IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received July 29, 2021, accepted August 7, 2021, date of publication August 18, 2021, date of current version September 3, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3106255

COVID-19 Diagnosis at Early Stage Based on Smartwatches and Machine Learning Techniques

JUSTYNA SKIBINSKA[1,2], RADIM BURGET[1], ASMA CHANNA[3,4], (Graduate Student Member, IEEE), NIRVANA POPESCU[3], AND YEVGENI KOUCHERYAVY[2]

[1]FEEC, Brno University of Technology, 61600 Brno, Czech Republic
[2]Unit of Electrical Engineering, Tampere University, 33100 Tampere, Finland
[3]Computer Science Department, University POLITEHNICA of Bucharest, 060042 Bucharest, Romania
[4]DIIES Department, University Mediterranea of Reggio Calabria, 89124 Reggio Calabria, Italy

Corresponding author: Justyna Skibinska (skibinska@feec.vutbr.cz)

This work was supported in part by the European Union's Horizon 2020 Research and Innovation Program under Marie Skłodowska Curie Grant 813278 (A-WEAR: A network for dynamic wearable applications with privacy constraints, http://www.a-wear.eu), and in part by the Project "Early COVID-19 Infection Detection System for the Safety of Vulnerable Groups Using Artificial Intelligence," Ministry of the Interior, Czech Republic, under Grant VI04000039.

ABSTRACT Early detection of COVID-19 positive people are now extremely needed and considered to be one of the most effective ways how to limit spreading the infection. Commonly used screening methods are reverse transcription polymerase chain reaction (RT-PCR) or antigen tests, which need to be periodically repeated. This paper proposes a methodology for detecting the disease in non-invasive way using wearable devices and for the analysis of bio-markers using artificial intelligence. This paper have reused a publicly available dataset containing COVID-19, influenza, and Healthy control data. In total 27 COVID-19 positive and 27 healthy control were pre-selected for the experiment, and several feature extraction methods were applied to the data. This paper have experimented with several machine learning methods, such as XGBoost, k-nearest neighbour k-NN, support vector machine, logistic regression, decision tree, and random forest, and statistically evaluated their perfomance using various metrics, including accuracy, sensitivity and specificity. The proposed experiment reached 78 % accuracy using the k-NN algorithm which is significantly higher than reported for state-of-the-art methods. For the cohort containing influenza, the accuracy was 73 % for k-NN. Additionally, we identified the most relevant features that could indicate the changes between the healthy and infected state. The proposed methodology can complement the existing RT-PCR or antigen screening tests, and it can help to limit the spreading of the viral diseases, not only COVID-19, in the non-invasive way.

INDEX TERMS Artificial intelligence, COVID, signal processing.

I. INTRODUCTION

The deadly coronavirus SARS CoV-2 belongs to the family of coronaviridae, which has two subfamilies: Coronavirid and Toroviridae. Those coronaviruses are known to be infectious to for example, birds, and mammals including humans [1]. The course of the disease varies significantly. It starts with completely asymptomatic courses to mild, moderate, or severe courses, which in many cases end in death. The spectrum of the symptoms of Coronavirus is broad and includes fever, cough, shortness of breath, hoarse voice, abdominal pain, or chest pain [2], and infected individuals may experience a rare loss of taste and smell [3]. The disease may end with long-term complications such as inter alia respiratory, neurological, cardiovascular problems, and many other potential issues that have yet to be fully described [4]–[6]. Some similarities in the development of the COVID-19 pandemic and epidemics of severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS) have been observed [7].

The average incubation period of COVID-19 currently ranges from 2 to 11 days [8]. The major disease transmission of the virus occurs through social contact, particularly through face-to-face exposure, coughing, sneezing, or during talking, [9]. The combination of the high reproduction number of the disease, high portion of asymptotic or

The associate editor coordinating the review of this manuscript and approving it for publication was Derek Abbott.

mild-symptom cases in the population, long prodromal stage, during which the patient is infectious but with no visible symptoms and infection fatality rate 1.04 % [10] has led to 108 million cases and 2.4 million deaths worldwide by 02/2021 [11].

According to the Johns Hopkins University, more than 2 million deaths occurred in the world since the outbreak of this pandemic [11]. The most accurate diagnostics for the disease are imaging technologies, which can achieve an accuracy of nearly 100 % [12]. The most popular diagnostic method (among others) is the reverse transcription-polymerase chain reaction (RT-PCR) [1]. PCR-tests showed a sensitivity of 77,7 % and specificity of 98,8 % [13]. Pharyngeal RT-PCR tests showed the sensitivity of 78.2 % and specificity of 98.8 %. Although these methods are relatively accurate, they are often used after the onset of the disease for confirmation. Unfortunately, the disease was proven to be the most communicable 2 days before and 1 day after the onset of the disease symptoms [14].

In [15], the authors presented the disease stages of COVID-19. They distinguished between the early stage of infection (stage I), the pulmonary phase (stage II), and the hyper-inflammation phase (stage III). The detection of the disease in the prodromal stage would have the most significant impact, since it reflects the phase in which the person is infectious but, she/he feels healthy, which leads to social contacts and spreading of the disease to other people [14]. The authors indicated that symptoms in this phase, included dry cough, fever, and mild constitutional symptoms. The detection of the virus in this stage minimise contagiousness, preventing the development of further complications and reducing the duration of the disease [15]. The risk factors for this disease are old age, civilization diseases, and renal and hepatic dysfunctions [8].

Vaccination is currently considered one of the most promising ways to fight the disease. Until this, the most effective methods to prevent the infections are mask wearing, social distancing, keeping hygienic practices, handwashing, applying quarantine, and disinfection. These methods are considered the most effective for fighting this mortal disease [7], [16].

Detection of the disease two days before the onset of the disease is very important. During those (on average) two days, a person is not aware of the infection which causes communication of the disease at the workplace, to family members, and to other possible social contacts. Unfortunately, such early detection is not an easy task since the disease symptoms are difficult to be recognized.

The common availability of wearables in the population makes them an ideal intelligent screening tool. These devices allow continuous monitoring of various physiological parameters, such as heart rate (HR), heart rate variability (HRV), resting heart rate (RHR), respiration rate (RR), skin temperature, and oxygen saturation (SpO2) [17].

This paper provides a comprehensive overview of recent progress in the area of wearable technologies for e-health and COVID-19 use cases. It also introduces a novel methodology based on artificial intelligence, with wearable sensor signal analyses, and proves that these devices can be used for early detection with interesting accuracy. For the statistical evaluation, 27 samples of COVID-positive cases and 27 healthy control cases were used. The paper also includes the results of the univariate tests conducted to identify the feature of importance. We tested a few classifiers in real time allowing us to identify those that would achieve the highest prediction. A significant advantage of this work is that it takes into consideration the important parameters related to the disease, that is, the incubation period and the highest contagiousness period thus ensuring, the reduction of disease contagiousness as much as possible.

- Creation of the methodology which is able to detect COVID-19 disease in the early stage
- The usage of wearable devices and machine learning allows created solution serve as potential screening test
- The consideration of the parameter such as incubation period and the highest contagiousness interval should reduce the most efficiently the number of new infected people
- Identification what kind of the features are indicators of being contagiousness in the early stage of the COVID-19 disease
- Extension of the original work into classification problem, with success of 78 % accuracy for k-NN

The rest of this paper is structured as follows. Section II describes related works. Section III - introduces the experiment, described the data, the methodology of feature extraction, metrics used for evaluation - based support system methodologies, and machine learning and statistical evaluation methodologies. Section IV presents the results and describes the research outcomes of this work. The discussion, which explains how to understand the results, is described in Section VII. The paper is concluded in Section VII.

II. RELATED WORKS

The coronavirus was announced by the WHO relatively recently, on 29 December 2019. Thus, resources amount the virus are still scant, and the behavior of this virus under various circumstances is still waiting exploration. Wearable sensors have seen significant improvements in their quality, accuracy, and reliability. Due to their broad availability in the population, they have opened new possibilities for their utilization.

Early detection of COVID-19 can significantly decrease the reproduction number and prevent spreading the infection to other people. Unfortunately, this very challenging, since, no visible symptoms are observed within approximately two days before onset the disease.

The most characteristic symptoms of COVID-19 after onset are currently considered a fever, fatigue, and dry cough [16]. Other symptoms, are nasal congestion, pains, aches, colds, dyspnoea, diarrhoea, sore throat, unusual loss of smell and taste, headaches, and sometimes also trembling. Unfortunately, these symptoms, especially visible ones, are useless for early detection. However, even slight changes in related physiological signals can be identified as for detection markers.

In [18], changes in several other biomarkers during COVID-19 were examined. For the evaluation - a wearable ring (in particular, Ourora) was used. The device measured inter alia the heart rate, heart rate variability, respiratory rate, and temperature. An interesting finding was that 38 out of 50 participants developed temperature anomalies, that were measurable before other symptoms. The temperature proved to be statistically more significant than other metrics. Some reports have proven a strong correlation between fever and cardiac rhythm [19]. Increase in HR has been reported as 8.5 beats per minute on average per 1 °C. This correlation is measurable especially during resting heart rate (RHR). A possible reason for this could be that the sensors of wearables are more accurate at resting time. This change was observed not just during the development of COVID-19 but also during the development of other influenza diseases. Unfortunately, other factors also have an impact on the increase in RHR, including short sleep.

In December 2020, a study was published [20] that confirmed that these correlations could also be valid for COVID-19. The authors tried to detect the disease in the pre-symptomatic stage using the smartwatches that are available on the market (only those gathered by Fitbit were selected). The data were collected on a per-minute basis. Anomaly detection was still relatively low (63 %) for COVID-19 cases. The results were validated on a database containing 114 samples, that were available online. The features were based on heart rate and the number of steps.

A few studies have also a tempted to detect infection from sleep activity patterns measured by wearable devices. The data for the experiment were collected from approximately 5300 subjects using smartwatches such as Fitbit, Apple Watch, or Garmica. Thirty-two of the participants were diagnosed with COVID-19. In 63 % of COVID-19 cases, some anomalies were found in the records, which was significantly more frequent than in the rest of the data. The authors proposed two approaches: the abnormal resting heart rate (RHR-Diff) - based approach and the heart rate-to-steps (HROS-AD algorithms) - based approaches. The RHR-Diff approach was based on standardized residuals and a 28-day sliding window. The second approach used, the HROS was computed by dividing the heart rate by the number of steps. One-hour intervals were compared using Gaussian density estimation (HROS-AD algorithm). The CuSum method was an online version based on the cumulative statistics. For this algorithm, a 28 day time-frame was adopted and the deviations of the elevated residual RHRs were evaluated. Unfortunately, the specificity of the methods was not provided, which is very important for this kind of study.

The above-described works focused on wearables at the person-level. Some studies have been built on these works, extending them to the analysis of the development of the pandemic trend, the so-called crowd-level analysis.

One such work is presented in [19]. The authors proposed a simple system for alerting physiological anomaly detection based on variance in sleep patterns and RHR computed from photoplethysmography (PPG) wearable data. The records of 1.3 million users were collected from Huami devices. The deviation from the mean and average values was taken into consideration for the analysis. Although, this system was not good enough for a common disease detection, a methodology for the pandemic was developed using this system. For this reason, CDNet architecture was used, which is a heterogeneous neural network regression model. CDNet consists of two neural networks: CatNN and DenNN. It contains sparse categorical features (holiday activity, season, and weather) and dense numerical features (historical physiological anomaly rate, historical officially reported COVID-19 rate, and active user density). The simulation was performed for North China, Central China, South China, and South-Central Europe. The detected physiological anomaly rate was compared with COVID-19 infection rate using Pearson's correlation. The highest correlation among the set of Chinese cities was observed for Foshan as 0.81. Average value among all cities was 0.68. Nevertheless, the capability to predict such disease shows some limits, in particular there is suspicion that to local events change people's common rhythm, and there were problems regarding individual variability.

In another work, the researchers analysed the influence of respiration rate together with oxygen saturation (SpO2) gathered from wearable smartwatches on COVID-19 [21]. The statistical analysis was evaluated with chi-square distribution and independent t-test based measures on pre-selected 208 cases. The authors demonstrated that there were no significant differences between gender and IoT factors based on chi-square distribution.

In [22], the methodology was based on the wearable device (Empatica E4) and on neural networks. The device allows analysis of a broader spectrum of signals, for example the galvanic skin response (GSR), the inter-beat interval (IBI), skin temperature, pulse oximeter, and blood pressure. A questionnaire with information about the occurrence of symptoms was also included to obtain. Parameters such as age, gender, weight, height, habits, and addiction to smoking and drinking. All data were used during the NN training [22]. The cohort included 87 subjects: 30 were HC, 27 asymptomatic COVID-19 cases and 30 symptomatic cases. The data were divided into 15 s windows. However, the details of the procedure are not provided in the article. The CovidDeep - a four-layer neural network, was introduced and used for the detection of COVID-19 cases. The architecture consists of data pre-processing, synthetic data generation with the TUTOR framework, architecture pre-training, grow-and-prune synthesis with a decision tree (DT) and random forest. This combination of data contains some hardly collectible data in reality. Common smartwatches do not provide as many spectra of data as Empatica offers, and, the number of its users is significantly lower than, that of Fitbit or Apple Watch [22]. In another study, symptoms such as respiration rate detected by wearables were analysed [23]. The cohort of 271 (81 positive cases and 190 negative cases) was evaluated with the WHOOP strap algorithm. Here, the median value of respiration per minute was collected during night which was regarded as the respiration rate. A total of 20 % of COVID-19 subjects were recognised 2 days before the clear onset of the disease, while 80 % of COVID-19 subjects were unfortunately detected 3 days after the onset. These results were achieved with the use of a gradient boosting classifier. The outcome of 80 % is a relatively high number; however, the most important in screening tests, are early symptoms - during the prodromal stage [23].

The variance in respiration rate was examined together with heart rate and heart rate variability in [24] using Fitbit. The methodology used achieved 0.77 ± 0.03 area under the curve (AUC) trained on NN. Nonetheless, the sensitivity was 47 % and specificity 95 %. The number of positive participants was 1181, and 13662 were negative. The following parameters were extracted: Shannon entropy of the nocturnal RR series, the mean nocturnal heart rate during non-Rapid Eye Movement (NREM) sleep, and the estimated mean respiration rate during deep sleep. During pre-processing, the physiological signals were normalised using z-score. A combination of parameters, including age, gender, and BMI, were fed into the convolutional neural network. It was found that the heart rate in combination with respiration rate increased during the illness, and the heart rate variability decreased.

An interesting work [25] studied differences between cases: the outbreak of the pandemic (6270), the actual COVID-19 cases (230), and influenza non-COVID-19 cases (426) by using smartwatches. The authors found a higher intensity and variety in symptoms in COVID-19 cases compared to normal influenza. A few symptoms, as chest pain, anosmia, and shortness of breath, were commonly observed in COVID-19 cases, of course, these are not early markers of the disease. Based on the wearables' analyses, a reduction in the number of daily steps for a longer period was treated as the symptom of "long COVID". Hence, the wearables could serve as devices for monitoring recovery after COVID-19.

Moreover, [26] some broader view of the usage of artificial intelligence (AI) and internet of things for creating the support system methodology in hospitals. The combination of these techniques could be useful in disease diagnosis, treatment, and management. During the COVID-19 pandemic, it will found application, for example, in data storing of PCR and tests, from imaging technologies. In this work, the authors used the machine learning algorithms, such as Random Forest (RF), Naive Bayes (NB), and Support Vector Machine (SVM). The dataset was reused and gathered under laboratory conditions in 2020 [27]. The data contained a record of 600 patients. Among these, 80 were diagnosed positively with COVID-19. The number of features was 18, and they included blood parameters. In this work, the authors achieved the best results for SVM: 95 % accuracy, 94 % F1, 95 % of precision, 95 % recall, and 95 % for the Area Under Curve. As the pre-processing step, normalisation and feature selection were used. Nevertheless, there is no mention of the stage of the disease at which the data were gathered.

To date, mainly smartwatches and smart masks [28] have been used for the early detection of COVID-19. Moreover, some new solutions for measuring physiological parameters have recently been introduced, which are described below.

The characteristics of the ear-related area have allowed researchers to create a photoplethysmography-based device for measuring oxygen saturation in-ear SpO2 [29]. It is as effective as finger pulse oximetry, with an even faster average response. This device is suitable for monitoring COVID-19 and post-COVID-19 cases such as the possible occurring of hypoxemia related to breathlessness declination [29]. Furthermore, the usage of a headset wearable device allowed coughing to be monitored, which, however is not a suitable symptom for early detection [30].

The Biovitals Sentinel platform is an academic project, that uses armband biosensors (Everion) to detect COVID-19. The monitored parameters are skin temperature, respiratory rate, blood pressure, pulse rate, blood oxygen saturation, and statistics regarding daily activities [31]. The introduced platform could serve as a source of data for potential support system methodologies.

In [32], the authors used a skin sensor that was equipped with an accelerometer and a temperature sensor, and can be placed on the throat. A remarkable advantage of this solution was that it could wirelessly register cough frequency, intensity, and duration, as well as the respiratory rate. The study also included heart rate measurements. This solution contributed to a continuous physiological monitoring system and to the data analytical part. Again, this system cannot be used to warn against infection in the early stages. Upon the onset of coughing, the person already knows about the infection and can change the behaviour to protect others.

In summary, there are currently several approaches that use wearables to detect COVID-19 and there is a big promise of that this technology can offer accurate early disease detection and screening to a broader population in an easy and non-invasive way. A methodology, based on a medical device - Empatica, that used advanced sensors was introduced. Unfortunately, this device is relatively expensive and is not widespread among the population.

Finding a methodology based on cheaper and commercially successful smartwatches, such as Fitbit or Apple Watch, could have a significantly higher impact. Several previous works suggest that the heart rate and activity can indicate the development of COVID-19 and flu in general in their early stage. It is noteworthy that the use of wearables for telemonitoring has both advantages and drawbacks. On the

TABLE 1. The summary of the state-of-the-art.

Citation	Main aim	Device	Kind of data gathered	Number of the dataset	Accuracy, efficiency	Machine learning method	Comments
[18]	Statistical analysis of daily temperature for COVID-19 disease and creating digital biomarkers	Ourora ring	Temperature	50 COVID-19 cases	38/50 patients exhibited some temperature anomalies before the onset of the disease	Threshold based on min/max temperature record after z-score, Statistical evaluation: nonparametric Kruskal Wallace test, with Tukey–Kramer post hoc comparison	More wearables should include temperature sensors
[19]	Predicting the epidemic trend including anomaly detection with COVID-19 infection rate	Huami (ACC, PPG)	Heart rate, sleep data	1.3 mln participants	The highest Pearson correlation for Chinese cities: Foshan 0.81, average 0.68	CDNet (CatNN, DenNN)	The simulation provided for North, Central, South China and South-Central Europe
[20]	Anomaly detection of COVID-19 disease in	Limited to Fitbit	Heart rate, sleep disorders, number of steps	73 HC, 32 COVID-19 cases, 15 Influenza	63 % anomaly detection in COVID-19 cases	Developed algorithms: RHR-Diff, HROS-AD, CuSum	It is anomaly detection evaluated on COVID-19 disease cases without considering classification problem
[21]	Correlation of wearables related data with gender and IoT factors	Lack of detailed informations	Respiration rate, oxygen saturation	208 cases	no significant differences between IoT factors and gender	Chi-Square distribution and independent measures t-Test	There should be o difference of future created support system methodologies between the population according to the analysed factors
[22]	Evaluation of COVID-19 disease based on Empatica device	Empatica E4	GSR, IBI, skin temperature, pulse oximeter, blood pressure monitor + questionnaire	24 HC, 46 COVID-19 cases (22 asymptomatic, 24 symptomatic)	98,1 \% accuracy	COVIDDeep	The data contains self-assement done by patients, the pre-processing step is not clear. The results are obtained with the rare using medical device – Empatica.
[23]	Detection of COVID-19 disease	WHOOP strap	Respiration rate	81 COVID-19 cases, 190 HC	20 % COVID-19 subjects recognised before the onset, 80 % cases 3 days after onset	Gradient Boosting classifier	80 % is well results of accuracy, however, the target is to detect disease before the clear onset.
[24]	Assessment of the need of hospitalization of COVID-19 patients based on respiration rate, heart rate, rate variability and also age, gender, BMI	Fitbit	Respiration rate, heart rate, heart rate variability	2754 COVID-19 cases	0.77 +/- 0.03 AUC, sensitivity 47 %, specificity 95 %	Computed parameters: Shannon entropy of the nocturnal RR series, the mean nocturnal heart rate during deep sleep, pre-processing: transformation into z-score, algorithm: CNN	Some extra parameters were provided during training – among others: age, gender, BMI. Heart rate together with respiration rate increasing during illness, heart rate variability is decreasing.
[25]	Comparision of COVID-19 disease in the early outbreak, later outbreak and also with Influezna	Fitbit	Self-report data, RHR, step counts, nightly sleep hours	41 COVID-19 cases 42685 self-reported flu, 1265 pre-pandemic COVID-19	Statistical differences in tests	Statistical evaluations	The authors demonstrate the higher intensity and variety in symptoms for COVID cases than for normal flu.

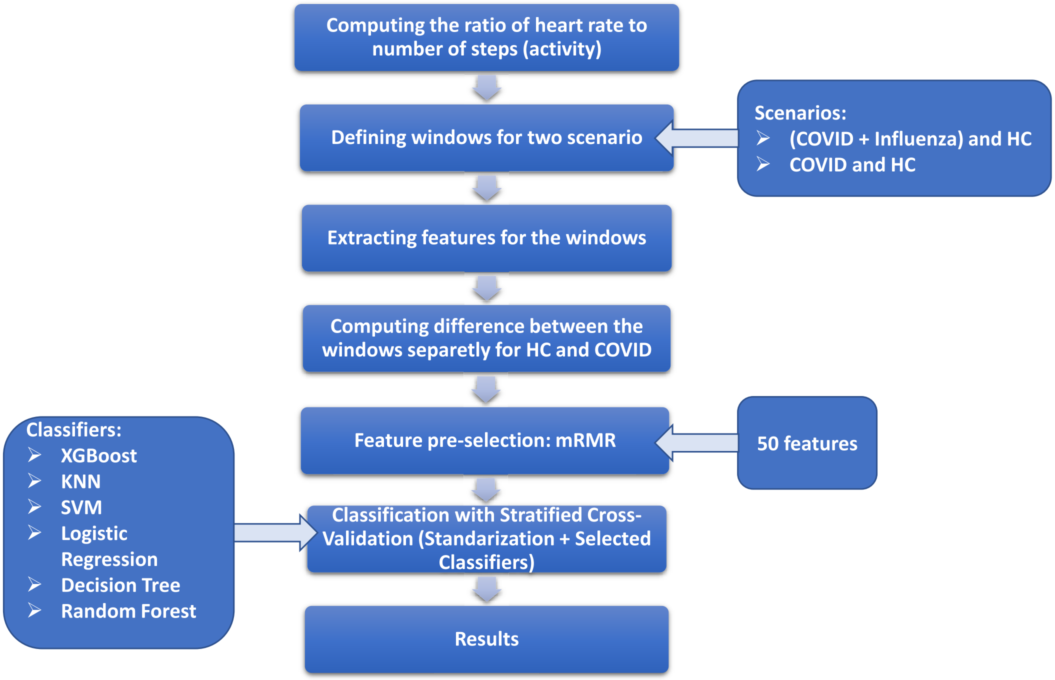

FIGURE 1. Scheme of the experiment.

one hand, they can present economic solutions for screening tests and continuous objective monitoring and they are easily accessible. On the other hand, they are not accurate as imaging technologies (X-ray or CT) or PCR testing technologies [33].

III. EXPERIMENT

The main objective of this research was to develop a methodology for the detection of COVID-19 in its early stages using wearable electronic devices. Notably, we developed s that considered the character of the development of the disease, that is, the incubation period and contagiousness of the disease. Moreover, the use of the most commonly gathered signals (heart rate and number of steps) by smartwatches allows us to potentially apply this solution as a broad screening test.

The research follows up mainly on paper [20] where regular fitness smartwatches were used for collecting data from 4642 volunteers in total. Of these, 114 of them were later diagnosed as COVID-19 positive. This experiment used the data from this paper [20]. In [20], the experiment consisted of three steps: (a) signal pre-processing of the data, (b) machine learning, and (c) statistical evaluation of the accuracy. The detailed scheme of the experiment is presented in Fig. 1. First, the ratio of the heart rate to the number of steps was computed. This was used as a pre-processing step to obtain a more informative feature. The experiment was designed for two scenarios: classification between COVID-19 and healthy controls (HC), and classification between COVID-19, HC, and influenza. Next, the feature extraction was carried out for two-time windows, and then the difference between them was computed. We applied the minimum redundancy maximum relevance (mRMR) based on the feature preselection. The number of features was set to 50 to select the most valuable of them and to facilitate the classifier's learning process. The next step was the classification with XGBoost, k-NN, SVM, logistic regression, decision tree, and random forest. The results were evaluated using metrics.

The data consisted of records of the number of steps and heart rate. These values were measured every minute. To partially suppress the effect of the person's activities, the heart rate value was first divided by the number of steps, and these normalised values were used for further processing.

The experiment was designed to compare values measured from sensors of COVID-19 positive persons and distinguish these values from healthy controllers. Based on previous works, there is an assumption that the infected human body has a different response in terms of heart rate in the prodromal stage of the diseases, even few days before the onset of the disease [18], [20], [24]. The highest changes in the elevated resting heart rate (RHR) are registered two days before the onset of the disease [25]. Some studies have also confirmed that, especially in the resting time, heart rate is slightly higher than in the case of a healthy person. Unfortunately, biological systems are not fully deterministic, and in particular the heart rate does not responding equally to the same activity or situation. For this reason, the data were averaged for a one minute period. Further, different persons have different responses of the body - for example, based on their fitness, age and weight, etc. Thus, we did compare absolute values between all persons, rather, we compared changes in each person during the healthy state and near the onset of the disease.

To construct a system that compares changes in the body responses of each person, we defined two time frames in the collected data. The detailed procedure of this step can be found in Subsection III-B. Next, the set of features was extracted for each window. Subsequently, the difference was calculated for the set of features extracted between the later and previous window.

Unfortunately, there is still a lack of data in this domain. Hence, we preferred to obtain insight into the data and use statistics to understand the features. The features used for the experiment were first pre-selected to suppress possible over-fitting. The method used for the selection of the features was the minimum redundancy maximum relevance (mRMR). The number of features was 50. To evaluate the resulting accuracy, we used the classification of COVID-19 positive or COVID-19 negative classes. For statistical evaluation, 10-fold Stratified Cross-Validation was used. We also included Standardization. For evaluation of the model, we compared results using different machine learning algorithms including XGBoost, k-Nearest Neighbour (k-NN), Support Vector Machines (SVM), Logistic Regression (LR), Decision Tree (DT), and Random Forest (RF). For further statistical evaluation, the Mann-Whitney U test was used. This method allows for checking whether is a statistical difference in the distribution of the analyzing groups. To minimize the number of type I errors, the false discovery rate (FDR) was calculated.

A. TRAINING DATA

The data used for this research cames originally from [20]. For the study, the data were gathered using the wearable devices and the application MyPHD app. The wearables used included Fitbit, Apple Watch, Garmin Watch, Oura Ring, BioStrap, Masimo Pulse Oximeter, Empatica, Motiv Ring, and others. Data regarding steps, heart rate, and partly sleep records were analysed. The data reused in this paper were the records of steps per minute and heart rate per second, while, the sleep data were omitted cause of their limited numbers. In this study, 5262 participants were enrolled. Among these, 114 participants were diagnosed with COVID-19. To balance the data, 34 HC and 27 COVID-19 cases, including 7 influenza cases, were utilized.

B. FEATURE EXTRACTION

The data gathered from wearable devices measure physiological data which are continuous time-series records. Although deep learning has shown great results in end-to-end learning, we have a relatively limited amount of training data. Thus, we manually extracted features characterising time-series signals [34]. The inspiration for the feature extraction of physiological signal features was taken from [34]–[36], and [37]. These covered the most frequently used features for physiological signals. The number of measured samples was still quite limited. These samples were records of a relatively long period, so we decided to use hand-crafted features for the feature extraction. To be sure about this decision, we evaluated also NNs, such as long-short term memory and 1-D convolution neural network. Nevertheless, as shown later, there were no observable significant results. Three types of features were extracted: (a) temporal, (b) statistical, and (c) in the spectral domain. For the temporal domain, auto-correlation, centroid, mean absolute differences, mean differences, median absolute differences, median differences, distance, the sum of absolute differences, total energy, entropy, peak-to-peak distance, the area under the curve, absolute energy, maximum peaks, minimum peaks, slope, and zero-crossing rate were taken into account. Subsequently, from the statistical domain, the following features were extracted: histogram, inter-quartile range, mean absolute deviation, median absolute deviation, root mean square, standard deviation, variance, empirical cumulative distribution function (ECDF) percentile count, ECDF slope, kurtosis, skewness, maximum, minimum, mean, median, ECDF, and ECDF Percentile. The spectral domain represents features such as - fast Fourier transform (FFT) mean coefficient, wavelet absolute mean, wavelet standard deviation, wavelet variance, spectral distance, fundamental frequency, maximum frequency, median frequency, spectral maximum peaks, maximum power spectrum, spectral centroid, decrease, kurtosis, skewness, spread, slope, variation, spectral roll-off, roll-on, human range entropy, Mel-frequency cepstral coefficients (MFCC), linear prediction cepstral coefficients (LPCC), power bandwidth, spectral entropy, wavelet entropy, and wavelet energy. The feature extraction was performed thanks using the Python package: tsfel [38]. By adding a few temporal features, the work builds on packages such as FATS, CESIUM, TSFRESH, and HCTSA.

A scheme for - feature extraction is presented in Fig. 2. For extracting the healthy control samples from the HC cohort, the set of features was computed for two windows (p_h - earlier and p_c - later window). The fixed-sized of the windows was set, and we specified the spacing between them. In the next step, the difference between a set of features for the aforementioned windows was calculated. Where:

- The vector of features extracted from the earlier healthy state is expressed as f_{HC1}
- The vector of features extracted from the later healthy state is expressed as f_{HC2}
- The final vector for HC is expressed as: $f = f_{HC2} - f_{HC1}$
- The end of earlier healthy state window is expressed as t_{HC1}
- The beginning of later healthy state window is expressed as t_{HC2}
- The Spacing between windows is expressed as: $Spacing = t_H C2 - t_H C1$

The scheme of the feature extraction for HC is presented in Fig. 2.

However, to extracting COVID-19 cases, a similar procedure was carried out. Otherness was represented by taking into consideration the onset of the disease. The shift in the computation of the windows was defined as the ability to detect the disease in the prodromal stage. This is because the

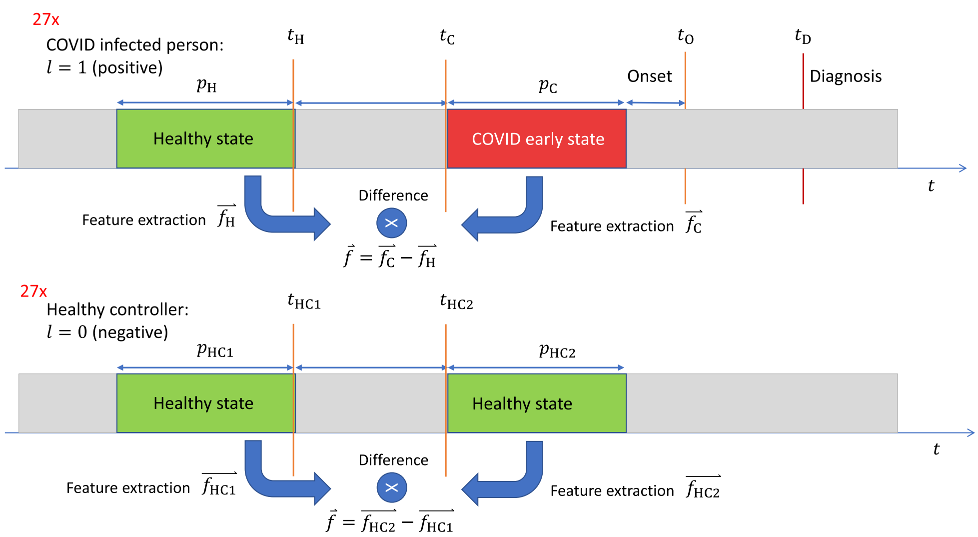

FIGURE 2. Scheme of feature extraction for HC and COVID-19 cases.

highest peak of contagiousness of this disease is registered two days before the onset of the disease. The next steps were the same as in the case of HC sample extraction. The scheme of the described steps is shown in Fig. 2.

Where:

- The vector of features extracted from the healthy state is expressed as f_H
- The vector of features extracted from COVID early state is expressed as f_C
- The final vector for COVID-19 case is expressed as: $f = f_C - f_H$
- The end of earlier healthy state window is expressed as t_H
- The beginning of COVID-19 window is expressed as t_C
- The beginning of the onset of symptoms is expressed as t_O
- The diagnosis of COVID-19 is expressed as t_D
- The Onset is expressed as: $Onset = t_C + p_C$
- The Spacing between windows is expressed as: $Spacing = t_C - t_H$

C. EVALUATION

To evaluate the quality of the algorithms several metrics were used: accuracy, sensitivity, specificity, and the Matthews correlation coefficient. From a clinical perspective, sensitivity, and specificity are only as important as the accuracy [39]. Sensitivity is defined as the ratio of positive cases regarded by the algorithm as positive cases to the whole set of real positive cases. Specificity is the ratio of negative cases classified as negative to the whole set of real negative cases [40]. The Matthews correlation coefficient is mostly dedicated to imbalanced datasets.

The equations for the metrics are presented below:

Accuracy:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{1}$$

Sensitivity:

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

Specificity:

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{3}$$

The Matthews correlation coefficient (MCC):

$$\text{MCC} = \frac{\text{TP} * \text{TN} - \text{FP} * \text{FN}}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}}$$

D. MACHINE LEARNING AND STATISTICAL EVALUATION

To identify the most informative features that could distinguish the disease, a univariate test was performed. In this case, a non-parametric Mann-Whitey U-test was used. For this purpose, we selected features, that were significantly different from each other. For the use of this test, it is not necessary to assume that the values are normally distributed. It is suitable for small datasets. One of the obstacles of the Mann-Whitney U-test is that it is endangered by error-type-I [41]. For this reason, the Benjamini-Hochberg procedure was used to controll the false discovery rate (FDR) [42]. However, FDR is a strong criterion for a small cohort, dedicated to multiple hypothesis testing.

1) MACHINE LEARNING ALGORITHMS

In this work, a few supervised machine learning algorithms were used. These were: SVM [43], logistic regression [43], k-NN [43], decision tree [43], random forest [43], and XGBoost [43]. The advantage of SVM is that it is generally robust to over-fitting [44] and has a relatively good generalization capability [45]. The principle of its operation is based on non-linearly transformed training data into a higher dimension. The decision boundary is chosen by the algorithm that converges to the best separating hyper-plane.

The hyper-plane is found by support vectors and determined by them, the so-called margins. The SVM tries to find the maximal marginal hyper-plane (MMH), which allows the correct classification of unseen data classification, correctly. The SVM works is based-on neural networks [44], [46].

The k-Nearest Neighbor classifier belongs to lazy and non-parametric classifiers. K is the number of the neighbours, and the test datasets are compared with the training dataset. This classifier could be vulnerable to over-fitting because of existing noise. Including match weighting, the attributes can solve this problem by considering a change in distance metrics [44] to calculate the distances between the sample and the neighbours.

One of the simplest supervised machine learning algorithms is logistic regression (LR), which is a classification algorithm. The output is interpreted as a probability [45]. Despite the non-complicated principle of working, it could achieve good results [46]. However, it could suffer from multicollinearity [45].

A decision trees, Random Forest, and XGBoost are a group of tree classifiers [44].

There are a few types of decision trees. The most commonly used are C4.5, ID3, and CART. In this study, CART was applied for the classification purposes [47]. CART is a non-parametric algorithm. Prominent advantages of CART are its ability to deal with missing values and use pruning (post pruning). The decision-making indicator uses the Gini diversity index [48]. Decision trees can deal with linearly inseparable data. The disadvantage of this algorithm is that it is difficult to manage high dimensional data, and is vulnerable to overfitting [45].

Random forest is an extended version of the decision tree. This algorithm trains many decision trees and defines classes based-on voting. The main advantage is that it is robust to noise and does not overfit. Hence, it can be regarded as a fast methodology. However, it slows down with an increase in the number of trees [45].

XGBoost belongs to the tree gradient boosting system. This algorithm can achieve state-of-the-art results for structured data. This algorithm characterises faster computing and also uses regularization techniques - XGBoost also uses shrinkage methodology and feature (column) sub-sampling to protect against overfitting.

The algorithm parameters were optimised using a grid search. For XGBoost, we optimized the following parameters: subsample ratio of the training instances, control the balance of positive and negative weights, minimum sum of instance weight (hessian) needed in a child, maximum depth of a tree, step size shrinkage used in the update to prevent overfitting, gamma minimum loss reduction required to make a further partition on a leaf node of the tree, set of parameters for subsampling of columns, and subsample ratio of columns for each level.

For k-NN, the parameters that were optimised were the: number of neighbours, type of metric used for the evaluation, and weight function used in prediction.

TABLE 2. The scenario of carried out experiments.

Cases	Len_window	SHIFT	Spacing
COVID+HC+Influenza	5	2	7
	7	2	7
	10	2	7
COVID+HC	5	2	7
	7	2	7
	10	2	7

SVM was evaluated using several values: C, regularization parameter, the degree of the polynomial kernel function, if chosen and gamma kernel coefficient for radial basis function, sigmoid, and polynomial. The kernel types used in the algorithm were: linear, polynomial, radial basis function, sigmoid, and precomputed.

The logistic regression was optimised by the parameters: C - the inverse of regularization strength, penalty normally used in the penalisation, and solver - a type of algorithm which is using in the optimization problem.

Whereas, the parameters which were used for optimising the Decision Tree was: the minimum number of samples necessary to split an internal node, the minimum number of samples necessary at a leaf node, the number of features taken into consideration for the best split, the maximum depth of the tree, types of weights during balancing the data.

The parameters that were used for the random forest were as follows: the number of trees in the forest, the minimum number of samples necessary to split an internal node, the minimum number of samples required to be at a leaf node, the number of features taken into consideration for the best split, the maximum depth of the tree, and types of weights during balancing the data.

IV. RESULTS

The machine learning algorithms were trained for two cases: A) for cohort containing COVID-19 cases and HC, and B) for cohort containing COVID-19 cases + Influenza and HC. The size of the spacing between windows was fixed to 7 days, and the SHIFT was equal 2 days (please, check the designation in Fig. 2). The summary of the parameters used for experiments is presented in Table 2. Based on the statistical evaluation of the extracted features, we conducted the Mann–Whitney U-test with FDR correction. The results for the two scenarios with Fig. 2 5-day windows are shown in Table 3 and Table 4. The distinction between those two groups is provided in the view of checking samples from two different distributions that is with pr without Influenza cases. The features distributions containing only people having COVID-19 versus HC should be different from the distribution of the features containing COVID-19 cases and influenza cases versus HC. The same dependency was compared. The number of extracted features was 381 and their descriptions can be found in [49].

The features, which passed the Mann-Whitney U test for the cohort containing 27 people suffering from COVID-19 disease and 27 HC are the following: sets of MFCC,

TABLE 3. Mann-Whitney U-test including also FDR correction for the cohort of COVID-19 disease and HC.

Features	pval	pval_FDR	Features	pval	pval_FDR
MFCC_11	0.0045	0.3546	Zero crossing rate	0.0283	0.3546
FFT mean coefficient_117	0.0070	0.3546	LPCC_3	0.0297	0.3546
FFT mean coefficient_189	0.0070	0.3546	LPCC_9	0.0297	0.3546
Spectral slope	0.0102	0.3546	Slope	0.0297	0.3546
FFT mean coefficient_43	0.0134	0.3546	Min	0.0304	0.3546
FFT mean coefficient_254	0.0134	0.3546	FFT mean coefficient_21	0.0309	0.3546
FFT mean coefficient_233	0.0140	0.3546	FFT mean coefficient_243	0.0321	0.3546
Maximum frequency	0.0160	0.3546	FFT mean coefficient_175	0.0333	0.3546
Spectral roll-off	0.0160	0.3546	FFT mean coefficient_144	0.0346	0.3546
FFT mean coefficient_130	0.0174	0.3546	MFCC_7	0.0346	0.3546
MFCC_0	0.0174	0.3546	FFT mean coefficient_163	0.0374	0.3546
FFT mean coefficient_49	0.0189	0.3546	Spectral centroid	0.0374	0.3546
FFT mean coefficient_149	0.0189	0.3546	LPCC_0	0.0388	0.3546
FFT mean coefficient_0	0.0206	0.3546	FFT mean coefficient_249	0.0403	0.3546
FFT mean coefficient_202	0.0215	0.3546	Median frequency	0.0403	0.3546
MFCC_2	0.0224	0.3546	ECDF Percentile_0	0.0409	0.3546
FFT mean coefficient_37	0.0243	0.3546	FFT mean coefficient_39	0.0418	0.3546
FFT mean coefficient_247	0.0243	0.3546	FFT mean coefficient_185	0.0418	0.3546
MFCC_9	0.0243	0.3546	FFT mean coefficient_242	0.0418	0.3546
FFT mean coefficient_167	0.0263	0.3546	FFT mean coefficient_6	0.0434	0.3546
FFT mean coefficient_188	0.0263	0.3546	FFT mean coefficient_57	0.0450	0.3546
Spectral kurtosis	0.0263	0.3546	Signal distance	0.0450	0.3546
Fundamental frequency	0.0274	0.3546	FFT mean coefficient_235	0.0467	0.3546
Spectral skewness	0.0274	0.3546	FFT mean coefficient_154	0.0484	0.3546
Histogram_5	0.0283	0.3546	FFT mean coefficient_194	0.0484	0.3546

TABLE 4. Mann-Whitney U-test including also FDR correction for the cohort of COVID-19 disease, Influenza and HC.

Features	pval	pval_FDR	Features	pval	pval_FDR
FFT mean coefficient_163	0.0024	0.2389	FFT mean coefficient_56	0.0212	0.2421
FFT mean coefficient_243	0.0031	0.2389	Min	0.0216	0.2421
FFT mean coefficient_189	0.0032	0.2389	FFT mean coefficient_236	0.0225	0.2421
FFT mean coefficient_202	0.0040	0.2389	FFT mean coefficient_53	0.0238	0.2421
FFT mean coefficient_149	0.0059	0.2389	FFT mean coefficient_29	0.0245	0.2421
FFT mean coefficient_242	0.0065	0.2389	FFT mean coefficient_15	0.0252	0.2421
Spectral kurtosis	0.0065	0.2389	FFT mean coefficient_165	0.0259	0.2421
FFT mean coefficient_182	0.0070	0.2389	FFT mean coefficient_247	0.0259	0.2421
FFT mean coefficient_167	0.0075	0.2389	FFT mean coefficient_152	0.0267	0.2421
Maximum frequency	0.0094	0.2389	FFT mean coefficient_185	0.0267	0.2421
Spectral roll-off	0.0094	0.2389	Spectral centroid	0.0267	0.2421
FFT mean coefficient_254	0.0101	0.2389	FFT mean coefficient_134	0.0299	0.2586
FFT mean coefficient_117	0.0118	0.2389	Slope	0.0299	0.2586
Histogram_5	0.0122	0.2389	Median frequency	0.0316	0.2615
FFT mean coefficient_25	0.0126	0.2389	Spectral spread	0.0316	0.2615
Zero crossing rate	0.0130	0.2389	FFT mean coefficient_125	0.0333	0.2647
FFT mean coefficient_233	0.0138	0.2389	FFT mean coefficient_155	0.0333	0.2647
FFT mean coefficient_194	0.0147	0.2389	FFT mean coefficient_250	0.0352	0.2738
MFCC_11	0.0152	0.2389	FFT mean coefficient_50	0.0372	0.2832
FFT mean coefficient_43	0.0157	0.2389	FFT mean coefficient_160	0.0382	0.2851
FFT mean coefficient_39	0.0162	0.2389	FFT mean coefficient_230	0.0392	0.2872
FFT mean coefficient_150	0.0162	0.2389	FFT mean coefficient_175	0.0413	0.2874
Spectral skewness	0.0162	0.2389	FFT mean coefficient_255	0.0413	0.2874
FFT mean coefficient_143	0.0167	0.2389	FFT mean coefficient_222	0.0424	0.2874
FFT mean coefficient_0	0.0172	0.2389	FFT mean coefficient_229	0.0447	0.2874
FFT mean coefficient_130	0.0172	0.2389	FFT mean coefficient_231	0.0458	0.2874
FFT mean coefficient_251	0.0172	0.2389	FFT mean coefficient_30	0.0470	0.2874
Spectral slope	0.0177	0.2389	Signal distance	0.0470	0.2874
FFT mean coefficient_235	0.0183	0.2389	FFT mean coefficient_180	0.0483	0.2874
FFT mean coefficient_207	0.0188	0.2389	FFT mean coefficient_196	0.0483	0.2874
FFT mean coefficient_48	0.0212	0.2421	FFT mean coefficient_187	0.0495	0.2874

FFT, mean coefficient, linear prediction cepstral coefficients (LPCC), spectral slope, maximum frequency, spectral roll-off, spectral kurtosis, fundamental frequency, spectral skewness, zero-crossing rate, slope, min, spectral centroid, median frequency, ECDF percentile, and signal distance. The features that passed the test, for the assumed confidence level $\alpha = 0.05$, are presented in Table 3. After applying the FDR correction, none of the checked features passed the test. Nevertheless, it is a strong criterion. The minimum obtained value was 0.3456. The p-value with FDR correction between the two scenarios. The minimum p-value of evaluated features for the scenario with COVID-19, influenza, and HC cases was lower than for the scenario with only COVID-19 and HC cases (0.2389).

TABLE 5. Results for detection of COVID-19 disease for 5-day windows (cohort: 27 HC, 27 COV).

Classifier	Accuracy	Sensitivity	Specificity	MCC
XGBoost	0.71	0.72	0.71	0.46
k-NN	**0.78**	**0.77**	**0.80**	**0.60**
SVM	0.65	0.66	0.65	0.33
Logistic Regression	0.69	0.69	0.69	0.41
Decision Tree	0.50	0.52	0.49	0.01
Random Forest	0.62	0.59	0.66	0.27

TABLE 6. Results for detection of COVID-19 disease for 7-day windows (cohort: 26 HC, 26 COV).

Classifier	Accuracy	Sensitivity	Specificity	MCC
XGBoost	0.67	0.68	**0.66**	0.35
k-NN	**0.68**	**0.73**	0.63	**0.37**
SVM	0.66	0.70	0.63	0.35
Logistic Regression	0.63	0.60	0.64	0.26
Decision Tree	0.54	0.50	0.57	0.08
Random Forest	0.59	0.56	0.61	0.18

For the second scenario, the features that passed the Mann-Whitney U test with the assumed confidence level were the following: the set of FFT mean coefficients, spectral kurtosis, maximum frequency, spectral roll-off, histogram, zero-crossing rate, spectral skewness, spectral slope, min, spectral centroid, slope, median frequency, spectral spread, and signal distance.

Subsequently, in this section, the classification results for the two cohorts are presented in Sections tables 5 to 10. In what follows, we explain the parameter selection should be explained. The shift (marked as Onset) was set to 2 days (Fig. 2). Due to registration, the highest contagiousness peak was registered exactly 2 days before the clear visibility of the patient's onset [14]. However, in view of the incubation period, the space between windows was set to 7 days. The incubation period was 2 to 11 days [8]. The selection of this parameter was set to 7 days, which indicates that the sum of the later windows (based on which the feaures were calculated) and spacing was longer than the registered maximum of the incubation period. The variable remained equivalent to the length of windows, that is, 5-, 7- and 10-days windows were tested. The results were achieved for the following classifiers: XGBoost, k-NN, SVM, logistic regression, decision tree, and random forest.

A. COVID-19 DETECTION

For the cohort containing COVID-19 cases and HC, the results of the classifications are presented in tables 5 to 7. For the 5-day windows (Table 5), the highest accuracy (0.78), specificity (0.77), sensitivity (0.80), and MCC (0.60) were registered for k-NN. We observed that high accuracies were achieved for XGBoost (0.71) and logistic regression (0.69). This was evaluated on 27 HC and 27 COVID-19 cases using stratified cross validation. We also optimised the machine learning models. For the best k-NN, the following parameters were registered as the most optimal: 11 nearest neighbors, as the distance metric was chosen as the Manhattan distance. Moreover, we used weight function, which was computed as weights points by the inverse of their distance.

TABLE 7. Results for detection COVID-19 disease for 10-days windows (cohort: 24 HC, 24 COV).

Classifier	Accuracy	Sensitivity	Specificity	MCC
XGBoost	0.65	0.65	0.67	0.34
k-NN	**0.71**	**0.84**	0.60	**0.46**
SVM	0.67	0.67	0.67	0.36
Logistic Regression	0.70	0.72	**0.68**	0.42
Decision Tree	0.53	0.61	0.46	0.07
Random Forest	0.58	0.53	0.63	0.18

TABLE 8. Results for detection of COVID-19 disease and including Influenza cases for 5-days windows (cohort: 34 HC, 27 COV, Influenza 7).

Classifier	Accuracy	Sensitivity	Specificity	MCC
XGBoost	0.66	0.68	0.65	0.35
k-NN	**0.73**	0.71	**0.76**	**0.49**
SVM	0.71	0.75	0.68	0.45
Logistic Regression	0.69	**0.76**	0.62	0.40
Decision Tree	0.52	0.50	0.55	0.05
Random Forest	0.56	0.56	0.56	0.13

The outcomes of the classification for the 7-day windows are shown in Table 6. The results obtained were lower in comparison to those from Table 6. The k-NN classifier had the highest accuracy (0.68), sensitivity (0.73), and MCC (0.37), whereas, the highest specificity was reported for XGBoost (0.66).

The results with a window of 10-days are presented in Tab. 7. Once again, the k-NN achieved the highest results in accuracy (0.71), sensitivity (0.84) and MCC (0.46). For the logistic regression, the highest results were obtained for specificity (0.68).

B. COVID-19 AND INFLUENZA DETECTION

The second dataset contained the COVID-19 cases, people with Influenza, and HC. COVID-19 disease and influenza were treated as one class and HC as the second. This time, the data were balanced. For the 5-day windows, the cohort contained 31 HC, 24 COVID, and 7 influenza cases (Table 8). The highest accuracy was obtained in this case for k-NN (0.73), and it also had, the best-recorded specificity (0.76) and MCC (0.49). The sensitivity was highest with logistic regression (0.76). The following parameters were identified as the most optimal for k-NN: three nearest neighbors, Euclidean distance as the best distance, and every point in the neighborhood were weighted equally. The most optimal parameters for logistic regression were the L2 penalty, and the "saga" algorithm for the optimization, and inverse of regularization strength was C=464. Furthermore, we evaluated the second dataset using a 7-day window length. This time, the best results were obtained with logistic regression (0.71). Specificity (0.68) and MCC (0.45) were also highest for this classifier. The sensitivity (0.89) was the best for logistic regression. Lastly, the best outcome in accuracy (0.73), sensitivity (0.82), and MCC (0.50) for the 10-day window length was achieved with k-NN. The specificity (0.66) was best with logistic regression.

V. EXPERIMENT

In this paper, we introduced a methodology for early COVID-19 detection in the prodromal phase based on records

TABLE 9. Results for detection of COVID-19 disease and including Influenza cases for 7-days windows (cohort: 33 HC, 26 COV, Influenza 7).

Classifier	Accuracy	Sensitivity	Specificity	MCC
XGBoost	0.63	0.64	0.62	0.28
k-NN	0.70	**0.89**	0.51	0.44
SVM	0.68	0.80	0.57	0.39
Logistic Regression	**0.71**	0.74	**0.68**	**0.45**
Decision Tree	0.54	0.50	0.57	0.08
Random Forest	0.60	0.55	0.65	0.21

TABLE 10. Results for detection of COVID-19 disease and including Influenza cases for 10-days windows (cohort: 31 HC, 24 COV, Influenza 7).

Classifier	Accuracy	Sensitivity	Specificity	MCC
XGBoost	0.63	0.63	0.62	0.27
k-NN	**0.73**	**0.82**	0.64	**0.50**
SVM	0.68	0.72	0.65	0.39
Logistic Regression	0.67	0.67	**0.66**	0.35
Decision Tree	0.52	0.54	0.50	0.04
Random Forest	0.59	0.58	0.59	0.18

from smartwatches. Machine learning techniques are used for the analysis of the signal. The data originated from paper [20], where only a portion of the data was used. We selected 27 COVID-19 and 7 influenza samples as a positive class, and 34 healthy controllers as a negative class. In the original paper, 32 COVID positive samples, 15 influenzas, and 72 healthy controllers were used. The reason for this is that some of the records missed some important data parts that were needed for the comparison of the signals. The size of the data is not very large, however, it is currently the largest public dataset of this kind available.

A few scenarios were tested with various fixed parameters, that is, shift between windows, spacing between windows, and changeable length of windows. These parameters were fixed by considering the incubation periods and the highest interval of contagiousness. A wrong selection can result in the risk of considering people already in quarantine. The purpose of this work was to concentrate on the prodromal stage analysis, and the parameters were selected accordingly.

The set of extracted features covers several domains: temporal, statistical, and spectral. Thanks to this, it could represent a larger range in the variability of the analysed signals. Based on the statistical analysis, the most informative features were those related to changes in frequency and the spectral. This could be explained by the fact that the final features were computed as the difference of the extracted features for two windows - later and earlier. The statistical evaluation showed that features such as MFCC, FFT, spectral-based, histogram, and LPCC differed significantly in the comparison of HC and COVID-19 cases. However, for the scenario with (COVID-19, Influenza) and HC, the following features were relevant: FFT, spectral-based, MFCC. These features seem to contradict. What is observable, under stricter requirements (after FDR correction), none of the p-values for the features were below 0.05. One of the most significant differences was in the p-value, as the p-value with the FDR correction was lower for the cohort with influenza. This could indicate differences between the intensity of the symptoms for influenza and COVID-19 cases, this same p-value should be various. This revealed changes in the patterns of some frequencies (activities) between the HC and COVID-19 cases. Besides an increase or a decrease in parameter detection, the results showed a higher accuracy registered for k-NN in most of the cases. Good outcomes were achieved also with the logistic regression and for cohort containing only COVID-19 and HC, as well also for XGBoost. The success of the logistic regression indicates the ease of the recognition of the samples containing disease cases and HC based on the extracted features. The best results obtained with k-NN suggest existing aggregations and sufficient boundaries between them. The results obtained with XGBoost may be reached due to the more complex nature of this classifier. Nevertheless, this classifier might have been slightly overfitted. In the context of analysing the length of the windows, cohorts, and obtained predictions, the best detection was obtained for 5-day windows and a dataset containing COVID-19 and HC. The accuracy was 0.77. The sensitivity (0.77) and specificity (0.80) were similar in value, which suggests that COVID-19 cases and HC were recognised on a comparable level. The results for detection of COVID-19 for 7-day windows and 10-day windows were also above 0.70. Nevertheless, the choice of shorter windows is much better in reality. The records for comparing HC and potentially ill windows do not need to be gathered from many days which is easier in the application - For all of these cases, the best-performing classifier was k-NN. Considering the datasets containing influenza cases, the best accuracies for each studied case were above 0.70. The highest sensitivity was recognised for 7-day windows, and with the use of k-NN, it was 0.89. Nonetheless, the specificity was low: at 0.51. When many positive cases were detected, the recognition of negative cases was random. The results between the two cohorts were at a similar level, although. The detection of only COVID-19 cases was slightly more accurate. This could indicate differences in the distribution of COVID-19 data and a dataset containing COVID-19 together with influenza. I has been reported that the symptoms of COVID-19 last longer than those of influenza. Similarly, they peak later after the illness onset [25]. Nonetheless, this statement is linked to symptoms in general. However, among symptoms onset, the increase in resting heart rate was reported to be higher for COVID-19 than for influenza. In the case of influenza, it was also visible, however, in a milder manifestation. These measurements were gathered with the smartwatches [25]. Due to the use of smartwatches, the results from this study could potentially be applied on a large scale. These devices are low-cost and can be used for limited screening tests [50]. The raw data used for this research could be distinguished into activity recognition (i.e. step counting) and heart rate measurements. Based on step counting, the most accurate device for this purpose is the pedometer. However, smartwatches are more comfortable for this purpose. They are non-invasive, user-friendly, and because of this suitable for this research. Another sensor, which is being used for activity recognition is an accelerometer. Some wearables are equipped also with sensors such as gyroscopes, magnetometers, barometers, and altimeters. They could increase the quality of activity recognition. However, the prices of these devices are higher: hence, they are often not included in standard smartwatches [51].

HR records were measured based on to the optical technique of PPG measurements. Unfortunately, optical sensing can cause some problems with the accuracy of this sensor, especially during activity. The signals gathered by PPG could be noised because of a movement, ambient light, and also tissue compression. The noise in the PPG signal can influence the correctness of computing the real value of HR [51]. Some devices use an accelerometer to increase the robustness of the HR estimation. Concerning these facts, the choice of a smartwatch could influence the robustness of the gathered HR values, which may limit the maximum accuracy of the created support system methodology. A solution could be combinations of more accurate pedometers for activity recognition, with smartwatches. With this combination, higher accuracy can be obtained, however, it would be more impractical as the use of these wearables should be as unobtrusive and discrete as possible [52]–[53].

Regarding the smartwatches, the most frequently used device for the original study was Fitbit [20]. The type of used device used was not provided in the database. Thus, it is challenging to claim the kind of device used for collecting the data in each case or this study. The most common fitness smartwatch - Fitbit, offers, among others, collecting data from a 3-axis accelerometer [54] - and PPG to extract the heart rate [55].

Some smartwatches can collect data more accurately and from extra sensors, such as - measuring skin conductance, skin temperature (gathered by infrared thermopile), BVP and HRV, and acceleration. Their disadvantage include their higher price is higher [56], [57]. The Empatica offers the ability to collect more accurate data and extended range of gathered data is regarded as medical device [58]. The experiment developed with the use of such a device could produce more accurate results for the created support system methodology. The increase in the number of modalities would bring more informative physiological data and this could also lead to a more robust model. Nevertheless, the use of Empatica for screening tests is difficult because of its higher price [22].

Finally, our findings must be compared with those of original study on which this study was based [20]. In this study, the same dataset was used as the one introduced in [20]. In the original research, 32 COVID-19 cases were analysed. In 25 cases some anomalies were detected, and 22 cases were in their early stage. Both works, this and the original focus on prodromal stage detection. First, this work did not treat outliers as anomaly detection issues, but rather as a classification problem. Hence, we increased parameters such as sensitivity and accuracy. The work also evaluated the data from several domains such frequency, spectral, and statistical aspects. The devices used for both studies were smartwatches, as mentioned before. The authors of the original work published offline and online algorithms for analysing the provided data.

Two types of data were used to detect the RHR and the ratio between the HR and the number of steps. The algorithm applied for analysing the RHR was the RHR difference (RHR-Diff) [20], was based on standardized residuals. We compared the 1-h resolution with the 28-day baseline. The second algorithm, the so-called HROS-AD, was built on heart rate over steps signals to detect anomalies. The values were compared for 1-hour with the rest of the record using Gaussian density estimation [20]. For RHR-Diff and HROS-AD, anomalies were detected in 22 COVID-19 cases. It should be emphasised that these methods are dedicated to anomaly detection, so they indicate some outliers. However, they do not clearly indicate wheter the person has or does not have COVID-19. Our methodology considers the classification problem.

The cumulative sum algorithm (CuSum) which works in real-time was introduced in [20]. In particular, it cumulated the deviations of the elevated residuals RHRs. The 28 days period was taken into consideration for this algorithm. Thus, 62,5 % of COVID-19 cases and viral infections were detected. Nevertheless, it should be treated as a specificity issue, rather than not accuracy issue. The difference between our paper and the original one is that we have created the support system methodology with a specificity higher than 77 % and an accuracy 78 % for 5-day windows for the COVID-19 detection. For the cohort containing influenza cases, the accuracy was 73 % and 71 % for the specificity. Moreover, our algorithm required a shorter period to detect the prodromal stage of the disease.

VI. LIMITATION OF THE WORK

A few limitations are evident in this study. Due to the reuse of the dataset from the original paper, some related problems could occur [20]. The previous study did not mention exactly where the data were gathered. The results obtained could be biased for a specific group of people. For example, race and ethnicity can influence the variance of cohort [25]. Further, the dataset is limited. This implies that the machine learning models created could be biased and overfitted. Future studies should extend the dataset to create a more robust algorithm.

Nevertheless, there could be also pointed out some challenges related to wearing wearable devices. It is the responsibility of the people to wear them all the time the device. In the real-world scenario, there is often a problem of a lack of data. This will result in the use of some pre-processing methodologies and will decrease the level of reliability of the machine learning model. Furthermore, this research focused on Fitbit; hence the solution may be limited to this device. Other smartwatches could use different pre-processing steps and approximations of the gathered signals [59]. Moreover, the gathered signals could be noisy [60]. People could wear the devices inappropriately. The signals could also be biased by the noise coming from various sensors. The common smartwatches are not certificated medical devices. Thus, future versions of the hardware could incorporate other parameters. It is also not clear what the responses of the human body of people with several comorbidities are. Moreover, the comorbidities could have an influence on the physiological parameters, and this could distort the pattern of the signal which could be typical for COVID-19 or HC cases.

VII. CONCLUSIONS

In this work, we introduced a methodology for the prodromal stage detection of COVID-19. The work contributes a methodology for the detection of COVID-19 in the early stages. This tool uses smartwatches (wearable devices) and reached interesting accuracy of 78 %. To limit the contagiousness of the disease, this approach is takes into consideration the character of the disease, that is, the incubation period and the highest contagiousness interval. This is a unique approach in comparison to the previous works. We evaluated three window lengths: 5, 7, and 10, and used features designed for biomarker analysis. The model based on 5-day windows allowed us to obtain the prediction with 78 % in accuracy. We tested a few sets of parameters. The most practical in reality will be the solution based on 5-day windows. This research was based on [20]. The biggest difference between our research and the original one was that this work focused on creating a classification algorithm, instead of an anomaly detection model, which has the potential to reach better accuracy and serve as a better diagnostic tool. Moreover, we provided an evaluation of the methodology expressed by more metrics and achieved better results for the models. The approach applied in this experiment could serve as a potential screening test based on smartwatches. The statistical evaluation based on the Mann-Whitney U test indicated which of the features differed most in the cohorts analyses. These were primarily from the statistical and spectral domains. We compared the model results trained on two different cohorts, COVID-19 and HC, and COVID-19, influenza, and HC. The results were quite similar for both cases, and slightly worse for the extended cohort with influenza. Considering the algorithms, simple classifiers such as k-NN and Logistic Regression, provided the best results which could indicate that a non complex dependency occurred in the datasets. In some cases, XGBoost achieved also good outcomes.

The major limitation of the study and the opportunity for future development is mainly the size of the training samples. It is, very difficult to obtain the data right at the time of disease onset, since usually the subjects are, not aware about the infection. Including parameters such as age, gender, weight, height, habits, and addiction to smoking and drinking could be worthwhile and could improve the accuracy of model [8], [22]. Another big advantage would be to extend the types and accuracy of the sensors used in wearable devices or optimally to use certified medical devices (e.g. Empatica). In the past, different sensor types were found across the same fitness products, which resulted in different measured values. Certified medical devices are not used very often because fitness devices are significantly cheaper. Not very practical but also possible, could be an experiment that combines several cheap fitness devices to cover more sensors at the same time. This work rely significantly on resting heart rate time; in the future, it could be interesting to identify and compare various activities (e.g. during the sleep). This might increase the accuracy and possibly shorten the required signal length which is currently 5 days. However, such an approach has a limitation - it cannot be broadly used for screening tests as standard smartwatches.

VIII. DATASET

The data were reused from the work [20] and could be found here: [1]

REFERENCES

[1] E. Sheikhzadeh, S. Eissa, A. Ismail, and M. Zourob, "Diagnostic techniques for COVID-19 and new developments," *Talanta*, vol. 220, Dec. 2020, Art. no. 121392.

[2] C. Menni, A. M. Valdes, M. B. Freidin, C. H. Sudre, L. H. Nguyen, D. A. Drew, S. Ganesh, T. Varsavsky, M. J. Cardoso, A. Visconti, P. Hysi, R. C. E. Bowyer, M. Mangino, M. Falchi, J. Wolf, S. Ourselin, A. T. Chan, C. J. Steves, and T. D. Spector, "Real-time tracking of self-reported symptoms to predict potential COVID-19," *Nature Med.*, vol. 26, no. 7, pp. 1037–1040, May 2020.

[3] T. Struyf, J. J. Deeks, J. Dinnes, Y. Takwoingi, C. Davenport, M. M. Leeflang, R. Spijker, L. Hooft, D. Emperador, S. Dittrich, J. Domen, S. R. A. Horn, and A. Van den Bruel, "Signs and symptoms to determine if a patient presenting in primary care or hospital outpatient settings has COVID-19," *Cochrane Database Syst. Rev.*, vol. 12, pp. 1–207, Jul. 2020.

[4] A. Pryce-Roberts, M. Talaei, and N. P. Robertson, "Neurological complications of COVID-19: A preliminary review," *J. Neurol.*, vol. 267, no. 6, pp. 1870–1873, Jun. 2020.

[5] S. Gupta and A. Mitra, "Challenge of post-COVID era: Management of cardiovascular complications in asymptomatic carriers of SARS-CoV-2," *Heart Failure Rev.*, vol. 11, pp. 1–11, Jan. 2021.

[6] E. Fraser, "Long term respiratory complications of COVID-19," *BMJ*, p. m3001, Aug. 2020.

[7] S. P. Adhikari, S. Meng, Y.-J. Wu, Y.-P. Mao, R.-X. Ye, Q.-Z. Wang, C. Sun, S. Sylvia, S. Rozelle, H. Raat, and H. Zhou, "Epidemiology, causes, clinical manifestation and diagnosis, prevention and control of coronavirus disease (COVID-19) during the early outbreak period: A scoping review," *Infectious Diseases Poverty*, vol. 9, no. 1, Mar. 2020, Art. no. 29.

[8] Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K. S. M. Leung, H. Y. Lau, J. Y. Wong, X. Xing, and N. Xiang, "Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia," *New England J. Med.*, vol. 382, pp. 1199–1207, Mar. 2020.

[9] W. J. Wiersinga, A. Rhodes, A. C. Cheng, S. J. Peacock, and H. C. Prescott, "Pathophysiology, transmission, diagnosis, and treatment of coronavirus disease 2019 (COVID-19): A review," *Jama*, vol. 324, pp. 782–793, Jul. 2020.

[10] R. Grewelle and G. De Leo, "Estimating the global infection fatality rate of COVID-19," *MedRxiv*, May 2020.

[11] *COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU)*. Accessed: Jan. 11, 2021. [Online]. Available: https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6

[12] R. C. Joshi, S. Yadav, V. K. Pathak, H. S. Malhotra, A. Parihar, N. Kohli, D. Himanshu, C. Alippi, and M. K. Dutta, "A deep learning-based COVID-19 automatic diagnostic framework using chest X-ray images," *Biocybern. Biomed. Eng.*, vol. 41, no. 1, pp. 239–254, Mar. 2021.

[1] https://storage.googleapis.com/gbsc-gcp-project-ipop_public/COVID-19/COVID-19-Wearables.zip

[13] N. S. Padhye, "Reconstructed diagnostic sensitivity and specificity of the RT-PCR test for COVID-19," *MedRxiv*, Feb. 2020.

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

[15] H. K. Siddiqi and M. R. Mehra, "COVID-19 illness in native and immunosuppressed states: A clinical–therapeutic staging proposal," *J. Heart Lung Transplantation*, vol. 39, p. 405, May 2020.

[16] A. G. Hadi, M. Kadhom, N. Hairunisa, E. Yousif, and S. A. Mohammed, "A review on COVID-19: Origin, spread, symptoms, treatment, and prevention," *Biointerface Res. Appl. Chem.*, vol. 10, no. 6, pp. 7234–7242, 2020.

[17] A. Ometov, V. Shubina, L. Klus, J. Skibinska, S. Saafi, P. Pascacio, L. Flueratoru, D. Q. Gaibor, N. Chukhno, O. Chukhno, and A. Ali, "A survey on wearable technology: History, state-of-the-art and current challenges," *Comput. Netw.*, vol. 193, Jul. 2021, Art. no. 108074.

[18] 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," *Sci. Rep.*, vol. 10, no. 1, Dec. 2020, Art. no. 21640.

[19] G. Zhu, J. Li, Z. Meng, Y. Yu, Y. Li, X. Tang, Y. Dong, G. Sun, R. Zhou, H. Wang, K. Wang, and W. Huang, "Learning from large-scale wearable device data for predicting the epidemic trend of COVID-19," *Discrete Dyn. Nature Soc.*, vol. 2020, pp. 1–8, May 2020.

[20] T. Mishra, M. Wang, A. A. Metwally, G. K. Bogu, A. W. Brooks, A. Bahmani, A. Alavi, A. Celli, E. Higgs, O. Dagan-Rosenfeld, and B. Fay, "Pre-symptomatic detection of COVID-19 from smartwatch data," *Nature Biomed. Eng.*, vol. 4, no. 12, pp. 1208–1220, Nov. 2020.

[21] R. Karthickraja, R. Kumar, S. Kirubakaran, and R. Manikandan, "COVID-19 prediction and symptom analysis using wearable sensors and IoT," *Int. J. Pervasive Comput. Commun.*, Nov. 2020.

[22] S. Hassantabar, N. Stefano, V. Ghanakota, A. Ferrari, G. N. Nicola, R. Bruno, I. R. Marino, K. Hamidouche, and N. K. Jha, "CovidDeep: SARS-CoV-2/COVID-19 test based on wearable medical sensors and efficient neural networks," 2020, *arXiv:2007.10497*. [Online]. Available: http://arxiv.org/abs/2007.10497

[23] D. J. Miller, J. V. Capodilupo, M. Lastella, C. Sargent, G. D. Roach, V. H. Lee, and E. R. Capodilupo, "Analyzing changes in respiratory rate to predict the risk of COVID-19 infection," *PLoS ONE*, vol. 15, no. 12, Dec. 2020, Art. no. e0243693.

[24] A. Natarajan, H.-W. Su, and C. Heneghan, "Assessment of physiological signs associated with COVID-19 measured using wearable devices," *NPJ Digit. Med.*, vol. 3, no. 1, Nov. 2020, Art. no. 156.

[25] 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, Jan. 2021, Art. no. 100188.

[26] K. H. Abdulkareem, M. A. Mohammed, A. Salim, M. Arif, O. Geman, D. Gupta, and A. Khanna, "Realizing an effective COVID-19 diagnosis system based on machine learning and IOT in smart hospital environment," *IEEE Internet Things J.*, early access, Jan. 11, 2021, doi: 10.1109/JIOT.2021.3050775.

[27] T. B. Alakus and I. Turkoglu, "Comparison of deep learning approaches to predict COVID-19 infection," *Chaos, Solitons Fractals*, vol. 140, Nov. 2020, Art. no. 110120.

[28] L. Pan, C. Wang, H. Jin, J. Li, L. Yang, Y. Zheng, Y. Wen, B. H. Tan, X. J. Loh, and X. Chen, "Lab-on-mask for remote respiratory monitoring," *ACS Mater. Lett.*, vol. 2, no. 9, pp. 1178–1181, Aug. 2020.

[29] H. J. Davies, I. Williams, N. S. Peters, and D. P. Mandic, "In-ear Spo_2: A tool for wearable, unobtrusive monitoring of hypoxaemia in COVID-19," *MedRxiv*, Jun. 2020.

[30] R. Stojanovic, A. Skraba, and B. Lutovac, "A headset like wearable device to track COVID-19 symptoms," in *Proc. 9th Medit. Conf. Embedded Comput. (MECO)*, Jun. 2020, pp. 1–4.

[31] C. K. Wong, D. T. Y. Ho, A. R. Tam, M. Zhou, Y. M. Lau, M. O. Y. Tang, R. C. F. Tong, K. S. Rajput, G. Chen, S. C. Chan, C. W. Siu, and I. F. N. Hung, "Artificial intelligence mobile health platform for early detection of COVID-19 in quarantine subjects using a wearable biosensor: Protocol for a randomised controlled trial," *BMJ Open*, vol. 10, no. 7, Jul. 2020, Art. no. e038555.

[32] H. Jeong, J. A. Rogers, and S. Xu, "Continuous on-body sensing for the COVID-19 pandemic: Gaps and opportunities," *Sci. Adv.*, vol. 6, no. 36, Sep. 2020, Art. no. eabd4794.

[33] R. J. Mobbs, D. Ho, W. J. Choy, C. Betteridge, and H. Lin, "COVID-19 is shifting the adoption of wearable monitoring and telemedicine (WearTel) in the delivery of healthcare: Opinion piece," *Ann. Transl. Med.*, vol. 8, no. 20, p. 1285, Oct. 2020.

[34] H. Banaee, M. U. Ahmed, and A. Loutfi, "Data mining for wearable sensors in health monitoring systems: A review of recent trends and challenges," *Sensors*, vol. 13, no. 12, pp. 17472–17500, 2013.

[35] P. Schmidt, A. Reiss, R. Dárichen, and K. V. Laerhoven, "Wearable-based affect recognition—A review," *Sensors*, vol. 19, no. 19, p. 4079, Sep. 2019.

[36] F. Li, K. Shirahama, M. A. Nisar, L. Köping, and M. Grzegorzek, "Comparison of feature learning methods for human activity recognition using wearable sensors," *Sensors*, vol. 18, no. 2, p. 679, 2018.

[37] D. J. Cook and N. C. Krishnan, *Activity Learning: Discovering, Recognizing, and Predicting Human Behavior From Sensor Data*. Hoboken, NJ, USA: Wiley, 2015.

[38] M. Barandas, D. Folgado, L. Fernandes, S. Santos, M. Abreu, P. Bota, H. Liu, T. Schultz, and H. Gamboa, "TSFEL: Time series feature extraction library," *SoftwareX*, vol. 11, Jan. 2020, Art. no. 100456.

[39] A. G. Lalkhen and A. McCluskey, "Clinical tests: Sensitivity and specificity," *Continuing Educ. Anaesthesia, Critical Care Pain*, vol. 8, no. 6, pp. 221–223, Dec. 2008.

[40] J. Skibińska and R. Burget, "Parkinson's disease detection based on changes of emotions during speech," in *Proc. 12th Int. Congr. Ultra Mod. Telecommun. Control Syst. Workshops (ICUMT)*, Oct. 2020, pp. 124–130.

[41] N. Nachar, "The Mann-Whitney U: A test for assessing whether two independent samples come from the same distribution," *Tuts. Quant. Methods Psychol.*, vol. 4, no. 1, pp. 13–20, Mar. 2008.

[42] D. Thissen, L. Steinberg, and D. Kuang, "Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons," *J. Educ. Behav. Statist.*, vol. 27, no. 1, pp. 77–83, Mar. 2002.

[43] C. M. Bishop, "Pattern recognition and machine learning," in *Information Science Statistics*. New York, NY, USA: Springer, 2013.

[44] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*. Princeton, NJ, USA: Kaufmann, 2006.

[45] A. Singh, N. Thakur, and A. Sharma, "A review of supervised machine learning algorithms," in *Proc. 3rd Int. Conf. Comput. Sustain. Global Develop. (INDIACom)*, Mar. 2016, pp. 1310–1315.

[46] R. Saravanan and P. Sujatha, "A state of art techniques on machine learning algorithms: A perspective of supervised learning approaches in data classification," in *Proc. 2nd Int. Conf. Intell. Comput. Control Syst. (ICICCS)*, Jun. 2018, pp. 945–949.

[47] A. Priyam, G. Abhijeeta, A. Rathee, and S. Srivastava, "Comparative analysis of decision tree classification algorithms," *Int. J. Current Eng. Technol.*, vol. 3, pp. 334–337, Jun. 2013.

[48] H. Sharma and S. Kumar, "A survey on decision tree algorithms of classification in data mining," *Int. J. Sci. Res.*, vol. 5, pp. 2094–2097, Apr. 2016.

[49] *Time Series Feature Extraction Library*. Accessed: Mar. 18, 2021. [Online]. Available: https://tsfel.readthedocs.io/en/latest/

[50] T.-C. Lu, C.-M. Fu, M. H.-M. Ma, C.-C. Fang, and A. M. Turner, "Healthcare applications of smart watches: A systematic review," *Appl. Clin. Informat.*, vol. 7, p. 850, Sep. 2016.

[51] A. Henriksen, M. Haugen Mikalsen, A. Z. Woldaregay, M. Muzny, G. Hartvigsen, L. A. Hopstock, and S. Grimsgaard, "Using fitness trackers and smartwatches to measure physical activity in research: Analysis of consumer wrist-worn wearables," *J. Med. Internet Res.*, vol. 20, no. 3, p. e110, Mar. 2018.

[52] T. G. Stavropoulos, A. Papastergiou, L. Mpaltadoros, S. Nikolopoulos, and I. Kompatsiaris, "IoT wearable sensors and devices in elderly care: A literature review," *Sensors*, vol. 20, no. 10, p. 2826, May 2020.

[53] Y. Guo, X. Liu, S. Peng, X. Jiang, K. Xu, C. Chen, Z. Wang, C. Dai, and W. Chen, "A review of wearable and unobtrusive sensing technologies for chronic disease management," *Comput. Biol. Med.*, vol. 129, Feb. 2021, Art. no. 104163.

[54] L. M. Feehan, J. Geldman, E. C. Sayre, C. Park, A. M. Ezzat, J. Y. Yoo, C. B. Hamilton, and L. C. Li, "Accuracy of fitbit devices: Systematic review and narrative syntheses of quantitative data," *JMIR mHealth uHealth*, vol. 6, no. 8, Aug. 2018, Art. no. e10527.

[55] Y. Bai, P. Hibbing, C. Mantis, and G. J. Welk, "Comparative evaluation of heart rate-based monitors: Apple watch vs fitbit charge HR," *J. Sports Sci.*, vol. 36, no. 15, pp. 1734–1741, Aug. 2018.

[56] U. Lee, K. Han, H. Cho, K.-M. Chung, H. Hong, S.-J. Lee, Y. Noh, S. Park, and J. M. Carroll, "Intelligent positive computing with mobile, wearable, and IoT devices: Literature review and research directions," *Ad Hoc Netw.*, vol. 83, pp. 8–24, Feb. 2019.

[57] A. Bizzego, G. Gabrieli, C. Furlanello, and G. Esposito, "Comparison of wearable and clinical devices for acquisition of peripheral nervous system signals," *Sensors*, vol. 20, no. 23, p. 6778, Nov. 2020.

[58] J. Chen, M. Abbod, and J.-S. Shieh, "Pain and stress detection using wearable sensors and devices—A review," *Sensors*, vol. 21, no. 4, p. 1030, Feb. 2021.

[59] G. Cosoli, S. Spinsante, and L. Scalise, "Wrist-worn and chest-strap wearable devices: Systematic review on accuracy and metrological characteristics," *Measurement*, vol. 159, Jul. 2020, Art. no. 107789.

[60] A. Rehman, M. Mustafa, N. Javaid, U. Qasim, and Z. A. Khan, "Analytical survey of wearable sensors," in *Proc. 7th Int. Conf. Broadband, Wireless Comput., Commun. Appl.*, Nov. 2012, pp. 408–413.

JUSTYNA SKIBINSKA was born in Poland. She received the bachelor's and master's degrees in biomedical engineering with a major in computer science and electronics in medicine from the University of Science and Technology, Kraków, Poland, in 2017 and 2018, respectively. She is an Early Stage Researcher and a Ph.D. student with Brno University of Technology, Czech Republic, and Tampere University, Finland, as a part of the H2020 MCSA ITN/EJD A-WEAR project.

RADIM BURGET is currently an Associate Professor with Brno University of Technology and is the Heading Signal Processing Program at the SIX Research Centre. He has been involved in the research of artificial intelligence for many years and in plenty of research projects which include projects funded on European level, national level, or privately funded projects. Companies he is cooperating with include Honeywell, Mitsubishi Electric, Rapidminer, and Konica-Minolta.

ASMA CHANNA (Graduate Student Member, IEEE) received the master's degree from Mehran University of Engineering and Technology Jamshoro, Pakistan, in 2018. She is currently pursuing the double Ph.D. degree with the University Politehnica of Bucharest, Romania, and Mediterranea University of Reggio Calabria, Italy, as a Marie Skłodowska-Curie Fellow in the European project A-WEAR. Her research interests include biomedical signal processing, wearables, e-health, and artificial intelligence.

NIRVANA POPESCU received the Ph.D. degree in computer science with a thesis called "Self-organizing intelligent fuzzy systems," from the University Politehnica of Bucharest (UPB), in 2003. She has been a Full Professor with the Computer Science Department, UPB, since 2014. Her research interests include neural networks, intelligent systems, fuzzy logic and control, e-health systems, cognitive and autonomous robots, and reconfigurable computers.

YEVGENI KOUCHERYAVY received the Ph.D. degree from TUT, in 2004. He is currently a Full Professor and the Laboratory Director with the Department of Electronics and Communications Engineering, Tampere University of Technology (TUT), Finland. He is the author of numerous publications in the field of advanced wired and wireless networking and communications. His current research interests include various aspects in heterogeneous wireless communication networks and systems, the Internet of Things and its standardization, and nanocommunications.