Severity Classification of Chronic Obstructive Pulmonary Disease and Asthma with Heart Rate and SpO_2 Sensors*

Tasnuba Siddiqui and Bashir I. Morshed, *Member, IEEE, EMBS*

***Abstract*— Asthma and Chronic Obstructive Pulmonary Disease are chronic and long-term lung diseases. Disease monitoring with minimal sensors with high efficacy can make the disease control simple and practical for patients. We propose a model for the severity assessment of the diseases through wearables and compatible with mobile health applications, using only heart rate and SpO_2 (from pulse oximeter sensor). Patient data were obtained from the MIMIC-III Waveform Database Matched Subset. The dataset consists of 158 subjects. Both heart rate signal and SpO_2 data of patients are analyzed via the proposed algorithm to classify the severity of the diseases. Strategically, a rule-based threshold approach in real time evaluation is considered for the categorization scheme. Furthermore, a method is proposed to assess severity as an Event of Interest (EOI) from the computed metrics in retrospective. This type of autonomous system for real-time evaluation of patient's condition has the potential to improve individual health through continual monitoring and self-management, as well as improve the health status of the overall Smart and Connected Community (SCC).**

I. Introduction

Asthma and Chronic Obstructive Pulmonary Disease (COPD) are characterized by respiratory symptoms due to airflow limitations and alveolar abnormalities in the presence of significant exposure to gas or noxious particles. COPD is currently the fourth most prevalent cause of death in the world and it is projected to be the third leading cause of death by 2020 [1]. Asthma and COPD are leading causes of morbidity and mortality, including economic and social burden. In the USA and European Union, COPD related management incurs from 38.6 billion Euros and 32 billion dollars loss, per year respectively [1-2]. The European Respiratory Society and American Thoracic Society Task Force consider asthma severity into mild, moderate, and severe groups excluding uncontrolled and refractory asthma based on pulmonary function tests along with treatments [2].

Sanchez-Morillo *et al.* suggested the scope for early detection of the severity of both COPD and asthma by different learning methods based on clinical measurements from various sensors [3]. According to the Global Initiative for Chronic Obstructive Disease (GOLD) guidelines, the severity of the patients with COPD is labeled in four classes according to the Forced Expiratory volume in 1 second (FEV1) values [1]. A decision support system with threshold has been adapted to detect the exacerbation with heart rate and oxygen saturation by a group of researchers [4], although their accuracy is high for predicting exacerbation, but they did not classify the severities. Heart rate and SpO_2 values are independent and no significant correlation was found during stable periods and severe periods COPD exacerbation [5]. The COPD and Asthma populations' distributions are compared with K-means clustering to set up correct model representations [6].

A smart and connected community (SCC) aims to use embedded sensors and computing for greater benefit to individuals, the community, and society with interlocking physical, social, behavioral, economic, and infrastructural challenges. Ability to detect and monitor progression of COPD and Asthma using physiological data such as heart rate and SpO_2 from wearables can lead to better self-management and improved overall health of SSC.

In this work, the primary focus is to classify the disease from normal to severe with the help of a model-based algorithm by analyzing time series signals captured using minimal sensors. In this scenario, Events of Interest (EOI) is the severities of asthma and COPD that are computed first individually, then combined, by using the peripheral oxygen saturation and heart rate signals captured from the pulse oximeter. Then, based on the EOI score, five distinct stages are categorized according to the GOLD guidelines.

II. Datasets

All the datasets used in this study are from the publicly available database from PhysioNet [7]. The total number of patients is 158, where 85 have COPD, 67 have Asthma, and 6 are healthy subjects as the control group in retrospect.

For asthma datasets, 5 subjects are taken from the MIMIC-II Waveform Database and the other 62 subjects from the closely related, recently released MIMIC-III Waveform Database Matched Subset. The subjects selected from the MIMIC-III are according to the International Classification of Diseases (ICD9) code 493 and only labeled 1 for the sequence number or the order of treatment. For COPD, all 85 subjects are from MIMIC-III Matched Waveform Dataset, according to their ICD9 code 496, and treatment order labeled as 1 or 2.

Healthy control groups are available from two separate databases; 5 subjects are from the CAP sleep database and one healthy subject is from ECG sleep apnea database. All the databases used here for the data collection are labeled as Class 3, standing for other contributed collections of data, including works in progress.

---

* This material is based upon work supported by the National Science Foundation under Grant No. 1637250.

Tasnuba Siddiqui and Bashir I. Morshed are with Electrical and Computer Engineering Department, University of Memphis, Memphis, TN 38152 USA. (Corresponding author - phone: 901-678-3650; e-mail: bmorshed@memphis.edu).

III. Methods

A. Data Collection

The heart rate and oxygen saturation values processed from the photoplethysmogram signal are captured by the pulse oximeter. Pulse oximeters are noninvasive portable medical devices and easily adaptable both in normal and hospital lives. In our case, we have used the dataset with one sample per minute collected continuously. In case of different sampling frequencies removing samples (under sampling) performed to match the samples per minutes.

B. Data Processing

Missing values are common occurrences in datasets. In this study, the mean is imputed by a continuous linear interpolation of neighboring, non-missing values. In Fig. 2, the raw signal and processed signal for both heart rate and oxygen saturation are shown. For heart rate signal, values outside 40 and 200 bpm are taken as outliers, as well as for oxygen saturation, values outside 20 and 100 are disregarded from consideration.

C. Feature Extractions

The mean and the standard deviation of heart rate and oxygen saturation are the features in the study of decision support system [4]. In this study for the proposed rule, features extracted from heart rate and SPO_2 are the percentages of the total duration of labeled minutes. To extract the features, the signals are divided into five zones based on threshold values like GOLD stages [1]. For heart rate, the classes are NOR (Normal) for less than 90, MIL (Mild) for 90-100, MOD (Moderate) for 100 – 110, SEV (Severe) for 110 – 120 and VES (Very Severe) for greater than 120 beats per minutes [8]. For peripheral oxygen saturation value, greater than 92% is considered as NOR, for stage MIL it is 90% - 92%, and for stage VES less than 80%. The stages MOD and SEV are differentiated by the clinicians based on the presence of only room air or external O_2 supply [8]. In this current study, stages MOD and SEV are taken as between 85% - 90% and 80% - 85% respectively. However, in GINA guideline asthma exacerbations management considers heart rate of 100 -120 bpm and SpO2 of 90 -95% mild or moderate asthma and heart rate greater than 120 bpm, and SpO_2 less than 90% as severe asthma.

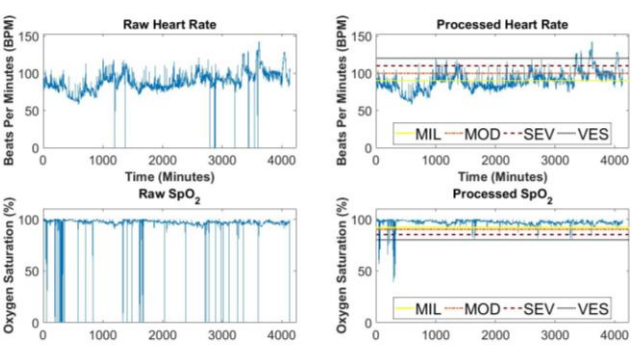

Figure 2: Raw and processed data for Heart Rate and SpO_2.

The processed signal on the right-hand side in Fig. 2, both heart rate and oxygen saturation signals are separated into stages as NOR, MIL, MOD, SEV, and VES minutes. Then, the minutes in each stage are summed and scaled in percentage compared to the total number of minutes. In this way, finally from heart rate and SpO_2, the percentage of minutes of five stages gives total ten extracted features.

D. EOI Calculation

The severity is computed as Events of Interest (EOI) [9] from the extracted features from the ten different percentages of minutes HR_i and SpO_{2i}. EOI_{HR} and EOI_{SpO2} values are calculated with α_i.

$$
\begin{gathered}
EOI_{HR} = \sum_{i=0}^{4} \alpha_i * HR_i \\
EOI_{SpO2} = \sum_{i=0}^{4} \alpha_i * SpO_{2_i} \\
EOI^{C} = \beta_{HR}^{C} * EOI_{HR} + \beta_{SpO2}^{C} * EOI_{SpO2} \\
EOI_N^C = \frac{EOI^C - E_{MIN}^C}{E_{MAX}^C - E_{MIN}^C} \\
EOI^{A} = \beta_{HR}^{A} * EOI_{HR} + \beta_{SpO2}^{A} * EOI_{SpO2} \\
EOI_N^A = \frac{EOI^A - E_{MIN}^A}{E_{MAX}^A - E_{MIN}^A}
\end{gathered} \tag{1}
$$

α_i = Weightage for Heart Rate (HR) or SpO_2 in stage *i*.
HR_i = Percentage of minutes for Heart Rate in stage *i*.
SpO_{2i} = Percentage of minutes for SpO_2 in stage *i*.
β_{HR} = Weightage for Heart Rate.
β_{SpO2} = Weightage for SpO_2.
E_{MIN} = the smallest value for variable EOI.
E_{MAX} = the largest value for variable EOI.

TABLE I. Mathematical functions for α Model

Models	Description	α Values
Cubic (CM)	$2^3, 3^3, 4^3, 5^3, 6^3$	[1.82, 6.14, 14.55, 28.41 49.1]
Tangent (TM)	tan(1°), tan(22.5°), tan(45°), tan(67.5°), tan(89°)	[0.03, 0.68, 1.64, 3.95, 93.71]

E. Model Selection

In the model-based approach, simple linear functions are explored to estimate the values of hyperparameter α_i by models in Table I. Different α values from the table are different weights for the distinct stages and making the mathematical model for calculating EOI. For stage NOR, it is expected to have low weight, for stages MIL, MOD and SEV the weights are gradually increased, and for stage VES the weight is expected to have the greatest value.

The cubic model (CM) has steady slope characteristics and all the severities except stage VES have greater values in a cubic model with comparison to the other tangent model (TM). Among the two models, the sharpest rising weight is the tangent model (TM) and this model gives the largest value to the stage VES compared to the other model (CM).

The hyper parameter β is expressed as a ratio of the weight of the heart rate to the oxygen saturation to calculate EOI. In the literature [4], Merone et. al., explored the ratio of heart rate and SpO_2 from 50:50 to 5:100, exploring equal weights to continuously increased weights to SpO_2. In this study, similarly, two β ratios are explored, Heart Rate: SpO_2 as 50:50 (β_E) and 40:60 (β_U).

F. Classification

Based on the EOI for COPD, Asthma, and healthy control groups, the stages are linearly separated into five separate groups. For COPD according to GOLD [1], the stages are mild, moderate, severe, and very severe. For asthma according to the GINA guidelines [2], the stages are mild intermittent, mild persistent, moderate persistent and severe persistent. For stage NOR classification, the EOI threshold for healthy to patient EOI (0-1) is taken. Then for MIL: EOI (2-9), MOD: EOI (10-19), SEV: EOI (20-49), VES: EOI (50-100) are used to calculate the distinct severities. This distribution gives highest agreement [6].

G. Severity Result Performance Evaluation

Some of the known and often used results are accuracies, sensitivity, specificity, harmonic mean (F1 Score), precision, positive predictive value, and negative predictive value. Based on the performances models are selected for calculating the severity of asthma and COPD.

IV. RESULTS

COPD, asthma, and healthy control groups showed the different percentage of time duration at various severity level for both signals. In Fig. 3, a representative distribution of the percentage of the duration for heart rate and SpO_2 signal are shown for subjects of the three groups. For COPD patients, heart rate signal is in stage NOR for 47% of the total duration. The other stages of heart rate are 23%, 22%, 4%, and 3% from low to high severe minutes. Similarly, SpO_2 of the COPD subject is in the stage NOR for 61% of the total duration spent. For the asthma subject, the percentage of minutes spent in each stage is 4%, 32%, 26%, 18% and 20 % gradually for heart rate. Similarly, SpO_2 percentage of minutes are 72% in stage 0, 9% in stage 1, 7% in stage 2, 3% in stage 3 and 8% in stage 4. For the healthy subject, heart rate stage NOR percentage of minutes is 85% and stage MIL is 15%. The healthy subject spent total time in SpO_2 stage NOR.

Table III delineates the total mean and standard error values for heart rate and SpO_2 among the COPD and asthma subjects. The standard error shows large values and a wide distribution of COPD and asthma subjects.

The final four different models consisted of combinations of two different α and β values are compared in the Fig. 4 to show the score distribution of the models by their relationships with EOI values. The vertical lines stand for the error bar. For the total samples, Fig. 4 depicts the mean and the standard error for all the subjects in the datasets. By only taking equal ratios (Heart Rate: SpO_2-50:50, β_E), asthma subjects showed greater EOI in compared to unequal ratios (Heart Rate: SpO_2-40:60, β_U) considering the same models. The overall EOI for asthma and COPD mean for same β values the cubic model (α_{CM}) showed greater mean compared to tangent model (α_{TM}).

The subjects from Fig. 3 for COPD, asthma, and healthy are shown with their calculated EOI values in Fig. 5. As the asthma subject spends severe heart rate minutes, so the EOI value reflects the high score compared to the COPD subject. The healthy subjects have the least EOI values in the figures. For the four models, the calculated EOI values are also plotted to compare their scores.

The main aim of this study is to make models to classify different severities for the diseases, but the severities of the subjects are not in the datasets. So, the performances of the models are reported based on calculating the accuracy, sensitivity, and area under the curve on the capability to distinguish the subjects from healthy to patients. A default EOI threshold scores (0-1) is set to avoid the overfitting. In Table IV, the four models are compared according to their performances. The best accuracy for COPD is 76%, and for asthma is 60%. The sensitivity is 45% for the model α_{TM} for COPD, for asthma 78%. Specificity is 100% for both diseases.

TABLE II. HEART RATE AND SpO_2 OF THE STUDIED GROUP

	SUBJECTS	NOR	MIL	MOD	SEV	VES
SpO_2 (%)	COPD	86.95± 18.29	4.45± 4.67	3.14± 4.81	1.13± 2.25	4.33± 13.29
	ASTHMA	89.88± 15.92	4.58± 7.14	2.82± 6.15	1.12± 3.18	1.61± 5.75
	HEALTHY	100	0	0	0	0
HR (BPM)	COPD	72.59± 30.89	15.22± 16.98	7.58± 11.62	2.65± 6.16	1.96± 6.91
	ASTHMA	48.47± 35.14	17.86± 16.04	16.48± 17.9	9.98± 15.52	7.21± 14.35
	HEALTHY	97.5± 6.12	2.5± 6.12	0	0	0

Values are presented as a mean ± standard deviation of the percentage of minutes.

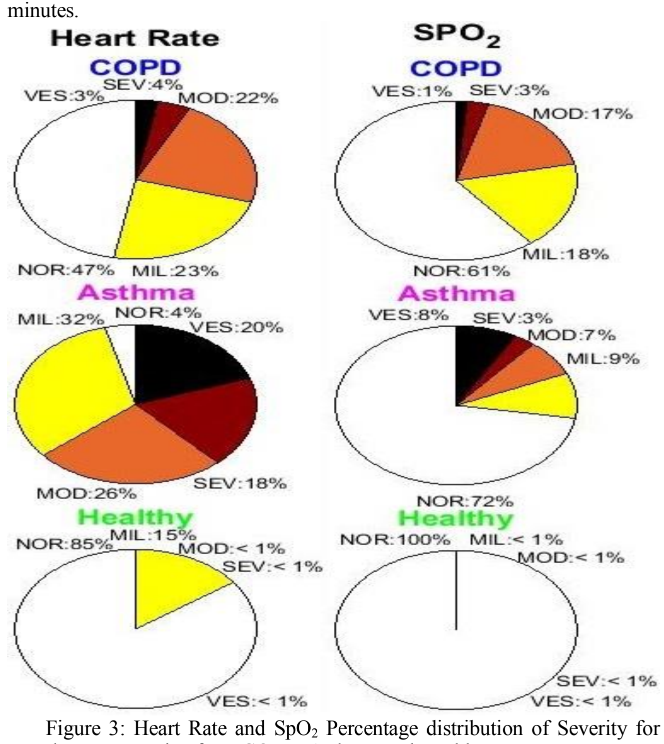

Figure 3: Heart Rate and SpO_2 Percentage distribution of Severity for sample representative from COPD, Asthma, and Healthy.

For both COPD and asthma, the tangent model (α_{TM}) have better accuracy, but the cubic models (α_{CM}) have better sensitivity results. For asthma, only the $\alpha_{TM}\beta_E$ performs well having maximum accuracy 60% and 78% sensitivity. Upon considering the accuracy results for COPD $\alpha_{TM}\beta_U$ (maximum accuracy 76%) and for asthma $\alpha_{TM}\beta_E$ are selected as the best models. The severities based on the generated EOI score are discretized into five stages. The distribution of the asthma subjects and COPD subjects with the healthy control subjects are in Table IV. The model exploits the observed data set variable dependencies as opposed to the fact of assumptions.

In this early stage work, the algorithm is developed by using ICU patients' information. However, this algorithm can be applicable for subjects beyond clinical settings. In future implementation, further evaluations of these settings need to be investigated. Furthermore, inclusion of higher number of healthy subjects and patients will further substantiate analysis in future.

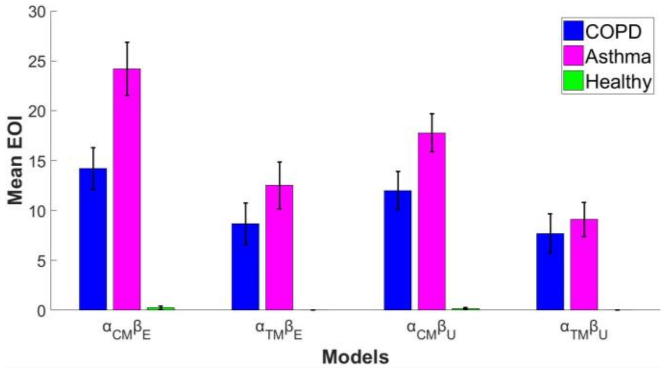

Figure 4. EOI Mean and error bar plots of four selected models for COPD, Asthma, and Healthy.

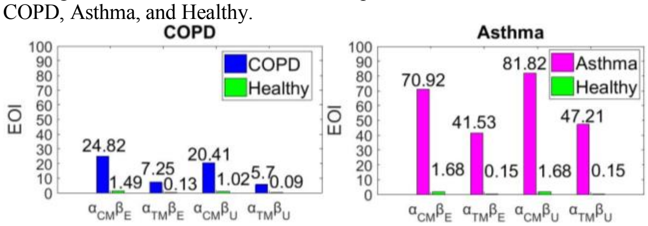

Figure 5. EOI validation plots for sample representative from COPD, Asthma, and Healthy.

TABLE III. THE PERFORMANCE METRICS

SUBJECT	MODEL	PERFORMANCE METRICS (%)						
		ACC	SEN	SPE	F1	PRE	PPV	NPV
COPD	$\alpha_{CM}\beta_E$	60	79	100	88	100	56	80
	$\alpha_{TM}\beta_E$	74	49	100	66	100	67	89
	$\alpha_{CM}\beta_U$	61	76	100	87	100	57	81
	$\alpha_{TM}\beta_U$	76	45	100	62	100	69	90
ASTHMA	$\alpha_{CM}\beta_E$	52	94	83	96	98	51	60
	$\alpha_{TM}\beta_E$	60	78	100	87	100	56	78
	$\alpha_{CM}\beta_U$	51	96	83	97	98	51	57
	$\alpha_{TM}\beta_U$	60	77	100	87	100	56	78

TABLE IV. RESULT OF FRQUENCY DISTRIBUTION

SUBJECT	MODEL	NOR	MIL	MOD	SEV	VES
COPD	$\alpha_{CM}\beta_E$	18	31	25	6	5
	$\alpha_{TM}\beta_E$	43	37	7	4	4
	$\alpha_{CM}\beta_U$	20	35	22	3	5
	$\alpha_{TM}\beta_U$	47	25	5	4	4
HEALTHY	$\alpha_{CM}\beta_E$	6	0	0	0	0
	$\alpha_{TM}\beta_E$	6	0	0	0	0
	$\alpha_{CM}\beta_U$	**6**	0	0	0	0
	$\alpha_{TM}\beta_U$	6	0	0	0	0
ASTHMA	$\alpha_{CM}\beta_E$	4	15	25	13	10
	$\alpha_{TM}\beta_E$	15	28	14	6	4
	$\alpha_{CM}\beta_U$	3	15	23	15	11
	$\alpha_{TM}\beta_U$	15	27	14	6	4
HEALTHY	$\alpha_{CM}\beta_E$	5	1	0	0	0
	$\alpha_{TM}\beta_E$	6	0	0	0	0
	$\alpha_{CM}\beta_U$	5	1	0	0	0
	$\alpha_{TM}\beta_U$	6	0	0	0	0

V. DISCUSSION

The proposed algorithm uses retrospective data from the publicly available databases and contribute to the advancement in the home management system for the patient well-being. In the absence of the annotation, in accordance to literature, we followed established clinical criteria and confirmed verifying by using k means clustering as an alternative method [6].

VI. CONCLUSION

Asthma and COPD are prevalent mostly among older adults and smokers. A severity classifier implementation will help to check the level of sickness for the patients suffering from the diseases. In this study, a simple approach is proposed based on the duration of the time passed for each vital sign. Accordingly, the proposed system of calculating severity as an Event of Interest (EOI) consists of a scaled value to inform the patient about the severity based on the score generated by the system. We have proposed flexible and reliable features and algorithms for both asthma and COPD to improve patient wellbeing.

The proposed solution can provide immediate feedback to the patient for effective self-management of diseases, which along with sharing of EOI with other community members can lead to an improvement of overall community SCC health.

REFERENCES

[1] *"Global Strategy for the Prevention, Diagnosis, and Management of COPD," GOLD 2018*. [Online]. Available: [http://goldcopd.org/gold-reports/]

[2] *"Global Strategy for Asthma Management and Prevention," 2018 GINA Report*. [Online]. Available: [http://ginasthma.org/2018-gina-report-global-strategy-for-asthma-management-and-prevention/]

[3] Sanchez-Morillo, Daniel, Miguel A. Fernandez-Granero, and Antonio Leon-Jimenez. "Use of Predictive Algorithms in-Home Monitoring of Chronic Obstructive Pulmonary Disease and Asthma A Systematic Review." *Chronic respiratory disease* : 1479972316642365, 2016.

[4] Merone, M., et al. "A Decision Support System for Tele-Monitoring COPD-Related Worrisome Events." *IEEE journal of biomedical and health informatics*, 2017.

[5] J. R. Hurst et al., "Domiciliary pulse-oximetry at exacerbation of chronic obstructive pulmonary disease: prospective pilot study", *BMC pulmonary medicine*, 2010.

[6] Siddiqui, Tasnuba and Bashir I Morshed. "Severity Exploratory Model Analysis of Chronic Obstructive Pulmonary Disease and Asthma with Heart Rate and SpO_2 Sensors " *Electro/Information Technology (EIT), 2018 IEEE International Conference on*. IEEE, 2018 (accepted).

[7] Goldberger AL, et al. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *Circulation* **101**(23):e215-e220 [Circulation Electronic Pages; http://circ.ahajournals.org/cgi/content/full/101/23/e215]; 2000.

[8] C. C. Bellos, et al. "Identification of COPD Patients Health Status using an Intelligent System in the CHRONIOUS Wearable Platform."*IEEE Journal of Biomedical and Health Informatics* 18.3: 731-8, 2014.

[9] Morshed, Bashir I., et al. "Inkjet Printed Fully-Passive Body-Worn Wireless Sensors for Smart and Connected Community (SCC)." *Journal of Low Power Electronics and Applications* 7.4 (2017): 26.