Wearable Analytics and Early Diagnostic of COVID-19 Based on Two Cohorts

Justyna Skibinska
Faculty of Electrical Engineering and Communication
Brno University of Technology
Unit of Electrical Engineering
Tampere University
Brno, Czechia
skibinska@vut.cz

Jiri Hosek
Faculty of Electrical Engineering and Communication
Brno University of Technology
Brno, Czechia
hosek@vut.cz

Asma Channa
Computer Science Department
University POLITEHNICA of Bucharest
DIIES Department
University Mediterranea of Reggio Calabria
Bucharest, Romania
asma.channa@stud.acs.upb.ro

Abstract—The outbreak of the COVID-19 pandemic forced a need to create screening tests to diagnose the disease. To answer this challenge, this paper introduces the support methodology for COVID-19 early detection based on wearable and machine learning likewise on two various cohorts. We compare the level of detection of the COVID-19 disease, Influenza, and Healthy Control (HC) thanks to the usage of machine learning classifiers likewise changes in heart rate and daily activity. The features obtained as the parameters of the ratio of heart rate to the variable of the number of steps proved to have the highest statistical importance. The COVID-19 cases versus HC were possible to be distinguished with 0.73 accuracy by the XGBoost algorithm, whereas COVID-19 cases, Influenza vs. HC were able to be differentiated on similar level of accuracy: in 0.72 by Support Vector Machine. The multiclass classification between the cases achieved a 0.57 F1-score for three classes by XGBoost. For early diagnosis, this solution could serve as an extra test for clinicians during the pandemic, and the result shows which metric could be useful for creating the machine learning model.

Index Terms—COVID-19, AI, wearable, machine learning

I. Introduction

The COVID-19 pandemic started in December 2019 [1] and reap the harvest of 6.39 milion deaths on the world so far according to Johns Hopkins University (JHU) [2]. This disease causes the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [1]. Regarding the contagiousness of the disease, the highest period is reported as 2 days before the symptom onset to 1 day after the visibility of the illness [3]. To the most recognised symptom of this disease belongs: fever, decrease in SpO_2, cough, changes in heart rate [4], fatigue, shortness of breath, hoarse voice, muscle pain, and headache [5], [6] likewise the loss of taste and smell [7]. The changes in heart rate are especially visible in COVID-19 cases and lasting longer than common Influenza, the resting heart rate is elevated nearly the symptom onset [4].

Moreover, deviations from the norm of heart rate were also observed during sleep for COVID-19 [8]. The researchers reported the elevation in the estimated mean respiration rate during deep sleep and the mean nocturnal heart rate during non-rapid eye movement (NREM) sleep. Whereas, the decrease was observed for the root mean square of successive differences (RMSSD) of the nocturnal RR series and the Shannon entropy of the nocturnal RR series. This was concluded based on Z-value analysis and 1257 participants wearing Fitbit devices.

The data could be collected in various ways for the screening test. The highest prediction of COVID-19 is achieved by X-ray [9]. Nevertheless, the fastest and cheapest way to allow the screening of a large population is the usage of wearable devices [10]. The data for remote tracking of symptoms are gathered thanks to the Internet of Medical Things [5]. The modalities which could be commonly analysed from wearable sensors to support society during a pandemic are cardiovascular strain, sleep parameters, activity levels, respirations variable, SpO_2 level, temperature, blood pressure, cough and sound monitoring, and humidity sensors [10], [11], [12]. The wearables on-body sensors could be placed in smartwatches, smart rings, headbands, sociometric badges, camera clips and sensors embedded in clothing [10]. The last step to create the support methodology is analysing the data. The big promise for the analysis of pandemic-related data brings Machine Learning [6], [12] including the novelty of usage Federated Learning [13].

In this paper, we proposed a support methodology for COVID-19 detection based on smartwatches and short-time

analysis. We analysed also Influenza cases, the machine learning model was trained on two datasets from Stanford University and Evidation 2019-2020 Participatory ILI Surveillance Program. We analysed the heart rate and steps taken during the day. We created features based on literature and proved the usability of the ratio of parameters heart rate related to a variable based on the number of steps. Our main contribution is the support methodology distinguishing COVID-19 cases from Healthy Control (HC) on the level of 0.73 accuracy for a balanced dataset likewise a combination of two datasets from different cohorts and analysis of the experiment results. Moreover, we proposed the classification approach which is an improvement and extension of the original paper which was introduced just anomaly detection.

The rest of the paper contains in Sec. II the state-of-the-art and Sec. III describes the experiment. In subsection III-A, it was marked out the experiment description, in subsection III-B, it was provided data description and subsection III-C brought the feature extraction closer. The results are described in Sec. IV likewise the discussion is provided in Sec. V. The conclusion is contained in the last section (Sec. VI).

II. RELATED WORKS

The research of this paper were based mostly on [14], [4], [6], [15]. In the first paper, the researchers collected the data thanks to the wearables, they analysed the data gathered by the Fitbit device. They enrolled circa 5,300 participants and among them, they had records of 32 COVID-19 cases, 15 Influenza, and 73 HC. They used in the experiment the heart rate, and the number of steps and collected also sleep patterns. They introduced 3 algorithms for anomaly detection: Resting Heart Rate (RHR) Difference (RHR-Diff) offline anomaly detection, the heart rate over steps anomaly detection (HROS-AD) offline anomaly detection, and CuSum online detection. HR-Diff offline anomaly detection used residuals standardization of RHR to detect anomalies in heart signals. They standardised a 1-hour signal based on an average of 28 days. If the time window was under the relevance of 0.05, it was detected anomaly. HROS-AD used a new metric HROS which was the ratio of heart rate to steps. This algorithm is the type of unsupervised anomaly detection and is using moving average, undersampling to one hour, and Z-score transformation. If it is found anomaly detection, the data are marked as outliers thanks to the Gaussian distribution analysis. Moreover, the author introduced an algorithm in real-time: the CuSum - the type of cumulative statistics which was summing the deviations of residuals of RHR. Under the considerations, it was taken earlier 28 days of records to recognise the anomaly. If the signal is suspicious for 24h, it is recognised as positive. 63% of COVID-19 cases by CuSum were detected as positive. Nevertheless, it was not mentioned the specificity of the algorithm by authors. Additionally, 9 of 15 Illness cases were recognised as ill during the beginning of Illness or before.

In [4], the authors used the data from 7000 people to detect the state between COVID-19 and 2 types of Influenza: Influenza before the main pandemic and Influenza in the middle of the pandemic. They were measuring the elevated resting heart rate of the infected people and the number of steps. The collected dataset included 41 COVID-19 cases, 85 Influenza at the time of the pandemic, and 1265 during pre-pandemic. The symptoms of the disease were longer for COVID-19, respectively 12 days vs. 9 and 7 for flu. Elevated resting heart rate occurs more frequently nearly the onset of the disease. The differences were confirmed by statistical tests. Whereas in [6], the authors analysed the data originally published in [14]. They presented the classification methodology in early stage of the disease when in the [14] was only applied and discussed the anomaly detection methodology.

In [6] was proposed the solution which used the wearable device (Fitbit), machine learning methodology, and creation of the features based on statistical analysis, spectral and frequency-related features. It was taken into consideration during choosing the time windows the incubation and contagiousness periods. For one sample, it was compared the difference between the features of a potentially ill state and healthy, and healthy vs. healthy. The best results were achieved for the 5-day windows and 7 days differences between the states. The COVID-19 vs. HC state was distinguishable in 78 % accuracy thanks to the k-NN and COVID-19, Influenza vs. HC in 73 % also for k-NN. Under analysis were taken 27 HC and 27 COVID-19 cases for the first scenario, and in the second scenario to balance the data, 24 COVID-9 cases, 7 Influenza, and 33 HC.

In [15], the authors differentiate the COVID-19 samples between Influenza in the middle of the pandemic and before the main pandemic with the usage of machine learning. They analysed data from [4] in the amount of 21 COVID-19 cases, 37 non-COVID-19 flu cases, and Pre-COVID-19. They used mRMR feature selection. For the feature pre-selection step were used the records of heart rate and number and isolated from the parameters such as: mean, standard deviation, maximum, minimum, range, variance, Shannon entropy, approximate entropy, slope, skew, and kurtosis. The authors were able to distinguish two Influenza cases in 80 % thanks to the XGBoost and Decision Tree algorithm. Moreover, they differentiated COVID-19 cases with people with flu in the middle of the pandemic in 73 % thanks to the k-NN algorithm likewise COVID-19 cases with Influenza before the main pandemic in 82 %. The multiclass classification allows distinguishing cases on the level of 0.64 F1-score for k-NN.

Additionally, there was also introduced neural network – PCovNet to analyse the data from [14] in [16]. PCovNet is a Long Short-term Memory Variantial Autoencoder (LSTM-VAE)-based anomaly detection solution dedicated to analysing RHR in the early stage of the disease. The evaluation shows that the detection was possible in 0.946 precision, and 0.234 recall. The authors also introduced F-beta metrics equal for this case 0.918, however, the parameter is unreliable. The network was trained on 25 COVID-19 cases. PCovNet recognised 100 % of the ill person based on RHR, nevertheless after recognised person as infected, so on quarantine. Additional approach of anomaly detection based on the dataset from [14]

was provided in [17]. They applied One Class-Support Vector Machine (OC-SVM) and they achieved better results than in [14]. They registered changes in signal behaviour 23,5 % - 40 % earlier in the comparison to [14] (12 h – 4 days) likewise they obtained the false positive rates. The optimal parameter of the time window for parameters of RHR was 300 and 350. The maximum of the found anomalies in signal was 21 among 29 for the COVID-19 for OC-SVM based on RHR (RHR-OC-SVM). Whereas based on HROS signal was found by OC-SVM (HROS-OC-SVM) 24 among 29 patients. Moreover, it has to be taken into consideration that the false positive rate, which was for RHR-OC-SVM lower than in the original paper, was 39.96 outlier for HC.

Subsequently, the authors in [18] and [19] proved the usability of the temperature for the COVID-19 diagnosis. In [18], they checked the changes in the area under the curve (AUC) after removing the temperature as one of the modalities with the classifier which analysed all the types of signal, i. e.: temperature, heart rate (HR), respiration rate (RR), heart rate variability and metabolic equivalents (MET). For the research, they enrolled and used data gathered by Oura Ring of 73 people with COVID-19 disease. As the classifier was chosen Random Forest and they achieved AUC = 0.819 for all modalities and AUC = 0.770 for the combination of the signals without temperature.

Moreover, it was also introduced the COVID-19 decompensation index (CDI) for monitoring the need for hospitalization in [20]. The wearable was placed on the chest for continuous measuring of the patients. Based on 308 negative patients (without the need for hospitalization) and 22 positive patients was achieved 0.84 AUC with the usage of the gradient boosted model.

What's more, the promising future direction of monitoring the COVID-19 disease by a wearable is the usage of smartwatches with the function of measuring electrocardiogram (ECG) like the Apple Watch [21].

III. EXPERIMENT

A. Experiment description

The main goal of this research was to compare classifications of the combined datasets gathered by the wearable device - Fitbit. Those data contain the COVID-19 cases, Influenza cases, and the HC group. The heart rate and the number of steps were analysed as likewise ratio of the parameters extracted from the heart rate to parameters obtained from the number of steps as the human body efficiency. Based on our previous research [15] and [6], we took into consideration the character of the disease, i.e. the highest contagiousness period and incubation period to limit the spreading of the disease by the early detection. The nature of the disease was also described in [16].

The dataset was originally published in [14] and [4] where participants were wearing the wearable device to collect the data which could be analysed to detect the anomaly in the human physiological signals. The difference between both datasets was in sampling rate and the demographic distribution

of the dataset. The dataset B was more homogenous and people had recognised obesity which could bias the research. Nevertheless, dataset B was collected in the USA. In the research, it was evaluated the pre-processing step of the data, statistical analysis likewise the machine learning approach to classify the data. The flow of the algorithm is visible in Fig. 1.
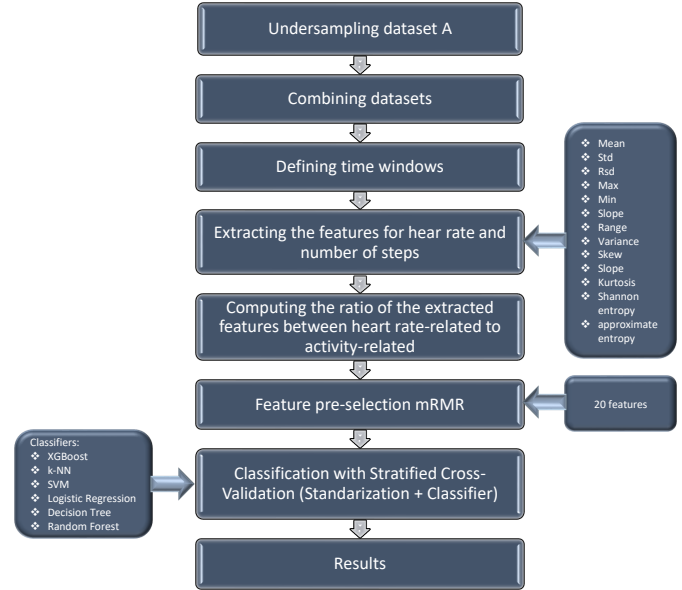

Fig. 1. Flow of the algorithm.

As the first step, it was undersampled the dataset B to be able to combine both datasets. The sampling rate was equal to 1 sample per day for heart rate and the number of steps. Next, the dataset was combined according to the scenario. 6 scenarios are visible in Table I. There are 6 types of data: COVID-19 A, COVID-19 B, Influenza A, Non-COVID-19 Flu, Pre-COVID-19 Flu, HC. There is also given information as to which class was assigned the data case.

TABLE I
SCENARIOS OF COMBINATION OF THE DATASET AND TYPES OF CLASS FOR EACH EXPERIMENT.

Types of data	Experiment 1a	Experiment 1b	Experiment 2	Experiment 3	Experiment 4	Experiment 5
COVID-19 A	1	1	1	1	1	2
COVID-19 B	1	1		1	1	2
Influenza A			1	0	1	1
Non-Covid-19 Flu				0	1	1
Pre-Covid-19 Flu				0	1	1
HC	0	0	0		0	0

In the next step, the time window was extracted with the respect to contagiousness period and incubation period. It was from 2 days before the onset to 7 days before the visibility of the symptoms. Based on those time frames, several parameters were computed for heart rate and the number of steps, i. e.: mean, standard deviations, maximum, minimum, slope, range, variance, skew, slope, kurtosis, Shannon entropy, and approximate entropy. Additionally, the ratios between the aforementioned parameters of heart rate and activity were computed. Among such prepared 35 features were selected

20 of them thanks to the Maximum Relevance Minimum Redundancy algorithm (mRMR). Because of the limited number of samples, it was as classification 10-fold Stratified Cross-Validation. The classifiers which were tested were XGBoost, Logistic Regression, Random Forest, Decision Tree, k-NN, and SVM. Additionally, one scenario was performed as multiclass classification and the rest were binary classification. To check the statistical difference between the datasets, we used the Mann-Whitney U test. This test does not require the assumption of the normal distribution of the data and it is suitable for small datasets. Additionally, to omit the curse of the error-type I, we used the Benjamini-Hochberg procedure [6]. The confidence level α was 0.05. We checked the scenarios for the COVID-19 (A and B dataset) vs. HC and COVID-19 (A and B dataset), Influenza vs. HC.

B. Data description

The created dataset was combination of two existing already dataset [14], [4]. The first of them (here so-called A) contained the COVID-19 cases, Influenza cases, and HC. The original data were collected with the usage of smartwatches and the publicly available dataset contains data gathered thanks to the Fitbit device. The dataset includes heart rate and activity expressed in the number of steps taken and lacking sleep records which contain information about the stage of the sleep and its duration. Originally, 5,262 participants were enrolled and among them, 3,325 wore Fitbit devices. The collected data were completed for 32 COVID-19 patients, 15 people with Influenza and 73 HC [14]. We choose for the research among them 27 COVID-19 cases, 7 Influenza cases, and also 73 HC. Moreover, we analysed heart rate and steps taken by participants. The sampling rate was heart rate per minute and steps taken per hour. Regarding the second dataset (here so-called B), there were representatives of three classes: COVID-19 cases (41 gathered with wearable data), people with flu before the pandemic (Pre-COVID-19 Flu, 1265 gathered with wearable), and people infected by Influenza during the pandemic (Non-COVID-19 Flu, 85 gathered with wearable data). The data analysed in the original paper were collected by the Fitbit device. Among the available dataset and found utility by us, we choose 21 COVID-19 cases, 675 Pre-COVID-19 Flu cases, and 37 Non-COVID-19 Flu cases to be used in the experiment. The sampling rate in both cases was 1 sample per day. Moreover, we used different combinations of the dataset to compare the possibility of the detection – distinction between groups. The datasets were balanced. The scenarios of the particular combinations could be found in the Table II. It is provided the number of samples per experiment.

TABLE II
SCENARIOS OF COMBINATION OF THE DATASET AND NUMBER OF SAMPLES FOR EACH EXPERIMENT.

Types of data	Experiment 1a	Experiment 1b	Experiment 2	Experiment 3	Experiment 4	Experiment 5
COVID-19 A	27	27	27	27	27	27
COVID-19 B	21	21	0	21	21	21
Influenza A	0	0	7	7	7	7
Non-Covid-19 Flu	0	0	0	19	10	20
Pre-Covid-19 Flu	0	0	0	20	9	21
HC	48	48	34	0	74	48

C. Feature extraction

The inspiration for extracting the features could be found in [15] and [6]. The process of extracting the features has a few steps. First of all, the idea was to extract useful information about the time series before the beginning of the visible symptoms, it is from -7 to -2 days of the disease. This is justified by the fact the highest contagiousness of the COVID-19 disease is regarded as the -2 days before the onset of the disease. Because, the dataset was limited in sampling numbers, i. e. the sampling rate was 1 per day of dataset B, we need to undersample dataset A. This step is visible in Fig. 2, where:

- t_B is the beginning of the considered period of the illness
- p_S is the length of the time window
- t_0 is the Onset of the disease $t_0 = t_B + p_S$
- t_D is the diagnosis of the disease $t_D = 2 + t_0$

As the next process of the data was chosen the time window of the aforementioned length. It was 5 days for each of the samples for the record of heart rate and the number of steps for both datasets A and B. From those fragments of the time series were computed statistical parameters, entropy values, and features based on the distribution of the time series, i. e.: mean, standard deviation (std), relative standard deviation (rsd), maximum (max), minimum, range, Shannon entropy, approximate entropy, variance, the slope of the time series, skewness and kurtosis. Additionally, we computed and inspired by the limit, we computed the ratio of the calculated 12 aforementioned parameters between parameters of heart rate and parameters of the activity.
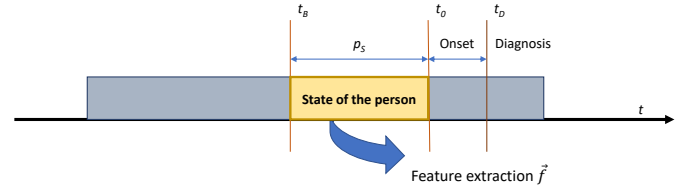

Fig. 2. Feature extraction.

IV. RESULTS

This section describes the statistical analysis for COVID-19 vs. HC likewise analysis of COVID-19, Influenza vs. HC and the outcomes of the classifications for various scenarios. The explanation of the combination of the scenarios is presented in Tables I and II. Here, there are pointed out the results in the tables for such cases:

- Table V: Classification of the COVID-19 (A and B dataset) vs. HC for all features
- Table VI: Classification of the COVID-19 (A and B dataset) vs. HC for chosen 20 features
- Table VII: Distinction between COVID-19 cases from dataset A, Influenza A vs. HC for 20 features
- Table VIII: Distinction between COVID-19 (A and dataset B) vs. Influenza (A, Influenza before the main pandemic and Influenza in the middle of pandemic)

TABLE III
STATISTICAL ANALYSIS THANKS TO THE MANN WHITNEY U-TEST
TOGETHER WITH FDR CORRECTION FOR COVID-19 CASES AND HC.

Feature	pval	pval_FDR
activ_std	0.00001	0.00014
activ_variance	0.00001	0.00014
activ_range	0.00001	0.00015
activ_rsd	0.00003	0.00027
activ_entropy_shannon	0.00092	0.00644
heart_std	0.00184	0.00921
heart_variance	0.00184	0.00921
heart_rsd	0.00291	0.01249
heart_range	0.00321	0.01249
steps_entropy_shannon	0.00750	0.02624
steps_range	0.01214	0.03540
activ_max	0.01138	0.03540
steps_slope	0.01593	0.04288
steps_std	0.02331	0.05439
steps_variance	0.02331	0.05439
steps_max	0.03409	0.07457
heart_max	0.03736	0.07692
heart_entropy_shannon	0.04967	0.09658
steps_skew	0.05699	0.10498
activ_mean	0.06626	0.11596
steps_kurtosis	0.09272	0.15453
activ_min	0.09860	0.15687
steps_mean	0.10960	0.16224
steps_min	0.11125	0.16224
heart_mean	0.12333	0.16602
activ_slope	0.12333	0.16602
steps_rsd	0.14433	0.18709
heart_min	0.24971	0.31214
activ_kurtosis	0.27791	0.33541
heart_slope	0.38690	0.45139
heart_skew	0.42406	0.47878
heart_kurtosis	0.64317	0.70347
activ_skew	0.92736	0.98357
steps_approx_entropy	1.00000	1.00000
heart_approx_entropy	1.00000	1.00000

TABLE IV
STATISTICAL ANALYSIS THANKS TO THE MANN WHITNEY U-TEST
TOGETHER WITH FDR CORRECTION FOR COVID-19 CASES, INFLUENZA
AND HC.

Feature	pval	pval_FDR
heart_std	0.00000	0.00002
heart_range	0.00000	0.00002
heart_variance	0.00000	0.00002
activ_std	0.00000	0.00002
activ_range	0.00000	0.00002
activ_entropy_shannon	0.00000	0.00002
activ_variance	0.00000	0.00002
heart_rsd	0.00002	0.00009
activ_rsd	0.00002	0.00009
heart_max	0.00005	0.00019
heart_entropy_shannon	0.00010	0.00031
heart_mean	0.00040	0.00118
steps_entropy_shannon	0.00303	0.00815
heart_min	0.00395	0.00987
activ_max	0.00952	0.02222
activ_min	0.01776	0.03884
activ_mean	0.01951	0.04017
steps_min	0.05476	0.10648
steps_max	0.10284	0.17997
steps_skew	0.09805	0.17997
steps_mean	0.11923	0.19872
steps_slope	0.13656	0.21726
steps_range	0.21119	0.32138
steps_std	0.26776	0.36045
steps_variance	0.26776	0.36045
steps_kurtosis	0.26611	0.36045
activ_kurtosis	0.44079	0.57139
heart_skew	0.57789	0.72237
steps_rsd	0.65872	0.75941
activ_slope	0.67262	0.75941
activ_skew	0.63672	0.75941
heart_slope	0.77282	0.84528
heart_kurtosis	0.88901	0.94289
steps_approx_entropy	1.00000	1.00000
heart_approx_entropy	1.00000	1.00000

- Table IX: Classification of COVID-19 cases (A and B dataset), Influenza (A, Influenza before the main pandemic and Influenza in the middle of pandemic) vs. HC
- Table X: Multiclass classification of COVID-19 (A and B dataset), Influenza (A, Influenza before the main pandemic and Influenza in the middle of pandemic), and HC

The results of the statistical analysis of the features are presented in Table III for COVID-19 (A and B dataset) vs. HC likewise for COVID-19 (A and B dataset), Influenza (A, Influenza in the middle of pandemic, Influenza before the

main pandemic) vs. HC in Table IV. The first scenario (Table III) shows that 13 features that are statistically significant according to the Mann-Whitney U test with the false discovery rate correction (FDR) thanks to the Benjamini-Hochberg procedure are: activ_std, activ_variance, activ_range, activ_rsd, activ_entropy_shannon, heart_std, heart_variance, heart_rsd, heart_range, steps_entropy_shannon, steps_range, activ_max, steps_slope. The 'activ' shortcut means the ratio of the created feature from the heart rate to the number of steps.

All of the above p-values with FDR correction were below 0.05. For the second scenario (Table IV), the 17 features which met the requirement of the confidence level $\alpha = 0.05$ are: heart_std, heart_range, heart_variance, activ_std, activ_range, activ_entropy_shannon, activ_variance, heart_rsd, activ_rsd, heart_max, heart_entropy_shannon, heart_mean, steps_entropy_shannon, heart_min, activ_max, activ_min, activ_mean.

For all of the features for classification COVID-19 for datasets A and B vs. HC achieved 0.73 accuracy and 0.75 sensitivity for XGBoost and the Matthews correlation coefficient (MCC) was also the highest for XGBoost, i. e. 0.48 (Table V). The best specificity was registered for k-NN: 0.77. This was achieved for 27 COVID-19 A cases, 21 COVID-19 B cases as 0 class, and 48 HC as 1 class. The data were balanced.

TABLE V
THE RESULTS OF CLASSIFICATION COVID-19 (A AND B DATASET) VS. HC (FOR ALL 36 FEATURES).

Classifier	Accuracy	Sensitivity	Specificity	MCC
XGBoost	**0.73 ± 0.14**	**0.75 ± 0.22**	0.75 ± 0.20	**0.48 ± 0.30**
k-NN	0.68 ± 0.13	0.60 ± 0.21	**0.77 ± 0.18**	0.39 ± 0.27
SVM	0.68 ± 0.15	0.61 ± 0.22	0.75 ± 0.22	0.38 ± 0.20
Logistic Regression	0.66 ± 0.16	0.62 ± 0.23	0.70 ± 0.21	0.33 ± 0.32
Decision Tree	0.66 ± 0.16	0.65 ± 0.22	0.66 ± 0.22	0.32 ± 0.33
Random Forest	0.70 ± 0.16	0.64 ± 0.24	0.76 ± 0.20	0.42 ± 0.33

Another analysis for provided for exactly this same data but after feature extraction with the usage of 20 features among 36 (Table VI). The results of the classification show that XGBoost achieved 0.73 accuracy, 0.71 sensitivity, and 0.48 MCC as the best outcome among other classifiers. The highest specificity has k-NN: 0.86.

TABLE VI
THE RESULTS OF CLASSIFICATION COVID-19 (A AND B DATASET) VS. HC (FOR 20 FEATURES).

Classifier	Accuracy	Sensitivity	Specificity	MCC
XGBoost	**0.73 ± 0.14**	**0.71 ± 0.22**	0.75 ± 0.19	**0.48 ± 0.29**
k-NN	0.72 ± 0.15	0.58 ± 0.24	**0.86 ± 0.16**	0.47 ± 0.30
SVM	0.67 ± 0.14	0.60 ± 0.22	0.73 ± 0.19	0.35 ± 0.29
Logistic Regression	0.65 ± 0.15	0.59 ± 0.23	0.70 ± 0.20	0.31 ± 0.31
Decision Tree	0.68 ± 0.15	0.64 ± 0.23	0.71 ± 0.18	0.37 ± 0.30
Random Forest	0.70 ± 0.16	0.64 ± 0.24	0.76 ± 0.20	0.42 ± 0.33

For the classification of the COVID-19 A, Influenza A vs. HC, the results are presented in Table VII. The distinction between the classes was for 27 COVID-19 A cases, 7 Influenza A cases, and 34 HC. The best accuracy obtained XGBoost: 0.63, and MCC was also the best for this classifier: 0.28. The highest sensitivity obtained SVM: 0.90 and the best specificity of 0.67 was recognised for Decision Tree.

The results of the classification of COVID-19 (A and B dataset) vs. Influenza (A, Influenza before the main pandemic and Influenza in the middle of a pandemic) are shown in Table VIII. Under the analysis was taken: 27 COVID-19 A, 21 COVID-19 B, 7 Influenza A, 19 Non-COVID-19 Flu and 20 Pre-COVID-19 Flu. The highest accuracy achieved XGBoost (0.67) and this classifier had 0.36 MCC. The Random Forest

TABLE VII
THE RESULTS OF CLASSIFICATION DATASET A BETWEEN COVID-19, INFLUENZA VS. HC.

Classifier	Accuracy	Sensitivity	Specificity	MCC
XGBoost	**0.63 ± 0.18**	0.65 ± 0.26	0.62 ± 0.28	**0.28 ± 0.38**
k-NN	0.56 ± 0.18	0.56 ± 0.27	0.58 ± 0.26	0.14 ± 0.39
SVM	0.57 ± 0.14	**0.90 ± 0.20**	0.24 ± 0.22	0.18 ± 0.31
Logistic Regression	0.49 ± 0.18	0.49 ± 0.28	0.48 ± 0.26	-0.03 ± 0.40
Decision Tree	0.57 ± 0.17	0.48 ± 0.25	**0.67 ± 0.30**	0.16 ± 0.37
Random Forest	0.54 ± 0.17	0.51 ± 0.33	0.57 ± 0.34	0.08 ± 0.36

had 0.66 sensitivity. The highest specificity was obtained for SVM (0.73).

TABLE VIII
THE RESULTS OF CLASSIFICATION COVID-19 (A AND B DATASET) VS. INFLUENZA (A AND INFLUENZA BEFORE THE MAIN PANDEMIC AND INFLUENZA IN THE MIDDLE OF PANDEMIC).

Classifier	Accuracy	Sensitivity	Specificity	MCC
XGBoost	**0.67 ± 0.13**	0.63 ±0. 21	0.71 ± 0.19	**0.36 ± 0.27**
k-NN	0.62 ± 0.14	0.54 ± 0.22	0.70 ± 0.22	0.25 ± 0.30
SVM	0.62 ± 0.14	0.52 ± 0.22	**0.73 ± 0.21**	0.26 ± 0.29
Logistic Regression	0.61 ± 0.15	0.58 ± 0.22	0.64 ± 0.21	0.24 ± 0.31
Decision Tree	0.61 ± 0.15	0.62 ± 0.25	0.61 ± 0.21	0.24 ± 0.31
Random Forest	0.63 ± 0.14	**0.66 ± 0.22**	0.61 ± 0.21	0.28 ± 0.30

The outcome of the distinction between COVID-19 cases (A and B dataset), Influenza (A, Influenza before the main pandemic, and Influenza in the middle of pandemic) vs. HC is shown in Table IX. Under consideration was taken 27 COVID-19 A cases, 21 COVID-19 B cases, 7 Influenza cases, 10 Non-COVID-19 Flu cases, 9 Pre-COVID-19 Flu cases and 74 HC. The best accuracy of 0.72 was registered for SVM. Sensitivity 0.66 was the highest for Logistic Regression. For Decision Tree was observed the highest specificity (0.90) and MCC (0.47).

TABLE IX
THE RESULTS OF CLASSIFICATION COVID-19 (A AND B DATASET), INFLUENZA (A AND INFLUENZA BEFORE THE MAIN PANDEMIC AND INFLUENZA IN THE MIDDLE OF PANDEMIC) VS. HC

Classifier	Accuracy	Sensitivity	Specificity	MCC
XGBoost	0.68 ± 0.12	0.62 ± 0.18	0.74 ± 0.16	0.37 ± 0.25
k-NN	0.70 ± 0.12	0.65 ± 0.17	0.75 ± 0.16	0.41 ± 0.24
SVM	**0.72 ±0.12**	0.61 ± 0.19	0.82 ± 0.13	0.45 ± 0.25
Logistic Regression	0.67 ± 0.12	**0.66 ± 0.17**	0.68 ± 0.17	0.35 ± 0.25
Decision Tree	0.71 ± 0.11	0.53 ± 0.18	**0.90 ± 0.11**	**0.47 ± 0.21**
Random Forest	0.67 ± 0.11	0.57 ± 0.18	0.78 ± 0.18	0.37 ± 0.24

Additionally, we performed multiclass classification between COVID-19 cases and Influenza, and HC for both datasets. 27 COVID-19 cases A, 21 COVID-19 B cases, 7 Influenza A cases, 20 Non-COVID-19 Flu, 21 Pre-COVID-19 Flu, and 48 HC were analysed. The highest 0.57 F1-score was achieved for XGBoost. Moreover, 0.58 accuracy was also the best for this same classifier likewise MCC 0.38 was the highest for XGBoost.

V. DISCUSSION

This research aimed to create a support system methodology to distinguish COVID-19 cases vs. HC thanks to wearable devices. Additionally, the group of people infected by a

Classifier	F1-score	Accuracy	MCC
XGBoost	**0.57 ± 0.13**	**0.58 ± 0.13**	**0.38 ± 0.20**
k-NN	0.56 ± 0.14	0.57 ± 0.13	0.37 ± 0.20
SVM	0.53 ± 0.13	0.55 ± 0.12	0.34 ± 0.19
Logistic Regression	0.52 ± 0.13	0.54 ± 0.13	0.32 ± 0.19
Decision Tree	0.40 ± 0.07	0.51 ± 0.09	0.31 ±0.15
Random Forest	0.53 ± 0.12	0.55 ± 0.11	0.34 ± 0.17

different kind of Influenza was taken into consideration during classification among others treated as one class with COVID-19 cases. We created simple methodology for early detection with those scenarios based on two datasets [14] and [4]. Naturally, those datasets were limited in the amount of the samples likewise we need to undersample the one dataset A to be able to combine A and B datasets. The extraction of the features took into consideration the contagiousness period and incubation period. Moreover, the features were created after the choice of the time windows based on statistics, distribution of the time series, and entropy of the signal. There were considered heart rate, number of steps, and activity of the person. For the statistical analysis, we performed two analyses. The first of them was the comparison between COVID-19 cases vs. HC and the second COVID-19 cases, Influenza cases vs. HC. The first analysis shows that one-third of the created features passed the Mann-Whitney test with FDR correction. The most important features are the standard deviation of the activity of the person, i. e. std of heart rate to std number of steps. Meaningful features were also variance, range, rsd, Shannon entropy, maximum of personal activity and std, variance, rsd, range heart rate likewise Shannon entropy, range, and slope of the number of steps during the day. The idea of creating the ratio of the measurements between the extracted features of heart rate to the number of steps occurs to be successful. For the second scenario, half of the features were statistically different, moreover, with the eliminating error-type I. The most powerful were the features obtained based on heart rate and human activity. According to classification, we performed 6 scenarios of the classification. The best results of accuracy for 5 out of 6 scenarios were achieved for the XGBoost. Additionally, the MCC was also in most cases the highest for XGBoost and once in the case of Decision Tree. F1-score was also the best in the case of XGBoost. The XGBoost is the most complex classifier among used algorithms, nevertheless, it could be also slightly overfitted. Regarding the comparison of the datasets, the highest accuracy of 0.73 was achieved for 2 scenarios, i. e.: for the distinction between COVID-19 (A and B dataset) vs. HC. The 0.72 accuracy was obtained for COVID-19 (A and B dataset), Influenza (A and Influenza before the main pandemic, and Influenza in the middle of a pandemic) vs. HC. It means that a similar result of the classification was achieved for the differentiation between COVID-19 cases vs. HC for both datasets and Illness (COVID-19 and Influenza in both

datasets) vs. HC. The combination of both datasets allows for creating simple support system detection in the early stages. It could be concluded based on the 0.57 F1-score that multiclass classification was somehow possible to differentiate between each other for HC, Influenza, and COVID-19 cases. Moreover, feature selection mRMR allows for obtaining higher outcomes of the classification than the usage of all the feature (Tables V and VI). The lower ability of distinction between cases for classifier was registered for COVID-19 cases vs. Influenza for both datasets (0.67 accuracy) likewise classification of COVID-19 A, Influenza A vs. HC (0.63 accuracy). The classi- fication of the proposed simpler methodology achieved lower values in comparison to the more extended approach based on extracting more advanced features including frequency and spectral features between two states of patients and HC on dataset A [6]. In this paper, the accuracy achieved 0.73 for such a scenario. Additionally, for the more advanced solution, the classification of the COVID-19 cases vs. HC gave higher results: 0.78 accuracy for k-NN, 5-days windows, and only dataset A. In this paper, we proposed simpler methodology which is more approachable for low-latency solutions together with compression of another dataset, nevertheless biased by people with obesity. What's more, we combined various types of Influenza which could be the advantage of the proposed support methodology. Subsequently, we used the methodology introduced in [15] which was applied to the dataset B. We extended the research for including more cases and HC thanks to the combination with the dataset A. Bigger and more diverse demographically dataset obtain lower results between distinction COVID-19 cases vs. Influenza cases (0.67 for new solution vs. 0.73 and 0.82 likewise 0.57 vs. 0.64 for multiclass classification) what it could statistically mean that people in the USA have different symptoms between the types of Influenza and COVID-19, or perhaps, it is easier to distinguish illness for them because of the obesity.

We assume that to succeed in higher accuracy, we would need more modalities like signals gathered from the gyroscope, magnetometers, altimeters, barometers [6] and temperature measurements, additionally to extend the database for low sampling rate. Another factor that could influence the quality of the data is how frequently and with responsible people are wearing the wearables, so the human factor.

VI. CONCLUSION

In our study, we created a support system methodology for detecting the COVID-19 disease. We analysed 6 scenarios that we have included in the machine learning model also with Influenza cases. We combined two different datasets to obtain more samples and various cohorts. First of all, we created the features for our model based on statistical parameters, distribution of the time series, and entropies. We included also the parameters based on the ratio between the heart rate-related variables to parameters based on the number of steps. The statistical analysis shows that the most informative features are those indicated on the activity of the human person. The machine learning models allow us to differentiate COVID-

19 samples vs. HC most precisely thanks to the XGBoost algorithm, with 0.73 accuracy. Almost this same accuracy (0.72) for an extended dataset including also Influenza cases allows us to distinguish people infected from HC by XGBoost. The multiclass classification shows that on the level of 0.57 F1-score was possible to find differences between the three classes. In the future direction, we could extend the dataset in the number of samples from various cohorts based on universal measurements from the Fitbit device.

REFERENCES

[1] M. Ciotti, M. Ciccozzi, A. Terrinoni, W.-C. Jiang, C.-B. Wang, and S. Bernardini, "The covid-19 pandemic," *Critical reviews in clinical laboratory sciences*, vol. 57, no. 6, pp. 365–388, 2020.

[2] "Global COVID-19 Statistics by JHU." https://www.arcgis.com/apps/dashboards/bda7594740fd40299423467b48e9ecf6. Accessed: 2022-07-27.

[3] X. He, E. H. Lau, P. Wu, X. Deng, J. Wang, X. Hao, Y. C. Lau, J. Y. Wong, Y. Guan, X. Tan, *et al.*, "Temporal dynamics in viral shedding and transmissibility of covid-19," *Nature medicine*, vol. 26, no. 5, pp. 672–675, 2020.

[4] A. Shapiro, N. Marinsek, I. Clay, B. Bradshaw, E. Ramirez, J. Min, A. Trister, Y. Wang, T. Althoff, and L. Foschini, "Characterizing covid-19 and influenza illnesses in the real world via person-generated health data," *Patterns*, vol. 2, no. 1, p. 100188, 2021.

[5] S. Mehrdad, Y. Wang, and S. F. Atashzar, "Perspective: Wearable internet of medical things for remote tracking of symptoms, prediction of health anomalies, implementation of preventative measures, and control of virus spread during the era of covid-19," *Frontiers in Robotics and AI*, vol. 8, p. 610653, 2021.

[6] J. Skibinska, R. Burget, A. Channa, N. Popescu, and Y. Koucheryavy, "Covid-19 diagnosis at early stage based on smartwatches and machine learning techniques," *IEEE Access*, vol. 9, pp. 119476–119491, 2021.

[7] T. Struyf, J. J. Deeks, J. Dinnes, Y. Takwoingi, C. Davenport, M. M. Leeflang, R. Spijker, L. Hooft, D. Emperador, J. Domen, *et al.*, "Signs and symptoms to determine if a patient presenting in primary care or hospital outpatient settings has covid-19," *Cochrane Database of Systematic Reviews*, no. 5, 2022.

[8] A. Natarajan, H.-W. Su, and C. Heneghan, "Assessment of physiological signs associated with covid-19 measured using wearable devices," *NPJ digital medicine*, vol. 3, no. 1, pp. 1–8, 2020.

[9] R. C. Joshi, S. Yadav, V. K. Pathak, H. S. Malhotra, H. V. S. Khokhar, A. Parihar, N. Kohli, D. Himanshu, R. K. Garg, M. L. B. Bhatt, *et al.*, "A deep learning-based covid-19 automatic diagnostic framework using chest x-ray images," *Biocybernetics and Biomedical Engineering*, vol. 41, no. 1, pp. 239–254, 2021.

[10] A. Channa, N. Popescu, J. Skibinska, and R. Burget, "The rise of wearable devices during the covid-19 pandemic: A systematic review," *Sensors*, vol. 21, no. 17, p. 5787, 2021.

[11] D. R. Seshadri, E. V. Davies, E. R. Harlow, J. J. Hsu, S. C. Knighton, T. A. Walker, J. E. Voos, and C. K. Drummond, "Wearable sensors for covid-19: a call to action to harness our digital infrastructure for remote patient monitoring and virtual assessments," *Frontiers in Digital Health*, p. 8, 2020.

[12] R. Krishnamurthi, D. Gopinathan, and A. Kumar, "Wearable devices and covid-19: state of the art, framework, and challenges," *Emerging Technologies for Battling Covid-19*, pp. 157–180, 2021.

[13] M. Hajder, P. Hajder, T. Gil, M. Krzywda, J. Kolbusz, and M. Liput, "Architecture and organization of a platform for diagnostics, therapy and post-covid complications using ai and mobile monitoring," *Procedia Computer Science*, vol. 192, pp. 3711–3721, 2021.

[14] T. Mishra, M. Wang, A. A. Metwally, G. K. Bogu, A. W. Brooks, A. Bahmani, A. Alavi, A. Celli, E. Higgs, O. Dagan-Rosenfeld, *et al.*, "Pre-symptomatic detection of covid-19 from smartwatch data," *Nature biomedical engineering*, vol. 4, no. 12, pp. 1208–1220, 2020.

[15] J. Skibinska and R. Burget, "Is it possible to distinguish covid-19 cases and influenza with wearable devices? analysis with machine learning," *Journal of Advances in Information Technology Vol*, vol. 13, no. 3, 2022.

[16] F. F. Abir, K. Alyafei, M. E. Chowdhury, A. Khandakar, R. Ahmed, M. M. Hossain, S. Mahmud, A. Rahman, T. O. Abbas, S. M. Zughaier, *et al.*, "Pcovnet: A presymptomatic covid-19 detection framework using deep learning model using wearables data," *Computers in biology and medicine*, vol. 147, p. 105682, 2022.

[17] H. R. Cho, J. H. Kim, H. R. Yoon, Y. S. Han, T. S. Kang, H. Choi, and S. Lee, "Machine learning-based optimization of pre-symptomatic covid-19 detection through smartwatch," *Scientific Reports*, vol. 12, no. 1, pp. 1–15, 2022.

[18] A. E. Mason, F. M. Hecht, S. K. Davis, J. L. Natale, W. Hartogensis, N. Damaso, K. T. Claypool, S. Dilchert, S. Dasgupta, S. Purawat, *et al.*, "Detection of covid-19 using multimodal data from a wearable device: results from the first tempredict study," *Scientific reports*, vol. 12, no. 1, pp. 1–15, 2022.

[19] B. L. Smarr, K. Aschbacher, S. M. Fisher, A. Chowdhary, S. Dilchert, K. Puldon, A. Rao, F. M. Hecht, and A. E. Mason, "Feasibility of continuous fever monitoring using wearable devices," *Scientific reports*, vol. 10, no. 1, pp. 1–11, 2020.

[20] D. M. Richards, M. J. Tweardy, S. R. Steinhubl, D. W. Chestek, T. L. V. Hoek, K. A. Larimer, and S. W. Wegerich, "Wearable sensor derived decompensation index for continuous remote monitoring of covid-19 diagnosed patients," *NPJ digital medicine*, vol. 4, no. 1, pp. 1–11, 2021.

[21] J. S. Chinitz, R. Goyal, D. C. Morales, M. Harding, S. Selim, and L. M. Epstein, "Use of a smartwatch for assessment of the qt interval in outpatients with coronavirus disease 2019," *The Journal of innovations in cardiac rhythm management*, vol. 11, no. 9, p. 4219, 2020.