Qeios

Assessment of COVID-19 from Features Extraction of Exhaled Breath Using Signal Processing Methods

Exhaled breath waveforms, utilized in ventilation monitoring, are aimed at upgrading into COVID-19 disease screening. An algorithm for valid exhaled breath waveform segmentation and feature computation is developed to identify COVID-19 infection using exhaled breath patterns for distinguishing a COVID and non-COVID condition. Two minutes of exhaled breath patterns were recorded using a device and a nasal cannula sampling tube, resulting in the collection of exhaled breath waveforms from each subject. The developed algorithm is utilized to evaluate the valid exhaled breath waveforms and compute the features classified to distinguish COVID and non-COVID conditions. Slope e_2, activity e_2, and intersection angle of expiration and inspiration phases showed *p*-values of 0.000, denoting the strong significant difference between COVID and non-COVID conditions. The statistical analyses revealed *p*-values of 0.039, 0.008, and 0.024 for area e_2, mobility of e_2, and complexity e_3, indicating their significance in differentiating the COVID-19 condition from the non-COVID condition. The slope, area, and intersection angle, as significant features, showed good predictive power for compliance with *p*-value analysis, with area under the receiver operating characteristic curves of 0.667, 0.693, and 0.775. The slope of e_2, the area of e_2, and the intersection angle of expiration and inspiration phases are identified as the promising features to be chosen in discriminating the COVID and non-COVID conditions.

M. B. Malarvili[1,*], **Santheraleka Ramanathan**[1], **Rishya Manikam**[2], **Ahmad Adib Ramli**[2], **Rafi' Uddin Radzi Ruslay**[2], **Anhar Kamarudin**[2], and **Aina Amalina Ahmad Tarmizi**[2]

[1] *Department of Biomedical Engineering and Health Sciences, Faculty of Electrical Engineering, Universiti Teknologi Malaysia, 81310, Skudai, Johor, Malaysia.*

[2] *University Malaya Specialist Centre, Pusat Perubatan Universiti Malaya, Lembah Pantai, 59100 Kuala Lumpur, Malaysia.*

[*]Corresponding author: malarvili@utm.my

Trial registration: Clinical trial approval by Medical Research and Ethics Committee (MREC), 53 Malaysia, NMRR-21-763-59692. Registered 9th June 2021 and valid till 8th June 2023.

Keywords: Exhaled breath; COVID-19; Algorithmic approach; Clinical decision making; SARS-CoV-2.

https://doi.org/10.32388/66LLGG

1. Introduction

COVID-19 disease is referred to as a respiratory illness that causes serious pulmonary complications and multi-organ infection. The severe effect caused by the spread of the SARS-CoV-2 virus, which resulted in an uncontrollable infectivity rate and a significant number of deaths (Seo et al., 2020; Taleghani and Taghipour, 2021). It has recapitulated the historical evidence caused by pneumonias like SARS and MERS (Ramanathan et al., 2022b; Wang et al., 2021). Medical professionals have been compelled to concentrate on the disease's virology, physiology, and immunology for clinical practices to stop the pandemic due to the widespread COVID-19 global pandemic. Both during and after their infection, the COVID-19-infected individuals reportedly displayed an acute abnormal breathing pattern (Ramanathan et al., 2022a; Zubieta-Calleja et al., 2023). Numerous studies have been done in this regard to identify COVID-19 illness using breath samples. Breath-based detection has attracted a lot of interest in clinical settings due to its significant reduction in workload and expense compared to invasive sampling methods like nasal swabs and other sophisticated sampling methods like RT-PCR, isothermal amplification techniques, and enzymatic assays (Kreyer et al., 2021; Maurya et al., 2022). Breath analysis is a quick and reliable process. It demonstrates massive potential for primary COVID-19 screening at the point of patient admission, which would aid in commencing preventive measures at clinical practices (Marshall et al., 2021; Shan et al., 2020). Over the years and to the present day, breath monitoring has become popular for the early detection of acute respiratory illnesses. Additionally, studies have demonstrated the ability to identify viral infections using breath samples. Given this, numerous studies have been conducted to determine the presence of COVID-19 transmission in breath samples (Godoy et al., 2023; Robba et al., 2020).

As critical care clinicians become more acquainted with the harmful course of COVID-19, experts in clinical practice suggest initiating invasive ventilation as soon as patients show a sudden surge in respiratory distress. One of the most important techniques for preventing fatalities is invasive ventilation, which was vital during the first COVID-19 pandemic outbreak (Carmichael et al., 2021; Mauer et al., 2021). To support patient survival as the virus's infectivity varies, most COVID-19 patients typically receive monitoring under mechanical ventilation for 14–20 days. To satisfy the demand for simultaneous life support in these circumstances, mechanical ventilators are in high demand (Stroh et al., 2023).

The current research aims at evaluating the CO_2 partial pressure variation in a COVID-19 patient, in contrast to other studies that focused on the biomolecules present in breath samples. For obtaining accurate breath cycles from patient-exhaled breath samples, a precise and targeted algorithm has been created. The core of the technique is to divide the exhaled breath waveform into five epochs, compute the features of each epoch (e), and assess the variance between the COVID and non-COVID conditions. The significant features on differentiating the COVID-19 condition based on the CO_2 gas threshold determined in the exhaled breath waveform were identified by a sequential statistical analysis. The proposed method of recognizing COVID-19 and the carefully constructed algorithmic simulation for exhaled breath cycle profiling is schematically illustrated in Fig. 1.

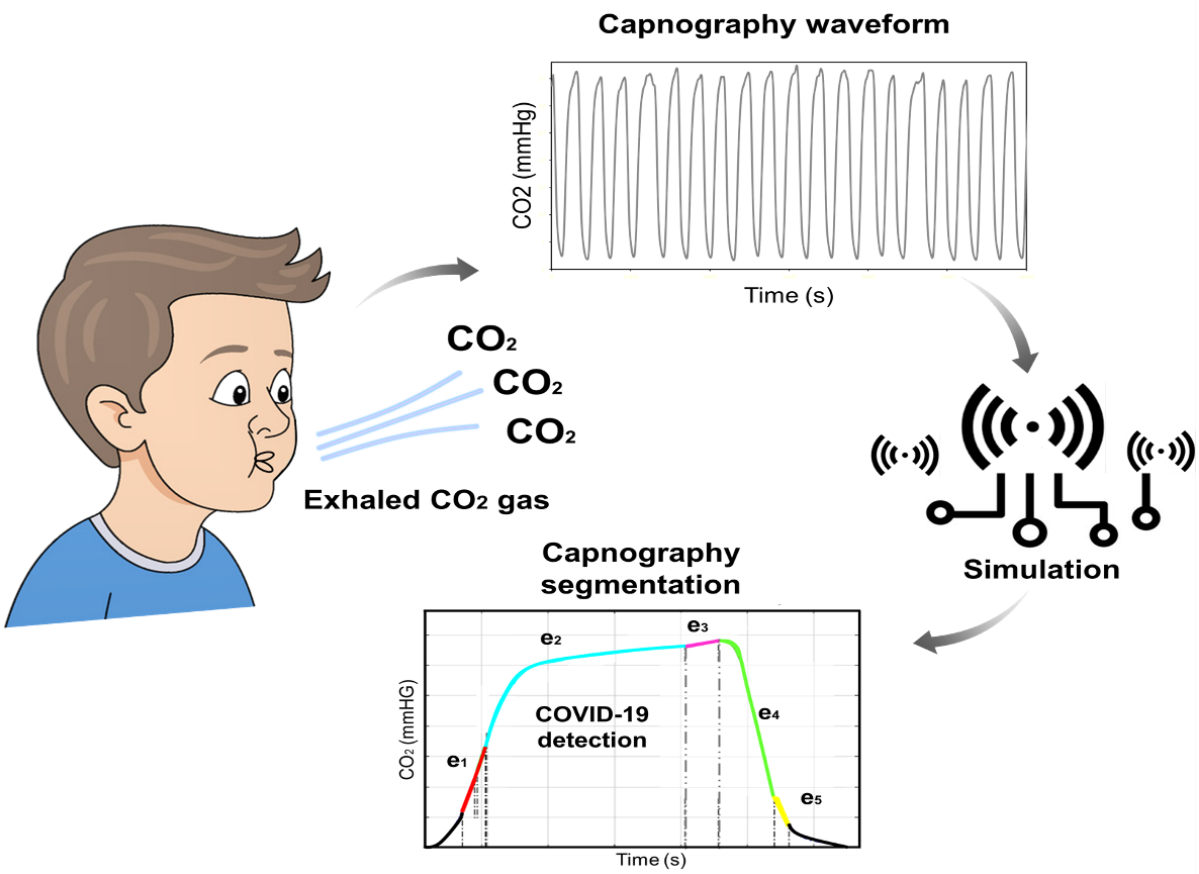

Fig. 1. Schematic illustration of screening for COVID-19 from the exhaled breath waveform.

2. Methods

2.1. Study Protocol

Clinical data collection was conducted at the emergency department of Hospital Pulau Pinang, located in the northern region of Malaysia, in June 2022. The Medical Research and Ethics Committee (MREC) approved the study. The clinical study is registered in the National Medical Research Register (NMRR) with the reference number (NMRR-21-763-59692). The clinical study was performed as it complies with the regulations and guidelines of Good Clinical Practice Malaysia 2018. The section below discusses the strategies conducted to evaluate the relationship between COVID-19 exhaled breath waveform features in developing an algorithmic-based COVID-19 diagnosis.

2.2. Target Population

The preliminary clinical research was focused on adult COVID-19 patients between the ages of 21 and 60 years old. Patients with only COVID-19 category 2 (CAT 2) infection stage were approached to participate in this study. Patients with a history of asthma, COPD, and pulmonary edema were excluded to ensure the exhaled CO_2 recording was not affected by other respiratory complications. The breath analysis was carried out with 40 participants, where the confirmed COVID-19 CAT 2 participants numbered 20. The study was evaluated with 50 non-COVID patients, who were negative COVID-19 participants.

2.3. Study Design

COVID-19 with CAT 2 infection stage was validated with an RTK test kit upon arrival at the administration of the emergency department. The patient was approached to request participation in the clinical study. As the patient agreed, the patient concern form was explained, and the patient signature was obtained prior to beginning the data acquisition. The patient was instructed to sit and lean against a chair. The patient was advised to stay relaxed and breathe calmly through the nasal cannula, which was used as the sampling tube and fixed at the patient's nose. The other end of the nasal cannula was locked at the device filter opening. The patient was instructed to breathe normally using the nasal cannula to ensure the patient was comfortable and, hence, the obtained data was reliable. During the measurement, the patient was advised to not speak or move body position as it may disrupt the flow of CO_2 gas in the sampling tube. Once a stable waveform was observed, the data was recorded for two minutes using a stopwatch. The recorded data is stored in comma-separated values (CSV) file format on an SD card, which was later extracted for further evaluation. As the measurement and recording process were completed, the nasal cannula was disconnected and removed from the patient's nose for disposal. The same study protocol was applied to negative COVID-19 patients, and the non-COVID data was acquired.

2.4. Signal Preprocessing

The filter implementation is designed using the Signal Processing Toolbox of MATLAB v2022a. A digital filter was implemented: a first-order FIR low-pass filter with a cut-off frequency of 10 Hz and a moving average filter. The FIR low-pass filter was implemented to restrict the bandwidth of the CO_2 signal to 10 Hz and focus on the important signal within the respiration signal, as specified in (Klco et al., 2018). The low-pass filtered data was used for the moving average filter. The filter was tested with varying span widths (8, 10, 18, 23, 30, 35, 50, 100, 150) to justify the suitable span width to be applied to the signal. The span width of 8 has been used in (Singh et al., 2018), and a span width of 13 has been utilized in (Balakrishnan et al., n.d.), while the additional widths were added for investigative purposes. The concept applied in the moving average filter is it moves a window of specified length down the data and calculates the averages of the data included in each window (Avuti et al., 2019; Kumar et al., 2019; Maramis and Delopoulos, 2011). The implementation is shown in equation 1.

$$y(n) = \frac{1}{span\ width}\Big(x(n) + x(n+1) + \cdots + x\big(n - (span\ width - 1)\big)\Big) \qquad (1)$$

Then, a moving average filter for the span width of 8 was implemented in the MATLAB code. First, the window width was selected, then the coefficient, depicted as 'A', was prepared following the equation. The function ones returns a 1 by 1 matrix, which was divided by the window span. Then, by using the filter function, it filters the input data with a preprogrammed rational transfer function defined by the coefficients set.

2.5. Feature Extraction Algorithms

A sorting hierarchical algorithm is developed to analyze the exhaled breath waveforms recorded. The algorithm for valid breath cycle selection is developed by assessing the minimum end-tidal carbon dioxide ($EtCO_2$) value, the peak-to-peak distance, and the pathologic properties of the exhaled breath waveforms. These assessments were validated in

previous studies by computing the mean, standard deviation, and its derivatives in each breath cycle recorded in the prototype (Malarvili et al., 2021). The selection criteria outlier exhalation breath cycles in the breathing pattern, which are designated as the valid CO_2 signals recorded from a patient. The core of the presented study is the algorithmic segmentation of exhaled breath. Fig. 2 shows the overview of the presented study of detecting COVID-19 from exhaled breath profile analysis. Each valid exhaled breath cycle extracted is divided into five epochs using a CO_2 partial pressure threshold value, as shown in Fig. 2.

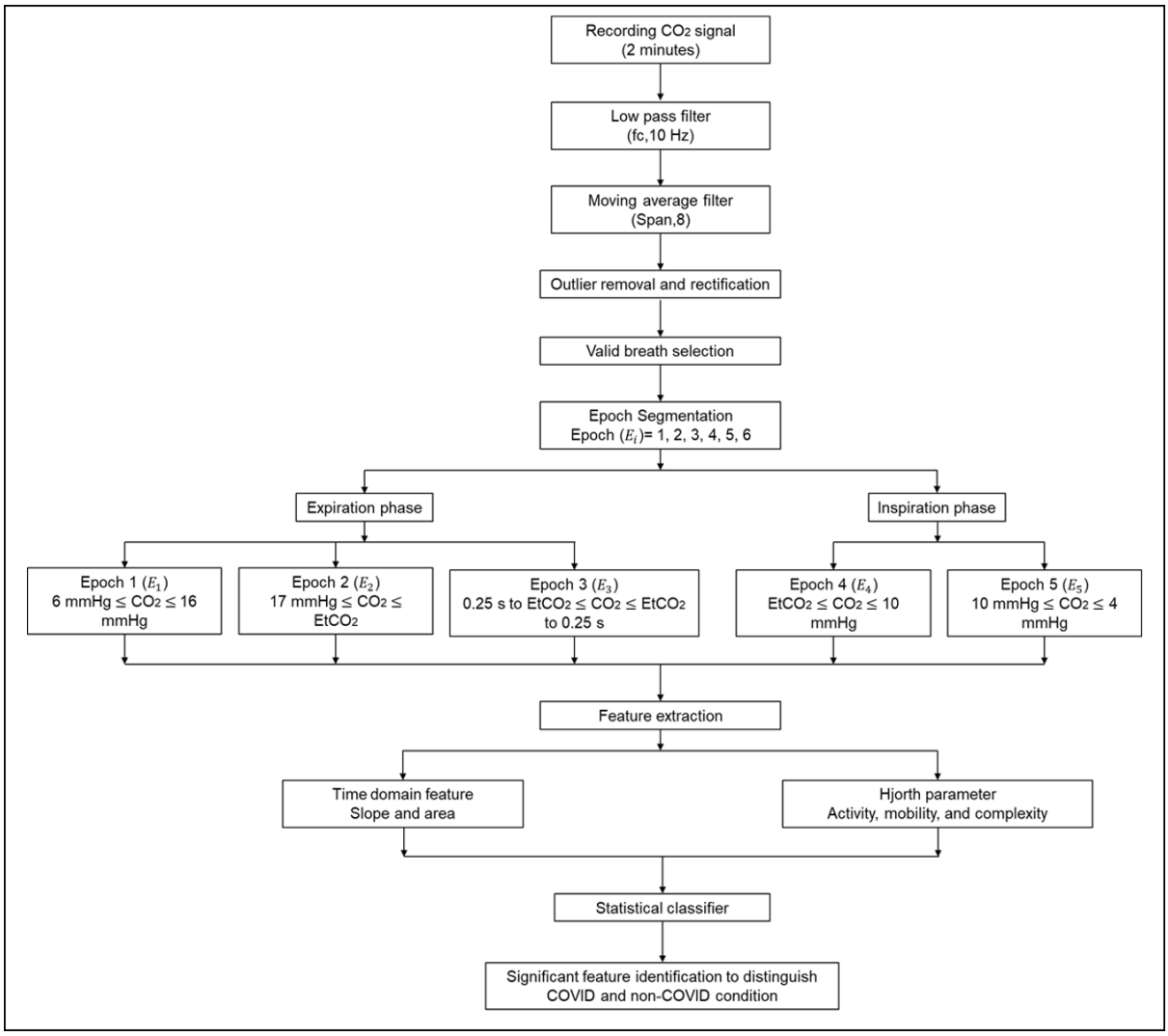

Fig. 2. Overview of the presented study of detecting COVID-19 from exhaled breath profile analysis.

These epochs are the original proposal of the study, which contrasts with the visual analysis of exhaled breath waveforms, where the alveolar plateau and $EtCO_2$ peak value were highly prioritized. These epochs were developed in contrast to the visual interpretation of the exhaled breath waveforms, which gave top priority to the alveolar plateau and $EtCO_2$ peak value. By considering the standard CO_2 partial pressure throughout a breath cycle, these portions were created. An exhaled breath waveform's e_1 and e_2 represent the expiration phase, whereas e_4 and e_5 stand for the inspiration phase. The end-tidal point of an exhaled breath waveform is shown as e_3, and the epoch is shown as being 0.25 seconds to and from the $EtCO_2$ peak. The trend-line-based analysis carried out in this study, which contributed to the epoch being created in this approach (Howe et al., 2011), and the developed algorithm, which

identifies each valid exhaled breath and segments it into five epochs, incorporate this information (Homayoonnia et al., 2021; van Bohemen et al., 2023). As established in previous studies, the slope and area indices of each epoch were computed using Equations 2 and 3, respectively (Malarvili et al., 2021). The least squares linear fitting method was used to compute the slopes of each sub-cycle. By lowering the residue in accordance with (3), it is possible to include the entire CO_2 signal while still computing the intercept and slope of the CO_2 waveform.

$$Slope\ (S_j) = \frac{1}{C} \sum_{j=0}^{C-1} b_j\ (M_j - S_j)^2 \quad (2)$$

$$AR_i = \frac{dt}{5} \sum_{j=0}^{i} (\ R_{j-1}(t) + 4R_j(t) + R_{j+1}(t)) \quad (3)$$

Where slope (*S*) length is defined by *C*, S_j defines the *j*th element, and the best fit and weight of the *j*th element are defined by M_j and b_j, respectively. The CO_2 signal and the sampling interval are defined by *R(t)* and *dt* in area (AR_i) computational formulas. In addition, the Hjorth parameters of each epoch were extracted. These parameters are based on the standard deviations of the amplitude of the signal and its derivatives (i.e., the first and the second derivative). Hjorth parameters are a set of three parameters, which are the activity, mobility, and complexity. These parameters provide information related to the amplitude, slopes of the signal, and the similarity between the shape of the signal under analysis and that of a pure sine wave. As the Hjorth parameters are highly important in differentiating the asthmatic condition, the parameters are included in the present study for evaluating the exhaled breath waveforms of COVID and non-COVID subjects. Hjorth activity indicates the signal's variability with respect to its mean value and provides a measure of the amplitude variance. When the variance is relatively high, it means that the signal's amplitude is widely distributed around the mean value. A slight difference, however, shows the reverse. Equation 4 defines the computational formula for Hjorth activity (variance) (El-Badawy et al., 2022).

$$\sigma_y^2 = \frac{1}{L} \sum_{l=0}^{L-1} (y(l) - \bar{y})^2 \quad (4)$$

Where, y(l) defines the CO_2 signal, and L indicates the data length. Mobility is the second Hjorth parameter, which indicates the ratio of the standard deviation of the signal's first derivative to that of the signal itself. The square root of the first derivative's standard deviation and the signal itself are both equally influenced by the signal's mean amplitude. As a result, the ratio will depend only on the signal's structure and be configured so that it measures the relative average slope. The computational formula for mobility is as follows, where y'(l) refers to the first derivative of the signal (Oh et al., n.d.).

$$Mobility[y(l)] = \frac{\sigma_{y\prime}}{\sigma_y} \quad (5)$$

The concept of complexity, which has no dimensions, describes how the shape of the signal under study resembles a sinusoidal waveform. When the signal's structure is more akin to a pure sine wave, the complexity value is near to one. The ratio of the mobility of the signal's first derivative to its own mobility is used to calculate complexity. Equation 6 represents the formula to calculate Hjorth complexity (Hjorth, n.d.).

$$Complexity[y(l)] = \frac{mobility\ [y'(l)]}{mobility\ [y(l)]} = \frac{\sigma_{y\prime\prime}/\sigma_{y\prime}}{\sigma_{y\prime}/\sigma_y} \quad (6)$$

The y′(l) and y′′(l) represent the first and second derivatives of the CO_2 signal. The following equations define the expressions for the derivatives.

$$y'(l) = \frac{1}{T_s}[y(l+1) - y(l)] \quad (7)$$

$$y''(l) = \frac{1}{T_s}[y'(l+1) - y'(l)] \quad (8)$$

Where Ts defines the sampling time interval (El-Badawy et al., 2022). Fig. 3 shows the flowchart of the algorithm for epoch segmentation and feature extraction. The overall results are typically impacted by the choice of data length, which can therefore have an impact on the findings' accuracy. In this study, an approximate 2-minute data sampling resulted in approximately 35 valid exhaled breath waveforms from a single patient. About 700 valid exhaled breath waveforms resulted in feature computation and interpretation for high-resolution prediction and the ability to detect even minute waveform changes. The $EtCO_2$ value and respiratory rate of a patient are computed from the exhaled breath waveform. The features and the indices for each patient were presented in a comma-separated values (CSV) Excel file. The features in digital and infographic forms are the output of the simulation.

2.6. Data analysis

A descriptive statistical analysis was performed to present the baseline data acquired as the outcome of the developed algorithm. The significance of features in each epoch in defining COVID and non-COVID conditions was evaluated using the significance *p* value from a paired sample T-test. Receiver operating characteristic (ROC) curves were generated for features and indices, where the sensitivity and specificity were quantified through the area under the curve (AUC) and its 95% confidence interval. The predicted values from the statistical analyses were interpreted to classify the presence of COVID and non-COVID conditions in a subject. A Bland-Altman plot was generated for significant features, and the data distribution was evaluated. The descriptive statistical analyses were performed using IBM SPSS (Version 25, USA), and the limit of statistical significance was fixed at $p = 0.05$.

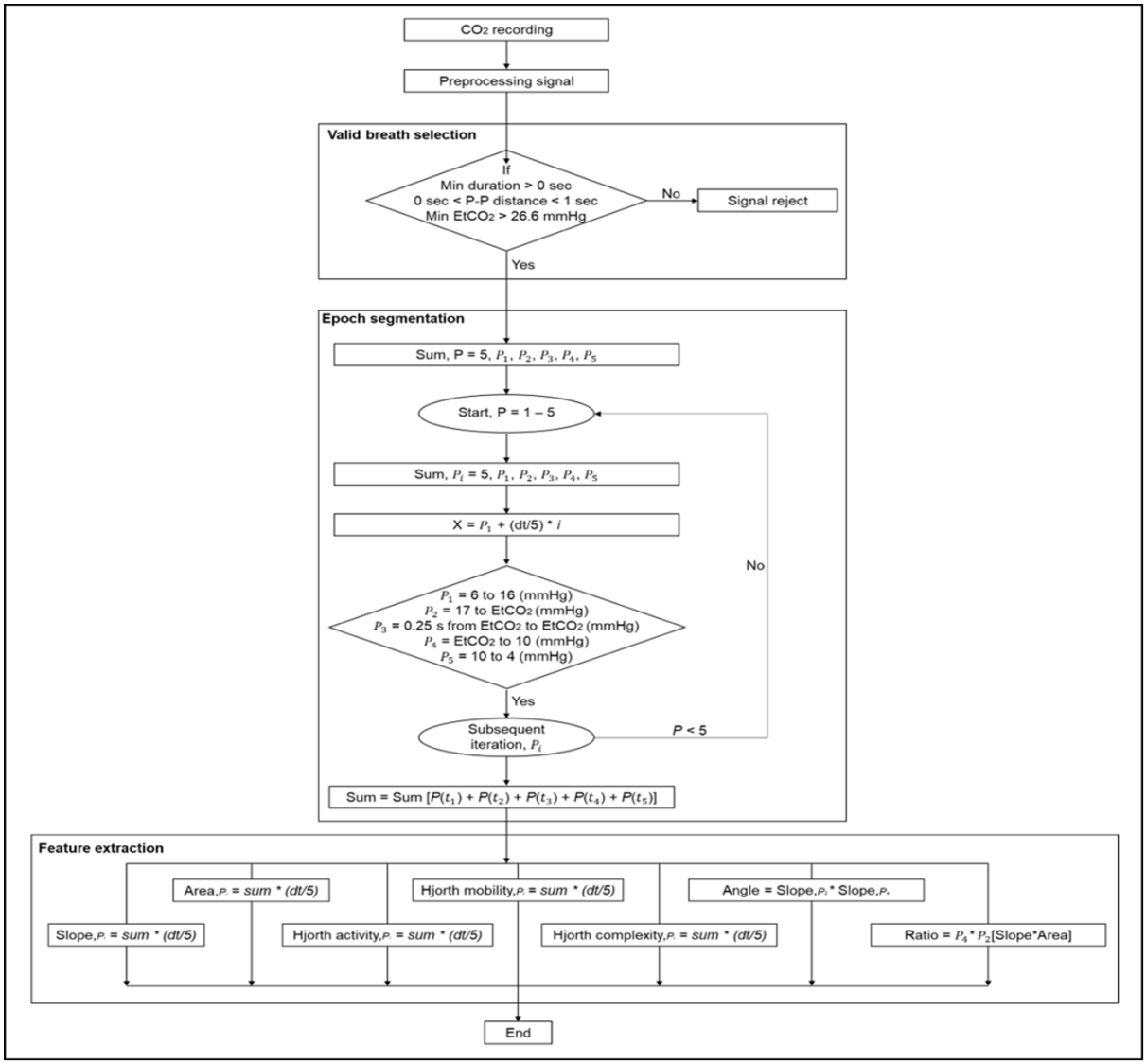

Fig. 3. Flowchart of a valid exhaled breath segmentation employed in the developed algorithm for detecting COVID-19 from exhaled breath profile analysis.

3. Results

3.1. Study Population

The CAT 2 infection stage of COVID-19 patients was validated through RTK-antigen rapid tests performed by the medical officers. The age of patients ranged from 21 to 60 years old, where 24 of them were male and the rest were female patients. The majority of 16 COVID-19 patients showed infection symptoms such as fever, sore throat, and high body temperature. About 4 patients showed symptomless CAT 2 COVID-19 condition. The 20 negative COVID-19 subjects were recruited equally in gender. The non-COVID condition of the subjects was validated through the RTK antigen rapid test. It was made sure that the tested patients had no history of breathing problems like asthma, chronic obstructive pulmonary disease, or airflow blockages. This was a significant input variable for the study to estimate the respiratory indices. Table 1 shows the clinical characteristics of subjects chosen in the study. The number of RTK-confirmed positive and negative CAT 2 COVID-19 patients is 20, respectively. The clinical features of the recruited subjects are shown in Table 1.

Characteristics	RTK-confirmed positive COVID-19 patients	RTK-confirmed negative COVID-19 patients	Total number
Age distribution			
0-20	0	3	3
21-40	7	11	18
41-60	13	6	19
61-80	-		
Gender distribution			
Male	14	10	24
Female	6	10	16
Symptomatic patients	16		
Asymptomatic patients	4		

Table 1. Clinical characteristics of subjects chosen in the study. The number of RTK-confirmed positive and negative CAT 2 COVID-19 patients is 20, respectively.

3.2. Exhaled Breath Cycle from Average CO_2 Partial Pressure

Fig. 4a shows the exhaled breath pattern recorded from a healthy subject. The figure indicates the breath pattern plotted using Microsoft Excel, based on the average CO_2 recorded using the nasal cannula. Prior to being used in the algorithmic simulation, the breath pattern recorded for two minutes was visually analyzed. Fig. 4b shows the expanded region of the exhaled breath waveform from the plot shown in Fig. 4a.

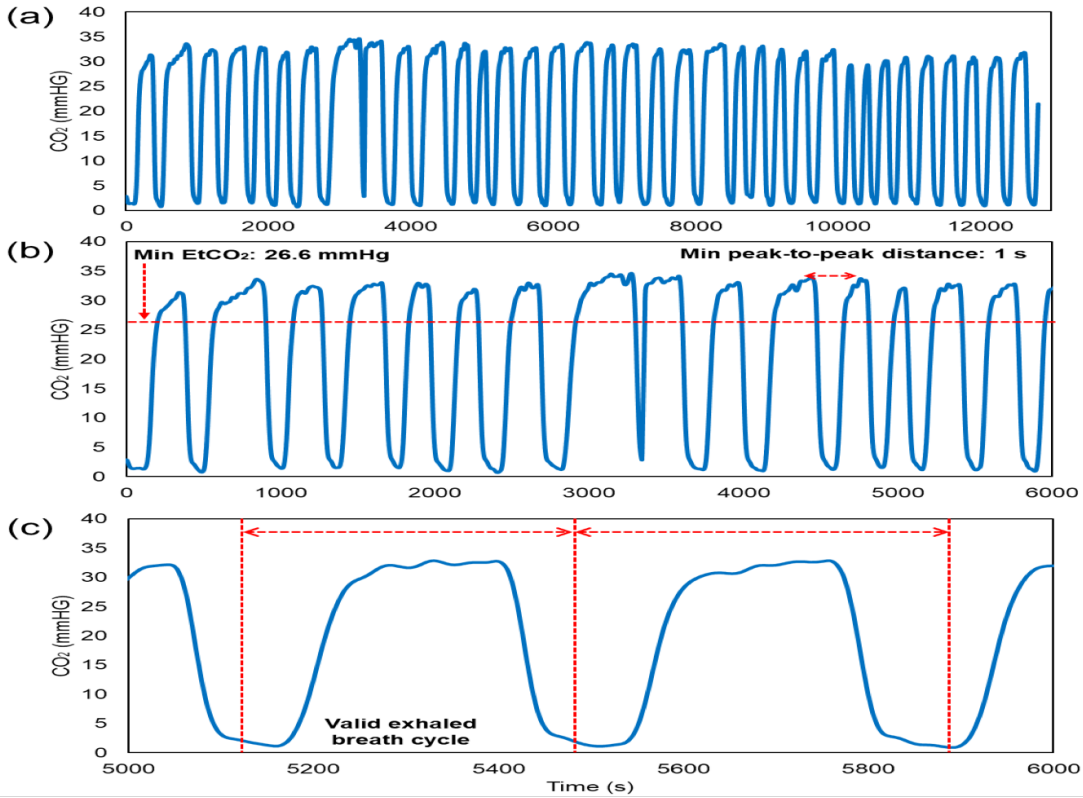

Fig. 4. Exhaled breath pattern; (a) breath pattern plotted using Microsoft Excel; (b) magnified vision of breath pattern, indicating the criteria set for valid breath cycle selection; (c) valid exhaled breath cycle, denoting the valid exhaled breath waveform.

The figure is remarked with the criteria for a valid exhaled breath cycle selection, which is the minimum $EtCO_2$ value at 26.6 mmHg and the minimum peak-to-peak distance at 1 s. These criteria were visually evaluated to detect the valid breath cycles recorded from the COVID and non-COVID subjects. The figure denotes that the minimum $EtCO_2$ and peak-to-peak distance are acceptable for selecting the valid exhaled breath cycles, which are the main input for the algorithm to produce valid exhaled breath waveforms. Fig. 4c shows the further expanded exhaled breath pattern, which denotes the valid exhaled breath cycles that are visualized from the average CO_2 plot. The shape of each valid exhaled breath cycle is close to the exhaled breath waveform, which is expected to be obtained in the designed simulation.

3.3. Algorithmic Exhaled Breath Classification

The developed sorting algorithm selects valid exhaled breath cycles and delivers the valid exhaled breath waveform. Fig. 5 shows the outcome of the simulation on selecting valid exhaled breath cycles from COVID and non-COVID subjects. Fig. 5a(i) shows the selection of a valid exhaled breath waveform pattern, whereas Fig. 5a(ii) shows a valid exhaled breath waveform extracted from a non-COVID subject. It is evident that the exhaled breath pattern of the non-COVID subject is nearly consistent with the similar shape and volume of the exhaled breath pattern. Fig. 5b(i) shows the exhaled breath cycle of a COVID subject. The green triangle on top and the red triangle at the bottom line of each peak indicate the selection of the exhaled breath cycle. Triangles with filled color indicate that the breath cycle meets the selection criteria and is extracted as a valid exhaled breath waveform. Inversely, triangles with unfilled color denote that the breath cycles do not meet the selection criteria and are eliminated from being extracted for a valid exhaled breath waveform. Fig. 5b(ii) shows a

valid exhaled breath waveform extracted from a COVID subject. The shape and volume of the exhaled breath waveform differs from the non-COVID waveform, which denotes the distinct difference between COVID and non-COVID subjects. The differences were further analyzed using a two-tail paired sample T-test.

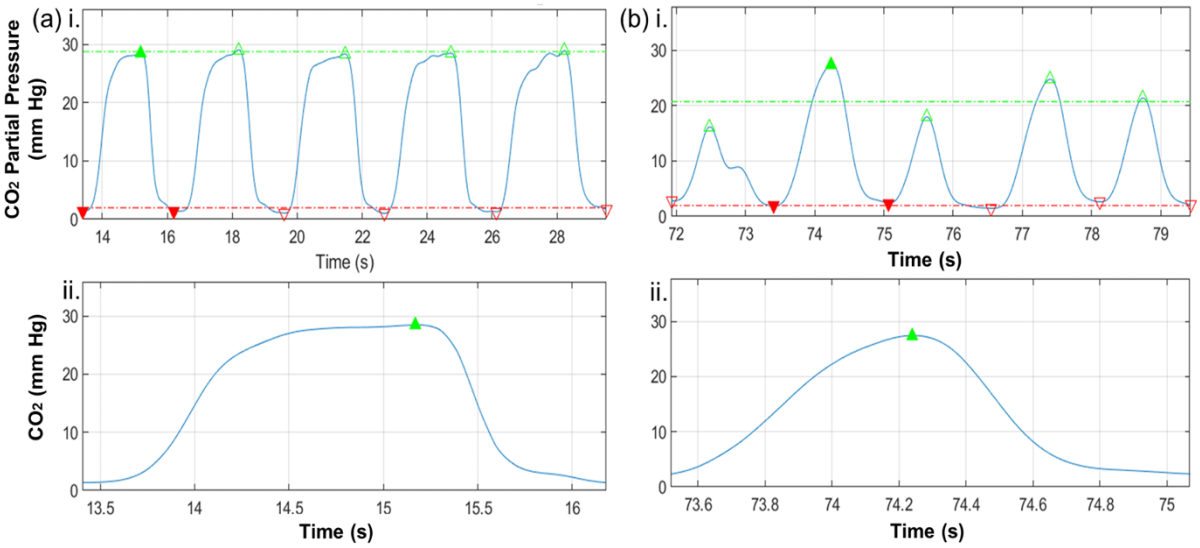

Fig. 5. Algorithmic simulation output on exhaled valid breath cycle selection; (a) i. Exhaled breath waveform of non-COVID subject. ii. Zoomed vision; (b) i. Exhaled breath waveform from COVID-19 subject. ii. Zoomed vision of valid exhaled breath cycle shows distinct difference between COVID and non-COVID waveforms.

3.4. Feature Interpretation

Fig. 6 shows the output of valid exhaled breath waveforms for COVID and non-COVID conditions. The figure indicates the five epochs identified and segmented in the simulation. Slope, areas, angle, activity, mobility, and complexity are the features evaluated in the study. Fig. 6a-6c shows the exhaled breath waveforms extracted from CAT 2 COVID subjects. The shape of the exhaled breath waveforms deviates from the standard shape. Deviation plays a key role in computing the algorithmic output of each epoch. Such variation is analyzed to determine the presence of COVID and non-COVID states in patients based on an algorithmic approach. Fig. 6d-6f shows the exhaled breath waveforms of non-COVID subjects.

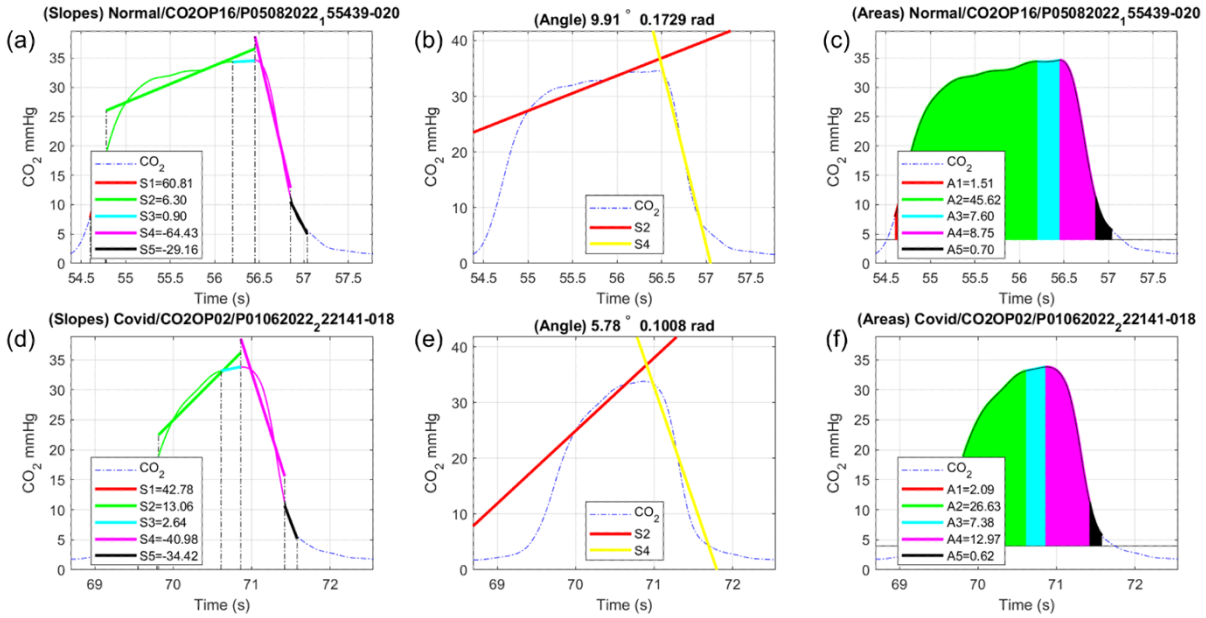

Fig. 6. Exhaled breath feature extraction; (a) – (b) show waveforms indicating slope, angle, and area of each epoch segmented from a COVID-19 subject breath profile, whereas (d) – (f) indicate segmented features for non-COVID.

3.5. Statistical Evaluation: Significant Features

The primary analyses of features in each epoch were presented as the mean with standard deviation (SD). Out of the possible features identified in the literature, 26 features were evaluated in this study. The slope e_2, activity e_2, and the intersection angle of the expiration and inspiration curves showed high significance between COVID and non-COVID conditions with 0.000 *p*-values. The area of e_2 showed high significance with $p = 0.039$. The mobility of e_2 showed $p = 0.008$, indicating significant variation to be considered for evaluating COVID conditions. The complexity feature showed significant variation for only e_3, with $p = 0.024$. The rest of the features of each epoch indicated insignificant variation between COVID and non-COVID conditions. The T-test outputs manifest good correlation between COVID and non-COVID conditions, with the *p* values exhibited by the significant features, which is highly approached for preliminary clinical decision making upon COVID-19 diagnosis.

3.6. Respiratory Indices: $EtCO_2$ and Respiratory Rate

The $EtCO_2$ and respiratory rate (RR) values were recorded to determine the discrepancy between COVID and non-COVID subjects recruited in this study. Fig. 7a shows the bar chart with error bars indicating the respiratory indices are variable in the range of 6 – 42 and 3 – 24, predicted for $EtCO_2$ and RR, respectively, within the COVID and non-COVID subjects. The RR did not reveal huge variation, with a 21.26 mean value and 3.57 SD, and with a 19.86 mean value and 4.72 SD for COVID and non-COVID subjects, respectively. However, the $EtCO_2$ mean for COVID subjects is 32.81 with an SD of 6.38, whereas the non-COVID subjects showed a 36.17 mean with 5.46 SD. The paired T-test analysis for both $EtCO_2$ and RR values showed no significant difference, with *p* values equal to 0.862 and 0.341, respectively. The results denote that $EtCO_2$ and RR indices did not show significant disparity in evaluating COVID and non-COVID conditions.

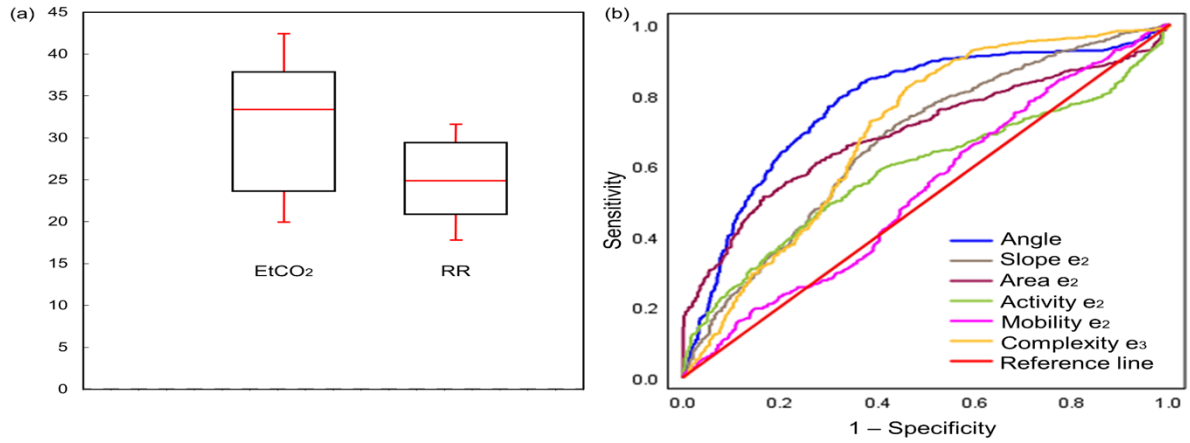

Fig. 7. (a) Correlation between $EtCO_2$ and RR; (b) Receiver operating characteristics (ROC) curves for the significant features indicating predictors of good agreement for differentiating COVID-19 condition.

3.7. Receiver Operating Characteristics (ROC) Analysis

The preliminary clinical studies revealed that receiver operating characteristics (ROC) showed good predictive values in comparing the significance of features for predicting COVID and non-COVID conditions. The area under the curve (AUC) values of the significant features denote the statically pertinent ability to detect the presence of COVID obstruction in the respiratory airway. The predictive value for the angle of intersection between the expiration and inspiration phases indicates a higher value than the other significant features, with a 0.775 AUC. The slope of e_2 and the area of e_2 showed moderately good ability for discriminating COVID and non-COVID conditions with AUC values of 0.667 and 0.693, respectively. Table 2 shows the statistical significance analysis performed for the extracted exhaled breath features between COVID and non-COVID patients. The significant Hjorth parameters of activity, mobility, and complexity showed fairly good compliance with AUC values between 0.500 and 0.550 (Fig. 7b, Table 2). The maximum Kolmogorov-Smirnov index was 0.471 and achieved a cut-off value of 8.30. At this cutoff value, the sensitivity obtained is 92.7%, and the specificity is 57%. The slopes and areas of e_2 revealed 89.8% and 80.1% sensitivity and specificities of 52.1% and 68.4%, respectively, for predicting good compliance with COVID and non-COVID conditions with the respective maximum cut-off values.

	Mean ± Standard deviation		AUC	*p* value	95% Confidence interval	
	COVID patients	Non-COVID patients			Lower limit	Upper limit
Slope e_2	10.52 ± 8.27	14.14 ± 8.26	0.667	0.000	-7.5036420	-5.5726855
Area e_2	37.57± 36.84	25.13 ± 14.09	0.693	0.039	20.7961909	28.1491159
Angle	11.60 ± 8.88	8.10 ± 7.00	0.775	0.000	6.1376349	8.2249380
Activity e_2	9.32 ± 5.30	13.33 ± 7.16	0.528	0.000	-5.4246899	-3.8987386
Mobility e_2	0.041 ± 0.015	0.043 ± 0.018	0.503	0.008	-0.0038495	0.0000657
Complexity e_3	10.82 ± 8.00	9.78 ± 7.73	0.603	0.024	-4.4076976	5.8604478

Table 2. Statistical significance analysis performed for the extracted exhaled breath features between COVID and non-COVID patients.

3.8. Bland-Altman Correlation

Fig. 8 shows the Bland-Altman plots generated for the readings of significant features obtained from the two-tailed T-test. The plot indicates the differences in a computed feature (y-axis) and the mean (x-axis) between COVID and non-COVID subjects. The analysis for the slope of the e_2 feature (Fig. 8a) showed a bias of 6.54 with 30.73 and -17.65 limits of agreement (LOA). The readings for the area of e_2 (Fig. 8b) showed a 24.45 bias with 116.58 and -67.65, lower and upper LOA, respectively. Fig. 8c shows a narrow distribution of angles with a 7.19 bias and 33.32 and -18.97 LOA. The mean difference of activity e_2 is highly scattered, showing a -4.66 bias and 14.45 and -23.18 LOA, as shown in Fig. 8d. The significant Hjorth parameters of mobility (Fig. 8e) and complexity (Fig. 8f) showed narrow distributions with -0.02 (-0.067 – 0.067) and 2.60 (-19.81 – 25), respectively. Although the data distributions are uneven, the linear regression coefficient computed between the dependent bias and the independent mean revealed a significant *p*-value for these features, as indicated in each plot of Fig. 8. It is evident that the slope, area, and the angle features show proportional distributions with significant *p* values equal to 0.027 for slope e_2, and 0.000 for area e_2 and angle. The significant Hjorth parameters showed insignificant correlation, denoted by the linear regression between bias and the mean of the features. The plots revealed the agreement between the significant features output of COVID and non-COVID by denoting the differences in a computed feature (y-axis) and the mean (x-axis).

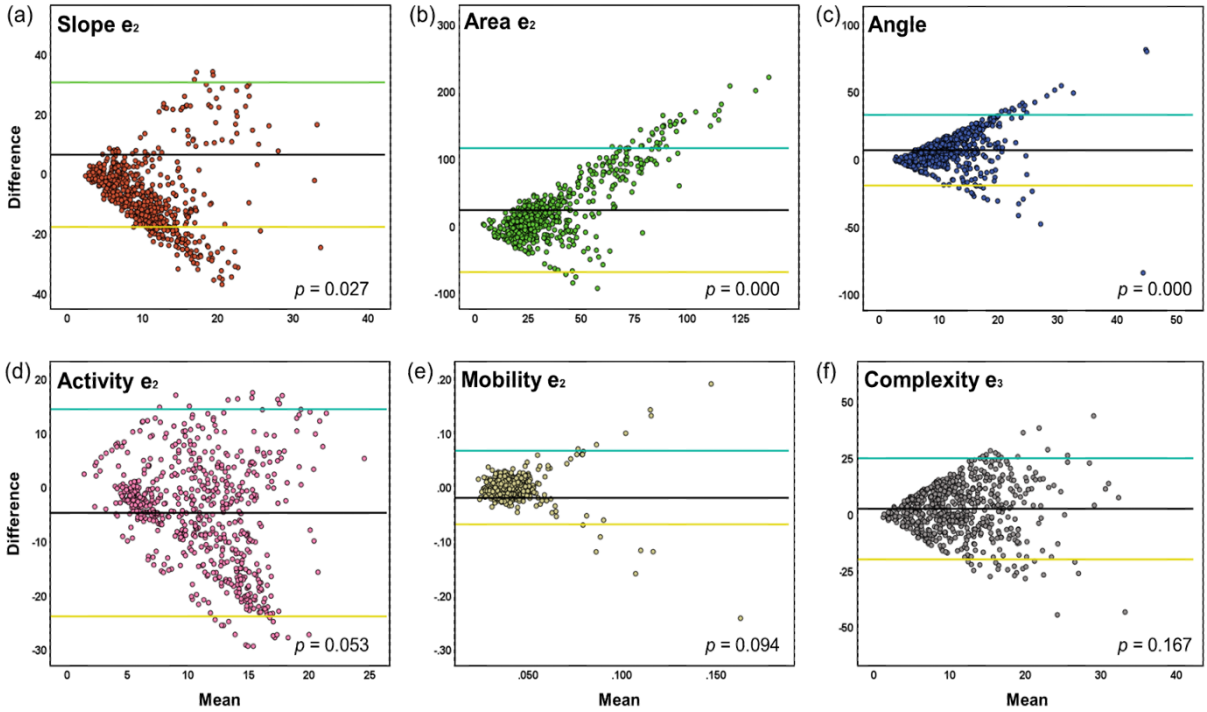

Fig. 8. Bland-Altman plots for (a) slope e_2, (b) area of e_2, (c) intersection angle between expiration and inspiration phases, (d) activity e_2, mobility e_2, and complexity e_3.

4. Discussions

The presented work demonstrates a preliminary approach to identifying the presence of COVID-19 infection through the exhaled breath-based waveform using the developed algorithm. The study was evaluated based on the statistical significance output score in distinguishing the state of COVID and non-COVID. The generation of the algorithm for

exhaled breath waveform segmentation was challenged by the variance that existed in COVID-19 patients and the many features associated with the segmented exhaled breath waveforms (Abbasi-Kesbi et al., 2020). Since the study demands the least variance in clinical data acquisition, the preliminary research was focused on only category 2 (CAT 2) COVID-19 patients. Out of 20 RTK-positive COVID-19 patients, only 4 showed asymptomatic infection. The rest of the patients showed symptoms of a runny nose, sore throat, loss of taste and smell, fever, severe cough, and general appearance of unwellness. However, the patients were approachable for declaring their interest in taking part in the pre-clinical study. The work that was displayed has significant medical applications. With the least amount of patient cooperation, the exhaled breath waveform recording can be performed. The data recording can therefore be repeated without being altered (Godoy et al., 2023; Tiller et al., 2021). Using a nasal cannula and monitoring equipment, the system is simple to set up and straightforward to carry out. It is acceptable for a non-clinician to do the assessment since the data collected is statistically analyzed. As COVID-19 disease infection begins at the upper respiratory tract, it is likely to be recognized through the breath pattern of infected individuals. One of the acceptable biomarkers for the detection of COVID-19 is volatile organic molecules (VOC). Nonetheless, VOCs are found in the breath of infected patients, and the infection state is assessed based on these findings. The statistical evaluation of the algorithmic outputs for COVID-19 infection detection was performed in this study.

A two-tailed paired sample T-test was performed as the primary statistical analysis to identify the features that show significant differences between the COVID and non-COVID conditions. Based on Table 2, the algorithmic simulation showed that the slope and area of e_2 are significant. It is highly expected that significant differences exist between the slope and area of e_2 as the e_2 indicates the exhalation region of the waveform. These results indicate that the slope and area features for e_2 are significant for use in distinguishing the COVID and non-COVID conditions. The developed algorithm was designed as it results in the intersection angle between the expiration and inspiration phases of a valid exhaled breath waveform. The angle targeted in this study has a huge difference from the alpha and beta angles of exhaled breath waveforms reported in the literature. The alpha and beta angles are prominently utilized for detecting respiratory diseases, which comes along with the simulation of computing the transition angles of phase II and phase III of a waveform. In this study, the intersection angle of expiration and inspiration is introduced as it simplifies the relationship between the alpha and beta angles (Peveling-Oberhag et al., 2020). In an aim of simplifying the use of these indexes, the intersection angle of the two phases is algorithmically obtained and analyzed. This feature showed a significant difference between the COVID and non-COVID conditions with p value equal to 0.000 with T-test analysis. This angle is expected to be highly employed in detecting airway obstruction, where inhomogeneity in ventilation is associated with the deviation of the slope and area of the expiration and inspiration regions. It is preliminarily determined by manual exhaled breath waveform inspection and helps for detecting malfunction in alveolar ventilation and the chambers in the bronchial region. In the presented study, the intersecting angle feature was statistically computed through the developed algorithm. It is justified to be significant in differentiating the COVID and non-COVID conditions. The activity and mobility features of e_2 showed significant differences. The computation of activity for one breath is proved to distinguish between asthmatic and non-asthmatic conditions in the stated report (Arjaria et al., 2021). Mobility is reported to be used in identifying transitional errors in electroencephalogram (EEG) signal evaluations. Based on p-value analysis, the activity and mobility of e_2 can be used in discriminating the COVID and non-COVID conditions. Besides, the complexity of $e_3$showed significant differences. The e_3 reveals the $EtCO_2$ peak variation

between the COVID and non-COVID conditions. The complexity feature aids in EEG signal interpretation and photoplethysmography examination (Grover and Turk, 2020; Rizal et al., 2019). As these features are highly welcomed in medical applications, they are further evaluated for identifying the COVID and non-COVID conditions. The significant features are further analyzed using ROC analysis. The area under the ROC curve is moderately greater than 0.500 for all significant features. The intersection angle showed a higher AUC than the rest, which emphasizes its importance in identifying the COVID and non-COVID conditions. The analyses on respiratory indexes revealed that $EtCO_2$ and RR computational analysis are insufficient to identify the COVID condition.

6. Conclusion

The work presented has provided support for an alternate method of identifying COVID-19 infection. According to the study, the developed algorithm divides a valid exhaled breath waveform into five epochs and computes its feature readings to provide the appropriate result. According to the statistical analyses, the slope of e_2, area of e_2, intersection angle of the expiration and inspiration phases, and exhaled breath features are the most promising for differentiating between COVID and non-COVID conditions. Prior to the algorithm developed in giving the accurate exhaled breath waveform features output, the preliminary clinical trial with 20 RTK-positive and 20 RTK-negative participants has proven that the exhaled breath waveform delivers an easy and speedy COVID-19 detection technique. The research has stressed the importance of exhaled breath analysis as a tool for algorithmic and statistical simulation that targets a particular threshold, in addition to aiding in the detection of diseases based on biomolecules. This will assist physicians in initiating preventative measures and reduce the workload associated with invasive sample methods.

Acknowledgements

This work was supported by "Pembangunan COVID-19 Screening and Monitoring Tool i-Breath" under Prototype Development Grant (R.J130000.7351.4B746) from the Ministry of Science, Technology, and Innovation, Malaysia.

Declarations

Competing Interests

The authors have no competing interests declared.

Ethical Approval and Consent to Participate

The study was approved by the Medical Research and Ethics Committee (MREC), Malaysia. The study protocol was registered in the National Medical Research Register (NMRR-21-763-59692).

Trial registration

Clinical trial approval by the Medical Research and Ethics Committee (MREC), 53 Malaysia, NMRR-21-763-59692. Registered 9th June 2021 and valid till 8th June 2023.

Authors Contributions

- M.B.M*: Conceived and designed the analysis, analysis, and interpretation of results.
- N.D.: Analysis and interpretation of results, manuscript preparation.
- S.R.: Performed the analysis and interpretation of results, manuscript preparation.
- N.A.M.: Contributed data or analysis tools
- R.M.: Contributed data or analysis tools
- A.A.R.: Collected the data
- R.R.R.: Collected the data
- A.K.: Collected the data
- A.A.A.T.: Collected the data

Funding

This work was supported by "Pembangunan COVID-19 Screening and Monitoring Tool i-Breath" under Prototype Development Grant (R.J130000.7351.4B746) from the Ministry of Science, Technology, and Innovation, Malaysia.

Conflict of Interests

The authors declare that they have no conflict of interest.

Availability of Data and Materials

The authors confirm that the data supporting the findings of this study are available within the article and its supplementary materials.

References

- Abbasi-Kesbi, R., Asadi, Z., Nikfarjam, A., 2020. Developing a wireless sensor network based on a proposed algorithm for healthcare purposes. Biomed Eng Lett 10, 163–170. https://doi.org/10.1007/s13534-019-00140-w
- Arjaria, S.K., Chaubey, G., Shukla, N., 2021. Hjorth Parameter based Seizure Diagnosis using Cluster Analysis. J Phys Conf Ser 1998. https://doi.org/10.1088/1742-6596/1998/1/012020
- Avuti, S.K., Bajaj, V., Kumar, A., Singh, G.K., 2019. A novel pectoral muscle segmentation from scanned mammograms using EMO algorithm. Biomed Eng Lett 9, 481–496. https://doi.org/10.1007/s13534-019-00135-7
- Balakrishnan, M., Kazemi, M., Mahmood, N.H., Malarvili, M.B., Humaimi Mahmood, N., n.d. Investigating Frequency Contents of Capnogram using Fast Fourier Transform (FFT) and Autoregressive Modeling (AR) Heart Rate Monitoring System View project Investigating Frequency Contents of Capnogram using Fast Fourier Transform (FFT) and Autoregressive Modeling (AR).
- Carmichael, H., Coquet, J., Sun, R., Sang, S., Groat, D., Asch, S.M., Bledsoe, J., Peltan, I.D., Jacobs, J.R., Hernandez-Boussard, T., 2021. Learning from past respiratory failure patients to triage COVID-19 patient ventilator needs: A multi-institutional study. J Biomed Inform 119, 103802. https://doi.org/10.1016/j.jbi.2021.103802

- El-Badawy, I.M., Omar, Z., Singh, O.P., 2022. An Effective Machine Learning Approach for Classifying Artefact-Free and Distorted Capnogram Segments Using Simple Time-Domain Features. IEEE Access 10, 8767–8778. https://doi.org/10.1109/ACCESS.2022.3143617
- Francescato, M.P., Thieschäfer, L., Cettolo, V., Hoffmann, U., 2019. Comparison of different breath-by-breath gas exchange algorithms using a gas exchange simulation system. Respir Physiol Neurobiol 266, 171–178. https://doi.org/10.1016/j.resp.2019.04.009
- Godoy, D.A., Longhitano, Y., Fazzini, B., Robba, C., Battaglini, D., 2023. High flow nasal oxygen and awake prone positioning – Two allies against COVID-19: A systematic review. Respir Physiol Neurobiol. https://doi.org/10.1016/j.resp.2023.104015
- Grover, C., Turk, N., 2020. Rolling Element Bearing Fault Diagnosis using Empirical Mode Decomposition and Hjorth Parameters. Procedia Comput Sci 167, 1484–1494. https://doi.org/10.1016/j.procs.2020.03.359
- Hjorth, B.O., n.d. TECHNICAL EEG ANALYSIS BASED CONTRIBUTIONS ON TIME DOMAIN PROPERTIES.
- Homayoonnia, S., Lee, Y., Andalib, D., Rahman, M.S., Shin, J., Kim, K., Kim, S., 2021. Micro/nanotechnology-inspired rapid diagnosis of respiratory infectious diseases. Biomed Eng Lett. https://doi.org/10.1007/s13534-021-00206-8
- Howe, T.A., Jaalam, K., Ahmad, R., Sheng, C.K., Nik Ab Rahman, N.H., 2011. The use of end-tidal capnography to monitor non-intubated patients presenting with acute exacerbation of asthma in the emergency department. Journal of Emergency Medicine 41, 581–589. https://doi.org/10.1016/j.jemermed.2008.10.017
- Klco, P., Kollarik, M., Tatar, M., 2018. Novel computer algorithm for cough monitoring based on octonions. Respir Physiol Neurobiol 257, 36–41. https://doi.org/10.1016/j.resp.2018.03.010
- Kreyer, S., Baker, W.L., Scaravilli, V., Linden, K., Belenkiy, S.M., Necsoiu, C., Muders, T., Putensen, C., Chung, K.K., Cancio, L.C., Batchinsky, A.I., 2021. Assessment of spontaneous breathing during pressure controlled ventilation with superimposed spontaneous breathing using respiratory flow signal analysis. J Clin Monit Comput 35, 859–868. https://doi.org/10.1007/s10877-020-00545-4
- Kumar, A., Ranganatham, R., Komaragiri, R., Kumar, M., 2019. Efficient QRS complex detection algorithm based on Fast Fourier Transform. Biomed Eng Lett 9, 145–151. https://doi.org/10.1007/s13534-018-0087-y
- Malarvili, M.B., Alexie, M., Dahari, N., Kamarudin, A., 2021. On Analyzing Capnogram as a Novel Method for Screening COVID-19 : A Review on Assessment Methods for COVID-19. Life Journal 1–26.
- Maramis, C.F., Delopoulos, A.N., 2011. A novel algorithm for restricting the complexity of virus typing via PCR-RFLP gel electrophoresis. Biomed Eng Lett 1, 239–246. https://doi.org/10.1007/s13534-011-0038-3
- Marshall, B.G., White, V., Loveridge, J., 2021. Breathlessness and cough in the acute setting. Medicine (United Kingdom) 49, 93–97. https://doi.org/10.1016/j.mpmed.2020.11.004
- Mauer, E., Lee, J., Choi, J., Zhang, H., Hoffman, K.L., Easthausen, I.J., Rajan, M., Weiner, M.G., Kaushal, R., Safford, M.M., Steel, P.A.D., Banerjee, S., 2021. A predictive model of clinical deterioration among hospitalized COVID-19 patients by

harnessing hospital course trajectories. J Biomed Inform 118, 103794. https://doi.org/10.1016/j.jbi.2021.103794
- Maurya, L., Zwiggelaar, R., Chawla, D., Mahapatra, P., 2022. Non-contact respiratory rate monitoring using thermal and visible imaging: a pilot study on neonates. J Clin Monit Comput. https://doi.org/10.1007/s10877-022-00945-8
- Oh, S.-H., Lee, Y.-R., Kim, H.-N., n.d. A Novel EEG Feature Extraction Method using Hjorth Parameter.
- Peveling-Oberhag, J., Michael, F., Tal, A., Welsch, C., Vermehren, J., Farnik, H., Grammatikos, G., Lange, C., Walter, D., Blumenstein, I., Filmann, N., Herrmann, E., Albert, J., Zeuzem, S., Bojunga, J., Friedrich-Rust, M., 2020. Capnography monitoring of non-anesthesiologist provided sedation during percutaneous endoscopic gastrostomy placement: A prospective, controlled, randomized trial. Journal of Gastroenterology and Hepatology (Australia) 35, 401–407. https://doi.org/10.1111/jgh.14760
- Ramanathan, S., Gopinath, S.C.B., Hilmi Ismail, Z., Subramaniam, S., 2022a. Nanodiamond conjugated SARS-CoV-2 spike protein: electrochemical impedance immunosensing on a gold microelectrode. Microchimica Acta 189, 1–14. https://doi.org/10.1007/s00604-022-05320-7
- Ramanathan, S., Gopinath, S.C.B., Ismail, Z.H., Md Arshad, M.K., Poopalan, P., 2022b. Aptasensing nucleocapsid protein on nanodiamond assembled gold interdigitated electrodes for impedimetric SARS-CoV-2 infectious disease assessment. Biosens Bioelectron 197, 113735. https://doi.org/10.1016/j.bios.2021.113735
- Rizal, A., Hidayat, R., Nugroho, H.A., 2019. Lung sound classification using hjorth descriptor measurement on wavelet sub-bands. Journal of Information Processing Systems 15, 1068–1081. https://doi.org/10.3745/JIPS.02.0116
- Robba, C., Battaglini, D., Ball, L., Patroniti, N., Loconte, M., Brunetti, I., Vena, A., Giacobbe, D.R., Bassetti, M., Rocco, P.R.M., Pelosi, P., 2020. Distinct phenotypes require distinct respiratory management strategies in severe COVID-19. Respir Physiol Neurobiol. https://doi.org/10.1016/j.resp.2020.103455
- Seo, G., Lee, G., Kim, M.J., Baek, S.H., Choi, M., Ku, K.B., Lee, C.S., Jun, S., Park, D., Kim, H.G., Kim, S.J., Lee, J.O., Kim, B.T., Park, E.C., Kim, S. Il, 2020. Rapid Detection of COVID-19 Causative Virus (SARS-CoV-2) in Human Nasopharyngeal Swab Specimens Using Field-Effect Transistor-Based Biosensor. ACS Nano 14, 5135–5142. https://doi.org/10.1021/acsnano.0c02823
- Shan, B., Broza, Y.Y., Li, Wenjuan, Wang, Y., Wu, S., Liu, Z., Wang, Jiong, Gui, S., Wang, L., Zhang, Z., Liu, W., Zhou, S., Jin, W., Zhang, Qianyu, Hu, D., Lin, L., Zhang, Qiujun, Li, Wenyu, Wang, Jinquan, Liu, H., Pan, Y., Haick, H., 2020. Multiplexed Nanomaterial-Based Sensor Array for Detection of COVID-19 in Exhaled Breath. ACS Nano 14, 12125–12132. https://doi.org/10.1021/acsnano.0c05657
- Singh, O.P., Howe, T.A., Malarvili, M.B., 2018. Real-time human respiration carbon dioxide measurement device for cardiorespiratory assessment. J Breath Res 12. https://doi.org/10.1088/1752-7163/aa8dbd
- Stroh, J.N., Smith, B.J., Sottile, P.D., Hripcsak, G., Albers, D.J., 2023. Hypothesis-driven modeling of the human lung–ventilator system: A characterization tool for Acute Respiratory Distress Syndrome research. J Biomed Inform 137, 104275. https://doi.org/10.1016/j.jbi.2022.104275

- Taleghani, N., Taghipour, F., 2021. Diagnosis of COVID-19 for controlling the pandemic: A review of the state-of-the-art. Biosens Bioelectron 174, 112830. https://doi.org/10.1016/j.bios.2020.112830
- Tiller, N.B., Turner, L.A., Hart, J., Casaburi, R., 2021. Airflow dynamics and exhaled-breath temperature following cold-water ingestion. Respir Physiol Neurobiol 284. https://doi.org/10.1016/j.resp.2020.103564
- van Bohemen, S.J., Rogers, J.M., Boughton, P.C., Clarke, J.L., Valderrama, J.T., Kyme, A.Z., 2023. Continuous non-invasive estimates of cerebral blood flow using electrocardiography signals: a feasibility study. Biomed Eng Lett. https://doi.org/10.1007/s13534-023-00265-z
- Wang, W., Alzate-Correa, D., Alves, M.J., Jones, M., Garcia, A.J., Zhao, J., Czeisler, C.M., Otero, J.J., 2021. Machine learning-based data analytic approaches for evaluating post-natal mouse respiratory physiological evolution. Respir Physiol Neurobiol 283. https://doi.org/10.1016/j.resp.2020.103558
- Zubieta-Calleja, G.R., Zubieta-DeUrioste, N., de Jesús Montelongo, F., Sanchez, M.G.R., Campoverdi, A.F., Rocco, P.R.M., Battaglini, D., Ball, L., Pelosi, P., 2023. Morphological and functional findings in COVID-19 lung disease as compared to Pneumonia, ARDS, and High-Altitude Pulmonary Edema. Respir Physiol Neurobiol. https://doi.org/10.1016/j.resp.2022.104000