Identification of Asthmatic Patient During Exercise Using Feature Extraction of Carbon Dioxide Waveform

Om Prakash Singh  
Bio-Signal Processing Research Group  
School of Biomedical Engineering and Health sciences  
Universiti Teknologi Malaysia  
Johor Bahru, Malaysia  
bioom85@yahoo.com

Rokini Kumarasamy  
Bio-Signal Processing Research Group  
School of Biomedical Engineering and Health sciences  
Universiti Teknologi Malaysia  
Johor Bahru, Malaysia  
rokini2009@gmail.com

MB Malarvili  
Bio-Signal Processing Research Group  
School of Biomedical Engineering and Health sciences  
Universiti Teknologi Malaysia  
Johor Bahru, Malaysia  
malarvili@biomedical.utm.my

Zahratun nur binti abd Hamid  
Bio-Signal Processing Research Group  
School of Biomedical Engineering and Health sciences  
Universiti Teknologi Malaysia  
Johor Bahru, Malaysia  
zahra_nur42@yahoo.com

Abstract— Asthmatic athletes encounter greater challenges while practicing sport in dealing with their disorder. Hence, this study explores the significance of features extracted from the different phases of carbon dioxide (CO$_2$) waveform morphology from an asthmatic patient during exercise. Herein, CO$_2$ data was collected using human respiration CO$_2$ measurement device from 9 asthmatic subjects in the stable mild state before and after exercise, aged between 20-25 years via convenience sampling method, chosen from the UTM Health Center. The subjects were asked to run on the medical treadmill (TMX428-15% elevation, 7.5 m/h) for 2 minutes. Thereafter, we automatically segmented each breath cycle into sub-cycles using threshold and computed $Area$ as a feature from each part using numerical methods compared with slope or derivative of the capnogram. We found that the feature ($Area$) for the segmented part of the mix of upward expiratory phase and alveolar phase possess higher area under curve (AUC) of 0.94 via receiver operating curve analysis. Further, the feature was sent to the classifier for the classification of athlete asthmatic condition before and after exercise. We found that the decision tree possesses higher accuracy (88.89%), sensitivity (100%) and specificity (77.78%) with an AUC of 0.85 than other classifiers. Thus, the incorporation of these features into a newly developed device could possibly allow asthmatic athletes to manage their condition during exercise and sports.

Keywords— Asthma, carbon dioxide, athletes, management, feature

I. INTRODUCTION

An athlete is a person who is good at sports and competes in one or more sports that involve physical strength, speed or endurance. It is common to have some breathing difficulties during or after the physical exertion of sports or other athletic events. According to a study [1], about 4 to 15% of the asthmatic athletes suffered from asthma or used anti-asthmatic medication during summer Olympic games. In addition, a workout may hyperventilate up to 200 l/min for short eras in speed and power athletes, and for longer eras in toleration athletes such as cross-country skiers and swimmers [2]. Growing numbers are atopic for young athletes. It means that they show signs of IgE-mediated allergy, which is a major risk factor for respiratory symptoms and asthma along with the sport event (toleration sport). Besides, asthma appears to be frequent in elite athletes due to the fact that athletes are extensively exposed to inhaled irritants and in the spring and summer to many pollen allergens [3]. Hence, there is a strong need for self-asthma management device.

The traditional device such as Peak flow meter (PFM) and the Spirometer are used as an objective assessment method for diagnosis and management of asthma. However, Both PFM and spirometer are patient dependent and require blowing of air forcefully and follow up of a set of instructions during the measurement [4-6]. Besides, patients who cannot liaise due to the sicknesses process or age are unable to make these assessments and may possibly provide unreliable results. Thus, capnography, which measures respired carbon dioxide (CO$_2$) waveform, has been proposed as a substitute for the assessment of asthma [7-9].

Capnography is a noninvasive method that analysis CO$_2$ concentration continuously in the respiratory cycle. The CO$_2$ concentration measured and displayed in terms of mmHg. Infrared technology is employed by capnography to sense and measure the carbon dioxide concentration from the air [10]. Non-elementary gases absorbed the infrared rays, which is the collection of dissimilar atoms (carbon dioxide gas combines 1 atom of carbon and 2 atoms of oxygen), while certain gases absorb specific wavelength revealing the absorption bands in the infrared
spectrums. The infrared intensity via radiation likely via a gas mixture consisting carbon dioxide is declined by absorption; this procedure permits the carbon dioxide absorption band to be accepted and is relative to the amount of carbon dioxide in the combination. Besides, carbon dioxide concentration strongly absorbed in the infrared region of 4.2µm [11]. With this carbon dioxide and infrared characteristic, infrared technology is employed in capnography to sense and measure the quantity of CO₂ concentration in breathing air.

Fig. 1. End of inspiration, beginning of exhalation, alveolar plateau, end of exhalation, and beginning of new breath. ST, end of new breath; PQ section, combination of alveolar gasses and dead space; QR segment, alveolar phase depicts the amount of gas delivered by alveoli; R indicates the end of exhalation; RS segment, initiation of new breath [12].

Fig. 1. depicts four phases in capnogram, phase 0, I, II, and III reflect the ST, PQ, QR, and ST segment [7]. A complete breath cycle is counted as the ending of inspiration and the beginning of new breath as ST-PQR-RS. Phase P-Q-R is considered the exhalation phase whereas RS is the beginning of inspiration. Phase ST indicates the removal of CO₂ free gas from the air tube, the apparatus dead and the anatomical space. Phase PQ represents a rapid increase in the tracing that is because of a combination of dead space with alveolar gas. Phase QR provide the information of alveolar plateau indicating CO₂ rich gas from the alveoli. It seems to have a positive slope, depicting increasing of CO₂ pressure. Phase RS shows the downward inspiration while inhaling air during respiration. The amount of CO₂ in expired gas is found to be higher than the inspired gas and estimated to be around 0.36 mmHg (inspiration) and 40 mmHg (expiration) of carbon dioxide. The highest value presents at the end of the expiration Phase P-QR reflects the end-tidal carbon dioxide (EtCO₂).

Alpha is the angle between PQ and QR in each breath [8]. Usually, the angle ranges from 100 and 110 degrees. The alpha angle raises due to an increase in the slope of Phase QR and illustrates. The change in the slope of Phase QR depends upon the ventilation/perfusion (V/Q) status of the lungs. Hence, computation of alpha angle can indirectly reveal V/Q status of lungs. Besides, the angle calculated between the Phase QR and the downward inspiratory phase is known as beta angle [11]. Usually, beta angle is used to be approximately 90 degrees and confirms the rebreathing status [7].

In this study, we computationally divided each breath cycle into different cycles. Thereafter, we extract the feature (Area) from the CO₂ waveform, collected pre- and post-exercise in order to discriminate asthma and non-asthma conditions. This paper is organized as follows. In section II we explained the methodology employed for the segmentation and feature extraction. In Section III we demonstrate the results and interpret the obtained results. Finally, section IV illustrates the conclusion and future scopes of the study.

II. METHOD

Nine asthmatic subjects in a stable mild state, aged between 20–25 years, took an interest in this study. The subjects were identified based on records in the medical case remarks at the time of hospital admission that do not require confirmatory documentary of Spirometer. An informed consent form was obtained in order to record their demographic details. Subjects were gone through the physical examination to assess the signs of asthma, including wheezing, air entry, and retractions. The subjects enrolled with asthma with a stable mild state were evaluated by an expert.

A. Data Collection Procedure

The data were collected using human respiration CO₂ measurement device, designed based on sidestream technology [12] as illustrated in Fig.2 (A and B) from the UTM health center using a convenience sampling method [13]. The device consists of four parts such as CO₂ recording unit, a controller unit, and a high-resolution display. This device acquires the CO₂ signal by employing the sampling tube and passes to the processing unit for estimation and transmission purpose. In addition, a serial communication was established in order to display the features on a high-resolution display.

![Fig. 2. Designed human respiration CO₂ measurement device (A) Nasal cannula was not fastened with subjects and (B) During the sampling tube fastened.](image-url)
The device was capable of recording the data into a secure digital (SD) card that is controlled by the processing unit as presented in [12]. Besides, the designed device was enough small in size that makes device handy as depicted in Fig.2, thereby, can be used by individuals outside of the hospital.

B. Data Processing and Partitioning

The collected CO₂ data were processed in order to remove the unwanted noise from the signal prior to feature extraction using (1). It eliminates the haphazard noise while maintaining the response as a sharp step that brands it as an optimal and leading filter from the programming point of view for time domain signal. The actual CO₂ signal is found to be inappropriate, have an uneven shape, and noisy, that may possibly deliver an imprecise result (i.e. features). On the other side, the smoothened CO₂ signal looks very crispy, having regular morphology, and smooth, while preserving the accuracy.

\[ y(n) = \frac{1}{2n+1} (y(n+m) + y(n+m-1) + \ldots + y(n-m)) \]  

Where, \( y(n) \) represents the smooth value for \( n \)-points, \( m \) indicates the neighboring data point on either side of \( y(n) \) and \( 2n+1 \) is the span width.

We decided to extract the Area as a feature because this feature is proven to be efficient in differentiating asthma and non-asthma conditions, a study conducted by You et.al. and Singh et.al. in their manual and automatic study. As well as, the feature shows good classification accuracies from asthma and normal CO₂ signal. On the other hand, the investigation was performed with healthy and acute asthmatic patients whereas this study focuses on asthmatic athletes who were in the controlled condition in order to prove the use of a device with an athlete as a device is small and handy.

D. Classifier

We employed the machine learning methods for the assessment of feature for the classification of data before and after exercise. Herein, decision tree (DT), support vector machine (SVM), and k-NN classifier were applied on the feature via 10-fold cross-validation. Each of the machine learning methods is described in subsequent sections.

Decision Tree (DT)

A divide-and-conquer way is utilized in a DT that consists of a set of tree-structured decision tests functioning. This method regularly divides the data set as per the standard that increases the parting of the data, subsequent in a tree-like structure [6,7]. The value of a target label is predicted based on several input features in a decision tree. The feature value couple with the most information gain is chosen for the division; it means that at every split, the reduction in entropy due to this divide is maximized. Every internal node (decision node) a test to be done on a sole feature and corresponds to a single of the input features. The outcome of all the likely results of the test of that input feature is signified by the edges to children and represents the path to be tailed. Leaf node represents the value of the target label, given the path from the root to the leaf reflects the values of the input features. A tree can be trained by dividing the basis set into subsets based on testing a sole feature. This process is repeated on each resulting subset in a recursive manner named recursive separating. The recursion is finished when the subgroup of the training patterns at an assumed node has a similar label or when excruciating no longer improves value to the forecasts. Besides, DT is simple to comprehend, perform well with hard data and is simply joint [15-17].

C. Feature Extraction

The feature was extracted from the morphology of CO₂ waveform using a numerical method by utilizing (2). The Area was computed from the segmented region of each breath cycle ranges 4-10 mmHg, 10-15 mmHg, 16-Unitil Maximum EtCO₂, 0.25Sec from EtCO₂, and 10-4 mmHg with a sampling interval of 0.01s.

\[ \text{Area} = \frac{dt}{6} \sum_{j=0}^{i} \left[ C_{j-1}(t) + 4C_j(t) + C_{j+1}(t) \right] \]

where \( dt \) and \( C(t) \) signify the sampling interval and CO₂ signal respectively.

We decided to extract the Area as a feature because this feature is proven to be efficient in differentiating asthma and non-asthma conditions, a study conducted by You et.al. and Singh et.al. in their manual and automatic study. As well as, the feature shows good classification accuracies from asthma and normal CO₂ signal. On the other hand, the investigation was performed with healthy and acute asthmatic patients whereas this study focuses on asthmatic athletes who were in the controlled condition in order to prove the use of a device with an athlete as a device is small and handy.
**Support Vector Machine (SVM)**

The support vector machine utilizes the linear Sequential Minimal Optimization (Simplified SMO) based on Linear Kernel. It employs the *divide and conquer* procedure to solve systematically a large quadratic programming (QP) optimization issue [18]. The advantages are several for a reduction in complexity: decreasing memory consumption and saving time processing. In addition, it has a very exciting implementation that can be used for pedagogical purposes, since the variables of the iterative method can be effortlessly interpreted and accessed in the learning protocol.

**K-Nearest Neighbor (K-NN)**

The K-Nearest Neighbor classifier is also known as the memory-based classifier is presumed one of the most direct and easiest methods for data mining [19]. In this algorithm, Euclidean distances employed to compute the distance between the features in order to work with discrete and continuous features. For instance, if the data is \((p_1, p_2, p_3, ..., p_n)\) and the second data \((q_1, q_2, q_3, ..., q_n)\), then the computation of the distance between the data is as follows:

\[
D(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + ... + (p_n - q_n)^2} \tag{3}
\]

where \(p\) and \(q\) are data to be compared with \(n\) characteristics.

**Performance evaluation**

The performance of the classifiers was assessed using confusion matrix by computing accuracy, sensitivity, specificity, and Area under curve (AUC) [20] using (4), (5), (6) and (7).

**Sensitivity**

\[
\text{Sensitivity} = \frac{TP}{TP + FN} \tag{4}
\]

**Specificity**

\[
\text{Specificity} = \frac{TN}{TN + FP} \tag{5}
\]

**Accuracy**

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FN + FP} \tag{6}
\]

**AUC**

\[
\text{AUC} = \frac{1}{2} \left( \frac{TP}{TP + FN} + \frac{TN}{TN + FP} \right) \tag{7}
\]

where TP, TN, FP, and FN indicate the true positive, true negative, false positive and false negative, respectively.

The signal processing algorithm was performed using Labview (Version 17) and a Notebook Intel (R) Core (TM) i3 CPU, 2 GHz, and OS Windows 10 (64 bit) environment. Besides, the features capability assessment before and after exercise were performed in SPSS (Version 24.0; SPSS Inc., Chicago, IL). Receiver operating characteristics (ROC) curve analysis was performed to verify the significance of features based on the area under curve (AUC) and standard error (SE) value computed using (7) and \(AUC > 0.5\) was presumed statistically significant [21, 22].
III. RESULTS AND DISCUSSIONS
This study explores the significance of feature extracted from the different phases of CO2 waveform morphology from the asthmatic patient during exercise in order to use develop a device with an asthmatic athlete. We considered extracting the Area as a feature as proven to be significant for the classification of asthma and non-asthma conditions. The feature was extracted from each breath cycle that was divided into five different phases. TABLE I. depicts the segmented part of each breath cycle, extracted feature, and performance of the feature.

TABLE I. STATISTICAL SIGNIFICANCE OF DISTINGUISHABLE AUC, STANDARD ERROR (SE) AND 95% CONFIDENCE INTERVAL (CI) OF THE FEATURES USED FOR THE CLASSIFICATION OF THE ASTHMATIC DURING EXERCISE

<table>
<thead>
<tr>
<th>Segmented Part</th>
<th>Feature Index</th>
<th>AUC</th>
<th>SE</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upward expiratory phase</td>
<td>4-10 mmHg</td>
<td>0.712</td>
<td>0.073</td>
<td>0.568 0.855</td>
</tr>
<tr>
<td>Upper-upward Expiratory Phase</td>
<td>11-15 mmHg</td>
<td>0.765</td>
<td>0.069</td>
<td>0.63 0.9</td>
</tr>
<tr>
<td>Upward Expiratory Phase + Alveolar Phase</td>
<td>16-Unitil EtCO2</td>
<td>0.947</td>
<td>0.034</td>
<td>0.875 1</td>
</tr>
<tr>
<td>Alveolar Phase</td>
<td>0.25Sec from EtCO2</td>
<td>0.62</td>
<td>0.091</td>
<td>0.442 0.798</td>
</tr>
<tr>
<td>Inspiratory Phase</td>
<td>10-4mmHg</td>
<td>0.70</td>
<td>0.070</td>
<td>0.517 0.819</td>
</tr>
</tbody>
</table>

The area under curve (AUC) and the p-value were computed for the feature (Area) extracted from the segmented regions. TABLE I lists the AUC, standard error (SE) and 95% confidence interval of AUC values analysis via ROC. The AUC and SE values aid in ranking the features, which assist in enhancing the performance of a classifier. The AUC and corresponding SE values indicate that part of upward expiratory phase and alveolar phase is the strongest in contrast to other segmented regions for the comparison of the asthmatic condition during pre- and post- exercise. The AUC and corresponding SE values indicate that the Area of the combined upward expiratory phase and alveolar phase (16-EtCO2; AUC = 0.947 with SE = 0.04), followed by upward expiratory phase (11-15 mmHg; AUC = 0.765 with SE = 0.07), and lower part of upward expiratory phase (4-10mmHg; AUC = 0.712 with SE = 0.073), are more highly statistically considerable than the rest of the features for the discrimination of asthmatic before and after exercise. Further, the performance of the features was assessed by applying machine learning method.

TABLE II. ILLUSTRATION OF FEATURE PERFORMANCE ANALYSIS USING DECISION TREE (DT), SUPPORT VECTOR MACHINE (SVM), AND K-NEAREST NEIGHBOR (K-NN)

| Features | Classifier | Performance analysis (Area) |
|----------|------------|-------------------------------|------------------|------------------|
| A1, A2, A3, A4, A5 | DT | 88.89 | 100 | 77.78 | 0.85 |
| SVM | 72.22 | 88.89 | 55.56 | 0.72 |
| K-NN | 77.77 | 88.89 | 66.67 | 0.75 |
| A1, A2, A3 | DT | 88.89 | 100 | 77.78 | 0.85 |
| SVM | 83.33 | 100 | 66.67 | 0.83 |
| K-NN | 77.78 | 88.89 | 66.67 | 0.75 |
| A3 | DT | 88.89 | 100 | 77.78 | 0.85 |
| SVM | 83.33 | 100 | 66.67 | 0.83 |
| K-NN | 83.33 | 77.78 | 88.89 | 0.85 |

The features were combined with a set of all, sets of three that has maximum AUC values, and sole features and fed to the classifier as a features vector. The set of features Upward expiratory phase (A1), upper-upward expiratory phase (A2), a combination of upward expiratory phase and alveolar phase (A3), alveolar phase (A4), and Inspiratory Phase (A5) were given to classifier via 10-fold cross-validation. Table 2 shows the performance analysis of the features for the classification of the asthmatic condition before and after exercise. It can be noticed that the DT classifier possesses higher accuracy (88.89%), sensitivity (100%) and specificity (77.78%) with an AUC of 0.85 than SVM and K-NN. In addition, DT was found to be consistent with all sets of a feature.

Hence, incorporation of the proposed feature into real- time device may possibly allow the management of asthma outside of hospital and during sports activity as well for the sportsperson asthmatic. In addition, it has been challenging for the sports trainers and athletes to identify the symptoms of asthma [2]. Further, according to a study [23], both a personalize asthma management device and reliever are needed for athlete asthmatic.
Therefore, inclusion of this feature into lately developed device [12] will allow the athlete asthmatic to manage their asthmatic condition by taking their medication at correct time.

IV. CONCLUSION AND FUTURE WORK

This study hypothesizes to test the *Area* as a feature before and after the exercise in order to use the newly developed handheld device to use with an athlete asthmatic. Herein, we segmented each breath cycle into five sub-cycles. The feature (*Area*) was extracted from each segmented part and area under curve was estimated. The AUC analysis reveals that the feature (*Area*) of the segmented part shows good AUC $>0.5$ however, a mix of upward expiratory phase and alveolar phase possess higher AUC of 0.94. Thereafter, set of features were combined and fed to the classifiers into three parts. Here, decision tree classifier has shown the higher accuracy (88.89%), sensitivity (100%) and specificity (77.78%) with an AUC of 0.85 than other classifiers. Thus, this preliminarily study shows the promising result and incorporation of this feature into newly developed may possibly allow the asthmatic athlete to manage asthmatic conditions during sports activities. Further, the device will be tested with a greater number of samples in order to generalize the finding of this research.

ACKNOWLEDGMENT

The authors gratefully acknowledge Universiti Teknologi Malaysia (UTM) for supporting and offering the required facilities and laboratory equipment. This study was conducted as a part of the Fundamental Research Grant Scheme by Ministry of Higher Education Malaysia, under Grant no.R.J130000.7745.4F943.

REFERENCES


