



HAL
open science

Respiratory Rate Prediction Algorithm based on Pulse Oximeter

Nurdan Cetinkaya, Sultan Turhan, Ozgun Pinarer

► **To cite this version:**

Nurdan Cetinkaya, Sultan Turhan, Ozgun Pinarer. Respiratory Rate Prediction Algorithm based on Pulse Oximeter. 2021 IEEE International Conference on Big Data (Big Data), Dec 2021, Orlando, United States. pp.4652-4658, 10.1109/BigData52589.2021.9671288 . hal-04355638

HAL Id: hal-04355638

<https://hal.science/hal-04355638>

Submitted on 3 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Respiratory Rate Prediction Algorithm based on Pulse Oximeter

Nurdan Cetinkaya
nurdancetinkya@gmail.com

Sultan Turhan
Galatasaray University
Dept. Computer Engineering
Istanbul, Turkey
sturhan@gsu.edu.tr
ORCID: 0000-0001-9763-0882

Ozgun Pinarer
Galatasaray University
Dept. Computer Engineering
Istanbul, Turkey
opinarer@gsu.edu.tr
ORCID: 0000-0002-0280-3689

Abstract—Respiratory rate (RR) is a physiological parameter typically used to monitor patient status in clinical settings. The goal of the Respiratory Rate Prediction Project is to use supervised machine learning techniques to estimate a person’s respiratory rate using real-time, continuous Photoplethysmogram (PPG) and Electrocardiogram (ECG) and oximeter data. In addition, it is also our goal to investigate the feasibility of using such data to improve diagnostic processes in healthcare. It consists of a series of studies of different algorithms for respiratory rate estimation from clinical data and is complemented by the provision of publicly available datasets and resources.

Index Terms—Respiratory Rate, Supervised Machine Learning, Photoplethysmogram, Electrocardiogram

I. INTRODUCTION

In 2020, the whole world faced a global epidemic caused by severe acute respiratory syndrome coronavirus (WHO, 2021). Although many atypical forms have been described in the literature [1]–[3], this virus commonly causes a disease in the form of a lower respiratory tract infection [4]. The most worrying aspect of the disease is its high contagiousness. Moreover, there is still no consensus on the primary contagion mechanisms of the disease. In this disease, which is described as “Mask-Distance-Cleaning” as the method of protection by the World Health Organization (WHO, 2021) and the health ministries of many states, early diagnosis has a significant impact on the healing process of the disease [5].

With the COVID-19 pandemic, the popularity and prevalence of wearable health devices has also increased. These devices have made it easy and possible for individuals to measure basic health metrics such as respiratory rate, heart rate, heart rate variability (HRV), daily step count and sleep duration in the home environment in order to diagnose COVID-19 and similar diseases early [6].

Respiratory Rate (RR) is the number of times a person breathes in one minute. The respiratory rate varies according to the age of the person. This number is 12–20 per minute in a healthy adult, 16–22 in children, and 18–24 in infants. The respiratory rate varies according to the health status of the person. For example, as a result of an increase in fever due to any disease, respiratory rate also increases [7]. Fever

is also commonly observed in lower respiratory tract infection caused by COVID-19.

It is an important parameter in the evaluation of patient status, along with basic health metrics such as respiratory rate, heart rate, and oxygen saturation [8]. Respiratory rate can be measured by counting manual breaths or using various devices and physiological principles.

Although manual measurement of respiratory rate may be inaccurate depending on the user, it is frequently used in clinical settings today [9]. In addition, devices such as spirometry, capnometry, electrocardiogram (ECG), photoplethysmography (PPG) are also used to measure respiratory rate. Recently, the use of electrocardiogram and photoplethysmography for respiration has received particular attention [10]. The electrocardiogram measures the electrical activity of the heart using electrodes attached to the skin. These electrodes are also used to determine the respiratory rate. As you breathe in, the volume of the chest increases and decreases. The flow resistance of the thorax increases with increasing volume and decreases during volume reduction. The fluctuations recorded during these changes are displayed on a monitor as a respiratory curve [11]. In this way, the respiratory rate is determined. Photoplethysmography is a method used to determine pulse rate and oxygen saturation. Respiration rate can also be extracted from these captured PPG signals using various algorithms [12].

Recently, different approaches have been developed to better predict respiratory rates by taking into account the unsteady dynamics in the electrocardiogram, photoplethysmography, and pulse oximetry [13].

The aim of this study is to decide on the best model that makes respiratory rate estimation using data such as heart rate, pulse rate, and oxygen saturation and to compare these data to determine whether the respiratory rate values of individuals are abnormal. Simultaneously, the study evaluates the feasibility of using such data to improve early detection processes in healthcare.

The ultimate goal of the project is to establish a smart system that evaluates people’s heart rate, pulse rate and oxygen saturation data, predicts respiratory rates and detects anomalies in the results obtained. With this purpose, in order to determine the estimation method to be used in the project within the

scope of this study, a comparison was made between the available data set and the prediction methods, and it was tried to determine the most suitable method for this intelligent decision support system.

The remainder of the paper is structured as follows: Section II discusses the theoretical background of the research. Section III presents the methods used and the work produced. In Section IV the obtained results will be presented. Finally, conclusion and discussion are given in Section V.

II. RELATED WORKS

Many studies have been carried out on respiratory rate estimation. One of these studies was published in 2017 by Marco A.F. It is the work of scientists Pimentel, Peter H. Charlton, and David A. Clifton. The study proposes an alternative method to increase the robustness of respiratory rate estimation with signals obtained from the Photoplethysmogram. The algorithm used in the study is a model known as autoregressive or AR model, which is observed at equal time intervals and applied to time series data. The algorithm includes 95 eight-minute Photoplethysmogram recordings obtained from both children and adults in different clinical settings. Respiratory rate estimation was made using these data with various algorithms. Using the results obtained, the performance of the model was compared with the existing ones. The proposed method has achieved an accuracy comparable to other studies in the literature. Considering the results obtained in the 32-second waveform with the mean absolute errors, it was concluded that a person breathes between 1.5 (0.3-3.3) and 4.0 (1.8-5.5) per minute. The analysis demonstrated the importance of using alternative datasets to evaluate the performance and capability of the proposed methods. Future studies will focus on using these raw data sources as a benchmark for comparison of new respiratory rate estimation approaches [14].

Another study is the work of Stephanie Baker and many scientists in 2021. In this study, the use of respiratory signal quality quantification and various neural network (NN) structures for improved respiratory rate estimation was investigated. Respiratory modulation signals were subtracted from the Electrocardiogram and Photoplethysmogram signals and a possible respiratory rate calculation was made from each subtracted signal. A simple and effective respiratory quality index (RQI) scheme has been developed that determines the quality of each respiratory signal extracted. Next, neural networks were developed for the estimation of respiratory rate using the estimated respiratory rate and its corresponding quality index as input features. It has been determined that the calculation of respiratory quality indices for modulated respiratory rates reduces the mean absolute error (MAE) of neural networks up to 38.17%. When trained and tested using 60 second waveform segments, the proposed scheme delivered 0.638 breaths per minute. Based on these results, it was concluded that it could easily be applied to non-invasive wearable devices for continuous respiratory rate measurement in many healthcare applications [15].

In 2019, a study was conducted by Maha Alafeef and Mohammad Fraiwan on smartphone-based respiratory rate estimation using Photoplethysmographic imaging and discrete wavelet transform. In this study, a smartphone-based method is proposed to accurately measure respiratory rate using video of the skin surface recorded by the smartphone's internal camera in the presence of flash light. Frame averaging was used to extract the photoplethysmographic signal of the red, green and blue channels from this input. Then, discrete wavelet transform was applied to the best representative photoplethysmographic signal for respiratory signal extraction and its velocity was estimated. 15 subjects participated in the test and evaluation. While the maximum absolute error was 0.67 bpm, the root mean Square error was 0.366 bpm. A comparison with other studies in the literature revealed superior performance in terms of accuracy, ease of use, and cost [16].

In another study conducted in 2020, respiratory rate and sleep position estimation were studied using a wearable accelerometer. In the study conducted by Doheny, Lowery, Russell, and Ryan, a triaxial accelerometer was attached to the chest and upper abdomen of 11 participants (9 men; age: 47.82 ± 14.14) undergoing clinical polysomnography (PSG). PSG flow and position data were used as benchmark data for respiratory rate (breaths per minute, bpm) and sleep position. Sleep position was classified using logistic regression with features derived from filtered acceleration and orientation. The accelerometer-derived respiratory rate was estimated for 30-second periods using a peak detection algorithm to identify individual breaths. Sensor derived and PSG respiration rates were then compared. While the mean absolute error (MAE) in respiratory rate did not differ between sensor positions (abdominal: 1.67 ± 0.37 beats; chest: 1.89 ± 0.53 beats; $p=0.52$), when participants were lying on their side ($1, 58 \pm 0.54$ bpm), decreased MAE was observed compared to supine (2.43 ± 0.95 bpm), $p < 0.01$. The position classifier distinguished between "supine and left/right" with an ROC AUC of 0.87 and "left and right" with an ROC AUC of 0.94. The proposed methods can provide a low-cost solution for prolonged sleep posture and respiratory monitoring at home [17].

A similar study was conducted in 2009 by Ki H Chon, Shishir Dash, and Kihwan Ju to estimate respiratory rate again. In this study, they presented a new method for estimating respiratory rate using pulse oximetry signal. In this method, the recently developed variable frequency complex demodulation estimation method (VFCDM) is used to describe the frequency modulation (FM) of the photoplethysmogram waveform. Frequency modulation has a measurable periodicity that provides an estimate of the respiratory period. The performance of the estimation method is compared with the continuous wavelet transform (CWT) and autoregressive (AR) model approaches. Both the CWT and AR models have been shown to predict well respiratory rates within the normal range (12-26 breaths/min). In this analysis of 15 healthy subjects, we used the VFCDM method to extract respiratory rates ranging from 12-36 breaths/min for accuracy, consistency (smaller interquartile range of median value), and computational efficiency (less

than 0.3 s in 1 minute of data). it has been shown to provide the best results [18].

In 2014, Yunyoung Nam et al. presented a method for estimating respiratory rate using the camera of a smartphone, MP3 player, and a tablet in South Korea. An attempt was made to estimate respiratory rate from the pulse signal obtained from a finger placed on the camera lens of iPhone 4S, iPad 2, iPod 5 and Galaxy S3 devices. Before estimating respiratory rates, the optimal signal quality of these 4 devices was investigated by dividing the resolution of the video camera into 12 different pixel regions. Then, the suitability of smartphones, iPod 5 and tablet for respiratory rate estimation was investigated using three different calculation methods. These methods are; autoregressive (AR) model, variable frequency complex demodulation (VFCDM) and continuous wavelet transform (CWT) approaches. In the study, data were collected from 10 healthy subjects to evaluate the performance of the three computational methods and pixel regions for optimum signal quality. The VFCDM method has been observed to provide good estimates for respiratory rates within the normal range (12-24 breaths/minute). While both the CWT and VFCDM methods provide reasonably good estimates for respiratory rates greater than 26 breaths/min, accuracies have been observed to decrease with increasing respiratory rates. Overall, the VFCDM method performs best for accuracy (smaller median error), consistency (smaller interquartile range of median value), and computational efficiency (less than 0.5 s in 1 minute of data using a MATLAB application) for extracting respiratory rates. to 36 breaths/minute [19].

In a study conducted by Vignesh Ravichandran and a group of friends in 2019, an end-to-end deep learning network called RespNet was proposed to perform the task of extracting the respiratory signal from a specific input PPG, unlike other studies. The proposed network is trained and tested on two different datasets using different modalities of the reference respiratory signal recordings. In addition, the similarity and performance of the proposed network versus two traditional signal processing approaches for respiration signal extraction are examined. The proposed method was tested on two independent datasets with a Mean Square Error of 0.262 and 0.145. The cross-correlation coefficients of the relevant data sets were found to be 0.933 and 0.931. Reported errors and similarity were found to be better than traditional approaches [20].

III. OUR APPROACH

With the emerging technologies, remote monitoring of individual's health data becomes possible. Thanks to smart watches, smartphone applications or special sensors, it is possible to monitor both the daily data of healthy people and patients' vital signs. Mobile health (m-health) is a rapidly expanding field in the digital health sector that provides healthcare support, delivery and intervention through mobile technologies such as smartphones, tablets and wearable devices [24]. Neither patients who are monitored at home nor healthy individuals who want to record their daily health data are willing to carry sensors on them for long periods of time.

The pulse oximeter is a simple medical device that is easily worn on the finger of the individual and has the ability to collect the desired data in a very short time [21].

During the ongoing coronavirus disease (COVID-19) pandemic, patients who are diagnosed with COVID-19 but not sick enough to warrant admission has been proposed to use pulse oximetry at home in order to monitor and spot deterioration [22]–[24].

A. Dataset

In order to choose the correct prediction method for Respiration Rate estimation, study is first carried out on the sample data set. The sample dataset used in the study was taken from PhysioNet, a web-based open source site designed to support new research in healthcare and studies of complex physiological and clinical data [25]. The dataset named “BIDMC PPG and Respiration” [14] includes data from a large group of patients admitted to medical and surgical intensive care units, randomly selected from a total of 53 adults, 32 females and 21 males, aged between 19 and 90+ years. In the dataset

- 1) Physiological signals data such as PPG, impedance respiratory signal and electrocardiogram (ECG)
- 2) Data on physiological parameters such as heart rate (HR), respiratory rate (RR) and blood oxygen saturation level (SpO2)
- 3) Age and gender data

are available. Heart rate, respiratory rate and blood oxygen saturation level parameters of this data set were used in the study.

B. Data Pre-Processing

The data preprocessing step was carried out before using the data within the regression models in order to compare them. In this context, fixed variables (age, gender) and respiratory rate (breaths per minute), heart rate (beats per minute derived from ECG), pulse rate (beats per minute derived from PPG) and blood oxygen saturation level of 53 patients are combined on a single source. This process was carried out on Jupiter Notebook with Python language. The general view of the data before the data preprocessing step is as in Figure 1.

There are 8 attributes in the data set. These are time (seconds), heart rate, pulse rate, respiratory rate, blood-oxygen saturation level, age, gender and key information of the patient, respectively. As to types of data, the time is numeric, the gender is categorical, and the other columns contain float values.

In data preprocessing, it is pivotal to identify and correctly handle the missing values in order to have accurate and adequate conclusions and inferences from the data. There are 25492 records in total in the data set. It has been observed that a certain number of data is missing for each attribute in these records. Figure 2 shows the number of missing data per attribute. Based on the total number of records available, the proportion of records with missing data is approximately 0.78%. Although this rate is too small to cause a serious error, we decided to keep them in order to ensure that the

	Time [s]	HR	PULSE	RESP	SpO2	Age	Gender	Key
0	0	94.0	93.0	25.0	97.0	88.0	M	1
1	1	94.0	93.0	25.0	97.0	88.0	M	1
2	2	94.0	93.0	25.0	97.0	88.0	M	1
3	3	92.0	93.0	26.0	97.0	88.0	M	1
4	4	93.0	93.0	26.0	97.0	88.0	M	1
...
25488	476	92.0	95.0	18.0	100.0	81.0	M	53
25489	477	90.0	95.0	19.0	100.0	81.0	M	53
25490	478	97.0	95.0	19.0	100.0	81.0	M	53
25491	479	99.0	95.0	19.0	100.0	81.0	M	53
25492	480	94.0	95.0	19.0	100.0	81.0	M	53

Fig. 1. Overview of the Dataset

performance evaluation of the methods will produce accurate results. Instead of deleting these records, the available data are filled with the average values of the attributes containing numerical values. For categorical attributes such as gender, the "LabelEncoder" module in the Python- Scikit Learn Library, which helps to label character labels with numbers, was used. The data of male, female and patients who did not want to specify gender were first converted to 0, 1, 2, and then the missing data were filled by calculating the mode values in these attributes.

Time [s]	0
HR	4
PULSE	128
RESP	153
SpO2	128
Age	481
Gender	481
Key	0

Fig. 2. Incomplete records counts for each attribute in the data set

The descriptive statistics results of the data set obtained after the completion of the missing data are displayed in Figure 3.

	Time [sec]	HR	PULSE	RESP	SpO2	Age	Gender	Key
count	25493.000000	25493.000000	25493.000000	25493.000000	25493.000000	25493.000000	25493.000000	25493.000000
mean	240.000000	89.180440	88.519401	17.562255	96.723823	64.730769	0.433952	27.000000
std	138.855163	13.325711	13.775857	3.131113	3.289710	17.439908	0.532340	15.297359
min	0.000000	44.000000	37.000000	1.000000	83.000000	19.000000	0.000000	1.000000
25%	120.000000	81.000000	80.000000	16.000000	95.000000	51.000000	0.000000	14.000000
50%	240.000000	89.000000	88.000000	18.000000	97.000000	67.000000	0.000000	27.000000
75%	360.000000	95.000000	95.000000	19.000000	99.000000	77.000000	1.000000	40.000000
max	480.000000	139.000000	129.000000	34.000000	100.000000	90.000000	2.000000	53.000000

Fig. 3. Descriptive Statistics results of completed data set

C. Applied Method

To determine the appropriate technique, seven methods are used to evaluate the dataset. These methods are:

- 1) Ordinary Least Squares: Ordinary least squares, or linear least squares, estimates the parameters in a regression model by minimizing the sum of the squared residuals. This method draws a line through the data points that minimizes the sum of the squared differences between the observed values and the corresponding fitted values [26].
- 2) Lasso Regression: It is a regression analysis method that performs both variable selection and regularization in order to enhance the prediction accuracy and interpretability of the resulting statistical model [27].
- 3) Elastic Net Regression: The elastic net is a regularization procedure for linear regression that also performs variable selection [28].
- 4) Ridge Regression: Ridge regression is a model tuning method that is used to analyse any data that suffers from multicollinearity. This method performs L2 regularization. When the issue of multicollinearity occurs, least-squares are unbiased, and variances are large, this results in predicted values to be far away from the actual values [28].
- 5) Bayesian Ridge Regression: It allows a natural mechanism to survive on scant data or poorly distributed data by formulating linear regression using probability distributors instead of point estimates. Output or response 'y' is assumed to be from a probability distribution rather than being estimated as a single value [28].
- 6) K-neighbors Regression: One of the most useful types of Bayesian regression is Bayesian Ridge regression, which estimates a probabilistic model of the regression problem [29].
- 7) Random Forest Regression: Random Forest Regression is a supervised learning algorithm that uses ensemble learning method for regression. Ensemble learning method is a technique that combines predictions from multiple machine learning algorithms to make a more accurate prediction than a single model [30].

Errors in each case are calculated and then tabulated. The model with the best performance is then chosen.

D. Comparison of the Methods

After the pre-processing of the data set, the above-mentioned models were defined and the comparison process was started. One of the main libraries used in this section is the library known as sklearn or scikit-learn, which has many algorithms for regression, clustering and classification [31]. The available data set is primarily partitioned into test and train datasets. Seventy percent of the available data is train data and thirty percent is test data (Figure 4). Then, with the help of the Sklearn library, the regression models mentioned above were tested on these data sets and the successes in estimation were compared.

```

from sklearn.preprocessing import StandardScaler
SS = StandardScaler()
X = df.drop("RESP",axis=1)
SS.fit(X,y=None)
y = df["RESP"].values
X = SS.transform(X)

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=10)

print(X_train.shape, X_test.shape)
print(y_train.shape, y_test.shape)

(17045, 7) (7648, 7)
(17045,) (7648,)

```

Fig. 4. Dividing the data set into train and test data sets.

In order to calculate the success of each model separately, a function is generated with Python (Figure 5). This function assesses the performance of each regression model. After executing this function with the relevant data set, for each model mean absolute error, R2 score, root mean square error and training-test time information are obtained. model is far ahead of other models with an R2 score of 0.98 and Mean Square Error (RMSE) and Average Absolute Errors of around 0.1.

```

models = {
    'OLS':LinearRegression(),
    'ElasticNet':ElasticNet(),
    'BayesianRidge':BayesianRidge(),
    'Lasso':Lasso(),
    'Ridge':Ridge(),
    'KNN':KNeighborsRegressor(),
    'RFF':RandomForestRegressor(),
    'Ada':AdaBoostRegressor(),
    'XGB':XGBRegressor()
}

def model_performance(X, y):
    times = []
    keys = []
    mean_squared_errors = []
    mean_abs_error = []
    r2_scores = []
    X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size = 0.3, random_state = 10)

    for k, v in tqdm(models.items()):
        model = v

        t0 = time.time()
        model.fit(X_train, y_train)
        train_time = time.time()-t0
        t1 = time.time()
        pred = model.predict(X_test)
        predict_time = time.time() - t1
        pred = pd.Series(pred)
        Time_total = train_time + predict_time
        times.append(Time_total)
        r2_scores.append(r2_score(y_test,pred))
        mean_squared_errors.append(mean_squared_error(y_test,pred))
        mean_abs_error.append(mean_absolute_error(y_test,pred))
        keys.append(k)

    table = pd.DataFrame({
        'model':keys,
        'RMSE':mean_squared_errors,
        'MAE':mean_abs_error,
        'R2 score':r2_scores,
        'time':times
    })

    return table

model_performance(X, y)

```

Fig. 5. Python Code for assessing the model performance

IV. RESULTS

In this study, seven different regression models were tested on the same data set. When the performance results of the models were compared, it was observed that the two models gave better results than the others.

The performances of the models Ordinary Least Squares, Lasso Regression, Elastic Net Regression, Ridge Regression, Bayesian Ridge, K-neighbors Regression, Random Forest Regression are evaluated with the criteria; mean absolute error, R2 score, root mean square error, and training and testing times. While the R2 score of the OLS model was 0.10, the Mean Square Error (RMSE) and Mean Absolute Error (MAE) values were 8.99 and 2.31, respectively. The estimation successes of Elastic Net Regression, Ridge Regression, Bayesian Ridge Regression models are also very close to these values. (In the range of RMSE: 8-10, around MAE: 2, R2: around 0.1) Compared to these results, it has been seen that the RFF model is far ahead of other models with an R2 score of 0.98 and Mean Square Error (RMSE) and Average Absolute Errors of around 0.1. All the results are given in Figure 6 and the success criteria curves are represented in Figure 7.

	model	RMSE	MAE	R2 score	time
0	OLS	8.991034	2.312735	0.106337	0.011971
1	ElasticNet	9.562305	2.330317	0.049555	0.002989
2	BayesianRidge	8.990809	2.312507	0.106359	0.005985
3	Lasso	10.061928	2.377057	-0.000105	0.001994
4	Ridge	8.991029	2.312730	0.106337	0.003989
5	KNN	0.354820	0.288502	0.964733	0.175529
6	RFF	0.178853	0.181566	0.982223	4.519999
7	Ada	5.386710	1.863174	0.464588	0.786867
8	XGB	0.435663	0.425621	0.956697	0.921523

Fig. 6. Performance Results of Regression models

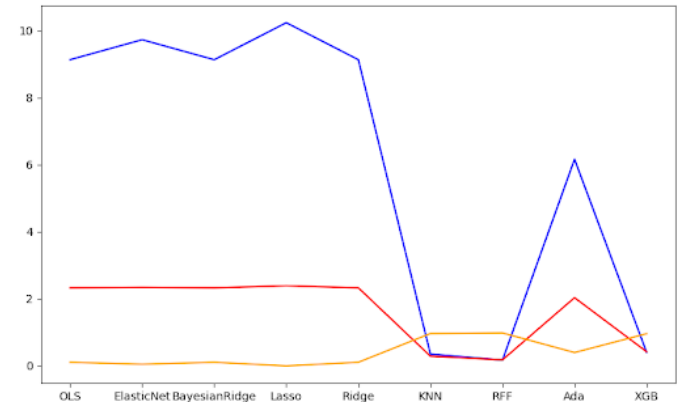


Fig. 7. Regression models' success criteria (blue: RMSE, red: MAE, orange: R2 score)

It was concluded that KNN Regression and Random Forest Regression model are the best 2 model for estimating respira-

tory rate. Among these two models, the model that made the best estimation is Random Forest Regression. Figure 8 shows the comparison of the estimated values calculated with the Random Forest Regression model with the actual values. As can be seen, the results are very close to each other.

	Real Values	Predicted Values
0	14.0	13.980000
1	18.0	17.870000
2	18.0	18.060000
3	7.0	8.352348
4	16.0	15.980000
...
7643	20.0	20.680000
7644	8.0	8.520000
7645	18.0	18.080000
7646	21.0	20.880000
7647	15.0	15.000000

7648 rows x 2 columns

Fig. 8. Comparison of predicted values calculated with Random Forest Regression model with actual real data

V. CONCLUSION AND FUTURE WORKS

The COVID-19 epidemic that we are experiencing has once again demonstrated the importance of mobile and remote health services. Especially thanks to the widespread use of wearable technological devices and smart phone applications, it has become much easier to provide such services. Respiratory rate measurement has an important place in the observation of both patient and healthy individuals. As a result of the comparison performed in the study, it was revealed that the model that gave the best result in estimating respiratory rate was Random Forest. It is clear that Random Forest Regression performs the best on dataset, with a low MSE and high R2 score, closely followed by KNN Regression.

Having established the validity of the Random Forest regression modelling approach, we plan to establish an end-to-end connected and intelligent decision support system which is able to monitor patient remotely in order to observe patients' vital signs nearly real time on streaming data and will enable to intervene by predicting a possible deterioration situation.

ACKNOWLEDGMENTS

This work is supported by Galatasaray University Research Foundation under the Grant No. FBA-2021-1041.

REFERENCES

- [1] M. Ekbatani, S. Hassani, L. Tahernia, B. Yaghmaei, S. Mahmoudi, A. Navaeian, M. Rostamyan, F. Zamani, and S. Mamishi, "Atypical and novel presentations of coronavirus disease 2019: a case series of three children," *British journal of biomedical science*, vol. 78, no. 1, pp. 47–52, 2021.
- [2] N. Singhania, S. Bansal, and G. Singhania, "An atypical presentation of novel coronavirus disease 2019 (covid-19)," *The American journal of medicine*, vol. 133, no. 7, p. e365, 2020.
- [3] S. L. Ng, Y. S. Ong, K. Y. Khaw, S. P. Teh, C. S. Tan, L. C. Ming, K.-G. Chan, L.-H. Lee, and B.-H. Goh, "Focused review: potential rare and atypical symptoms as indicator for targeted covid-19 screening," *Medicina*, vol. 57, no. 2, p. 189, 2021.
- [4] K. Habas, C. Nganwuchu, F. Shahzad, R. Gopalan, M. Haque, S. Rahman, A. A. Majumder, and T. Nasim, "Resolution of coronavirus disease 2019 (covid-19)," *Expert review of anti-infective therapy*, vol. 18, no. 12, pp. 1201–1211, 2020.
- [5] S. Sun, A. A. Folarin, Y. Ranjan, Z. Rashid, P. Conde, C. Stewart, N. Cummins, F. Matcham, G. Dalla Costa, S. Simblett *et al.*, "Using smartphones and wearable devices to monitor behavioral changes during covid-19," *Journal of Medical Internet Research*, vol. 22, no. 9, p. e19992, 2020.
- [6] G. Quer, J. M. Radin, M. Gadaleta, K. Baca-Motes, L. Ariniello, E. Ramos, V. Kheterpal, E. J. Topol, and S. R. Steinhubl, "Wearable sensor data and self-reported symptoms for covid-19 detection," *Nature Medicine*, vol. 27, no. 1, pp. 73–77, 2021.
- [7] M. M. Jensen, J. G. Kellett, P. Hallas, and M. Brabrand, "Fever increases heart rate and respiratory rate; a prospective observational study of acutely admitted medical patients." *Acute medicine*, vol. 18, no. 3, pp. 141–143, 2019.
- [8] T. Flenady, T. Dwyer, and J. Applegarth, "Explaining transgression in respiratory rate observation methods in the emergency department: A classic grounded theory analysis," *International journal of nursing studies*, vol. 74, pp. 67–75, 2017.
- [9] C. Kelly, "Respiratory rate 1: why measurement and recording are crucial," *Nursing times*, vol. 114, no. 4, pp. 23–24, 2018.
- [10] M. A. Pimentel, P. H. Charlton, and D. A. Clifton, "Probabilistic estimation of respiratory rate from wearable sensors," in *Wearable electronics sensors*. Springer, 2015, pp. 241–262.
- [11] H. Liu, J. Allen, D. Zheng, and F. Chen, "Recent development of respiratory rate measurement technologies," *Physiological measurement*, vol. 40, no. 7, p. 07TR01, 2019.
- [12] D. J. Meredith, D. Clifton, P. Charlton, J. Brooks, C. Pugh, and L. Tarassenko, "Photoplethysmographic derivation of respiratory rate: a review of relevant physiology," *Journal of medical engineering & technology*, vol. 36, no. 1, pp. 1–7, 2012.
- [13] J. Lee, J. P. Florian, and K. H. Chon, "Respiratory rate extraction from pulse oximeter and electrocardiographic recordings," *Physiological measurement*, vol. 32, no. 11, p. 1763, 2011.
- [14] M. A. Pimentel, A. E. Johnson, P. H. Charlton, D. Birrenkott, P. J. Watkinson, L. Tarassenko, and D. A. Clifton, "Toward a robust estimation of respiratory rate from pulse oximeters," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 8, pp. 1914–1923, 2016.
- [15] S. Baker, W. Xiang, and I. Atkinson, "Determining respiratory rate from photoplethysmogram and electrocardiogram signals using respiratory quality indices and neural networks," *Plos one*, vol. 16, no. 4, p. e0249843, 2021.
- [16] M. Alafeef and M. Fraiwan, "Smartphone-based respiratory rate estimation using photoplethysmographic imaging and discrete wavelet transform," *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 2, pp. 693–703, 2020.
- [17] E. P. Doheny, M. M. Lowery, A. Russell, and S. Ryan, "Estimation of respiration rate and sleeping position using a wearable accelerometer," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2020, pp. 4668–4671.
- [18] K. H. Chon, S. Dash, and K. Ju, "Estimation of respiratory rate from photoplethysmogram data using time–frequency spectral estimation," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 8, pp. 2054–2063, 2009.
- [19] Y. Nam, J. Lee, and K. H. Chon, "Respiratory rate estimation from the built-in cameras of smartphones and tablets," *Annals of biomedical engineering*, vol. 42, no. 4, pp. 885–898, 2014.
- [20] V. Ravichandran, B. Murugesan, V. Balakarthikeyan, K. Ram, S. Preejith, J. Joseph, and M. Sivaprakasam, "Respnet: A deep learning model for extraction of respiration from photoplethysmogram," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2019, pp. 5556–5559.
- [21] M. T. Petterson, V. L. Begnoche, and J. M. Graybeal, "The effect of motion on pulse oximetry and its clinical significance," *Anesthesia & Analgesia*, vol. 105, no. 6, pp. S78–S84, 2007.

- [22] A. M. Luks and E. R. Swenson, "Pulse oximetry for monitoring patients with covid-19 at home. potential pitfalls and practical guidance," *Annals of the American Thoracic Society*, vol. 17, no. 9, pp. 1040–1046, 2020.
- [23] N. Starr, D. Rebollo, Y. M. Asemu, L. Akalu, H. A. Mohammed, M. W. Menchamo, E. Melese, S. Bitew, I. Wilson, M. Tadesse *et al.*, "Pulse oximetry in low-resource settings during the covid-19 pandemic," *The Lancet Global Health*, vol. 8, no. 9, pp. e1121–e1122, 2020.
- [24] S. Shah, K. Majmudar, A. Stein, N. Gupta, S. Suppes, M. Karamanis, J. Capannari, S. Sethi, and C. Patte, "Novel use of home pulse oximetry monitoring in covid-19 patients discharged from the emergency department identifies need for hospitalization," *Academic Emergency Medicine*, vol. 27, no. 8, pp. 681–692, 2020.
- [25] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "Physiobank, physiotookit, and physionet: components of a new research resource for complex physiologic signals," *circulation*, vol. 101, no. 23, pp. e215–e220, 2000.
- [26] G. D. Hutcheson, "Ordinary least-squares regression," *L. Moutinho and GD Hutcheson, The SAGE dictionary of quantitative management research*, pp. 224–228, 2011.
- [27] J. Ranstam and J. Cook, "Lasso regression," *Journal of British Surgery*, vol. 105, no. 10, pp. 1348–1348, 2018.
- [28] J. O. Ogutu, T. Schulz-Streeck, and H.-P. Piepho, "Genomic selection using regularized linear regression models: ridge regression, lasso, elastic net and their extensions," in *BMC proceedings*, vol. 6, no. 2. Springer, 2012, pp. 1–6.
- [29] R. MurtiRawat, S. Panchal, V. K. Singh, and Y. Panchal, "Breast cancer detection using k-nearest neighbors, logistic regression and ensemble learning," in *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)*. IEEE, 2020, pp. 534–540.
- [30] A. Liaw, M. Wiener *et al.*, "Classification and regression by randomforest," *R news*, vol. 2, no. 3, pp. 18–22, 2002.
- [31] P. Developer, "Bitdegree,2020,python kütüphaneleri," <https://tr.bitdegree.org/tutorial/python-kutuphaneleri/>, 2020, last updated 2021-10-18 UTC.