GITA	100%

Mehmet et al. used the pre-trained models ResNet-18, ResNet-50, and ResNet-101, as well as the LSTM as a classifier, in conjunction with the Mel-spectrograms obtained from the denoised signal via variational mode decomposition. The findings demonstrate that the ResNet-101 outperforms the others with an accuracy of up to 98.61% (Er et al., 2021). Karan et al. used a combination of VMD and HAS to study tremor in Parkinson's disease patients, extracting a novel HCC feature and using it as input for the three machine learning classifiers SVM, RF, and MLP, the latter of which has a greater accuracy of 91% (Karan & Sekhar Sahu, 2021). Karan et al utilize instantaneous energy deviation Cepstral coefficient (IEDCC) and SVM as a classifier to predict speech deficiencies in people with Parkinson's disease, and their outcomes were 82% accurate (Karan, Sahu, Orozco-Arroyave, et al., 2020). Karan et al. proposed a new characteristic based on Empirical Mode Decomposition (EMD) called Intrinsic Mode Function Cepstral Coefficient Extraction (IMFCC), to effectively represent Parkinson's speech properties. These IMFCC features provide 96% accuracy by the SVM classifier (Karan, Sahu, & Mahto, 2020). Zahid et al proposed three methods, the second one is based on the assessment of deep characteristics extracted from speech spectrograms using machine learning classifiers, the outcome is 99.7% by the MLP classifier (Zahid et al., 2020). Lopez et al. suggested the Hilbert-Huang Transform (HHT) and Mel-Frequency Cepstral Coefficients (MFCCs) to model an automatic discrimination between PD and HC subjects with accuracy up to 72% by RBF-SVM (López-Pabón et al., 2020). Boualoulou et al. applied a method based on delta delta MFCC coefficients and different types of discrete wavelet transforms, with the decision tree classifier in order to choose the appropriate wavelet to discriminate PD patients from healthy patients, for the vowel /a/ the accuracy reached 97.5% for Discrete Meyer wavelet (Nouhaila et al., 2022). Zayrit et al. proposed a novel combination of genetic algorithm and SVM classifier to distinguish people with Parkinson's disease from healthy people, based on Debauchies wavelets and MFCCs as extracted features, with the obtained accuracy reaching 91.18% (Soumaya et al., 2021). Belhoussine et al. suggested a new approach relying on the extracted MFCC features by replacing the filter bank with a DWT and an SVM classifier for classification, this method generated an accuracy of 81.75% (Taoufiq et al., 2022).

The results demonstrated that the GTCC is more accurate and appropriate for PD prediction. Thus, there is a substantial gain in accuracy over the Mel Frequency Cepstral Coefficient (MFCC).

4. CONCLUSION AND FUTURE WORK

The work carried out is a part of the analysis of the voice signal to achieve the goal of an efficient prediction system for diagnosing PD. In this document, the authors suggested a diagnostic model founded on two acoustic analysis methods dedicated to the automatic evaluation of the voice produced by various people (normal subjects and subjects suffering from Parkinson's disease) pronouncing the sustained vowel /a/. The first DWT-EMD

estimation method consists of decomposing the speech signals by DWT precisely Daubechies wavelet at level two in the third scale, followed by a decomposition of the approximation a3 by EMD, finally extracting the GTCC and MFCC cepstral coefficients from each IMF. The second estimation method EMD-DWT operates on the use of the empirical mode decomposition algorithm to decompose the speech signals into Intrinsic Mode Functions (IMFs), followed by a decomposition of the IMFs by DWT, then the extraction of the GTCC and MFCC coefficients from the approximation a3. These cepstral coefficients are employed as an input in the classification step by applying the CNN and LSTM. As a result, it is pointed out that the GTCC produce optimal efficiency for detecting PDs than MFCC for both DWT-EMD and EMD-DWT methods and for both Datasets. The comparison of performance was performed using different evaluation metrics, including accuracy. The GTCC coefficients produced higher results than the MFCC coefficients in the Sakar dataset, with 100% accuracy for the EMD-DWT and DWT-EMD methods using the CNN classifier, including higher results for the PC-GITA dataset, with accuracy up to 96.55% for EMD-DWT and 100% for DWT-EMD using the CNN classifier. Hence, the conclusion is that the GTCC coefficients are more accurate than the MFCC coefficients and they have an interesting effect on the evaluation of PD patients. This approach is aimed at being expanded in the future by using them for other types of signals, such as gait signals and writing tasks, which are affected in patients with Parkinson's disease.

REFERENCES

Ali, Z., Elamvazuthi, I., Alsulaiman, M., & Muhammad, G. (2016). Automatic Voice Pathology Detection With Running Speech by Using Estimation of Auditory Spectrum and Cepstral Coefficients Based on the All-Pole Model. *Journal of Voice*, *30*(6), 757.e7-757.e19. https://doi.org/10.1016/j.jvoice.2015.08.010

Altuve, M., Suárez, L., & Ardila, J. (2020). Fundamental heart sounds analysis using improved complete ensemble EMD with adaptive noise. *Biocybernetics and Biomedical Engineering*, *40*(1), 426–439. https://doi.org/10.1016/j.bbe.2019.12.007

Dash, T. K., Mishra, S., Panda, G., & Satapathy, S. Ch. (2021). Detection of COVID-19 from speech signal using bio-inspired based cepstral features. *Pattern Recognition*, *117*, 107999. https://doi.org/10.1016/j.patcog.2021.107999

Davis, S., & Mermelstein, P. (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *28*(4), 357–366. https://doi.org/10.1109/TASSP.1980.1163420

Demir, F., Siddique, K., Alswaitti, M., Demir, K., & Sengur, A. (2022). A Simple and Effective Approach Based on a Multi-Level Feature Selection for Automated Parkinson's Disease Detection. *Journal of Personalized Medicine*, *12*(1), 55. https://doi.org/10.3390/jpm12010055

Drissi, T. B., Zayrit, S., Nsiri, B., & Ammoummou, A. (2019). Diagnosis of Parkinson's disease based on wavelet transform and Mel Frequency Cepstral Coefficients. *International Journal of Advanced Computer Science and Applications*, *10*(3), 125–132. https://doi.org/10.14569/IJACSA.2019.0100315

Er, M. B., Isik, E., & Isik, I. (2021). Parkinson's detection based on combined CNN and LSTM using enhanced speech signals with Variational mode decomposition. *Biomedical Signal Processing and Control*, *70*, 103006. https://doi.org/10.1016/j.bspc.2021.103006

Grossmann, A., Morlet, J., & Paul, T. (1985). Transforms associated to square integrable group representations. I. General results. *Journal of Mathematical Physics*, *26*(10), 2473–2479. https://doi.org/10.1063/1.526761

Hammami, I., Salhi, L., & Labidi, S. (2020). Voice Pathologies Classification and Detection Using EMD-DWT Analysis Based on Higher Order Statistic Features. *IRBM*, *41*(3), 161–171. https://doi.org/10.1016/j.irbm.2019.11.004

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N.-C., Tung, C. C., & Liu, H. H. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, *454*(1971), 903–995. https://doi.org/10.1098/rspa.1998.0193

Karan, B., Sahu, S. S., & Mahto, K. (2020). Parkinson disease prediction using intrinsic mode function based features from speech signal. *Biocybernetics and Biomedical Engineering*, *40*(1), 249–264. https://doi.org/10.1016/j.bbe.2019.05.005

Karan, B., Sahu, S. S., Orozco-Arroyave, J. R., & Mahto, K. (2020). Hilbert spectrum analysis for automatic detection and evaluation of Parkinson's speech. *Biomedical Signal Processing and Control*, *61*, 102050. https://doi.org/10.1016/j.bspc.2020.102050

Karan, B., & Sekhar Sahu, S. (2021). An improved framework for Parkinson's disease prediction using Variational Mode Decomposition-Hilbert spectrum of speech signal. *Biocybernetics and Biomedical Engineering*, *41*(2), 717–732. https://doi.org/10.1016/j.bbe.2021.04.014

Kethireddy, R., Kadiri, S. R., & Gangashetty, S. V. (2022). Exploration of temporal dynamics of frequency domain linear prediction cepstral coefficients for dialect classification. *Applied Acoustics*, *188*, 108553. https://doi.org/10.1016/j.apacoust.2021.108553

López-Pabón, F. O., Arias-Vergara, T., & Orozco-Arroyave, J. R. (2020). Cepstral Analysis and Hilbert-Huang Transform for Automatic Detection of Parkinson's Disease. *TecnoLógicas*, *23*(47), 93–108. https://doi.org/10.22430/22565337.1401

Mondal, A., Banerjee, P., & Tang, H. (2018). A novel feature extraction technique for pulmonary sound analysis based on EMD. *Computer Methods and Programs in Biomedicine*, *159*, 199–209. https://doi.org/10.1016/j.cmpb.2018.03.016

Moro-Velázquez, L., Gómez-García, J. A., & Godino-Llorente, J. I. (2016). Voice pathology detection using modulation spectrum-optimized metrics. *Frontiers in Bioengineering and Biotechnology*, *4*. https://doi.org/10.3389/fbioe.2016.00001

Nagarajan, S., Nettimi, S. S. S., Kumar, L. S., Nath, M. K., & Kanhe, A. (2020). Speech emotion recognition using cepstral features extracted with novel triangular filter banks based on bark and ERB frequency scales. *Digital Signal Processing*, *104*, 102763. https://doi.org/10.1016/j.dsp.2020.102763

Najnin, S., & Banerjee, B. (2019). Speech recognition using cepstral articulatory features. *Speech Communication*, *107*, 26–37. https://doi.org/10.1016/j.specom.2019.01.002

Nouhaila, B., Taoufiq, B. D., & Benayad, N. (2022). 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*, *13*(4), 562–571. https://doi.org/10.14569/IJACSA.2022.0130466

Orozco-Arroyave, J. R., Arias-Londõno, J. D., Vargas-Bonilla, J. F., González-Rátiva, M. C., & Nöth, E. (2014.). New Spanish speech corpus database for the analysis of people suffering from Parkinson's disease. *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)* (pp. 342-347). European Language Resources Association (ELRA).

Oyinloye, B. E., Iwaloye, O., & Ajiboye, B. O. (2021). Polypharmacology of Gongronema latifolium leaf secondary metabolites against protein kinases implicated in Parkinson's disease and Alzheimer's disease. *Scientific African*, *12*. https://doi.org/10.1016/j.sciaf.2021.e00826

Qin, J., Liu, T., Wang, Z., Zou, Q., Chen, L., & Hong, Ch. (2022). Speech Recognition for Parkinson's Disease Based on Improved Genetic Algorithm and Data Enhancement Technology. In Wang, Y., Zhu, G., Han, Q., Wang, H., Song, X., & Lu, Z. *Communications in Computer and Information Science*, (vol. 1628, pp. 273–286).Springer. https://doi.org/10.1007/978-981-19-5194-7_21

Quan, Ch., Ren, K., Luo, Z., Chen, Z., & Ling, Y. (2022). End-to-end deep learning approach for Parkinson's disease detection from speech signals. *Biocybernetics and Biomedical Engineering*, *42*(2), 556–574. https://doi.org/10.1016/j.bbe.2022.04.002

Sakar, B. E., Isenkul, M. E., Sakar, C. O., Sertbas, A., Gurgen, F., Delil, S., Apaydin, H., & Kursun, O. (2013). Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings. *IEEE*

Journal of Biomedical and Health Informatics, *17*(4), 828–834. https://doi.org/10.1109/JBHI.2013.2245674

Sakar, C. O., Serbes, G., Gunduz, A., Tunc, H. C., Nizam, H., Sakar, B. E., Tutuncu, M., Aydin, T., Isenkul, M. E., & Apaydin, H. (2019). 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 Journal*, *74*, 255–263. https://doi.org/10.1016/j.asoc.2018.10.022

Soumaya, Z., Drissi Taoufiq, B., Benayad, N., Yunus, K., & Abdelkrim, A. (2021). The detection of Parkinson disease using the genetic algorithm and SVM classifier. *Applied Acoustics*, *171*, 107528. https://doi.org/10.1016/j.apacoust.2020.107528

Soumaya, Z., Taoufiq, B., Benayad, N., Achraf, B., & Ammoumou, A. (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 & Sensors*, *10*(1), 60. https://doi.org/10.4103/jmss.JMSS_61_18

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *In Journal of Machine Learning Research*, *15*(56), 1929-1958.

Taoufiq, B. D., Soumaya, Z., Benayad, N., & Nouhaila, B. (2022). Cepstral Coefficient Extraction using the MFCC with the Discrete Wavelet Transform for the Parkinson's Disease Diagnosis. *International Journal of Engineering Trends and Technology*, *70*(7), 283–290. https://doi.org/10.14445/22315381/IJETT-V70I7P229

Terriza, M., Navarro, J., Retuerta, I., Alfageme, N., San-Segundo, R., Kontaxakis, G., Garcia-Martin, E., Marijuan, P. C., & Panetsos, F. (2022). Use of Laughter for the Detection of Parkinson's Disease: Feasibility Study for Clinical Decision Support Systems, Based on Speech Recognition and Automatic Classification Techniques. *International Journal of Environmental Research and Public Health*, *19*(17) 10884. https://doi.org/10.3390/ijerph191710884

Valero, X., & Alias, F. (2012). Gammatone Cepstral Coefficients: Biologically Inspired Features for Non-Speech Audio Classification. *IEEE Transactions on Multimedia*, *14*(6), 1684–1689. https://doi.org/10.1109/TMM.2012.2199972

Yagnavajjula, M. K., Alku, P., Rao, K. S., & Mitra, P. (2022). Detection of Neurogenic Voice Disorders Using the Fisher Vector Representation of Cepstral Features. *Journal of Voice*. https://doi.org/10.1016/j.jvoice.2022.10.016

Zahid, L., Maqsood, M., Durrani, M. Y., Bakhtyar, M., Baber, J., Jamal, H., Mehmood, I., & Song, O.-Y. (2020). A Spectrogram-Based Deep Feature Assisted Computer-Aided Diagnostic System for Parkinson's Disease. *IEEE Access*, *8*, 35482–35495. https://doi.org/10.1109/ACCESS.2020.2974008

Zhang, T., Zhang, Y., Sun, H., & Shan, H. (2021). Parkinson disease detection using energy direction features based on EMD from voice signal. *Biocybernetics and Biomedical Engineering*, *41*(1), 127–141. https://doi.org/10.1016/j.bbe.2020.12.009