Comprehensive Assessment of Neuromuscular Dysfunction in Parkinson's Disease Using Multi-Axial Accelerometers

Rima Touahria[1], Abdenour Hacine Gharbi[2], Philippe Ravier[3], Karim Abed-Meraim[3] and Olivier Buttelli[3]

[1]LMSE laboratory, University Mohamed El-Bachir El-Ibrahimi of bordj Bou Arreridj Algeria
[2]LMSE laboratory, University Mohamed El-Bachir El-Ibrahimi of bordj Bou Arreridj Algeria
[3]PRISME laboratory, University of Orleans, 12 rue de Blois, 45067 Orleans, France
touahria.rimaa@gmail.com, gharbi07@yahoo.fr, philippe.ravier,karim.abed-meraim,olivier.buttelli@univ-orleans.fr

Abstract. This study investigates the potential of multi-axial accelerometers (ACC) for a comprehensive assessment of neuromuscular dysfunction in Parkinson's disease (PD) patients. Emphasizing the importance of utilizing data from three axes (ACC_X,ACC_Y,and ACC_Z) to discriminate the movement characteristics, the proposed method leverages Wavelet Cepstral Coefficients (WCC) extracted from ACC data. A k-Nearest Neighbour (KNN) classifier is employed to differentiate between normal and Parkinson. Feature selection based on mutual information criteria further optimizes the classification process in terms of complexity and accuracy. Evaluation using the ECOTECH project database yielded a promising accuracy of vectors of 98.91% with 100% of signals classification rate(CRs) using only two features selected using the Conditional Infomax Feature Extraction filter selection strategy were sufficient to explain the two classes (N) and (P). These findings suggest that ACC_Y data combined with WCC feature extraction and KNN classification holds promise as efficient tool for PD diagnosis and potentially monitoring disease progression.

Keywords: ACC Signal, Features Extraction, WCC, K-nearest Neighbours, Features Selection, Mutual Information, Classification, Parkinson's Disease (PD).

1 Introduction

Parkinson's disease, often shortened to PD, is a neurodegenerative disorder. This means it involves the progressive loss of nerve cells in the brain. In PD, this loss primarily occurs in a specific area called the substantia nigra, which plays a crucial role in movement control [1].

The diagnosis of Parkinson's disease (PD) involves identifying both motor and non-motor symptoms, which can be challenging due to the progressive nature of the disease. Accurate and early diagnosis is crucial for effective management and treatment. Traditional diagnosis methods are often subjective, leading to a growing need for objective tools. Recent advancements in signal processing and machine learning have shown high potential in improving the accuracy of PD diagnosis. This study focuses on utilizing accelerometer (ACC) signals and advanced feature extraction techniques to develop a robust classification system for distinguishing between healthy individuals and those with PD [2].

Accelerometer signals (ACC) with X, Y, and Z axes provide crucial insights into movement dynamics and orientation tracking. An accelerometer detects acceleration, which is the rate of change in velocity, along three orthogonal axes [3]:

- X-axis: Represents forward-backward movement. Positive values indicate forward acceleration, while negative values indicate backward acceleration.
- Y-axis: Reflects side-to-side movement. Positive values denote rightward acceleration, while negative values denote leftward acceleration.
- Z -axis: Corresponds to up-down movement. Positive values indicate upward acceleration, and negative values indicate downward acceleration.

Several research works are carried out for the analysis, evaluation and identification of PD using ACC signals. Veltink et al [1] investigated the use of uniaxial accelerometers to detect both static (stationary) and dynamic (moving) activities. Key features extracted in this approach include acceleration peaks, duration of movements, and transitions between different types of activities. Mathie et al [4] emphasized the use of accelerometery for

continuous, long-term monitoring of human movement, extracting features such as amplitude, frequency, and variability of movement, which provide detailed and sustained data about a person's physical activity. Hossen et al. [5] presented a method for distinguishing between Parkinsonian tremor (PT) and essential tremor (ET) using statistical signal characterization of accelerometer signals. Features such as power spectral density, signal entropy, and time-frequency representations are crucial for this classification. Electromyogram (EMG) and ACC signals continue to be used in the classification of neuromuscular dysfunctions [6]. In the study developed by Ghassemi et al, [7], the authors explored the combined use of accelerometer and electromyography analysis to differentiate between ET and PD. Aich et al [8] focused on using a wearable accelerometer to detect and potentially predict freezing of gait (FoG) in PD patients. Important features included stride length, stride time, gait speed, and variability in gait patterns, which are indicative of gait disturbances in PD. Aich et al [9] also demonstrated that stride interval fluctuations are increased in PD and correlate with the severity of the disease, using features such as stride interval variability, mean stride interval, and the coefficient of variation to quantify these fluctuations.

The challenge of classifying neurodegenerative disorders, such as Parkinson's disease, is closely related to feature extraction and classification approaches, which were primarily developed in the context of gait analysis [10]. Rastegari et al [11] used machine learning to classify Parkinson's disease, extracting informative features from accelerometer data related to gait and employing algorithms like Support Vector Machine (SVM) and Random Forests to distinguish PD patients from healthy subjects. In [12], specific features from acceleration signals have been extracted and used with an SVM to classify them as representing either normal gait or freezing of gait (FOG). The authors of [13] proposed a multiclass classification to determine the severity level of PD (mild, moderate, severe) using both empirical wavelet packet transform (EWPT) and empirical wavelet transform (EWT) on movement and audio signals.

The principal aim of the present work is **to** implement a high-performance diagnostic system for Parkinson's disease (PD). This involves the method for extracting WCC parameters. The work also focuses on studying the importance of each axis of the ACC signal for diagnostic purposes and the fusion of information from different ACC signal axes by applying parameter selection strategies based on mutual information.

The organization of this study is as follows. Section 2 outlines the proposed classification system, including the feature extraction and selection steps. Section 3 presents the experiments and their results. Section 4 concludes the paper with suggestions for future research.

2 Classification of ACC Signal

2.1 Dataset

The French national research project ECOTECH database [14] is a comprehensive resource used for the study of neuromuscular dysfunction, particularly in the context of Parkinson's disease (PD). This database includes detailed recordings of electromyographic (EMG) and accelerometer (ACC) signals from six different muscles. The muscles typically studied in this database are selected to provide a thorough understanding of movement and muscle activity in both healthy individuals and those with PD. In this particular research, we focused on the biceps muscle of the ACC signal with X, Y, and Z. Data from this muscle were used to analyze movement patterns and muscle activity in eight Parkinsonian patients and nine healthy subjects. The tibialis muscle was chosen due to its significant role in arm movements, which are often affected in PD. Accelerometric signals, derived from tri-axial sensors (X, Y, Z), provide precise measurements of body movements by capturing acceleration variations across three dimensions. measure accelerations at sampling frequencies ranging from 50 Hz to 500 Hz, with resolutions of 8 to 16 bits and typical sensitivities of ±2 g to ±16 g. Raw recorded signals, which may include noise artifacts caused by external vibrations, are generally filtered (e.g., using low-pass or band-pass filters) to isolate relevant frequencies, typically between 0.1 Hz and 20 Hz. Tables 1 and 2 present a description of the data base.

Table 1. Distribution of data in testing and training dataset.

Subjects		Signal duration of ACC (second)
Data base for training phase	$Control_1$	2.6069
	$Control_2$	1.1213
	$Control_3$	1.4400
	$Control_4$	1.4744
Data base for testing phase	$Control_5$	1.1164
	$Control_6$	1.4346
	$Control_7$	0.6530
	$Control_8$	0.7712
	$Control_9$	2.8571

Table 2. Distribution of data in testing and training dataset.

Subjects		Signal duration of ACC (second)
Data base for training phase	$Park_1$	0.3777
	$Park_2$	0.9132
	$Park_3$	0.4866
	$Park_4$	3.9616
Data base for testing phase	$Park_5$	1.1288
	$Park_6$	0.8915
	$Park_7$	0.4474
	$Park_8$	0.4649

2.2 Proposed Feature Extraction Method

A robust classification system relies on effective feature extraction. In this process, each signal is initially transformed into a sequence of overlapping windowed signals, and each of these windowed signals is then converted into a feature vector. As a result, the ACC signal is represented as a sequence of these feature vectors. For this study, we utilized the descriptors Discrete Wavelet Energy (DWE), Log Wavelet Energy (LWE), and Wavelet Cepstral Coefficients (WCC) [15]. The choice of this features as descriptors is justified by their ability to capture detailed and discriminative information from accelerometer signals. WCC combines both time and frequency analyses, allowing for better distinction of subtle movement characteristics essential for identifying neuromuscular dysfunctions such as those associated with Parkinson's disease. Their robustness against noise and individual variations makes them a superior option for precise and reliable classification Figure 1 illustrates the procedure for extracting these attributes.

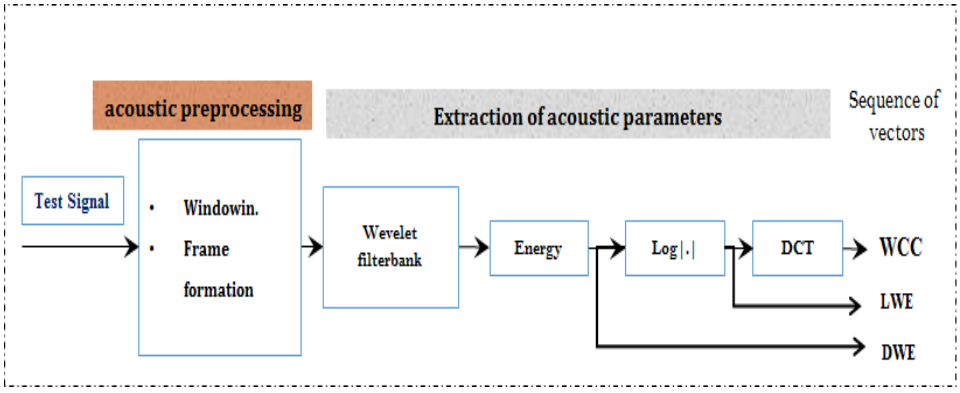

Fig. 1. block diagram illustrating the process of DWE, LWE and WCC features extraction.

2.3 Classification System

The proposed classification system comprises learning and testing phases. In the learning phase, features are extracted and models for two classes, P and N, are constructed using the KNN method. The testing phase involves extracting features (Discrete Wavelet Energy (DWE), Log Wavelet Energy (LWE), and Wavelet Cepstral Coefficients (WCC)). The feature selection step in this work involves choosing the relevant features by maximizing mutual information. This process is conducted using various feature selection strategies, such as JMI [16], ICAP [17], CIFE [18], MRMR [19], and CMI [20]. Optimizing the KNN algorithm's performance involves identifying the most relevant features and classifying each feature vector using a KNN classifier. Subsequently, a voting rule is applied to classify sequences of feature vectors from the test database to determine the predominant class. The ACC signal database used originates from the French national project ECOTECH [14] and includes signals collected from multiple subjects, including 9 healthy individuals and 8 subjects with Parkinson's disease.

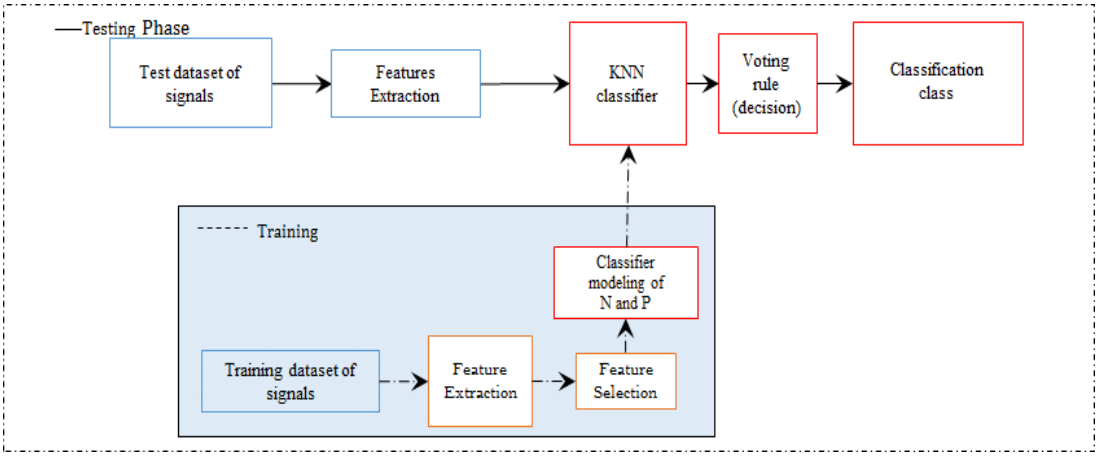

Fig. 2. Flowchart of the proposed automated classification system.

The classification performance of the ACC signal classification systems is evaluated using two metrics: the classification rate of signals (CRs) and the averaged accuracy of vectors (CRv).

The CRv is calculated by concatenating all signals and computing the ratio of correct decisions to the total number of input feature vectors, as defined by:

$$CRv(\%) = \frac{number\ of\ recognised\ signals}{\text{total}\ number\ of\ tested\ signals} \quad (1)$$

Additionally, the classification rate of signals (CRs) is determined using the majority voting rule, which assigns a class label based on the highest number of votes. This voting rule is applied separately for ACC signals. The CRs metric represents the ratio of correctly classified signals to the total number of signals and is defined as:

$$CRs = \frac{the\ number\ of\ correct\ decisions\ made\ for\ the\ signals}{total\ number\ of\ signals} \quad (2)$$

2.4 Features Selection

In large dimensional problems, dimensionality reduction of the number of features is often a necessity. The reduction can be achieved either by transforming the features and by selecting them. The solution to this problem consists of two methods. The first method is transforming the features of an initial set $\mathcal{F}$ of i features $\{N_1, N_2, \dots, N_i\}$ in a low dimensional subset of k features. But this approach requires calculating all the features

as well as choosing a suitable criterion for defining transformation, which is not a straightforward task. The second method consists in selecting the *k* most relevant features $\{N_{P_1}, N_{P_2}, \dots, N_{P_k}\}$ from the set *M* which form the subset $\mathcal{S}_{opt}$. This approach will be preferred; because it needs only the *k* selected features to be computed for the classification task in the testing phase. In opposite to the previous approach.

The feature selection procedure uses an information measure of a subset of features useful for a classification task. $\mathcal{S}_{opt}$ is an optimal subset of features if its information is maximum for the classification task . Mutual Information (MI) is often used as a measure of the quantity of information because of its ability of assessing the nonlinear statistical dependency between variables. So, the subset $\mathcal{S}_{opt}$ is chosen in such a way that the MI between $\mathcal{S}_{opt}$ and the class label C is maximized:

$$\mathcal{S}_{opt} = arg \max_{S \subset M} I(C;S) \tag{3}$$

However, the number of combinations of features for exhaustively constructing the sets $\mathcal{S}$ to be tested rapidly becomes prohibitive when the size of $\mathcal{S}$ grows. To circumvent this problem, ‘‘greedy forward” search strategy can be employed. The search is a one-by-one selection procedure that gives at each step *j* the best feature N_{P_j} from the unselected features set. This new selected feature N_{P_j} grows the already selected subset $\mathcal{S}_{j-1}$ by appending it as $\mathcal{S}_j = \mathcal{S}_{p_j} \cup \mathcal{S}_{j-1}$:

$$N_{p_j} = arg \max_{N_i \in \mathcal{F} - \mathcal{S}_{j-1}} \left[I\left(C; N_i, \mathcal{S}_{j-1}\right) \right] \tag{4}$$

Since $I\left(C; N_i, \mathcal{S}_{j-1}\right) = I\left(C; \mathcal{S}_{j-1}\right) + I\left(C; N_i \{ \mathcal{S}_{j-1}\right)$ [21], (4) can be reduced to:

$$N_{p_j} = arg \max_{N_i \in \mathcal{F} - \mathcal{S}_{j-1}} \left[I\left(C; N_i \setminus \mathcal{S}_{j-1}\right) \right] \tag{5}$$

Most of the algorithms propose a simplification of (4) following different strategies like MIFS, MRMR, JMI, CIFE, CMIM, MIM, and ICAP ([22]; [23] [24]. The derivation of (4) is given below, for three selected strategies.

- ICAP (Interaction Capping) [17]

$$N_{P_j} = arg \max_{N_i \in F - \mathcal{S}_{j-1}} \left[I(C;N_i) - \sum_{k=1}^{j-1} max[I\left(C;N_i;N_{s_k}\right), 0] \right] \tag{6}$$

- CIFE (Conditional Infomax Feature Extraction) [18]

$$N_{P_j} = arg \max_{N_i \in F - \mathcal{S}_{j-1}} \left[I(C;N_i) \sum_{k=1}^{j-1} I(C;N_i;N_k) \right] \tag{7}$$

- JMI (Joint Mutual Information) [16]

$$N_{P_j} = arg \max_{N_i \in F - \mathcal{S}_{j-1}} \left[I(C;N_i) - \frac{1}{j-1} \sum_{k=1}^{j-1} I(C;N_i;N_{\mathrm{Pk}}) \right] \tag{8}$$

- CMIM (Conditional Mutual Information Maximization) [20]

$$N_{p_{j+1}} = arg \max_{N_i \in \mathcal{F} - \mathcal{S}_j} \left[\min_{N_{p_k} \in \mathcal{S}_j} \left[I\left(C; N_i \setminus N_{p_k}\right) \right] \right] \tag{9}$$

- MIFS (Mutual Information Feature Selection) [25]

$$N_{p_j} = arg \max_{N_i \in \mathcal{F} \setminus \mathcal{S}_{j-1}} I(C; N_i) \quad (10)$$

- MIM (Mutual Information Maximization) [20]

$$N_{p_j} = arg \max_{N \in F - \mathcal{S}_j} [IM(C; N_i)] \quad (11)$$

- MRMR (Maximum-Relevance Minimum Redundancy) [19]

$$N_{p_j} = arg \max_{N_i \in F - \mathcal{S}_{j-1}} \left[I(C; N_i) - \frac{1}{j-1} \sum_{k=1}^{j-1} I(N_i; N_{p_k}) \right] \quad (12)$$

3 Experimental Results

3.1 Experiments

In this section, we outline the various experiments we conducted to evaluate the performance of our algorithm using features extracted from the ACC signals of the ECOTECH dataset [14] . These experiments utilized a KNN classifier. The tests included the following:

- Explored the optimal feature configuration by evaluating different feature types (DWE, LWE, and WCC) based on the classification rate for ACC signals from various channels (X, Y, and Z)
- Examined the optimal signal channel by comparing the classification rates of ACC signals across the X, Y, and Z axes.
- Applied a mutual information (MI)-based feature selection method using multiple strategies MIFS, ICAP, JMI, CIFE, MIM, MRMR, and CMIM to identify the optimal signal features.
- Used the KNN classifier to validate the relevance of the selected features at iteration j.

3.2 Results and Discussion

Comparative Study between Different Features Type

We carried out comparative experiments with respect to different parameters (DWE, LWE and WCC) for different modality ACC_X, ACC_Y , ACC_Z *and* ACC_{XYZ} to investigate the influence of these factors on the classification performance. For these descriptors, we consider the mother wavelet Coif 5 for a decomposition level L_{decomp} = 4 for analysis frame duration equal to 132.91*ms* as reference features and using the KNN classifier (taking the optimal K).

ACC_X signal

The results of a classification task, highlight the critical importance of feature. Table 3 compares the CRv and CRs achieved for an $\boldsymbol{ACC_X}$ signal using different features. This table suggests that using five-dimensional feature vectors derived from WLE or WCC features enables achieving an optimal CRv of 91.62 % and CRs of 100% with optimal number of nearest neighbors K equal to 22. This indicates that models employing LWE or WCC features outperform those using other wavelet-based features (such as DWE) due to their ability to analyze

frequency content and temporal dynamics. This finding underscores the importance of carefully selecting features to optimize classification performance.

Table 3. Comparison of CRv (%) and CRs(%) for different feature types with optimal number of nearest neighbors K using Coiflets (coif 5) at level 4 *[15]* for the modality ACC_X.

ACC_X	DWE	LWE	WCC
$K_{optimal}$	42	22	22
CR(%)	56.88 $_{(5)}$	91.62$_{(5)}$	91.62$_{(5)}$
CRs(%)	100 $_{(5)}$	100 $_{(5)}$	100 $_{(5)}$

ACC_Y signal

Table 4 presents the classification outcomes for the ACC_Y signal using different feature types: DWE, LWE, and WCC. The table underscores the importance of feature optimization in achieving optimal classification performance. Specifically, LWE and WCC features deliver superior performance, achieving a validation CRv of 96.93% with an optimal number of nearest neighbors ($K_{optimal}$) equal to 5. In contrast, DWE features result in a considerably lower CRv of 67.25% at k=6 clusters. Moreover, the CRs is uniformly 100% for all three feature types. These results highlight the impact of feature selection on classification accuracy and suggest that LWE and WCC offer a more effective approach compared to DWE.

Table 4. Comparison of CRv (%) and CRs(%) for different feature types with optimal number of nearest neighbors K using Coiflets (coif 5) at level 4 *[15]* for the modality ACC_Y.

ACC_Y	DWE	LWE	WCC
$K_{optimal}$	6	5	5
CRv(%)	67.25 $_{(5)}$	96.93 $_{(5)}$	96.93 $_{(5)}$
CRs(%)	100 $_{(5)}$	100 $_{(5)}$	100 $_{(5)}$

ACC_Z signal

As in the previous cases, Table 5 presents the CRv and CRs for the ACC_Zsignal using three different feature types: DWE, LWE, and WCC. For DWE, a CRv of 61.31% and a CRs of 100% are achieved at k = 10, whereas both LWE and WCC obtain an impressive CRv of 94.28% and a perfect CRs of 100% when k is set to 39. These results further confirm the superior performance of LWE and WCC over DWE for the classification of the ACC_Zsignal.

Table 5 . Comparison of CRv (%) and of CRs(%) for different feature types with optimal number of nearest neighbors *K* using Coiflets (coif 5) at level 4 [15] for the modality ACC_Z.

ACC_Z	DWE	LWE	WCC
$K_{optimal}$	10	39	39
CRv(%)	61.31 $_{(5)}$	94.28$_{(5)}$	94.28$_{(5)}$
CRs(%)	100 $_{(5)}$	100 $_{(5)}$	100 $_{(5)}$

ACC_{XYZ} signal

In this case, all three signals are collected and analyzed. Table 6 presents the classification results for the ACC_{XYZ} (accelerometer data for X, Y, and Z) signal using three different feature types: DWE, LWE, and WCC. The optimal number of nearest neighbors $K_{optimal}$is 5 for DWE, whereas LWE and WCC both achieve optimal performance with K set to 27. In terms of CRv and CRs, DWE reaches 62.81% at k=5, while LWE and WCC significantly outperform it, both achieving identical CRv values of 94.82%. This result is obtained when K is set to 27. Moreover, every feature type consistently achieves a CRs of 100%.

These findings highlight the superior effectiveness of LWE and WCC over DWE for the ACC_{XYZ}signal in this classification task

Table 6. Comparison of CRv (%) and $CR\,s(\%)$ for different feature types with optimal number of nearest neighbors *K* using Coiflets (coif 5) at level 4 [15] for the modality $\boldsymbol{ACC_{XYZ}}$.

$\mathbf{ACC_{XYZ}}$	DWE	LWE	WCC
$\mathbf{K_{optimal}}$	5	27	27
CRv(%)	$62.81_{(15)}$	$94.82_{(15)}$	$94.82_{(15)}$
CRs(%)	$100_{(15)}$	$100_{(15)}$	$100_{(15)}$

In the next sections, the value of nearest neighbor's *K* will be set to 27 be chosen.

Feature selection

This section addresses the challenge of feature selection for the classification task, aiming to identify the most relevant features. Features are deemed relevant if they effectively differentiate between the two classes, normal and Parkinson. Various selection algorithms employed in this experiment differ in terms of the extent to which they consider redundancy between a subset of already selected variables and those yet to be chosen. The algorithms under consideration include MRMR, JMI, CMIM, ICAP, MIFS or CIFE [22]. These algorithms are grounded in the principle of maximizing mutual information, as described in detail in Section 2.4.

In this section, we consider two criteria to determine the optimal number of relevant features after the selection process. For the first criterion (CRT1), the ideal number is the dimension of the smallest subset of selected features that achieves a classification rate (CR) equal to or higher than the CR of the full feature set. For the second criterion (CRT2), the ideal number is the dimension of the subset of features that yields the highest CR.

The classification results using different feature selection strategies and the two stopping criteria (CRT1 and CRT2) are presented in Table 7. The key observations from these results are as follows:

- **Under Criterion 1 (CR ≥ CR_end):**
  - The JMI strategy achieves a CRv of 94.89% using a subset of 3 features, which is comparable to the MRMRR strategy.
  - The CIFE strategy, however, achieves the highest CRv of 98.91% while selecting only 2 features, indicating the most efficient dimensionality reduction.
  - The ICAP method selects a slightly larger subset (4 features) but still achieves a high CRv of 98.56%.
- **Under Criterion 2 (CR = max(CR)):**
  - The CIFE and ICAP strategies both achieve the highest CRv, but CIFE requires only 2 features with CRv of 98.91%, making it the most effective for dimensionality reduction.
  - The MIFS, CMIM, and MIM methods all achieve a CRv of 98.09% but differ in the number of selected features, with MIFS selecting the largest subset (7 features).
  - The MRMR strategy consistently selects the largest subset of features under both criteria.

Table 7. CR and number of relevant features by using MRMR,CMIM, CIFE, JMI,MIFS and ICAP strategies with selection criteria CRT1 and CRT2 [24].

	Criterion 1 CR>=CR(end)	Criterion 2 CR==max(CR)

	Number of relevant features	CRv (%)	CRs(%)	Number of relevant features	CRv (%)	CRs(%)
MRMR	3	94.89	100	5	97.75	100
CMIM	2	96.66	100	5	98.09	100
JMI	3	94.89	100	5	97.00	100
MIFS	2	97.27	100	7	98.09	100
MIM	2	96.66	100	4	98.09	100
ICAP	4	98.56	100	4	98.56	100
CIFE	2	98.91	100	2	98.91	100

More information about the first ten selected features, as determined by all feature selection strategies, is shown in Table 8.

Table 8. The first ten selected features obtained for different MI strategies for the WCC descriptor of the three modalities ACC_{XYZ}.

	Subset of selected features
MRMR	$ACC_{Y;8}^{(WCC)}; ACC_{Z;11}^{(WCC)}; ACC_{Y;7}^{(WCC)}; ACC_{Y;10}^{(WCC)}; ACC_{X;1}^{(WCC)}; ACC_{X;2}^{(WCC)}; ACC_{Z;12}^{(WCC)}; ACC_{X;4}^{(WCC)}; ACC_{Z;14}^{(WCC)}; ACC_{Z;15}^{(WCC)}; ACC_{Z;11}^{(WCC)};$
CMIM	$ACC_{Y;8}^{(WCC)}; ACC_{Y;7}^{(WCC)}; ACC_{Z;15}^{(WCC)}; ACC_{Y;5}^{(WCC)}; ACC_{X;1}^{(WCC)}; ACC_{X;3}^{(WCC)}; ACC_{Z;12}^{(WCC)}; ACC_{Y;6}^{(WCC)}; ACC_{Z;14}^{(WCC)}; ; ACC_{X;3}^{(WCC)};$
JMI	$ACC_{Y;8}^{(WCC)}; ACC_{X;5}^{(WCC)}; ACC_{Z;15}^{(WCC)}; ACC_{Y;5}^{(WCC)}; ACC_{Z;11}^{(WCC)}; ACC_{Y;10}^{(WCC)}; ACC_{Y;6}^{(WCC)}; ACC_{Y;9}^{(WCC)}; ACC_{Z;14}^{(WCC)}; ACC_{X;3}^{(WCC)};$
MIFS	$ACC_{Y;8}^{(WCC)}; ACC_{Y;7}^{(WCC)}; ACC_{X;1}^{(WCC)}; ACC_{Z;15}^{(WCC)}; ACC_{X;5}^{(WCC)}; ACC_{y;6}^{(WCC)}; ACC_{X;3}^{(WCC)}; ACC_{x;4}^{(WCC)}; ACC_{Y;11}^{(WCC)}; ACC_{Y;10}^{(WCC)};$
MIM	$ACC_{Y;8}^{(WCC)}; ACC_{Y;7}^{(WCC)}; ACC_{X;1}^{(WCC)}; ACC_{Z;15}^{(WCC)}; ACC_{X;4}^{(WCC)}; ACC_{X;5}^{(WCC)}; ACC_{Y;6}^{(WCC)}; ACC_{Z;11}^{(WCC)}; ACC_{Y;9}^{(WCC)}; ACC_{Z;13}^{(WCC)}$
ICPAP	$ACC_{Y;8}^{(WCC)}; ACC_{Y;7}^{(WCC)}; ACC_{X;2}^{(WCC)}; ACC_{Z;13}^{(WCC)}; ACC_{Z;12}^{(WCC)}; ACC_{X;3}^{(WCC)}; ACC_{Y;9}^{(WCC)}; ACC_{Z;11}^{(WCC)}; ACC_{Z;14}^{(WCC)}; ACC_{Z;15}^{(WCC)}$
CIFE	$ACC_{Y;8}^{(WCC)}; ACC_{Y;9}^{(WCC)}; ACC_{X;1}^{(WCC)}; ACC_{Z;15}^{(WCC)}; ACC_{Y;6}^{(WCC)}; ACC_{X;4}^{(WCC)}; ACC_{X;5}^{(WCC)}; ACC_{Z;11}^{(WCC)}; ACC_{Y;7}^{(WCC)}; ACC_{Z;13}^{(WCC)}$

The table's results indicate that the ACC_Y feature is more relevant compared to the others. Through this experiment, we notice that the selected features of the ACC are of great importance in this task. Notably, all methods consistently identified the same initial feature $\mathrm{ACC}_{Y;8}^{(WCC)}$.

Figure 3 illustrates how the classification rate varies with the number of selected features, evaluated using seven feature selection strategies: CMIM, JMI, ICAP, CIFE, MRMR, MIM, and MIFS. The analysis reveals that the best performance a CRv of 98.91% and a CRs of 100% is achieved with only two features selected by the CIFE strategy. This feature selection process effectively reduces the number of features needed for classification. Additionally, the graphs highlight the occurrence of the peaking phenomenon.

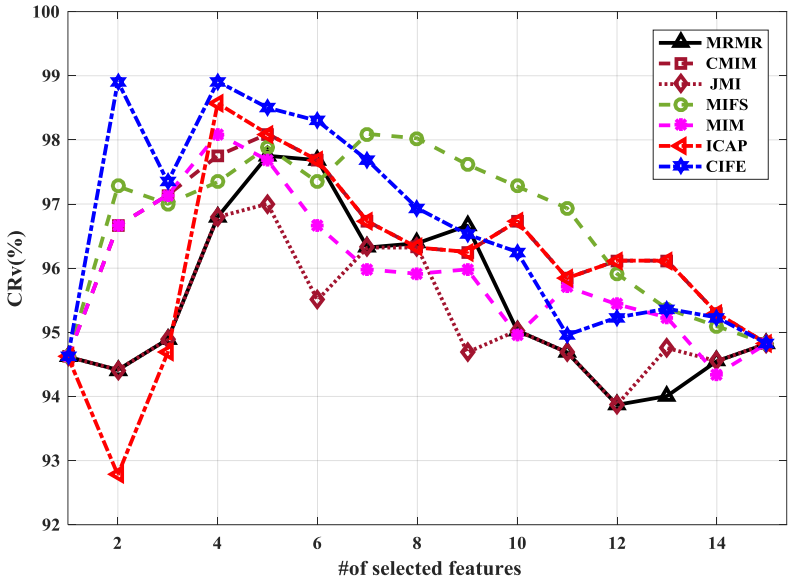

Fig. 3 CRv(%) as a function of the number of selected features with the five feature selection strategies.

4 Conclusions

The classification and diagnosis of diseases affecting muscle activity has important clinical application. The present paper describes new approach to deal with Parkinson's disease classification based on Accelerometer ($ACC_X; ACC_Y; ACC_Z; ACC_{XYZ}$) signal. The method employs a k-NN classifier combined with a voting role. The proposed methodology is tested on a subset of (ACC_{XYZ}) signals obtained from the French national research project ECOTECH, dataset.

The proposed method of feature extraction of the signals has been applied on DWE, LWE, WCC descriptors. The results firstly illustrate the significance of the LWE and WCC descriptors in comparison to other descriptors whether the type of the signals.

To reduce dimensionality, we applied several feature selection strategies based on the MI maximization criterion to the feature set. These strategies effectively reduced the number of features while achieving higher classification rate values than those obtained with the full feature set. Results indicate that the CIFE strategy yielded the highest accuracy of vectors of 98.91% with 100% of signals classification rate (CRs) with the fewest features selected (two features) using the Conditional Infomax Feature Extraction filter selection strategy.

Additionally, our findings reveal that features from the $\mathrm{ACC}_{Y;8}^{(WCC)}$. modality predominantly contribute to classification. Across all results, the first relevant feature consistently came from the ACC_Y modality, regardless of the selection strategy or descriptor type.

REFERENCES

[1] P. Veltink, H. Bussmann, W. Vries, W. Martens and R. van Lummel, "Detection of static and dynamic activities using uniaxial accelerometers," 1996 .

[2] M. Alissa, M. Lones, J. Cosgrove, J. Alty, S. Jamieson, S. Smith and M. Vallejo, "Parkinson's disease diagnosis using convolutional neural networks and figure-copying task," 2021.

[3] R. San-Segundo, A. Zhang, A. Cebulla, S. Panev, G. Tabor, K. Stebbins, R. Massa, A. Whitford, F. Torre and J. Hodgins, “Parkinson’s Disease Tremor Detection in the Wild Using Wearable Accelerometers,” ; Sensors 2020, 20, 5817 doi:10.3390/s20205817 .(2020).

[4] M. Mathie, N. Coster, N. Lovell and B. Celler, “Accelerometry: Providing an integrated, practical method for long-term,ambulatory monitoring of human movement,” . Meas., vol. 25, pp. R1–R20.,(2004) .

[5] A. Hossen , M. Muthuraman , J. Raethjen , G. Deuschl and U. Heute , “Discrimination of Parkinsonian tremor from essential tremor by implementation of a wavelet-based soft-decision technique on EMG and accelerometer signals.,” . Biomed Signal Process Control. ; 5: 181-188,(2010).

[6] A. Hossen , M. Muthuraman , Z. Al-Hakim, J. Raethjen , G. Deuschl and U. Heute , “Discrimination of Parkinsonian tremor from essential tremor using statistical signal characterization of the spectrum of accelerometer signal.,” ; Bio-Med Mater Eng. ; 23: 513-531;(2013).

[7] H. Ghassemi, F. Marxreiter, F. Pasluosta, P. Kugler, J. Schlachetzki, A. Schlachetzki and M. Eskofier, “Combined Accelerometer and EMG Analysis to Differentiate Essential Tremor from Parkinson’s Disease,” ; IEEE, EMBS, Jochen Klucken 978-1-4577-0220-4/16/$31.00 ©European Union.; (2016) .

[8] S. Aich , P. Pradhan , J. Park , N. Sethi , V. Vathsa and H. Kim , “A validation study of freezing of gait (fog) detection and machine-learning-based fog prediction using estimated gait characteristics with a wearable accelerometer.,”. Sensors 18, 3287. 10.3390/s18103287.(2018)

[9] S. Aich, P. Pradhan, S. Chakraborty, H. Kim, H. Lee, I. Kim, M. Joo and S. Seong, “Design of a Machine Learning-Assisted Wearable Accelerometer-Based Automated System for Studying the Effect of Dopaminergic Medicine on Gait Characteristics of Parkinson's Patient,” Hindawi Journal of Healthcare Engineering Volume , Article ID 1823268, 11 page https://doi.org/10.1155/2020/1823268? (2020).

[10] T. Carletti, D. Fanelli and A. Guarino, “A new route to non invasive diagnosis in neurodegenerative diseases ?,” Neuroscience letters, 394(3) :252–255Neuroscience letters, 394(3) :252–255.; (2006).

[11] E. Rastegari, A. Sasan and A. Hesham , “Machine Learning and Similarity Network Approaches to Support Automatic Classification of Parkinson’s Diseases Using Accelerometer-based Gait Analysis,” Proceedings of the 52nd Hawaii International Conference on System Sciences.(2019).

[12] D. Rodríguez-Martín and C. Pérez-López , “Detecting freezing of gait with a tri-axial accelerometer in parkinson's disease patients.,”, Med. Biol. Eng. Comput. 54, 223–233. 10.1007/s11517-015-1395-3. (2016).

[13] Q. Oung, H. Muthusamy, S. Basah, H. Lee and V. Vijean, “Empirical wavelet transform based features for classification ofparkinson’s disease severity,” Journal of medical systems, 42(2) :29,(2018).

[14] O. Buttelli, “Agence nationale de la recherche.,” http://www.agence-nationale-recherche.fr/ProjetANR-12-TECS-0020 ;(2012).

[15] H. Bengacemi, A. Hacine-Gharbi and P. Ravier, “Surface EMG Signal Classification for Parkinson’s Disease using WCC Descriptor and ANN Classifier,”, 10th International Conference on Pattern Recognition Applications and Methods. (2021).

[16] H. Young, I. Moody, . J. Feature selection based on joint mutual information. Intelligent Data Analysis (AIDA) and Computational Intelligent Methods and Application (CIMA).,”(1999).

[17] A. Jakulin, “Learning based on attribute interactions.,”; Computational Statistics and Data Analysis,Vol. 49, No. 4, pp.1205–1227.; (2005).

[18] I. Kojadinovic, “Relevance measures for subset variable selection in regression problems based on k-additive mutualinformation,” Computational Statistics and Data Analysis,Vol. 49, No. 4, pp.1205–1227.,(2005).

[19] H. L. F. a. D. C. Peng, “Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy,” ; IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 27, No. 8, pp. 1226–1238;(2005).

[20] F. Fleuret, “Fast binary feature selection with conditional mutual information,” Journal of Machine Learning Research, 5:1531–1555; (2004).

[21] T. Cover and J. Thomas, “Elements of Information Theory.,” New York, New York: Wiley Series In Telecommunications.. (1991).

[22] G. Brown, A. Pocock, M.-J. Zhao and M. Lujan, “Conditional Likelihood Maximisation A UnifyingFramework for information theoretic featur selection.,” . Journal of Machine Learning Research, 13(1), pp.27-66.;(2012).

[23] A. Hacine-Gharbi, M. Deriche, P. Ravier, R. Harba and T. Mohamadi, “A new histogram-based estimation technique of entropy and mutual information using mean squared error minimization.,” n. Comput. Electr. Eng. 39 (3), 918–933. (2013).

[24] R. Touahria, A. Hacine- Gharbi, P. Ravier and M. Mostefai, “Relevant features selection of multi domains based on mutual information for heart sound classification,”; International Conference on Pattern Recognition Applications and Methods. (2024).

[25] R. Battiti, “Using mutual information for selecting features in supervised neural net learning,” IEEE Trans. Neural Networks, vol. 5, no. 4, pp. 537–550; (1994)..