Severity Exploratory Model Analysis of Chronic Obstructive Pulmonary Disease and Asthma with Heart Rate and SpO_2

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

***Abstract*—Managing Asthma and chronic obstructive pulmonary disease (COPD) at home have immense potential to progress in healthcare. At home measurements of severities of the chronic diseases by telehealth technologies enables effective self-management and can reduce sudden deterioration. The number of COPD, asthma, and healthy subjects collected from MIMIC III databases are 168. Time series signals and principal components are explored as predictive features from the heart rate and SpO_2 collected via pulse oximeter. The goal of the work is to develop models, then to select the best predictive model and to compare the severity groups with clustering of the diseases. Severity estimation algorithms are built with mathematical functions. Finally, four hyper-parameter models and K Means Clustering algorithms are implemented for the distributions of severities of the diseases. The COPD and asthma severity models are capable to classify the subjects according to the GOLD and GINA severity stages respectively. The proposed solution can provide immediate feedback to the patient for self-awareness of the disease and sharing of the severity score with other community members can lead to an improvement of overall community health towards a Smart and Connected Community (SCC).**

I. Introduction

Asthma and chronic obstructive pulmonary diseases (COPD) need continuous attention and care to control the diseases. With the advent of telemonitoring, real-time data used for the assessment of the condition of the diseases, the patient can check their disease without wasting time in an unnecessary and regular visit to the hospital. Hence, continuous real-time monitoring of the disease condition might aid the patient to improve their health, as well as improve the health status of the Smart and Connected Community (SCC).

Sanchez-Morillo *et al.* describes the scenario in the review of both asthma and COPD in the telehealth literature [1]. In the study conducted by Christos C. Bellos *et al.* [2] several sensors have been used for various parameters such as heart rate, systolic and diastolic pressure, oxygen saturation (SpO_2), respiratory frequency, inspiratory and expiratory time, and tidal volume and they have presented a hybrid classification system with Support Vector Machine and Random Forest. The severity of COPD was classified according to both clinician's labeling and Global Initiative for Chronic Obstructive Disease (GOLD) guidelines. Alternatively, a decision support system with threshold has been adapted to detect the exacerbation with heart rate and oxygen saturation by a group of researchers [3]. Another study focused on the risk of exacerbation results by the Classification and Regression Tree (CART) classifier, by using a few sensor measurements: Forced Expiratory Volume in One Second (FEV1), respiratory rate, oxygen saturation, heart rate, body weight, and body temperature [4]. Heart rate and SpO_2 values are independent and no significant correlation was found during stable periods and periods with severe COPD exacerbation [5]. Established combinations of models for COPD and Asthma, and measurement of the classification performances for different models and parameters are assessed [6].

In this study, modeling for the COPD and asthma severity traits and machine learning algorithms are analyzed and compared to assess their severity group distributions. In clinical practices, physiological features like heart rate and SpO_2 are used to measure the exacerbation of both diseases. The proposed method has the potential for real-time monitoring of asthma and COPD at home with wearables.

II. Methods

The method for severity calculation is divided into the six following steps in Figure 1. Initially, the data is to be collected from the wearable sensors. For the data analysis discussed here, heart rate and SpO_2 signals are collected from MIMIC II, MIMIC III, and CAP sleep datasets, from openly available PhysioNet database [7] and Sleep Heart Health datasets. The total number of COPD, asthma and healthy subjects are 85, 67 and 16 respectively.

Most of the heart rate and SpO_2 signals were processed over one-minute time duration. In few cases where the sample frequency was not one minute, the mean value over one-minute time duration are taken into consideration. Missing values of the signals are prevalent in the datasets. Christos C. Bellos *et al.* investigated list wise detection, mean/mode imputation, K-nearest neighbor estimation to take care of the missing values [2].

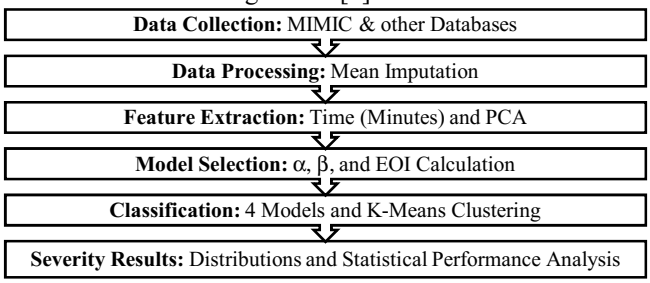

Figure 1: Working Flowchart for Asthma and COPD level of severity detection.

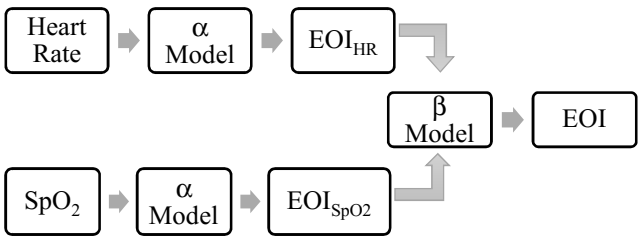

Figure 2: Modeling of Asthma and COPD for severity detection.

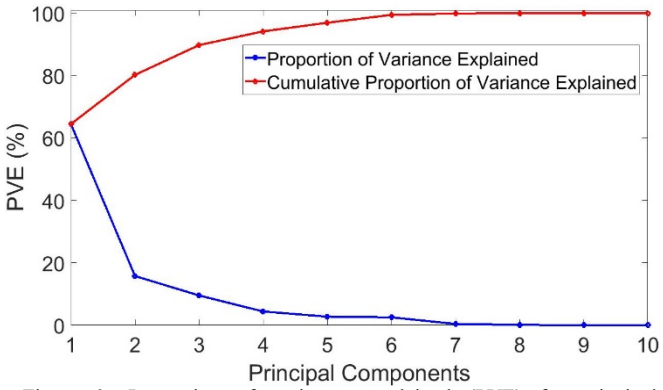

Figure 3: Proportion of variances explained (PVE) for principal components.

According to Global Initiative for Chronic Obstructive Disease (GOLD), the groups are mild, moderate, severe, and very severe for COPD. According to the Global Initiative for Asthma (GINA) guidelines, the groups are mild intermittent, mild persistent, moderate persistent and severe persistent for asthma. During data processing, mean is imputed from neighboring non-missing values by linear interpolations. After that, the heart rate signals are split into five stages based on heart rate (NOR: <90, MIL: 90-100, MOD: 100-110, SEV: 110-120, and VES: >120) BPM. Similarly, SpO_2 signals are split into five stages based on SpO_2 (NOR: >92%, MIL: 90-92%, MOD: 85-90%, SEV: 80-85%, and VES: <80%) [2]. Finally, for each individual subject, the time spent in each stage per signal is calculated. In this way, total ten features are taken from heart rate and SpO_2. For the modeling, the relationship of all the ten features with hyperparameters α and β are shown in Figure 2. After that, Events of Interests (EOI) [8] are scaled over the combined EOIs of heart rate (EOI_{HR}) and SpO_2 (EOI_{SpO2}) for each of the disease individually.

After principal component analysis (PCA), the principal components are taken from all the ten features of both the signals. In Fig. 3, the Proportion of Variances Explained (PVE) captured after each principal component are sketched. From the cumulative PVE, more than 80% variance of all the 10 features is within the first 2 components. Then, 5 clusters are formed by the K-Means clustering for the five severities. Finally, the selected model and K-Means Clustering classification algorithms' results are compared to see the similarity of the severities between them.

III. Modeling

During modeling, two hyperparameters α and β are explored to represent the models [6]. For the α modeling, Severity NOR is expected to have the smallest values and all the other severities should have significant weight increased proportional to their intensity.

In Table I, total seven mathematical functions and their corresponding weights for the distinct severities are tabulated. In the Fig. 4 the percentage weights after scaling are plotted. Finally, as the tangent model stands for the sharpest rise and cubic model stands for the most gradual rise, both models are selected for comparison.

β is the ratio of the weights for heart rate and SpO_2. Heart rate and SpO_2 ratios are designated 50: 50 as β_E, 40: 60 as β_U, 20: 80 as β_O and they are compared in Figure 5. COPD is in the 1st row and Asthma is in the 2nd row. The lines are the fitted curve with 1 degree of freedom. For each subject, x-axis represents the calculated EOI score for individual model, while y-axis depicts the corresponding percentage of minutes in each severity. The lines are showing the relationship between the EOI score and the features or the percentage of minutes. The columns are stage NOR and Stage VES to depict their relationships with the β values. The rest of the stages MIL, MOD and SEV, are similar for all β. So, β is selected from the two extreme severities.

The relationship of EOI with severity NOR and severity VES are expected to be inversely proportional and proportional respectively. But for β_O the expected relationships are not found. So, for final model selections β choice is limited to β_E and β_U.

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]
Exponential (EM)	e^2, e^3, e^4, e^5, e^6	[1.17,3.17,8.61, 23.41, 63.64]
Square root (SM)	$\sqrt{0.01}, \sqrt{0.1}, \sqrt{1}, \sqrt{10}, \sqrt{100}$	[0.69,2.17,6.86, 21.69, 68.59]
Four Power (FM)	$4^2, 4^3, 4^4, 4^5, 4^6$	[0.29,1.17,4.69, 18.77, 75.07]
Power Ten (PM)	$2^{10}, 3^{10}, 4^{10}, 5^{10}, 6^{10}$	[0.02,0.08,1.47, 13.69, 84.76]
Inverse Tenth (IM)	0.01, 0.1, 1, 10, 100	[0.009, 0.09, 0.9, 9, 90]
Tangent (TM)	tan (1°), tan (22.5°), tan (45°), tan (67.5°), tan (89°)	[0.03, 0.68, 1.64, 3.95, 93.71]

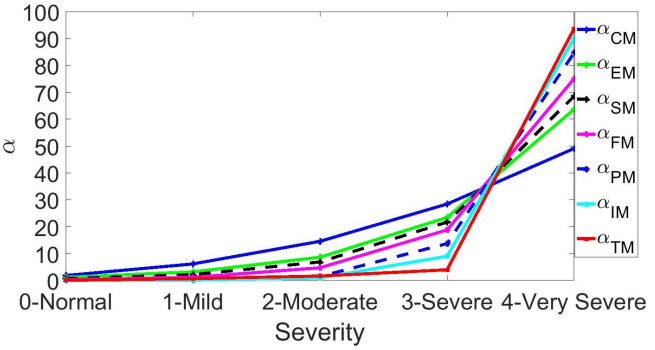

Figure 4. The α plot as Percentage and for various levels of severities.

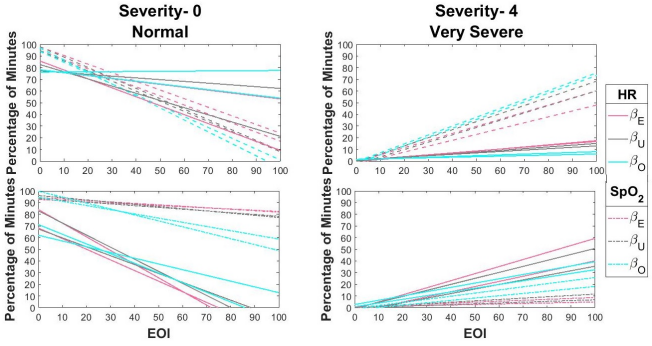

Figure 5. The β plot for various levels of severities.

IV. Results

COPD, asthma, and healthy control groups show different percentages of time duration at various severity stages for

both signals. In Fig. 6 the median distribution of the percentage of total minutes compared among the three groups are shown along with the 25th and 75th percentiles for SpO_2 and heart rate respectively. In case of SpO_2, the median value of the all the severities time duration for COPD is like asthma. The 1st quartile and 3rd quartile are very close to the median values for both diseases.

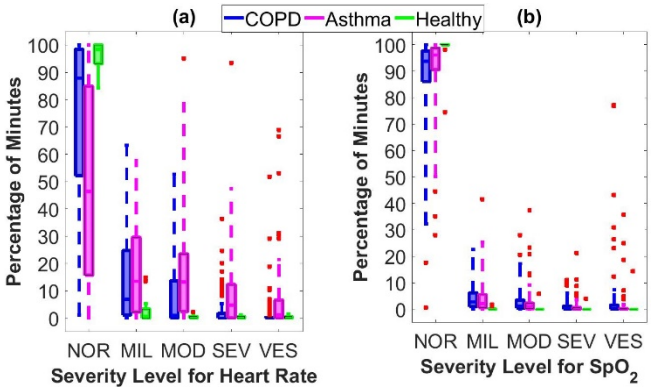

Figure 6. Box plot for Percentage of minutes for distinct levels of a) Heart Rate b) Peripheral Oxygen Saturation (SpO_2).

Considering heart rate, in Fig. 6 shows median of stage NOR is more than 80% for COPD, on the other hand for asthma the median is more than 40%. So, the asthma subject group tend to spend most of the times in stage MIL to SEV. However, the SpO_2 stage NOR is 100% and severity MIL to VES are almost null for the control group.

The 1st row stands for COPD and the 2nd row for Asthma in Fig. 7. The columns are for different severities. The $\alpha_{CM}\beta_E$ has the sharpest inverse ratio Stage NOR with the increasing EOI value in case of heart rate. All the models give a similar relationship with EOI for the VES. So, $\alpha_{CM}\beta_E$ performs best considering only heart rate. But, $\alpha_{CM}\beta_E$ does not fit SpO_2 in any way. So, β_U is chosen for COPD models.

SpO_2 at all the stages vary highly with increasing EOI, as a result, give best results with $\alpha_{TM}\beta_U$ for COPD. But for the other stages, $\alpha_{CM}\beta_U$ performs the best. Considering higher priorities to the VES, $\alpha_{TM}\beta_U$ is finally selected for severity classification for COPD. β dominates the decision a lot for asthma. According to the observations, Severity MIL to VES are important and the best performer among the four is considered to represent the characters of α and β well.

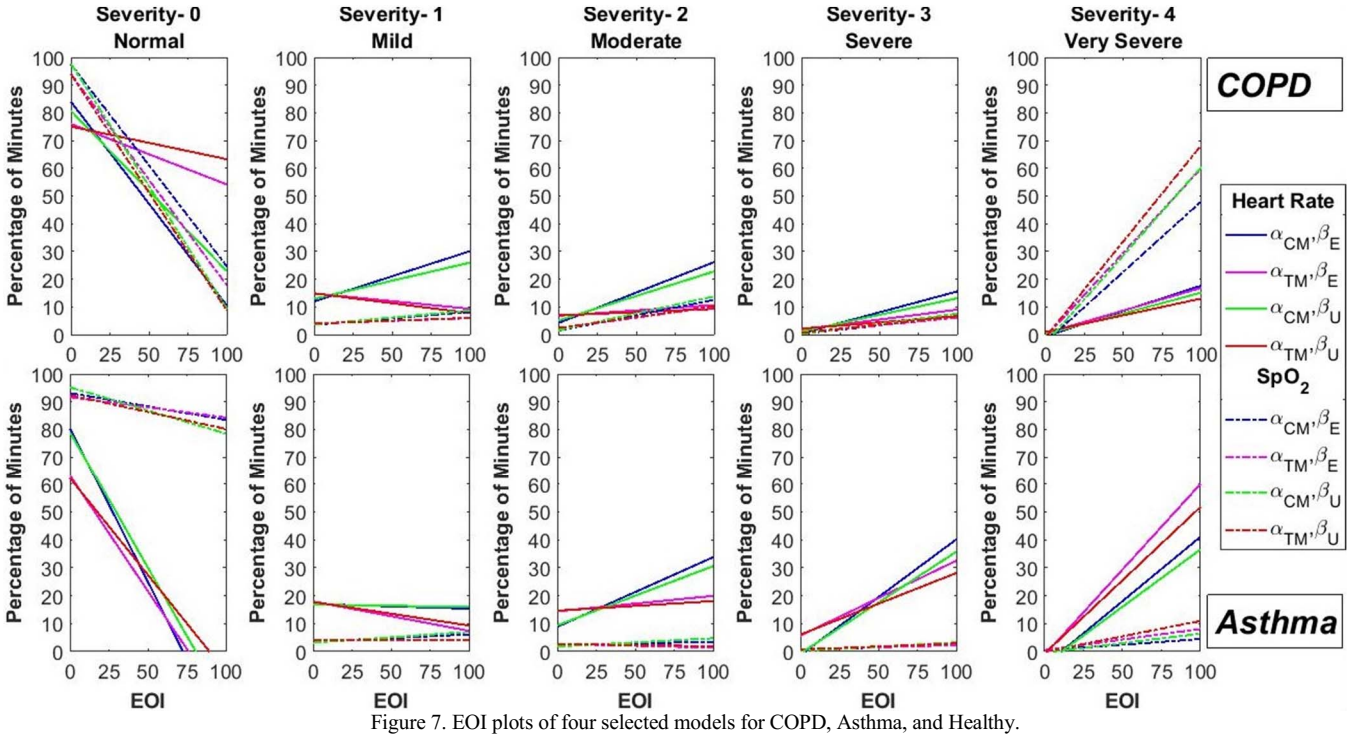

Figure 7. EOI plots of four selected models for COPD, Asthma, and Healthy.

The cubic model $\alpha_{CM}\beta_E$ and $\alpha_{CM}\beta_U$ shows a directly proportional relationship while the $\alpha_{TM}\beta_E$ and $\alpha_{TM}\beta_U$ show the inversely proportional relationship in case of heart rate for the stage MIL and MOD. So, for heart rate, the cubic model performance is good for Stage NOR and SEV. But, with $\alpha_{TM}\beta_E$ VES varies highly with increasing EOI, as a result, gives best results. Also, in case of the SpO_2, the $\alpha_{TM}\beta_E$ performs well for the VES. So, $\alpha_{TM}\beta_E$ is selected for Asthma.

A default EOI threshold scores (0-1) is set to avoid the overfitting, the other stages are separated as 10, 30 and 50 based on the generated EOI score. The distribution of the asthma subjects and COPD subjects with the healthy control subjects are in the Fig. 8. The stages are separated by the vertical lines.

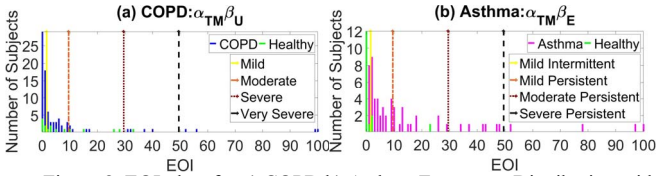

Figure 8. EOI plots for a) COPD b) Asthma Frequency Distribution with Healthy.

According to the different severity separation, the number of subjects is shown in Table 2 and 3. In the Asthma distribution, the number of subjects by the $\alpha_{TM}\beta_E$ Model in Severity NOR to VES is 15, 28, 14, 6 and 4 respectively. For COPD, the distribution for the stages NOR, MIL, MOD, SEV, and VES by $\alpha_{TM}\beta_U$ is 47, 25, 5, 4 and 4 respectively.

Now, with the help of the K- Means Clustering in Figure 9 and 12, five different clusters are plotted with principal component 1 in the x-axis and principal component 2 in the y-axis. The principal components 1 and 2 are taken for cluster visualization. The clusters are calculated from the total 10 features of the Heart Rate and SpO_2. For the stages of COPD 43, 17, 4, 17 and 4 and Asthma 17, 10, 28, 4, and 8 are the distributions by the clusters. The confusion matrix for two results are described in Table II and III. For all the stages the exact assignments of groups are 22, 6, 1, 1 and 0 for COPD and 12, 8, 12, 0, and 3 for asthma respectively.

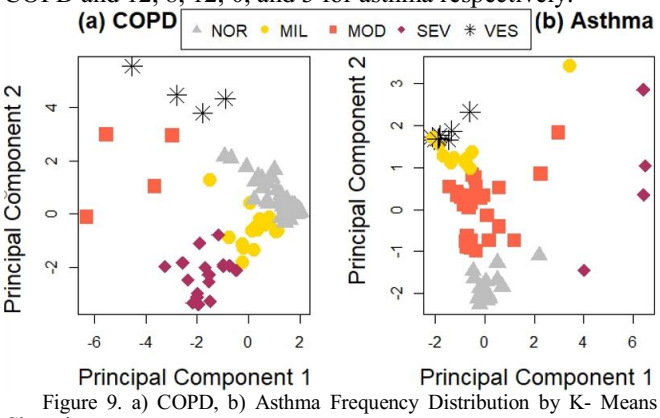

Figure 9. a) COPD, b) Asthma Frequency Distribution by K- Means Clustering.

TABLE II. COPD CONFUSION MATRIX FOR MODEL AND K-MEANS CLUSTERING

$\alpha_{TM}\beta_U$ Model / K- Means	Sev 0	Sev 1	Sev 2	Sev 3	Sev 4	Total
Sev 0	22	10	1	11	3	**47**
Sev 1	12	6	1	5	1	**25**
Sev 2	3	1	1	0	0	**5**
Sev 3	3	0	0	1	0	**4**
Sev 4	3	0	1	0	0	**4**
Total	**43**	**17**	**4**	**17**	**4**	**85**

TABLE III. ASTHMA CONFUSION MATRIX FOR MODEL AND K-MEANS CLUSTERING

$\alpha_{TM}\beta_E$ Model / K- Means	Sev 0	Sev 1	Sev 2	Sev 3	Sev 4	Total
Sev 0	12	0	3	0	0	**15**
Sev 1	5	8	11	3	1	**28**
Sev 2	0	1	12	0	1	**14**
Sev 3	0	1	2	0	3	**6**
Sev 4	0	0	0	1	3	**4**
Total	**17**	**10**	**28**	**4**	**8**	**67**

V. DISCUSSION

We designed and evaluated several models to develop a clinical decision support system to aid in the diagnosis of severities in patients with asthma and COPD. Because there are many asthmatics and COPD patients and huge economic costs associated with the diseases, simple but low levels of efficacy can contribute to improving patient health via self-management. The EOI value depends upon both of SpO_2 and heart rate, so the fluctuations of heart rate or SpO_2 because of reasons other than the severity of the diseases will not cause a rise of false positives. Moreover, the number of sensors used for the assessment is only the pulse oximeter which is easy to deploy and manage. In another work, we reported different models and determined their performances [6].

The number of healthy subject is low compared with COPD and Asthma patients and there is probability of irregular representations which is not prevalent here for the classification. The collected datasets are from ICU patients, so the algorithm needs revaluation for home management with data collected in home settings. The results are not directly validated because the "True" values of estimated parameters are not included in the databases.

VI. CONCLUSION

In this work, we show that COPD and asthma can be monitored with only heart rate and SpO_2, both of which can be obtained from a single pulse oximetry sensor. Predictive models are developed for severity estimation of both COPD and Asthma in-home telemonitoring interventions. Additionally, the proposed algorithm is simple and easy to implement that can allowing patient self-management of asthma or COPD condition. More heathy subjects and future validation of the algorithm with labelled dataset in home settings can be done in further work. This can improve patient health status, reduce the burden on hospitals and improve the overall community health of SCC.

ACKNOWLEDGMENT

This material is based upon work supported by the National Science Foundation under Grant No. 1637250. The authors gratefully acknowledge use of the travel funds of the Institute for Intelligent Systems Student Organization (IISSO) at the University of Memphis, funded by the Institute for Intelligent Systems (IIS).

REFERENCES

[1] 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.

[2] 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.

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

[4] Mohktar, Mas S., et al. "Predicting the Risk of Exacerbation in Patients with Chronic Obstructive Pulmonary Disease using Home Telehealth Measurement Data." *Artificial Intelligence in Medicine* 63.1: 51-9, 2015.

[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 Classification of Chronic Obstructive Pulmonary Disease and Asthma with Heart Rate and SpO_2 Sensors." *Engineering in Medicine and Biology Society (EMBC), 40th Annual International Conference of the IEEE*. 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] 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.