Biomedical Signal Processing and Control 87 (2024) 105558

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

Machine learning-based cardiac activity non-linear analysis for discriminating COVID-19 patients with different degrees of severity

Pedro Ribeiro [a], João Alexandre Lobo Marques [b], Daniel Pordeus [c], Laíla Zacarias [d], Camila Ferreira Leite [d], Manoel Alves Sobreira-Neto [e], Arnaldo Aires Peixoto Jr [e], Adriel de Oliveira [f], João Paulo do Vale Madeiro [c], Pedro Miguel Rodrigues [a,*]

[a] Universidade Católica Portuguesa, CBQF–Centro de Biotecnologia e Química Fina–Laboratório Associado, Escola Superior de Biotecnologia, Rua de Diogo Botelho 1327, Porto, 4169-005, Portugal
[b] University of Saint Joseph, Laboratory of Applied Neurosciences, 999078, Macao Special Administrative Region of China
[c] Federal University of Ceará, Department of Computing, Fortaleza, Ceará, Brazil
[d] Federal University of Ceará, Graduate Program in Cardiovascular Sciences, Fortaleza, Ceará, Brazil
[e] Federal University of Ceará, Department of Clinical Medicine, Faculty of Medicine, Fortaleza, Ceará, Brazil
[f] University for the International Integration of the Afro–Brazilian Lusophony, Fortaleza, Ceará, Brazil

ARTICLE INFO

Keywords:
COVID-19
ECG signals
Non-linear analysis
Machine learning classifiers
Accuracy
F1–Score

ABSTRACT

Objective: This study highlights the potential of an Electrocardiogram (ECG) as a powerful tool for early diagnosis of COVID-19 in critically ill patients with limited access to CT–Scan rooms.
Methods: In this investigation, 3 categories of patient status were considered: Low, Moderate, and Severe. For each patient, 2 different body positions have been used to collect 2 ECG signals. Then, from each collected signal, 10 non-linear features (Energy, Approximate Entropy, Logarithmic Entropy, Shannon Entropy, Hurst Exponent, Lyapunov Exponent, Higuchi Fractal Dimension, Katz Fractal Dimension, Correlation Dimension and Detrended Fluctuation Analysis) were extracted every 1s ECG time-series length to serve as entries for 19 Machine learning classifiers within a leave-one-out cross-validation procedure. Four different classification scenarios were tested: Low vs. Moderate, Low vs. Severe, Moderate vs. Severe and one Multi-class comparison (All vs. All).
Results: The classification report results were: (1) Low vs. Moderate - 100% of Accuracy and 100% of $F1–Score$; (2) Low vs. Severe - Accuracy of 91.67% and an $F1–Score$ of 94.92%; (3) Moderate vs. Severe - Accuracy of 94.12% and an $F1–Score$ of 96.43%; and (4) All vs All - 78.57% of Accuracy and 84.75% of $F1–Score$.
Conclusion: The results indicate that the applied methodology could be considered a good tool for distinguishing COVID-19's different severity stages using ECG signals.
Significance: The findings highlight the potential of ECG as a fast and effective tool for COVID-19 examination. In comparison to previous studies using the same database, this study shows a 7.57% improvement in diagnostic accuracy for the All vs All comparison.

1. Introduction

The outbreak of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in 2019 caused a pandemic, resulting in a significant increase in the number of confirmed cases and deaths worldwide. Effective and efficient diagnosis of COVID-19 is crucial to control its spread and minimize its impact on public health [1–3].

There are several ways to detect COVID-19. The most widely used method is the RT-PCR (Reverse Transcription Polymerase Chain Reaction) test, which is time-consuming, requires specialized laboratory equipment, and has sensitivity and specificity limitations. Another test is the Rapid Antigen Test, which detects specific proteins on the virus surface and provides quick results but is less accurate than the RT-PCR test. Antibody tests check for the presence of antibodies in the blood, which are produced by the body in response to an infection and are

* Correspondence to: Universidade Católica Portuguesa, CBQF–Centro de Biotecnologia e Química Fina, Escola Superior de Biotecnologia, Office EBI112, Rua Diogo Botelho 1327, 4169-005 Porto, Portugal.
E-mail address: pmrodrigues@ucp.pt (P.M. Rodrigues).

https://doi.org/10.1016/j.bspc.2023.105558
Received 22 May 2023; Received in revised form 20 August 2023; Accepted 30 September 2023
Available online 7 October 2023

Table 1
State-of-the-art literature report on COVID-19 severity activity discrimination area with information about the source, comparison groups, used classifiers, and classification metrics.

Ref	Year	Source	Comparison group	Classifier	Accuracy	$F1-Score$
[19]	2021	X-ray	High vs Moderate vs Mild Severity	XGBoost	97.00%	96.00%
[20]	2020	X-ray	Severe vs Non-severe	DenseNet-201	95.34%	–
[21]	2021	X-ray	Mild vs Moderate vs Severe vs Critical	CNN	95.52%	97.00%
[22]	2020	X-ray	Severe vs Non-severe	SVM	81.48%	–
[14]	2020	Laboratory parameters	Survival vs Non-survival	SVM	91.25%	94.74%
[23]	2023	ECG (previous study from same authors)	All vs All	LinearSVC	71.00%	76.00%

used to determine if a person has been previously infected [4]. Chest X-ray or CT Scans can help to identify lung damage caused by COVID-19 [4]. It is important to note that no single test is 100% accurate and to address these limitations, there has been a growing interest in developing alternative diagnostic methods based on image and signal processing [5].

Due to the nature of the virus, COVID-19 will present a higher impact on the respiratory system. However, The efferent activities of the Autonomic nervous system (ANS) directed to the sinus node are synchronized with each cardiac cycle, being modulated both by the respiratory center and by the respiratory movement [6]. Thus, the influence of the ANS on the heart is dependent on changes in the respiratory system, specifically when considering the high frequency (HF) component, with a variation of 0.15 to 0.4 Hz, indicating the action of the vagus nerve on the heart [7]. This close bond between the two systems can cause a variety of cardiac complications, with some studies showing the appearance of myocardial injury, heart failure, cardiogenic shock, and cardiac arrhythmias [8,9].

These complications are worrisome because they can take the patient's life. The use of ECG signals to detect COVID-19 came with some interesting findings. Several studies have found an increase in abnormal ECG recordings. The most common findings on the signal were ST-elevation/depression with in some cases the exam presenting a T-wave inversion [10].

To detect these changes in the signal, some studies took advantage of pre-trained Machine Learning Classifiers and applied Transfer Learning to those classifiers. Feature extraction and classification are applied by signal processing algorithms and classifiers. In some cases, this process has been done by using Convolution Neural Networks, a Deep Learning algorithm that is normally used in the classification of images [11–13].

The detection of COVID-19 is an important step in controlling the disease but prognosis prediction can counteract the terrible complications [14,15].

The use of a signal such as the ECG presents a good opportunity to predict the prognosis of COVID-19 because multiple studies showed that the visualization of an abnormal T wave or the presence of S-T segment elevation/depression could be a good prognostic indicator in predicting the mortality of COVID-19 patients [16–18].

Table 1 presents the current state-of-the-art COVID-19 severity activity discrimination methods results.

The main contributions and goals of the present work are twofold: (1) Propose new nonlinear features to characterize COVID-19 clinical severity stages; (2) Evaluate the performance of a set of Machine Learning classifiers with the proposed features as inputs for discrimination.

2. Methodology

This proposed methodology, illustrated in Fig. 1, is divided into 4 parts:

- Data Collection and Processing
- Feature Extraction
- Data Compressor
- Machine Learning Classification

2.1. Experimental setup

This work was done using 2 different coding languages: MATLAB and Python. MATLAB, version R2022a, was used to remove the ECG signal artifacts, extract the ECG non-linear features, and compress and organize data for classification. Python (version 3.9.12) was used for designing, performing, and getting discrimination reports from a set of Machine Learning models.

All work has been performed with a Macbook Pro 14 with an M1 Pro (8-Core CPU and 14-Core GPU) and 16 GB of RAM.

2.2. Data collection and processing

The database is a cross-sectional descriptive study of quantitative nature carried out at Hospital Universitário Walter Cantídio (HUWC) and Hospital Estadual Leonardo da Vinci (HELV), Ceará, Brazil, from May 2021 to January 2022, corresponding to the second and third wave of COVID-19 in Brazil.

The Ethics Committee approved the study for Research on Humans (CAAE - HELV: 47229221.9.3001.5684; CAAE - HUWC: 472292 21.9.0000.5045).

The database has 8 patients with low severity, 10 with moderate severity, and 33 with a severe stage of the disease. The youngest patient was 18 years old, and the oldest patient was 88 years old, with an average age of 54.59 and a standard deviation of 17.23.

The severity classification of COVID-19 was established according to the following criteria: Low (signal portion example in Fig. 2(a)) - mild clinical symptoms and no signs of pneumonia on imaging examination, Moderate (Fig. 2(b)) - the presence of fever and respiratory symptoms with radiological evidence of pneumonia, Severe (Fig. 2(c)) - cases meeting any of the following criteria: (1) respiratory distress (> 30 breaths/min); (2) oxygen saturation <93% at rest; (3) arterial partial pressure of oxygen (PaO2)/ fraction of inspired oxygen (FiO2)<300 mmHg and cases with chest imaging that showed evident lesion progression of >50% within 24-48 h, critical — cases that met any of the following criteria: (1) respiratory failure and requiring mechanical ventilation, (2) shock, (3) other organ failure requiring ICU care [24].

The ECG recordings were performed with the participants being placed in two different body positions, the supine (*Down*) position and the orthostasis (*Up*) position. After a 5-min period of resting in the supine position and stabilization of the signal [25], recording of heartbeats was started. During a 5-minute interval, the volunteers were instructed to remain in the supine position, motionless and silent. After this period, the subjects were submitted to the active postural manoeuvre (APM) and asked to stand up suddenly, in the shortest time

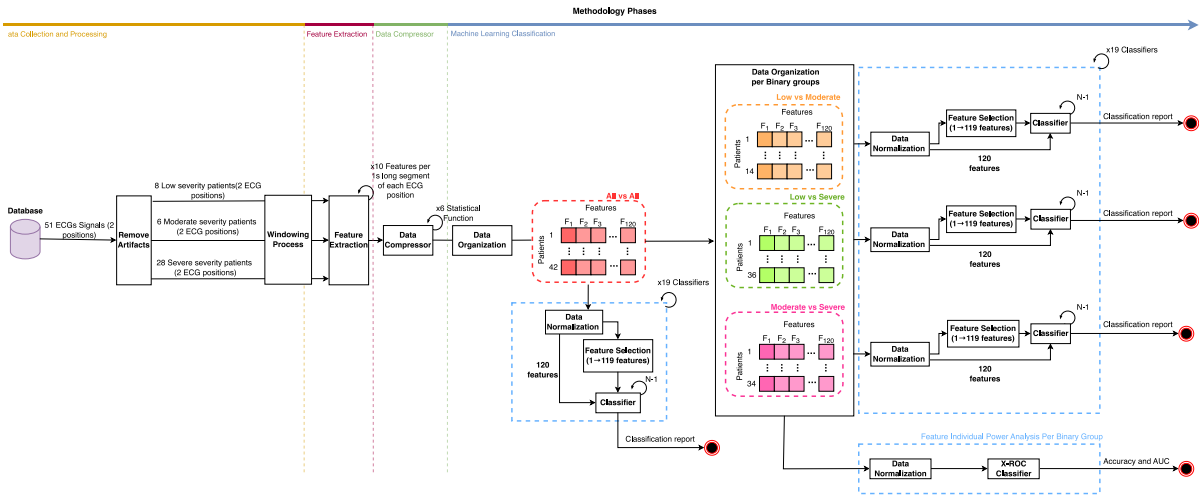

Fig. 1. Methodology phases and workflow diagram.

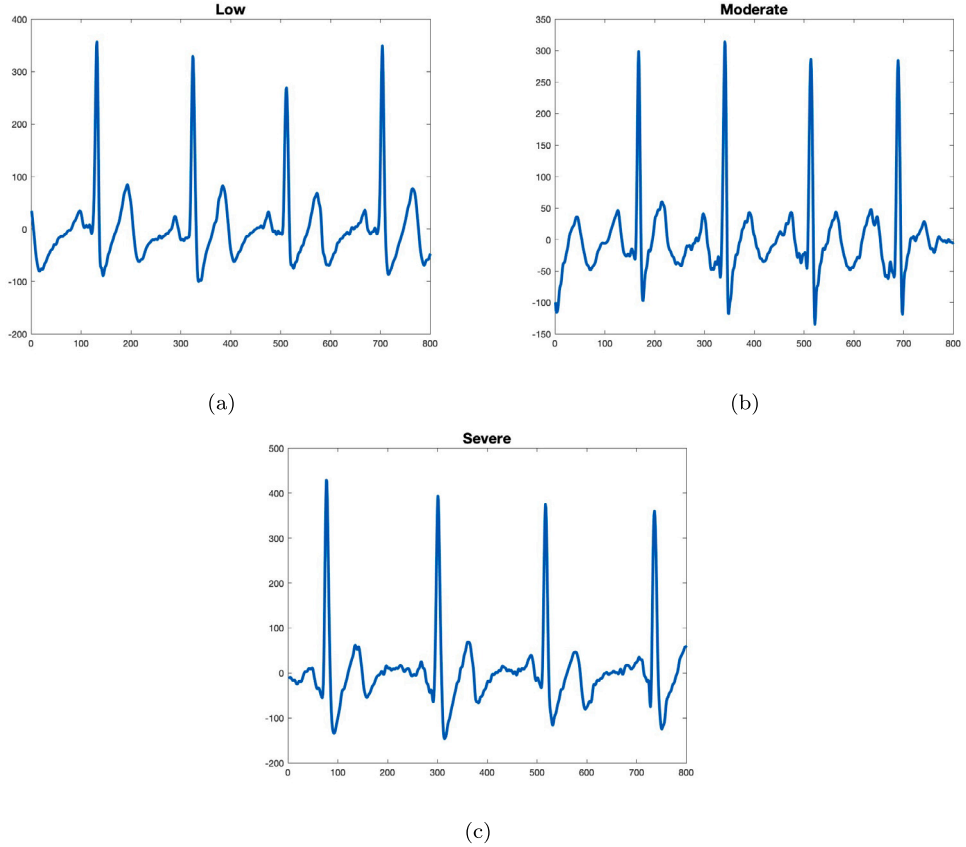

Fig. 2. ECG signal portion example for each class: (a) Low, (b) Moderate and (c) Severe.

possible, seeking the orthostatic position and remain there, without moving, for a 5-minute interval [26]. Soon after, they were instructed to return to the supine position, and then the electrocardiographic recording was completed and the electrodes were removed. Heartbeat recordings were made continuously during the entire maneuver. During the analysis of the records, the data obtained at the time of postural changes were excluded because of signal instability.

The frequency sample was 256 Hz, and the participants were monitored, in spontaneous breathing, in the modified Lead II.

Table 2 shows the Continuous Measure and Categorical Measure of the data present in the previously mentioned database.

2.3. Artifacts removal

The raw ECG signals data were analyzed by a Medical Doctor to identify artifacts. Subsequently, one of two actions was performed for each signal: (1) complete signal deletion if a significant portion of the signal was affected by artifacts or (2) removal of specific signal segments affected by artifacts, while retaining the portions with acceptable quality, as illustrated in Fig. 3.

In the beginning, the database had signals from 51 patients. After the artifacts removal stage, the number of available signals in the database for the following tasks was reduced to 42 patients (8 patients

Table 2
Socio-demographic and clinical characteristics of the database.

	Continuous measure	Min	Max	AVG	SD
Age		18	88	54.59	17.23
	Categorical measure			%	
Gender					
	Male			60	
	Female			40	
Clinical information					
	Low			16	
	Moderate			18	
	Severe			66	

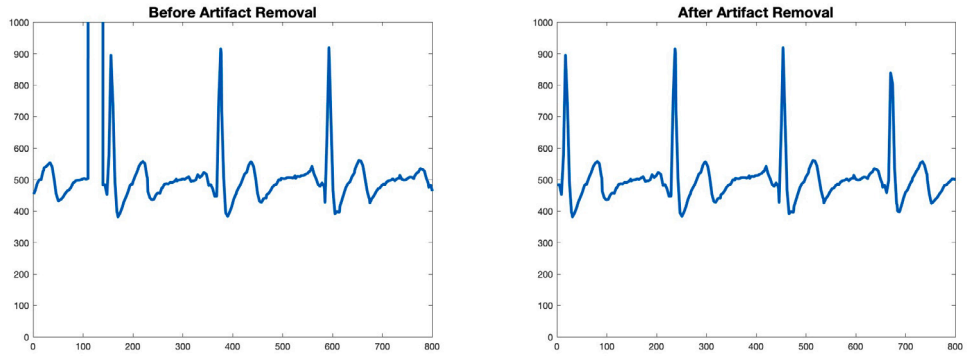

Fig. 3. Example of artifacts removal: (a) Before artifact removal, (b) After artifact removal.

with low severity, 6 patients with moderate severity, and 28 patients with a severe stage of the disease).

2.4. ECG non-linear analysis and feature extraction

The heart is a nonlinear system, and its electrical activity is reflected in the ECG signal [27]. This means that linear analysis methods may not be able to accurately capture the dynamics of the ECG signal. To address this, we opted to analyze the ECG in a nonlinear way, which is a more complex approach that takes into account the nonlinear nature of the heart.

In each patient ECG recording, 10 non-linear features were extracted from every 1 s-long segment within 6 ways of different windowing signal analysis crops: Rectangular non-overlapping windowing (Rec), Rectangular 0.5s overlapping windowing ($RecO$), Hamming non-overlapping windowing (Hm), Hamming 0.5s overlapping windowing (HmO), Hanning non-overlapping windowing (Hn) and Hanning 0.5s overlapping windowing (HnO). The ten extracted features were:

- **Energy** (En) of discrete signal $x(n)$ is defined as:

$$En = \sum_{n=0}^{N-1} \mid x(n) \mid^2 \tag{1}$$

The En, which represents the system's capacity to perform work, is derived by the sum of each squared signal component [28].

- **Entropy** is a metric that evaluates the energy content of a complex system, it can also be employed to quantify the masked information within a signal, characterizing the irregular and unpredictable nature of the pathological signals [28]. For this research, three different entropies are considered: Approximate (EA), Logarithmic ($ELog$), and Shannon ($ESha$). The three types of entropies can be defined as follows:

$$EA(m,r) = \lim_{N\to\infty} \Theta^m(r) - \Theta^{m+1}(r), \tag{2}$$

$$ELog = \sum_{n=1}^{N} \log[\mid x(n) \mid^2] \tag{3}$$

and

$$ESha = -\sum_{n=1}^{N} \mid x(n) \mid^2 \log[\mid x(n) \mid^2] \tag{4}$$

where N is the vector length, r denotes a tolerance value, Θ is the Heaviside step function and m is the dimension [29].

- **Hurst Exponent** (EH) is used to quantify how chaotic or unpredictable a time series is [30]. The equation can be defined as:

$$K_q(\tau) \sim (\frac{\tau}{\nu})^{qEH(q)}, \tag{5}$$

with

$$K_q(\tau) = \frac{((\mid X(t+\tau) - X(t)) \mid^q)}{(\mid X(t) \mid^q)}, \tag{6}$$

where q is the order moments of the distribution increments, ν is the time resolution and t is the period of a given time series signal $X(t)$ [30].

- **Lyapunov Exponent** ($ELya$) is a metric to evaluate the system predictability and sensitivity for changing [31].

$$ELay(x_0) = \lim_{n\to\infty} \frac{\sum_{k=1}^{n} \ln \mid f'(x_k - 1) \mid}{n}, \tag{7}$$

where f' is the integrator function f derivative [32].

- **Higuchi Fractal Dimension** (H) estimates the fractal dimension of a time series signal. The equation can be illustrated as:

$$H = \frac{\ln(L(k))}{\ln(\frac{1}{k})}, \tag{8}$$

where k is a number of composed sub-series and $L(k)$ is the averaged curve length.

- **Katz Fractal Dimension** (K) estimates the fractal dimensions through a waveform analysis of a time series [30]. The equation is defined as:

$$K = \frac{\log(n)}{\log(n) + \log(\frac{max_n(\sqrt{(n-1)^2+(X(n)-X(1))^2})}{\sum_{n=2}^{N}\sqrt{1+(X(n-1)-X(n))^2}})}, \quad (9)$$

where $X(n)$ is the signal.

- **Correlation Dimension** (CD) is used to measure self-similarity. A higher value of CD means an elevated degree of complexity and less-similarity [29].

$$CD = \lim_{M\to\infty} \frac{2\sum_{i=1}^{M-k}\sum_{j=i+k}^{M}\Theta(l \mid X_i - X_j \mid)}{M^2}, \quad (10)$$

where $\Theta(x)$ is the Heaviside step function, the X_i and X_j are the position vectors on attractor, l is the distance under consideration, k is the summation offset and M is the reconstructed vectors numbers from the original signal [29].

- **Detrended Fluctuation Analysis** (DFA), used to address non-stationary time-series, is a technique for measuring the power scaling observed through R/S analysis [33]. The DFA can be calculated as

$$DFA(n) = \sqrt{\frac{\sum_{k=1}^{N}[y(k) - y_n(k)]^2}{N}}, \quad (11)$$

where n is the length, $y_n(k)$ is the local trend and $y(k)$ is define as

$$y(k) = \sum_{i=1}^{k}[x(i) - \overline{x}], \quad (12)$$

with $x(i)$ as the inter-beat interval and $\overline{x}$ as its average [34].

2.5. Data compressor

Taking into consideration the 10 extracted features of all segments as time series distributions per different windowing analyses, 6 distinct statistical functions acted like data compressors over time. This process reduces the dimensionality problem and ensures that the number of metrics per subject is equal. The six statistical functions were:

- **Average** (Avg) is calculated by the sum of values in the vector and then divided by the length of the vector [35]. The equation is defined as

$$Avg = \frac{\sum_{n=1}^{N} x(n)}{N}, \quad (13)$$

where N is the length of the vector, and x is the vector.

- **Standard Deviation** (Std) is a measure of how dispersed the data is relative to the mean [35] and can be defined as

$$Std = \sqrt{\frac{\sum_{n=1}^{N} \mid x(n) - Avg \mid^2}{N-1}} \quad (14)$$

- **95th Percentile** ($P95$) is the number that is equal or grater than 95% of the values present in a vector [36]. $P95$ is obtained by ordering the values in increasing order and using the formula

$$n_{P95} = 0.95 \times N, \quad (15)$$

where n_{P95} is the $P95$ position in a vector and N is the number of samples.

- **Variance** (Var) is the spread of numbers in a vector [36] and is defined as

$$Var = \frac{\sum_{n=1}^{N} \mid x(n) - Avg \mid^2}{N-1} \quad (16)$$

- **Median** (Med) is the value separating the highest values in half from the lowest values. Med is obtained by arranging the values in ascending order and using the equation

$$n_{Med} = 0.5 \times N, \quad (17)$$

where n_{Med} is the Med position in a array and N is the number of samples [36].

- **Kurtosis** (Kur) measures how a distribution differs from a normal distribution. The equation is defined as

$$Kur = \frac{\sum_{n=1}^{N} \mid x(n) - Avg \mid^4}{Std^4} \quad (18)$$

2.6. Data driven analysis by classifiers

2.6.1. Feature individual power analysis per binary groups

The assessment of each feature distribution individual power for discriminate between each pair of study groups (*Low vs. Moderate*, *Low vs. Severe*, *Moderate vs. Severe*) was performed by X-ROC classifier [37], a binary classifier. A total of 720 features (time series of 10 non-linear features compressed by (×) 6 statistics × 6 types of windowing analysis × 2 modalities per participant) have been individually assessed for each pair of study groups to quantify their ability to discriminate between them. It should be noted that each feature vector per study group suffered a normalization by $z - score$ [35] before classifying. Finally, X-ROC classifier leverages the squared distance between distribution means to identify an optimal threshold for classification. Moreover, the classification results were evaluated by using the receiver operating characteristic (ROC) curve.

2.6.2. Features combined power assessment for groups discrimination using machine learning tools

In this task, we analyzed the combined power of features to discriminate between study groups (*Low vs. Moderate*, *Low vs. Severe*, *Moderate vs. Severe*, *All vs. All*) with the helplessness of 19 Sci-learn classifiers. The next list shows individually each used classifier plus its hyper-parameters:

- AdaBoostClassifier (AdaBoost) - Default parameters;
- BaggingClassifier (BaggC) - Default parameters;
- DecisionTreeClassifier (DeTreeC) - max_depth was set to 5;
- ExtraTreesClassifier (ExTreeC) - n_estimators was set to 300;
- GaussianNB (GauNB) - Default parameters;
- GaussianProcessClassifier (GauPro) - 1.0*RBF(1.0);
- GradientBoostingClassifier (GradBoost) - Default parameters;
- KNearestNeighborsClassifier (KNN) - default parameters;
- LinearDiscriminantAnalysis (LinDis) - Default parameters;
- LinearSVC (LinSVC) - Default parameters;
- LogisticRegression (LogReg) - solver was set to “lbfgs”;
- LogisticRegressionCV (LogRegCV) - cv was set to 3;
- MLPClassifier (MLP) - alpha was set to 1 and max_iter was equal to 1000;
- OneVsRestClassifier (OvsR) - random_state was equal to 0;
- QuadraticDiscriminantAnalysis (QuadDis) - Default parameters;
- RandomForestClassifier (RF) - the max_depth, n_estimators and max_features were set to 5, 300 and 1, respectively;
- SGDClassifier (SGD) - max_iter was 100 and tol was 0.001;
- SGDClassifier (SGDC) - Default parameters;
- Support-vector Machines (SVC) - Gamma was set to “auto”;

Each time-series feature vector per study group comparative analysis suffered a normalization by $z-score$. Then, the model’s performance has been evaluated by feeding the previously 19 designed and selected machine learning models with different combinations of features, from 1 until the maximum of 120 features (meaning no feature selection — time series of 10 non-linear features compressed by (×) 6 statistics × 2 modalities per participant) selected by f-score [38], per comparative

analysis, within a leave-one-out-cross-validation procedure. This procedure is well-known for allowing the use of the whole dataset for testing, without leakage between train and test sets.

2.6.3. Classification metrics

To evaluate the models' performance, 4 metrics have been used: $Accuracy$, $F1-Score$, $Precision$, and $Recall$.

$Accuracy$ means how many cases did we correctly label out of all the cases, and it is defined as,

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (19)$$

where, a TP (True Positives) is an outcome where the MP model correctly predicts a positive class, a TN (True Negatives) is an outcome where the model correctly predicts the negative class, a FP (False Positives) is an outcome where the model incorrectly predicts the positive class, and, finally, FN (False Negatives) is an outcome where the model incorrectly predicts the negative class [39].

$Precision$ and $Recall$ are performance metrics that are measured with TP under the spotlight. A metric such as the $Precision$ brings to light what was the proportion of positive values that were correctly identified, with the $Recall$ we visualize the proportion of TP that were correctly classified [39]. Both equations are defined as:

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (20)$$

and

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (21)$$

The $F1-Score$ is the harmonic mean of $Precision$ and $Recall$ and it can be defined as,

$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \times 100\% \quad (22)$$

The Area Under the Curve (AUC) of the Receiver Operating Characteristic Curve (ROC) is a metric that evaluates how well a model can distinguish between positive and negative classes. It does this by comparing the TP rate against the FP rate at different classification thresholds. The value of AUC ranges between 0 and 1, with higher values indicating better performance. A perfect classifier has an AUC of 1, while a random classifier has an AUC of 0.5. Using AUC allows for a single-value measure of the model's performance. This is especially useful for comparing models and assessing performance in scenarios where there is an imbalance between classes [40].

3. Results and discussion

The X-ROC classifier has been used to analyze the individual power of each extracted feature compressed by different statistical functions and extracted within different windowing analyses (72 features have been generated for classification from each one of the 10 originally defined features, as each originally defined feature has been extracted within 6 types of windowing analysis for the 2 modalities per participant and time-serially compressed by 6 statistics, respectively). Fig. 4 displays the results of this analysis. The discriminating accuracy of each feature was further examined in three binary comparisons: $Low\ vs\ Moderate$, $Low\ vs\ Severe$, and $Moderate\ vs\ Severe$, as shown in Figs. 4(a), 4(b), and 4(c), respectively. These bar graphs indicate how many times each feature achieved a discriminating accuracy above 70%, 80%, and 90%. Table 3 lists the feature that provides the best individual discriminant power for each binary comparison group. In the $Low\ vs\ Moderate$ comparison group (Fig. 4(a)), features En, $ESha$, $ELya$, and K displayed an accuracy value of 80% or lower. The feature EH achieved an accuracy discriminating value above 70% for nearly all of the features generated by this particular kind of feature (60 in a total of 72 features). The feature EA achieved the highest accuracy value of 92.86%, as demonstrated in Table 3. In the $Low\ vs\ Severe$ comparison group (Fig. 4(b)), all features demonstrated a discrimination accuracy value between 70% and 90%. According to Table 3, features $ELog$, $Elya$, CD, and $ESha$ provided the highest accuracy value of 86.11%. Finally, in the $Moderate\ vs\ Severe$ comparison group (Fig. 4(c)), all features displayed an accuracy value above 80% and below 90%, except the feature EA that over-passed the 90% with a discrimination accuracy of 91.18%, as indicated in Table 3. When examining all three graphs simultaneously, it becomes clear that the feature originally generated through EA (both differing only on the window type used for analysis) from the Up body position modality provided the highest $Accuracy$ values when the $Moderate$ class is present in the classification pair.

Figs. 5 and 6 present the Heat-maps with classification metrics report for the comparison groups, with and without the feature selection step. Each heat-map uses a colorbar of greens to clearly highlight the Accuracy, Recall, Precision, and F1-score discrimination power of the method for each comparison analysis. Lighter greens indicate lower power, while darker greens indicate higher power. By comparing the results with and without feature selection, it can be seen that a feature selection process can boost the models' performance for all comparison groups, as expected, except just in the case of the $Recall$ metric with slightly lower values for $All\ vs\ All$.

Concerning Fig. 5, where we avoid the feature selection step, we can see that the windowing process with Hm and Hn, with Overlap, allows for finding the best results. The comparison group $Moderate\ vs\ Severe$ provided the best table results with an $Accuracy$ of 85.29% and a $F1-Score$ of 91.23%. The $All\ vs\ All$ group even though it reached a 100% $Recall$ value, the model presented some difficulty in correctly classifying the FP, with the $Precision$ rounding 68.29%.

Regarding the analysis of Fig. 6, with feature selection, the Rec and the HmO windows provided the best results. The $Low\ vs\ Moderate$ group reached the best performance among the others, with an $Accuracy$ and a $F1-Score$ reaching 100%. The $All\ vs\ All$ classification results improved, compared with the ones achieved without feature selection, with an $Accuracy$ of 78.57% and a $F1-Score$ of 84.75%.

Fig. 7 shows the best performance results for each classifier without taking into account if the feature selection stage has been made or not. The ROC Curves and the Confusion Matrices of the best results presented in Fig. 7 are shown in Figs. 8 and 9, respectively.

For the binary comparison $Low\ vs\ Moderate$ it can be seen in Fig. 7 that the best classifiers were $QuadDis$ and $OvsR$ with 100% for the $Accuracy$, $Recall$, $Precision$, and $F1-Score$ and with AUC of 1.0. Among the 6 types of windowing analysis, the Hm window achieved the highest quantity of good results for each classifier classification report. The Fig. 8(a) displays the ROC of the top classification results. The confusion matrix (Fig. 9(a) - Binary group $Low\ vs\ Moderate$) indicates 100% of discrimination accuracy for both Low class (8/8) and $Moderate$ class (6/6).

On Fig. 7, for the binary comparison $Low\ vs\ Severe$, the classifier that provided the best results was MLP with 91.67% of $Accuracy$, 100% for the $Recall$, 90.32% of $Precision$, an $F1-Score$ of 94.92% and a AUC of 0.81. The windowing analyses with the best results have been done with the window HnO. The ROC graph in Fig. 8(b) highlights difficulty in distinguishing one of the classes. Fig. 9(b) displays a Confusion Matrix for the binary group ($Low\ vs\ Severe$). Results show 100% of discrimination accuracy for the $Severe$ class (28/28), but only 62.5% for the Low class samples (5/8).

For the comparison $Moderate\ vs\ Severe$ (Fig. 7), we can see the classifier was $LinDis$ with the best performance, achieving an $Accuracy$ of 94.12%, a $Recall$ of 96.43%, a $Precision$ of 96.43% and an $F1-Score$ of 96.43%. The most represented windowing analysis was associated with the Rec window. The ROC in Fig. 8(c) has an associated AUC result of 0.90, indicating a great capacity to discriminate between the two classes. Additionally, the Confusion Matrix displayed in Fig. 9(c) shows a 96.43% correct prediction rate for the $Severe$ class (27/28) and an 83.33% (5/6) correct prediction rate for the $Moderate$ class.

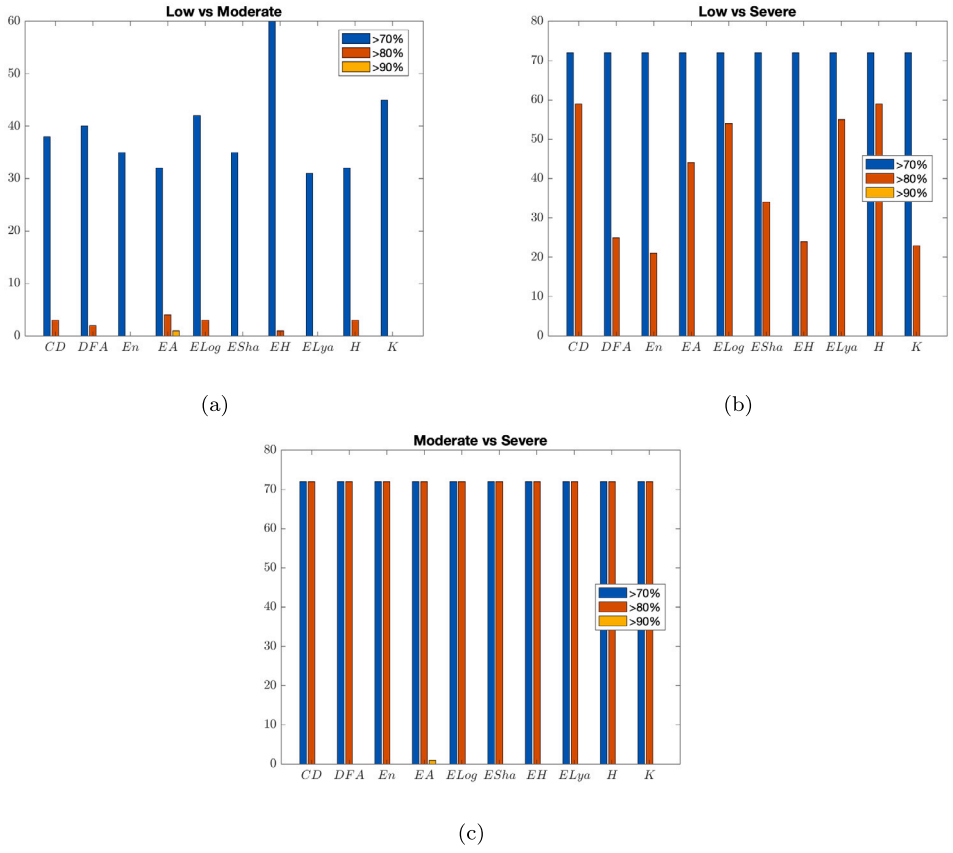

Fig. 4. X-ROC Accuracy per individual feature: (a) Low vs Moderate, (b) Low vs Severe and (c) Moderate vs Severe.

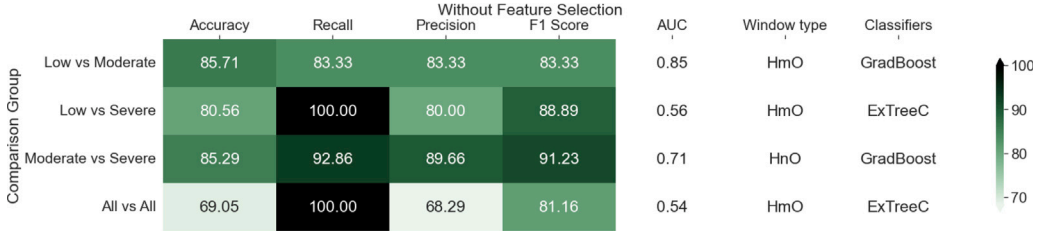

Fig. 5. Heat-map classification report – classification without Feature Selection step – best Accuracy, Recall, Precision, $F1 - Score$, and AUC results for each comparison group plus the information of classifier and the window type applied for signal analysis.

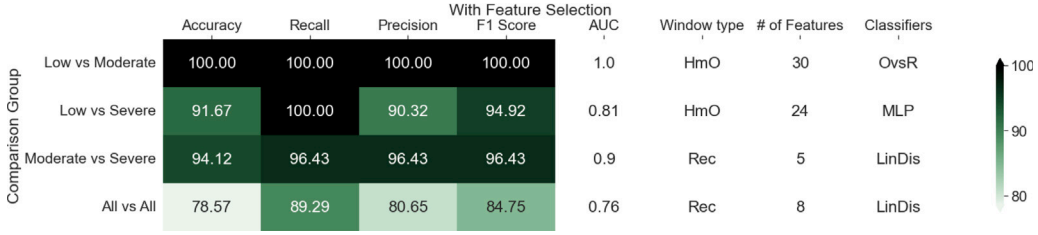

Fig. 6. Heat-map classification report – classification with Feature Selection step – best Accuracy, Recall, Precision, $F1 - Score$ and AUC results for each comparison group plus the information of classifier, the window type applied and the number of selected features for classification.

Table 3
X-ROC individual feature highest accuracy per binary group.

Binary group	Feature	Window type	Body position	Data compressor	Accuracy	AUC
Low vs Moderate	*EA*	*HmO*	*Up*	*Kur*	92.86%	0.94
Low vs Severe	*ELog*	*Hm, Hn* and *Rec*	*Up*	*Kur*	86.11%	0.81
	ELya	*HmO* and *RecO*	*Up* and *Down*	*Kur, Std* and *Var*		
	CD	*HnO*	*Down*	*Std and Var*		
	ESha	*RecO*	*Up* and *Down*	*P95*		
Moderate vs Severe	*EA*	*HnO*	*Up*	*Kur*	91.18%	0.79

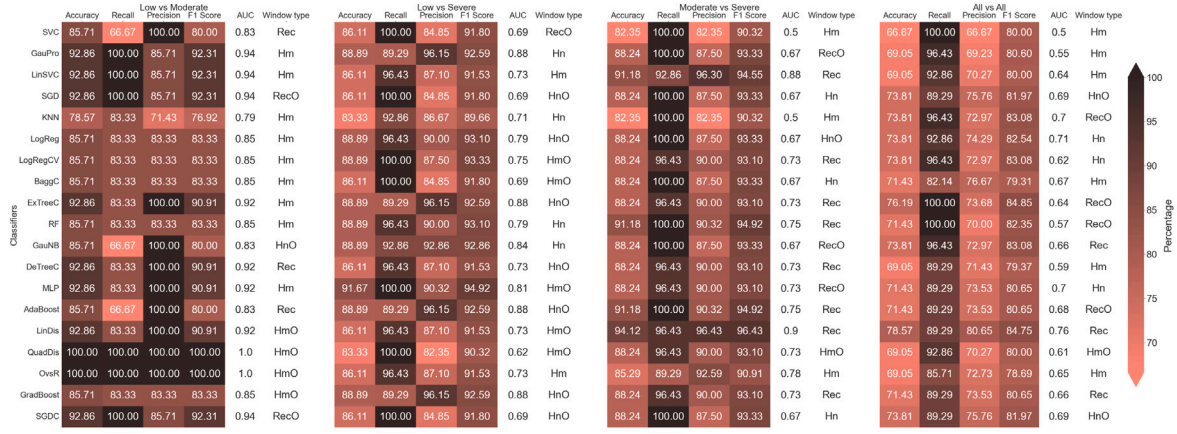

Fig. 7. Heat-map classification report for every comparison group with the best *Accuracy*, *Recall*, *Precision*, *F1 – Score* and AUC results for each one of the used classifiers plus the information of the window type applied on signal analysis.

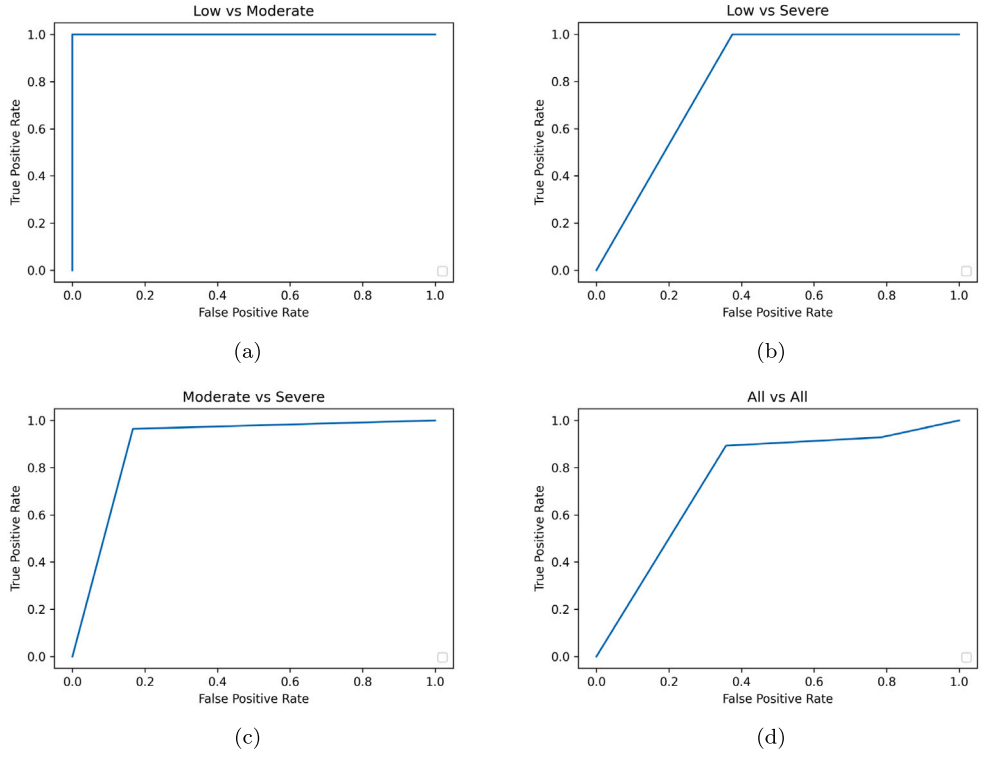

Fig. 8. ROC Curves for best classification results per study group: (a) Low vs Moderate, (b) Low vs Severe, (c) Moderate vs Severe and (d) All vs All.

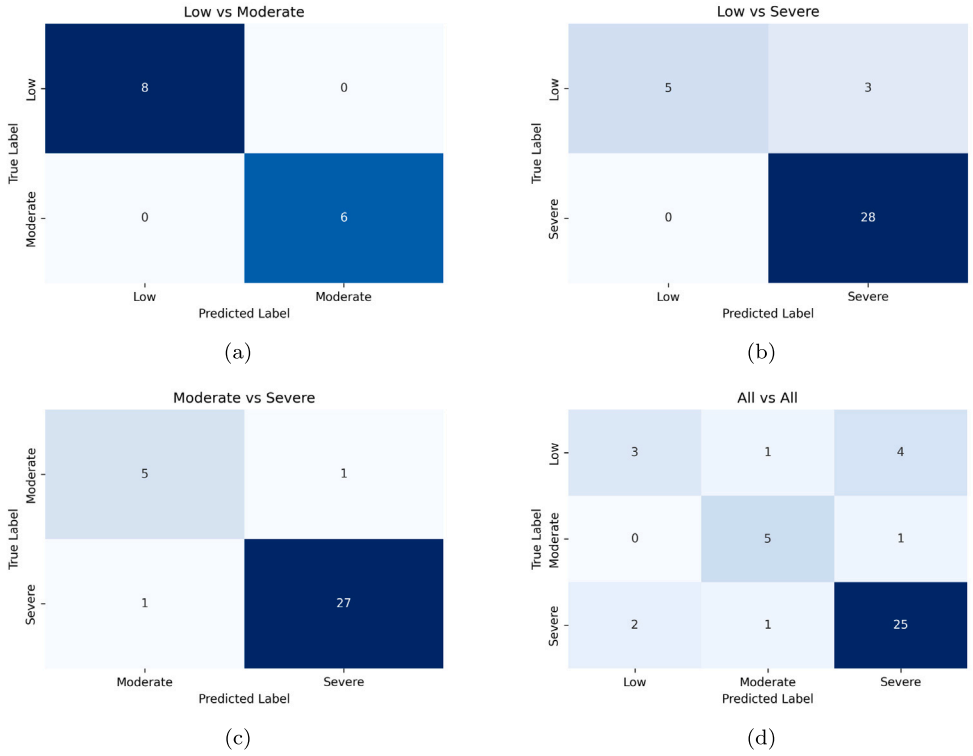

Fig. 9. Confusion Matrix for each comparison group: (a) Low vs Moderate, (b) Low vs Severe, (c) Moderate vs Severe and (d) All vs All.

At last, the multi-class comparison *All vs All* showed on Fig. 7 that the classifier that obtained the best performance report was *LinDis* with 78.57% of *Accuracy*, 89.29% for *Recall*, a *Precision* of 80.65%, and an $F1 - Score$ of 84.75%. The windowing process with the highest amount of excellent results was related to the window *Hm*. The Fig. 8(d) shows a huge difficulty to distinguish the Multiclass group with a corresponding *AUC* value of 0.76. The Fig. 9(d), illustrates the Confusion Matrix for this comparison group, where we can see a 37.50% correct prediction for the *Low* class (3/8), an 83.33% correct prediction for the *Moderate* class (5/6) and a correct prediction of 89.29% for the *Severe* class (25/28).

Making a comparison between the discriminatory power of individual features (Table 3) and that of combined features (Fig. 7), we can observe an improvement in discrimination accuracy when we are using features combination for discrimination. For Low vs Moderate, the accuracy improved by 7.14%. For Low vs Severe, it improved by 5.56%. And for Moderate vs Severe, it improved by 3.39%. We also observed improved AUC values, except for the Low vs Severe comparison where the AUC values remained unchanged between both analyses.

Globally, the comparison *Low vs Moderate* had the highest overall results, the *LinDis*, *MLP* and *OvsR* are strictly connected with the best-achieved overall results and the *HmO* windowing process appeared the most as the selected one for signal analysis.

Finally, a comparison between the present work and the previous one with the same ECG database is presented in Table 4. As can be seen, in the presented work, we use a different approach based on a non-linear analysis over heart rate variability (HRV) and QRS complex analyses. This new approach allows us to outperform the previous work All vs All discrimination results in 7.57%. Furthermore, we cannot compare the other achieved binary comparison results with the previous work as we did not perform binary classifications between study pairs in that work [23]. Due to the high discriminant capacity of the utilized non-linear features, the present work represents an improvement in the classification results.

Lastly, looking at Table 1, we can see that our results are in line with those found in the state-of-the-art, giving us a good perspective of improvising the robustness shortly. In addition, ECG presents some key advantages when compared to other sources usually used on the state-of-the-art for COVID-19 discrimination propose, as ECG is a low-cost, non-invasive and easy-to-operate exam, with the possibility to perform the test on patients that are incapable of being outside of the hospital's bed.

4. Conclusion

For this research, 10 non-linear features (*Energy*, *Approximate Entropy*, *Logarithmic Entropy*, *Shannon Entropy*, *Hurst Exponent*, *Lyapunov Exponent*, *Higuchi Fractal Dimension*, *Katz Fractal Dimension*, *Correlation Dimension* and *Detrended Fluctuation Analysis*) were extracted from ECG signals, collected from 2 different patient's body positions. Each signal went under a windowing process, with 1 s of length windows, the process was applied in 6 ways of windowing signal analysis crops and Machine Learning classifiers were employed to predict the different evolution stages in COVID-19 (*Low vs. Moderate*, *Low vs. Severe*, *Moderate vs. Severe*) and 1 Multi-class comparison (*All vs. All*).

The best *Accuracy* levels ranged between 78.57% and 100%. The *OvsR* with *HmO*, the *MLP* with *HmO* or the *LinDis* with *Rec* were the classifiers that obtained the best results with their corresponding window types. These results showed the applied methodology could be considered a good tool for distinguishing the different stages of COVID-19 through the use of ECG signals. Additionally, compared to prior research utilizing the same database, this study demonstrates a 7.57% increase in diagnostic precision for the All vs All comparison.

For future research, extracting more features to refine and improve the system will be important. One possible way of accomplishing that would be to analyze the ECG signals by time–frequency domain transforms. Additionally, it would be interesting to analyze also ECG data collected from subjects with the same severity of COVID-19, with guaranteed ventilatory control (i.e., a standard respiratory rate was maintained during collection). The results should also be updated with a larger and more balanced population to ensure a more consistent generalization.

Table 4
Comparison between present study and a previous work with the same ECG database.

Ref.	Summary	Extracted features	Classification accuracy			
			Low vs Moderate	Low vs Severe	Moderate vs Severe	All vs All
[23]	An automated machine learning system was employed to classify COVID-19 severity. ECG signals were processed to obtain QRS complexes and HRV metrics combined to serve as entries for several machine-learning algorithms for All vs All classification, with LinearSVC performing the best.	ApEn MSE 1, ApEn MSE 2, ApEn MSE 3, ApEn MSE 4, ApEn MSE 5, Approximate Entropy, DET, ENT, Heartbeats, HRV index, LF/HF, Lmax, L and Vmax, NN50, pNN50, Peak HF, Peak LF, Peak VLF, Power HF, Power LF, Power NU/HF, Power NU/LF, Power VLF, PRSA AC, PRSA DC, RMSSD, RR, SampEn MSE 1, SampEn MSE 2, Sample Entropy, SD1 Poincare, SD2 Poincare, SDANN, SDNN index, Stdev HR, Stdev NN, TINN, Average HR, Average NN.	–	–	–	71.00% (LinearSVC)
Present work	An automated machine learning system was developed to classify the severity of COVID-19 based on 10 non-linear features extracted from ECG signals. The features were collected from 2 different patient positions and analyzed using 6 different windowing signal analyses. The features' individual and combined power for discriminating between different degrees of COVID-19 severity were evaluated by several machine learning approaches, with LinDis, MLP, and QuadDis producing the best evaluation metrics (with combined features on their entries) for Low vs Moderate, Low vs Severe, Moderate vs Severe and All vs All analyses.	En, EA, $ELog$, $ESha$, EH, $ELya$, H, K, CD, and DFA	100% (QuadDis)	91.67% (MLP)	94.57% (LinDis)	78.57% (LinDis)

CRediT authorship contribution statement

Pedro Ribeiro: Conceptualization, Investigation, Writing – original draft, Writing – review & editing. **João Alexandre Lobo Marques:** Conceptualization, Validation, Investigation, Writing – original draft, Writing – review & editing, Supervision, Funding acquisition. **Daniel Pordeus:** Validation, Writing – review. **Laíla Zacarias:** Validation, Writing – review. **Camila Ferreira Leite:** Validation, Writing – review. **Manoel Alves Sobreira-Neto:** Validation, Writing – review. **Arnaldo Aires Peixoto Jr:** Validation, Writing – review. **Adriel de Oliveira:** Validation, Writing – review. **João Paulo do Vale Madeiro:** Validation, Writing – review. **Pedro Miguel Rodrigues:** Conceptualization, Validation, Investigation, Writing – original draft, Writing – review & editing, Supervision, Funding acquisition.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgments

This work was supported by National Funds from FCT - Fundação para a Ciência e a Tecnologia, Portugal through project UIDB/50016/2020.

João Paulo Madeiro thanks for the financial support of the Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico - Funcap, Brazil, Process PS1-0186-00439.01.00/21.

References

[1] D. Yesudhas, A. Srivastava, M.M. Gromiha, COVID-19 outbreak: History, mechanism, transmission, structural studies and therapeutics, Infection 49 (2) (2020) 199–213, http://dx.doi.org/10.1007/s15010-020-01516-2.

[2] E.S. Hosseini, N.R. Kashani, H. Nikzad, J. Azadbakht, H.H. Bafrani, H.H. Kashani, The novel Coronavirus disease-2019 (COVID-19): Mechanism of action, detection and recent therapeutic strategies, Virology 551 (2020) 1–9, http://dx.doi.org/10.1016/j.virol.2020.08.011.

[3] M. Ciotti, M. Ciccozzi, A. Terrinoni, W.-C. Jiang, C.-B. Wang, S. Bernardini, The COVID-19 pandemic, Crit. Rev. Clin. Lab. Sci. 57 (6) (2020) 365–388, http://dx.doi.org/10.1080/10408363.2020.1783198.

[4] P. Ribeiro, J.A.L. Marques, P.M. Rodrigues, COVID-19 detection by means of ECG, voice, and X-ray computerized systems: A review, Bioengineering 10 (2) (2023) 198, http://dx.doi.org/10.3390/bioengineering10020198, URL http://dx.doi.org/10.3390/bioengineering10020198.

[5] C.K.C. Lai, W. Lam, Laboratory testing for the diagnosis of COVID-19, Biochem. Biophys. Res. Commun. 538 (2021) 226–230, http://dx.doi.org/10.1016/j.bbrc.2020.10.069.

[6] Heart rate variability. Standards of measurement, physiological interpretation, and clinical use. Task force of the European society of cardiology and the North American society of pacing and electrophysiology, Circulation 93 (5) (1996) 1043–1065.

[7] L.C.M. Vanderlei, C.M. Pastre, R.A. Hoshi, T.D. de Carvalho, M.F. de Godoy, Noções básicas de variabilidade da frequência cardíaca e sua aplicabilidade clínica, Revista Brasileira de Cirurgia Cardiovasc. 24 (2) (2009) 205–217, http://dx.doi.org/10.1590/s0102-76382009000200018.

[8] B. Long, W.J. Brady, R.E. Bridwell, M. Ramzy, T. Montrief, M. Singh, M. Gottlieb, Electrocardiographic manifestations of COVID-19, Am. J. Emerg. Med. 41 (2021) 96–103, http://dx.doi.org/10.1016/j.ajem.2020.12.060.

[9] F. Farshidfar, N. Koleini, H. Ardehali, Cardiovascular complications of COVID-19, JCI Insight 6 (13) (2021) http://dx.doi.org/10.1172/jci.insight.148980.

[10] Ştefania Teodora Duca, A. Chetran, R. Ştefan Miftode, O. Mitu, A.D. Costache, A. Nicolae, D. Iliescu-Halițchi, C.-O. Halițchi-Iliescu, F. Mitu, I.I. Costache, Myocardial ischemia in patients with COVID-19 infection: Between pathophysiological mechanisms and electrocardiographic findings, Life 12 (7) (2022) 1015, http://dx.doi.org/10.3390/life12071015.

[11] M.M. Bassiouni, I. Hegazy, N. Rizk, E.-S.A. El-Dahshan, A.M. Salem, Automated detection of COVID-19 using deep learning approaches with paper-based ECG reports, Circuits Systems Signal Process. 41 (10) (2022) 5535–5577, http://dx.doi.org/10.1007/s00034-022-02035-1.

[12] T. Nguyen, H.H. Pham, H.K. Le, A.T. Nguyen, N.T. Thanh, C. Do, Detecting COVID-19 from digitized ECG printouts using 1D convolutional neural networks, 2022, http://dx.doi.org/10.48550/ARXIV.2208.05433, arXiv.

[13] O. Attallah, ECG-BiCoNet: An ECG-based pipeline for COVID-19 diagnosis using bi-layers of deep features integration, Comput. Biol. Med. 142 (2022) 105210, http://dx.doi.org/10.1016/j.compbiomed.2022.105210.

[14] A.L. Booth, E. Abels, P. McCaffrey, Development of a prognostic model for mortality in COVID-19 infection using machine learning, Mod. Pathol. 34 (3) (2020) 522–531, http://dx.doi.org/10.1038/s41379-020-00700-x.

[15] S. Subudhi, A. Verma, A.B. Patel, Prognostic machine learning models for COVID-19 to facilitate decision making, Int. J. Clin. Prac. 74 (12) (2020) http://dx.doi.org/10.1111/ijcp.13685.

[16] M.H. Aghajani, A. Toloui, M. Aghamohammadi, A. Pourhoseingholi, N. Taherpour, M. Sistanizad, A.M. Neishaboori, Z. Asadpoordezaki, R. Miri, Electrocardiographic findings and in-hospital mortality of COVID-19 patients; A retrospective cohort study, Arch. Acad. Emerg. Med. 9 (2021) e45, http://dx.doi.org/10.22037/aaem.v9i1.1250.

[17] D. Kaliyaperumal, K. Bhargavi, K. Ramaraju, K.S. Nair, S. Ramalingam, M. Alagesan, Electrocardiographic changes in COVID-19 patients: A hospital-based descriptive study, Indian J. Crit. Care Med. 26 (1) (2022) 43–48, http://dx.doi.org/10.5005/jp-journals-10071-24045.

[18] S. Wang, Y. Zha, W. Li, Q. Wu, X. Li, M. Niu, M. Wang, X. Qiu, H. Li, H. Yu, W. Gong, Y. Bai, L. Li, Y. Zhu, L. Wang, J. Tian, A fully automatic deep learning system for COVID-19 diagnostic and prognostic analysis, Eur. Respir. J. 56 (2) (2020) 2000775, http://dx.doi.org/10.1183/13993003.00775-2020.

[19] S.A.-F. Sayed, A.M. Elkorany, S.S. Mohammad, Applying different machine learning techniques for prediction of COVID-19 severity, Ieee Access 9 (2021) 135697–135707, http://dx.doi.org/10.1109/access.2021.3116067.

[20] Z. Yu, X. Li, H. Sun, J. Wang, T. Zhao, H. Chen, Y. Ma, S. Zhu, Z. Xie, Rapid identification of COVID-19 severity in CT scans through classification of deep features, BioMed. Eng. OnLine 19 (1) (2020) http://dx.doi.org/10.1186/s12938-020-00807-x.

[21] E. Irmak, COVID-19 disease severity assessment using CNN model, IET Image Process. 15 (8) (2021) 1814–1824, http://dx.doi.org/10.1049/ipr2.12153.

[22] H. Yao, N. Zhang, R. Zhang, M. Duan, T. Xie, J. Pan, E. Peng, J. Huang, Y. Zhang, X. Xu, H. Xu, F. Zhou, G. Wang, Severity detection for the Coronavirus disease 2019 (COVID-19) patients using a machine learning model based on the blood and urine tests, Front. Cell Dev. Biol. 8 (2020) http://dx.doi.org/10.3389/fcell.2020.00683.

[23] D. Pordeus, P. Ribeiro, L. Zacarias, J.P. Madeiro, J.A.L. Marques, P.M. Rodrigues, C. Leite, M.A. Neto, A.A. Peixoto Jr., A. de Oliveira, Classification of severity of COVID-19 patients based on the heart rate variability, in: Computerized Systems for Diagnosis and Treatment of COVID-19, Springer International Publishing, 2023, pp. 155–177, http://dx.doi.org/10.1007/978-3-031-30788-1_10.

[24] X. Yan, X. Han, D. Peng, Y. Fan, Z. Fang, D. Long, Y. Xie, S. Zhu, F. Chen, W. Lin, Y. Zhu, Clinical characteristics and prognosis of 218 patients with COVID-19: A retrospective study based on clinical classification, Front. Med. 7 (2020) http://dx.doi.org/10.3389/fmed.2020.00485.

[25] S. Brown, J. Brown, Resting and postexercise cardiac autonomic control in trained masters athletes, J. Physiol. Sci. 57 (1) (2007) 23–29, http://dx.doi.org/10.2170/physiolsci.rp012306.

[26] B. Shah, S. Kunal, A. Bansal, J. Jain, S. Poundrik, M.K. Shetty, V. Batra, V. Chaturvedi, J. Yusuf, S. Mukhopadhyay, S. Tyagi, G.M. Palleda, A. Gupta, M.D. Gupta, Heart rate variability as a marker of cardiovascular dysautonomia in post-COVID-19 syndrome using artificial intelligence, Indian Pacing Electrophysiol. J. 22 (2) (2022) 70–76, http://dx.doi.org/10.1016/j.ipej.2022.01.004.

[27] Z. Qu, G. Hu, A. Garfinkel, J.N. Weiss, Nonlinear and stochastic dynamics in the heart, Phys. Rep. 543 (2) (2014) 61–162, http://dx.doi.org/10.1016/j.physrep.2014.05.002.

[28] D. Sundararajan, Discrete Wavelet Transform a Signal Processing Approach, first ed., John Wiley & Sons, New Jersey, 2015.

[29] W. Caesarendra, B. Kosasih, K. Tieu, C.A.S. Moodie, An application of nonlinear feature extraction - a case study for low speed slewing bearing condition monitoring and prognosis, in: 2013 IEEE/ASME International Conference on Advanced Intelligent Mechatronics, IEEE, Wollongong, 2013, pp. 1713–1718, http://dx.doi.org/10.1109/aim.2013.6584344.

[30] A.A.T. Garcia, C.A.R. Garcia, L. Villasenor-Pineda, O.M. Montoya, Biosignal Processing and Classification Using Computational Learning and Intelligence Principles, Algorithms, and Applications, Academic Press, London, 2022, pp. 59–91.

[31] M. Rüdisüli, T. Schildhauer, S. Biollaz, J.V. Ommen, Measurement, monitoring and control of fluidized bed combustion and gasification, in: Fluidized Bed Technologies for Near-Zero Emission Combustion and Gasification, Elsevier, Cambridge, 2013, pp. 813–864, http://dx.doi.org/10.1533/9780857098801.3.813.

[32] G. Silva, P. Batista, P.M. Rodrigues, COVID-19 activity screening by a smart-data-driven multi-band voice analysis, J. Voice (2022) http://dx.doi.org/10.1016/j.jvoice.2022.11.008.

[33] R.M. Bryce, K.B. Sprague, Revisiting detrended fluctuation analysis, Sci. Rep. 2 (1) (2012) http://dx.doi.org/10.1038/srep00315.

[34] R. Hardstone, S.-S. Poil, G. Schiavone, R. Jansen, V.V. Nikulin, H.D. Mansvelder, K. Linkenkaer-Hansen, Detrended fluctuation analysis: A scale-free view on neuronal oscillations, Front. Physiol. 3 (2012) http://dx.doi.org/10.3389/fphys.2012.00450.

[35] R. Peck, C. Olsen, J.L. Devore, Introduction to Statistics and Data Analysis, Cengage Learning, Belmont, 2008, p. 880.

[36] B. Rosner, Fundamentals of Biostatistics, Duxbury Press, Boston, 2010.

[37] C.T. Nakas, C.T. Yiannoutsos, Ordered multiple-class ROC analysis with continuous measurements, Stat. Med. 23 (22) (2004) 3437–3449, http://dx.doi.org/10.1002/sim.1917, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/sim.1917.

[38] N. Sevani, I. Hermawan, W. Jatmiko, Feature selection based on F-score for enhancing CTG data classification, in: 2019 IEEE International Conference on Cybernetics and Computational Intelligence, CyberneticsCom, IEEE, Indonesia, 2019, pp. 18–22, http://dx.doi.org/10.1109/cyberneticscom.2019.8875656.

[39] O. Doğan, Data linkage methods for big data management in industry 4.0, in: Optimizing Big Data Management and Industrial Systems with Intelligent Techniques, IGI Global, Hershey, 2019, pp. 108–127, http://dx.doi.org/10.4018/978-1-5225-5137-9.ch005.

[40] F.S. Nahm, Receiver operating characteristic curve: Overview and practical use for clinicians, Korean J. Anesthesiol. 75 (1) (2022) 25–36, http://dx.doi.org/10.4097/kja.21209.