Journal Pre-proof

Few-shot meta-learning for pre-symptomatic detection of Covid-19 from limited health tracker data

Atifa Sarwar, Abdulsalam Almadani, Emmanuel O. Agu

PII: S2352-6483(24)00015-1
DOI: https://doi.org/10.1016/j.smhl.2024.100459
Reference: SMHL 100459

To appear in: *Smart Health*

Received date : 1 October 2023
Revised date : 22 January 2024
Accepted date : 21 February 2024

Please cite this article as: A. Sarwar, A. Almadani and E.O. Agu, Few-shot meta-learning for pre-symptomatic detection of Covid-19 from limited health tracker data. *Smart Health* (2024), doi: https://doi.org/10.1016/j.smhl.2024.100459.

Smart Health (2024)

Contents lists available at ScienceDirect

Smart Health

journal homepage: www.elsevier.com/locate/smhl

Few-Shot Meta-Learning for Pre-symptomatic Detection of Covid-19 from Limited Health Tracker Data

Atifa Sarwar[a,*], Abdulsalam Almadani[a], Emmanuel O. Agu[a]

[a] *Worcester Polytechnic Institute, Worcester, MA, USA*

ARTICLE INFO

Communicated by S. Sarkar

Keywords:
Meta-Learning
Few-Shot Learning
Covid-19
Consumer-grade health tracker
Deep networks
Vital signs variability

ABSTRACT

Detecting (or screening for) Covid-19 even before symptoms fully manifest, could enable patients to receive timely and life-saving treatment. Prior work has demonstrated that heart rate and step data from low-end wearables analyzed using deep learning models can detect Covid-19 reliably. However, significant individual differences in vital sign manifestation (high inter-subject variability) present a challenge to the generalization of deep learning models across diverse users. The limited amount of data in many medical scenarios further exacerbates this issue. Consequently, neural network models that can learn from limited vital sign data and varied inter-subject patterns are compelling. Meta-learning has emerged as a powerful technique for tackling various machine learning challenges, including insufficient data, domain shifts across datasets, and issues with generalization. This study proposes *MetaCovid*, a deep adaptation framework that employs meta-learning to address the variability of vital sign manifestation between subjects using only two days of data in order to detect Covid-19 before symptoms manifest. *MetaCovid* leverages heart rate and step measurements collected from consumer-grade health trackers over the preceding 2 days, extracts 45 digital bio-markers (features), which along with raw data, are fed into a deep GRU-based network with an attention mechanism, followed by uncertainty filtering. *MetaCovid* is trained using OC-MAML, a one-class few-shot MAML variant that adapts to the target distribution/user using only samples from the majority class. *MetaCovid* generalized well across two relatively small, publicly available Covid-19 datasets, achieving a recall of 0.81 and 0.92, and detecting 61% (14 out of 23) and 50% (17 out of 34) of users infected with Covid-19 before symptom onset. When OC-MAML was excluded from *MetaCovid* in an ablation study, the F_2 score dropped by 36%, highlighting that meta-learning indeed facilitates adaptation of deep sensing models to varying vital sign patterns. Notably, *MetaCovid* outperforms the current state-of-art method by predicting Covid-19 early on day *N* using heart rate and step measurements from only the preceding 2 days compared to 28 days, reducing data requirements by 93%. To the best of our knowledge, our study is the first to propose utilizing meta-learning to mitigate vital sign variability with limited data for Covid-19 screening. We believe that *MetaCovid* will pave the way for innovative Covid-19 interventions that are accurate even with limited data and help contain the spread of infectious diseases in the future.

1. Introduction

Covid-19, a disease caused by the coronavirus virus, was first discovered in Wuhan, China, and was declared a global pandemic on March 11, 2020. Covid-19 vaccinations were introduced in early 2021 but have faced some adoption challenges. Some subjects are unwilling to receive

*Corresponding author: Email address: asarwar@wpi.edu (Atifa Sarwar);

http://dx.doi.org/10.1016/j.smhl.xxxx.xx.xxx
Received 1 May 2013; Received in final form 10 May 2013; Accepted 13 May 2013; Available online 15 May 2013

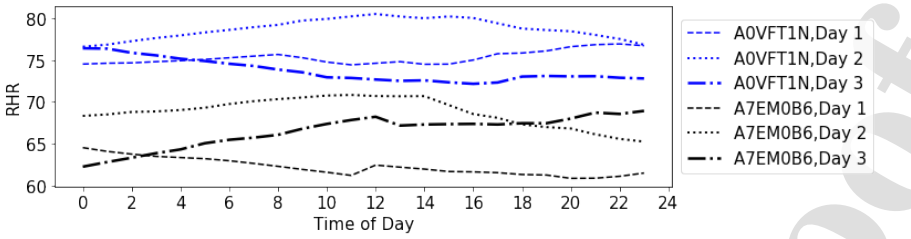


Fig. 1. Resting Heart rate of two subjects over three healthy days. A7EM0B6 has a mean RHR of 66 [Max: 71, Min: 61], while A0VFT1N exhibits higher RHR with a mean of 76 [Max: 80, Min: 72], illustrating both intra and inter-subject vital signs variability.

it due to concerns about its safety Karlsson et al. (2021) and religious beliefs Olagoke et al. (2021). Due to its high transmission rate and the difficulty of containing the virus' spread early, Covid-19 has significantly impacted societies, economies, and healthcare systems worldwide. This has underscored the importance of screening infectious diseases to enable timely public health interventions, clinical management, and disease containment.

Mobile sensing has solved several important real world problems including Human Activity Recognition (HAR) Mim et al. (2023), disease prediction Breda et al. (2023), and transportation and urban planning Liono et al. (2018). In 2022, about 216.43 million people worldwide were reported to wear smart wearables Ruby (2023). These wearables can passively monitor physiological signs such as heart rate, respiration rate, and physical activity, and can be utilized to passively assess a broad range of diseases. Physiologically, a 1°C increase in body temperature increases the heart rate by 8.5 beats per minute (bpm) Karjalainen & Viitasalo (1986). Using smart wearables to monitor heart rate continuously can detect such deviations, facilitating early detection of illness. Numerous studies have explored using abnormalities in physiological signs for early disease detection including paroxysmal atrial fibrillation Narin et al. (2018), cardiovascular disease Zhang et al. (2020), and Covid-19 Mishra et al. (2020)Abir et al. (2022).

While mobile sensing offers significant potential for early disease detection, it faces certain challenges. People exhibit substantial ***individual differences*** in the manifestation of vital signs. For instance, adults' heart rate can range from 60 to 100 beats per minute Pulse & HR (2022). Various factors contribute to the variability of physiological signs, including age, gender, stress levels, work routines, and medication usage Physiopedia. Figure 1 depicts the resting heart rate (RHR) of two healthy subjects over three days, highlighting significant variability within and across subjects over time. Such wide range of variations can introduce a covariate shift in data distributions, which present a challenge to effectively transferring knowledge across subjects and confound machine learning models. Ultimately, performance degradation results when models trained on such data are deployed in real-world settings.

To mitigate vital sign variability, prior work on predicting Covid-19 from physiological signs have employed techniques such as personalized training Mishra et al. (2020) Cho et al. (2022), training on only a subset of data Sarwar et al. (2023), and estimating individual patients' baseline using extended historical data Chung et al. (2023)Liu et al. (2022). While these approaches achieved some encouraging results, additional issues remain including: i) Personalized training requires substantial data collection per subject, which is an expensive and time-consuming process, ii) Training model only on a subset of data may omit vital information and hinder generalization, and iii) Estimating baseline using extended past data requires a longer subject participation period before a prediction can be made, which can result in delaying or even completely missing disease detection. Since increasing the efficiency of diagnoses can decrease new infections by up to 88.8% Rong et al. (2020), conversely, delayed prediction can endanger a patient's life and increase the likelihood of transmission of the virus to others, which in turn leads to more infections and potential outbreaks. In contrast to previous research, our goal is to develop a single model that can learn from varying vital signs with only few data samples for timely and accurate Covid-19 detection. Thus, the overarching goal of this study is to answer the following research question: *"Is there a generalizable neural networks framework for pre-symptomatic screening of Covid-19 from physiological signs collected using consumer-grade health trackers, which can overcome high inter-subject vital signs variability in limited data scenarios?*

Meta-learning, or learning to learn, is a branch of machine learning that tries to adapt a model previously trained on a given task, to new tasks with minimal additional training. To achieve this, the model is trained on a set of known tasks during the training phase in such a way that the trained model can quickly adapt to new tasks. Although meta-learning has been applied to various domains, it is currently under-utilized in healthcare. Only Banluesombatkul et al. Banluesombatkul et al. (2020) and Liu et al.Liu et al. (2021) have utilized MAML for handling data variations for sleep stage classification and video-based vital sign measurement respectively. However, both methods require the network to be pre-trained on a large dataset, which: i) does not address our overarching goal of overcoming variability in scenarios with limited training data, and ii) we speculate that pre-training enables their proposed models to learn and adapt to varying distributions in advance. To the best of our knowledge, our work is the first to propose employing meta-learning for Covid-19 screening, where each subject is treated as a new task to overcome high inter-subject variability in the physiological sign without requiring any data beforehand. We employed the idea of few-shot learning, facilitated by meta-learning, which adapts to new subjects/distributions with only a few samples per subject.

This paper proposes *MetaCovid*, a deep meta-learning-based adaptation framework for pre-symptomatic detection of Covid-19 from physiological signs (heart rate and step) collected passively using consumer-grade wearables. Using a 2-day sliding window (1-day offset), we extracted a combination of 45 known as well as novel digital biomarkers, characterizing circadian rhythms and physical activity. Circadian rhythms are 24-hour cycles that are part of the human body's internal clock, running in the background to carry out essential physiological functions and processes, such as performance, sleep, rest-activity cycles, and mood Vitaterna et al. (2001). The extracted features and raw data were input into a deep GRU-based neural network to identify the subject as either healthy or infected. Since Covid-19 has a mean incubation period of 6.38 days

Elias et al. (2021), the goal of *MetaCovid* is to detect Covid-19 infection accurately from D_{-6} to D_{+6} where D_0 refers to symptom onset day. To address the aforementioned challenge regarding variability in patient data, we trained *MetaCovid* using One Class MAML (OC-MAML) Frikha et al. (2021), a few-shot one class classification variant of MAML Finn et al. (2017). OC-MAML modifies the data sampling technique of MAML during meta-training such that the model adapts to a new task using K samples from the majority class, while the performance of the adapted model is evaluated using a class-balanced set. Our results demonstrate that *MetaCovid* outperforms all the baselines for the task of early Covid-19 prediction.

Our key contributions are as follows:

1. We propose *MetaCovid*, an innovative deep meta-learning-based adaptation framework for pre-symptomatic detection of Covid-19 from heart rate and step data collected passively using consumer-grade health trackers. To the best of our knowledge, our work is the first to mitigate inter-subject variability inherent in patients' physiological signs using meta-learning to accurately detect pre-symptomatic Covid-19, i.e., between 6 days before to 6 day after symptom onset.
2. Our rigorous evaluation of *MetaCovid* demonstrates that it outperformed all baselines achieving a recall and F_2 score of 0.81, precision of 0.82, NPV of 0.76, and AUC-ROC of 0.78 with only two shots (2 samples or days of data) per subject. It generalized well on two publicly available Covid-19 datasets, identifying 61% (14 out of 23) and 50% (17 out of 34) subjects before symptom onset. Additionally, *MetaCovid* outperformed state-of-art methods by predicting Covid-19 early on day N using heart rate and step data from the preceding 2 days compared to 28 days, reducing data requirement by 93%. It will enable the detection of Covid-19 induced vital sign abnormalities sooner, facilitating timely medical intervention, and potentially reducing infections by up to 88.8% Rong et al. (2020).
3. Our approach highlighted that meta-learning indeed facilitates overcoming inter-subject vital signs variability because without meta-learning, *MetaCovid* had a 36% drop in F_2 score.
4. *MetaCovid* works with only two shots and is not affected by imbalanced class distributions, significantly reducing the laborious task of collecting class-balanced data for each user while still achieving promising performance.

The rest of the paper is as follows: Section 2 outlines related work. Section 3 describes our methodology including an overview of our dataset, pre-processing techniques, feature extraction, and our proposed framework, MetaCovid. Section 4 discusses our experimental setup and validation protocols, section 5 evaluates our approach with section 6 highlighting important findings and limitations of the study. Finally, section 7 concludes the paper.

2. Related Work

This section reviews prior work on pre-symptomatic detection of Covid-19 based on physiological signs collected using consumer-grade health trackers. In summary, prior work is limited to personalized training, computing an individual's baseline using prolonged historical data, or training on a subset of available data. In contrast, our work employed the concept of meta-learning-based personalized models to mitigate high inter-subject variability in the manifestation of Covid-19 physiological signs, while significantly reducing the amount of data required.

2.1. Personalized Training:

Bogu et al. Bogu & Snyder (2021) proposed a Long Short Term Memory (LSTM) cell-based Autoencoder for Anomaly Detection (LAAD) to predict Covid-19 infection, which analyzed the user's Resting Heart Rate (RHR) relative to its baseline values. The study detected 56% of Covid-19 cases in their pre-symptomatic phases and 36% after the onset of symptoms. *Cho et al. Cho et al. (2022)* employed OC-SVM on physiological data of 29 Fitbit users for pre-symptomatic detection of Covid-19. The Resting Heart rate (RHR) and Heart Rate over Steps (HROS) were computed from the heart rate and steps data and passed to OC-SVM to detect anomalous samples. The authors reported that their method is better in three aspects: earlier and more detection and fewer false positives. *Abir at el. Abir et al. (2022)* proposes PCovNet, a Long Short-term Memory Variational Autoencoder (LSTM-VAE)-based anomaly detection framework, to detect Covid-19 infection in the pre-symptomatic stage from the Resting Heart Rate (RHR) derived from a wearable device. The proposed method identified 44% subjects before symptom onset. However, to dampen the impact of inter-subject variability in the data, these approaches utilized personalized training to train the model. However, personalized training requires a massive amount of data per subject and a model trained for a given subject cannot be reused for another subject, making it impractical to deploy in real-life settings.

2.2. DNNs Requiring Prolonged Historical data for Prediction:

Chung et al. Chung et al. (2023) proposed a transformer model, which learns HR variability patterns for pre-symptomatic prediction of Covid-19. The proposed method first calculates Resting Heart Rate (RHR) by removing HR values with step values greater than 0. Thereafter, the RHR values on day N are normalized using the mean and standard deviation from the preceding 28 days and fed into a transformer for predicting them as either healthy or infected. The proposed approach obtained a sensitivity of 0.84 and an AUC-ROC of 88%. *Abir et al.* proposed PCovNet+ Abir et al. (2023), a combination of a CNN-based Variational AutoEncoder and an LSTM network to detect anomalous RHR considering 16 days of past RHR data. The proposed model was initially pretrained on data from a healthy population, and subsequently fine-tuned using RHR data from Covid-positive individuals to achieve a personalized version for each subject. PCovNet+ successfully detected 47% of the subjects in the pre-symptomatic period. However, it is important to note that both of these approaches require extended historical data (28 and 16 days), which is a long participation period before a prediction can be made.

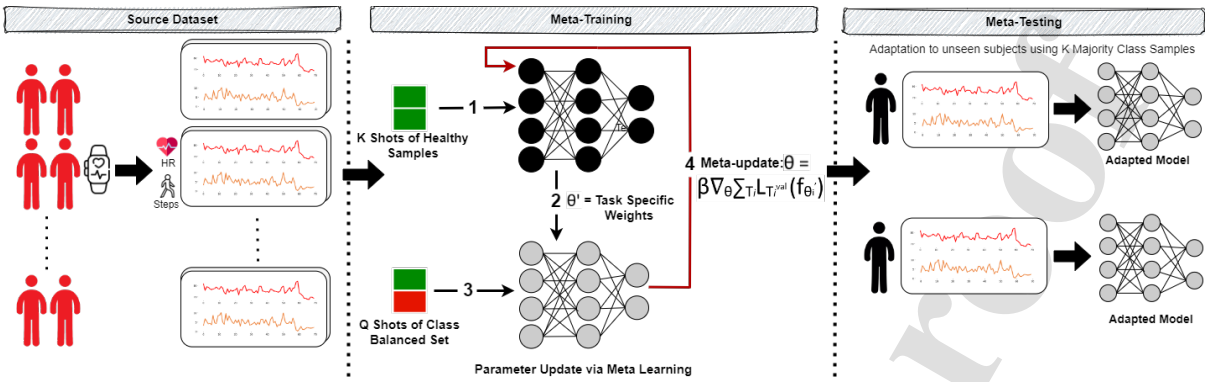


Fig. 2. An Overview of *MetaCovid*, our Proposed Approach for Pre-symptomatic Detection of Covid-19 using Consumer-grade wearables and Meta-learning. The source dataset comprises the heart rate and steps of Covid-positive subjects. Each subject's data is divided into a support set (K-shots of healthy samples) and a query set (Q samples of Class-Balanced set). First, the support set is fed into the base model (step 1) to learn the subject-specific weights (step 2), followed by loss computation using the query set (step 3), and updating the base model (step 4). Once trained, the base model can be adapted to unseen subjects using only K-shots of the healthy samples.

2.3. DNN Training on a Subset of Data:

Mayer et al. Mayer et al. (2022) decomposed HR collected using the Fitbit health tracker into six parameters: basal heart rate, autocorrelated noise, HR per step residual, circadian phase uncertainty, amplitude, and uncorrelated noise. The features extracted were utilized to train a Support Vector Machine (SVM) classifier, achieving an AUC-ROC of 0.76. The study labeled samples from days D_{-5} to D_{-1} as infected, and data from days D_{-10} to D_{-6} as early pre-symptomatic where D_0 refers to the symptom onset day. Apart from data gathered on the above days, data from other days were discarded. *Sarwar et al. Sarwar et al. (2023)* proposed a Gated Recurrent Unit (GRU) Network with Multi-Head Self Attention (MHSA) to predict Covid-19 one day before symptom onset using biobehavioral rhythmic dysregulation. To create a labeled dataset, the study randomly labeled a 24-hour interval during which the subject was healthy as a healthy sample, labeled data collected one day before symptom onset date as an infected sample and discarded all the remaining data. The proposed method achieved a sensitivity of 0.69 and a specificity of 0.89. Despite the encouraging performance, using only a subset of data for model training has two shortcomings. First, it may lead to not including some important information during model training. Second, it raises questions about the generalizability of these approaches as the model's performance is highly dependent on the specific instances utilized during training.

3. Material and Methods

This section presents an overview of the dataset, pre-processing techniques, and feature extraction employed by our study. We then detail, *MetaCovid*, our proposed deep meta-learning-based model for pre-symptomatic detection of Covid-19 using consumer-grade health trackers and Meta-learning. Fig 2 presents a general overview of our proposed approach.

3.1. Covid-19 Dataset

MetaCovid was trained using a dataset previously gathered by Mishra et al., a study that previously explored pre-symptomatic detection of Covid-19 from physiological data (heart rate and step) gathered from a smartwatch Mishra et al. (2020). The study enrolled 5,262 participants, of which 3,325 used a Fitbit, 984 used an Apple device, 428 reported using a Garmin device Garmin (2023), and the remaining used other devices, including the Oura Ring Oura (2023), chest BioStrap Biostrap (2023), and the Empatica wristband Empatica (2023). During enrollment, participants were asked to provide: i) demographic information, such as age, sex, and weight, ii) Medical history, and iii) Any previous Covid-19 illness status, i.e., confirmed or suspected, and if tested, the test date, results, and symptom onset date. Furthermore, in order to track symptoms experienced and their severity, and to discover any new Covid-19 tests or diagnoses, test results, and recovery dates, all participants were asked to complete a daily survey. Since most enrolled participants used a Fitbit device, the study gathered data from only Fitbit users and had 73 healthy, 15 non-Covid-19 illnesses, and 32 Covid-positive individuals. As our work focuses on pre-symptomatic detection of Covid-19, we utilized only the data of 25 Covid-positive participants that also had heart rate and steps data available between 20 days prior to symptom onset and up to 21 days afterwards. Table 1 summarizes the demographics and health characteristics of the Covid-positive cohort. To create a labeled dataset, we labeled days D_{-6} to D_{+6} as infected, records prior to D_{-6} as healthy, and discarded all data after D_{+6} where D_0 refers to symptom onset day. In contrast to prior work that discarded data occurring before the disease manifests, we specifically excluded data only after D_{+6} as it belongs to either the infectious or recovery period Cevik et al. (2021) He et al. (2020) Since the primary focus of this study is early Covid-19 detection, we believe that excluding prolonged infectious or the recovery period makes sense because in these periods, although physiological signs return to baseline values, there is the possibility of viral shedding and prolonged manifestation of abnormal values of physiological signs. Excluding data from these periods ensures that *MetaCovid* concentrates on and learns meaningful patterns from periods relevant to early detection and generalizes well, while avoiding noise from irrelevant data.

Table 1. Demographic and Health Characteristics of 32 Covid-Positive Participants from Dataset by Mishra et al.Mishra et al. (2020)

Demographics					
Mean Age	47 (27-67)	Gender	Female: 25 (78.1%)		
Ethnicity					
European	27 (84.4%)	Mixed/Other	5 (15.6%)		
Input Devices					
Fitbit Ionic	Fitbit Charge 3, Charge 4				
Self-Reported Health					
Respiratory Lung Disease	6 (18.8%)	High Blood Pressure	4(12.5%)	High Cholesterol	4 (12.5%)
Heart Disease	1 (3.1%)	Psychiatric Illness	4 (12.5%)	Gastrointestinal or Digestive System Disease	5 (15.6%)
Allergy/Immune Disease	7 (21.9%)	Unknown	1 (3.1%)		
Baseline Body Mass Index					
< 25	12 (37.5%)	25 to <30	7 (21.9%)	30 or Higher	6 (18.8%)
Unknown	7 (21.9%)				
Sensor Readings					
Heart Rate	After every 15 Seconds		Steps	After every 1 Minute	

3.2. Pre-Processing

Raw data available in the Mishra et al. dataset were pre-processed using four main steps:

1. **Outlier Removal:** Outliers were removed from the heart rate signal by dropping all records in which the HR was < 30 or > 200 Beats Per Minute (BPM).
2. **Synchronization:** Despite gathering the data from Fitbit devices only, some wearable data had different timestamps and frequencies. To synchronize HR and step input streams, we resampled HR to a one-minute resolution, which was then aggregated with the step data stream using the HR timestamp.
3. **Data Imputation:** To reduce the impact of missing values on the final prediction, the aggregated data was resampled to a 1-hour resolution and days during which the heart rate and step values were absent for more than 12 hours (50% of the day) were dropped. The remaining missing values were filled with the last observed value (Last Observed Carried Forward or LOCF algorithm), which was found to be the most effective imputation method for accurate detection of Covid-19 (Discussed further in Section 5).
4. **Segmentation:** The raw sensor data was segmented into overlapping 2-day windows with a 1-day offset between the consecutive windows. The impact of various window lengths on performance will be studied in section 5. With regards to labeling, the window was labeled as infected if the 2-day interval overlapped with the period between D_{-6} and D_{-1} where D_0 refers to the symptom onset day.

3.3. Feature Extraction

Motivated by the work of Rykov et al. Rykov et al. (2021), a range of digital bio-markers (features) characterizing physical activity, circadian rhythm, and physiological parameters were extracted. This set of digital biomarkers has previously proven effective in depression screening, with our work being the first to explore them for Covid-19 detection. Table 2 is a detailed description of the extracted features and their mathematical expressions. All features were extracted over a window length of 2 days. To extract the parametric rhythmic features, we employed the Cosinor Halberg et al. (1967), a method of obtaining an estimate of the Midline Statistic of Rhythm (MESOR), the amplitude, a measure of phase (acrophase) for the chosen period using cosinor curve fitting. The Cosinor is based on a trigonometric regression model. When the period is known, the model is defined as:

$$y(t) = \sum_{i=1}^{N} (A_{i,1} * sin(\frac{t}{P/i}.2\pi) + A_{i,2} * cos(\frac{t}{P/i}.2\pi)) + M + e(t)$$

where t corresponds to the time points to be observed within the time series, N is the number of components, M is MESOR, P is the observed period, $e(t)$ is the error term, and $A_{i,1}$ and $A_{i,2}$ are the parameters of the model. We employed CosinorPY Moškon (2020), a publicly available Python implementation of cosinor-based methods for rhythmicity detection.

3.4. MetaCovid: Proposed Framework

This section presents technical details of *MetaCovid*, our proposed framework, which consists of three parts: 1) a deep GRU-based attention network for sample-level Covid-19 infection prediction, 2) uncertainty estimation in order to improve performance by removing predictions that the model is uncertain about, and 3) a meta-learning algorithm for overcoming vital signs variability with minimal data. Figure 3 is a high overview of our proposed framework.

Table 2. List of Digital Biomarkers (Features) with Definition

Category	Feature	Description
Physical Activity (PA)	Daily Steps	Average of daily sum of steps over the observation period.
	Sedentary Time	Average daily time when no activity is performed.
	Light Intensity PA (LPA) - mean	Average daily time of physical activity with steps < 6000 steps per hour O'Brien et al. (2018).
	Moderate Intensity PA (MPA) - mean	Average daily time of physical activity with steps > 6000 but < 7920 per hour O'Brien et al. (2018).
	Vigorous Intensity PA (VPA) - mean	Average daily time of physical activity with steps > 7920 per hour O'Brien et al. (2018).
Heart rate	HR - mean, SD, CV	Average, SD, and CV of HR, reflecting the extent of stability/variability in HR.
	Daytime HR - mean, SD, CV	HR at day-time (14:00 - 16:00)
	Nighttime HR - mean, SD, CV	HR at night-time intervals (00:00 - 2:00, 2:00 - 4:00, 4:00 - 6:00)
	RMSSD	Root mean square of successive differences of HR characterizes the sharpness of successive HR deviations and can be interpreted as a proxy measure of ECG-based HR variability.
Parametric Rhythmic Features*	Acrophase	The time of day time when the rhythm reaches its maximal value for the first time in the cycle.
	Mesor	The Midline Estimating Statistic of Rhythm refers to the mean daily activity.
	Amplitude	The difference between the maximum value of the fitted cosine curve and MESOR. The lower amplitude indicates a more dampened rhythm.
Non-Parametic Rhythmic Features*	Inter-daily Stability (IS)	Measures stability/regularity of circadian rhythm over a series of 24h cycles, and is defined as: $IS = \frac{N \sum_{h=1}^{p} (x_h - \overline{x})^2}{p \sum_{i=1}^{N} (x_i - \overline{x})^2}$ where N is the total number of data items (48 in our case), p is the number of data items per day (24 in our case), x_h corresponds to each hour of the mean profile, while x_i represents each given hour of raw data, and $\overline{x}$ is the average of all data.
	Intra-daily Variability (IV)	Quantifies the fragmentation of periods of activity from periods of rest within a 24-h cycle and is given as: $IV = \frac{N \sum_{i=2}^{N} (x_i - x_{i-1})^2}{(N-1) \sum_{i=1}^{N} (x_i - \overline{x})^2}$ where N is the total number of data items (48 in our case), x_i represents each given hour of raw data, and $\overline{x}$ is the average of all data.
	M10	Diurnal activity, the mean activity of the ten consecutive most active hours of the average daily activity profile.
	L5	Nocturnal activity, the mean activity of the five consecutive least active hours of the average daily activity profile.
	Relative Amplitude	Represents daytime activity and is given as the difference between M10 and L5 divided by the sum of M10 and L5, i.e., $\frac{M10-L5}{M10+L5}$
	Inter-daily Coefficient of variation (ICV)	The 24h mean of by-hour coefficients of variation (CV), where CV is the ratio of SD to average in each hour between days. $ICV = \frac{1}{p} \sum_{h=1}^{p} \frac{\sqrt{\frac{\sum_{i=1}^{N} (x_i - x_h)^2}{N}}}{x_h}$ where p is the total number of data items per day (24 in our case), x_i represents values corresponding to each hour from all days, x_h represents values from each hour from the mean 24h profile.
	Rhythm autocorrelation (AC)	The autocorrelation of time series with a day-length lag, another alternative measure of a rhythm stability; higher values indicate higher stability/similarity of data patterns across days. $AC = \frac{\sum_{i=1}^{N-k} (x_i - \overline{x})(x_{i+k} - \overline{x})}{\sum_{i=1}^{N} (x_i - \overline{x})^2}$ where k is a day-length lag (24 in our case), x_i represents value of each interval, $\overline{x}$ is the average of all data, and N is total number of data points.
Peak detection*	Daily peaks - mean, and SD	The number of peaks per day in time series. Robust peak detection algorithm was used to identify peaks (following algorithm parameters were set for steps data: lag = 10, threshold = 10, influence = 0; for HR data: lag = 10, threshold = 2, influence =0.25)

*These metrics are computed seperately for both heart rate, and steps values over the observation period.

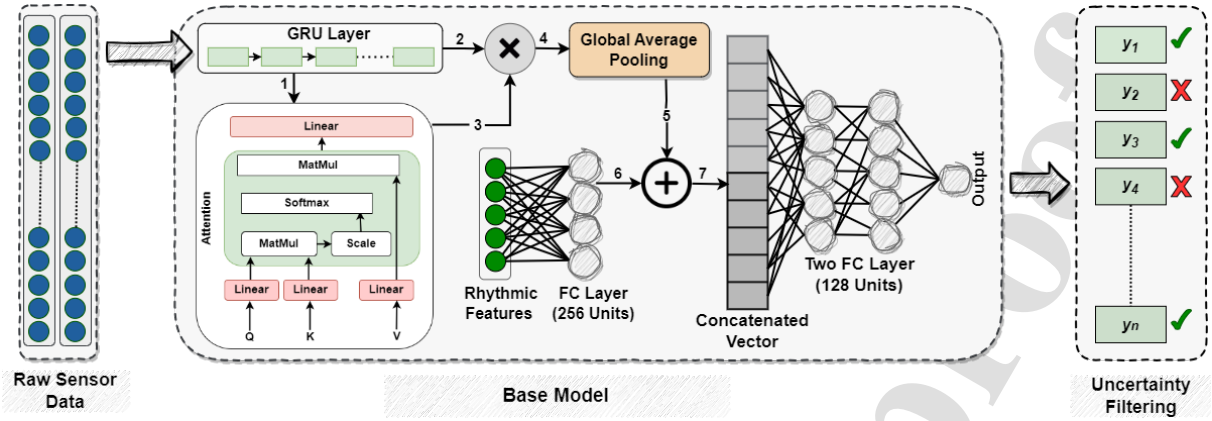


Fig. 3. Architecture of *MetaCovid*. Raw Sensor Data (Heart rate and Step) collected over 2 days (48X2) is fed into a GRU Layer, followed by Self-Attention (Step 1), and element-wise multiplication (Step 2 and 3). The output then passes through a Global Average Pooling layer (Step 4), and is concatenated with Rhythmic Features (Step 5 and 6). Subsequently, it traverses two fully connected layers with the sigmoid activation applied at the end for final prediction (Step 7). Finally, labels predicted with low confidence are removed by applying Entropy-based Uncertainty Quantification.

3.4.1. ***Deep GRU-based Attention Network***

Inspired by prior work of Sarwar et al. (2023), our proposed deep learning model comprises of five layers: a single Gated recurrent layer, an attention layer, and three fully connected layers. Each fully connected layer is followed by a dropout layer, with a sigmoid function at the last layer. These design choices were made as: i) GRU Cho et al. (2014) is able to learn rarely occurring events such as abnormal physiological signs better than other temporal neural networks architectures (LSTM, RNN) and can outperform them in terms of the speed of convergence and generalization Chung et al. (2014), ii) In disease diagnosis, medical practitioners generally examine only critical measurements corresponding to vital signs, as not all the records contribute equally to estimating one's health status. We employ attention, first introduced by Bahdanau et al.Bahdanau et al. (2014), which is a mechanism for determining the importance of a word in a given sentence or more generally tries to discover and weight the most predictive parts of input data more. Formally, attention is expressed as:

$$Q = XW_Q,\ V = XW_V,\ K = XW_K$$
$$score = softmax(\frac{QK^T}{\sqrt{d_k}})$$
$$Attention(Q, K, V) = score * V$$

where X refers to an input sequence, and Query(Q), Value(V), and Key(K) to the linear projection of X. *MetaCovid* employs self-attention, which weights various positions in data in order to compute the representation of the same sequence. This is achieved by setting query, key, and value to the input sequence itself. With regards to data flow through the network, the model analyzes a tensor of shape 48X2, i.e., raw heart rate and step values over the preceding 2 days. The input is passed from a single Gated Recurrent Unit (GRU) to a self-attention layer, followed by element-wise multiplication to learn the weighted representation of the input. The resultant output is fed into the Global Average Pooling layer and concatenated with rhythmic features (1X45). Subsequently, it is traversed through two fully connected layers, each followed by a dropout with a rate of 0.25. Finally, the sigmoid is used to produce the final output: whether the subject is infected with Covid-19 or not.

3.4.2. ***Entropy-Based Uncertainty Filtering***

Uncertainty is a significant barrier to the deployment of deep learning models in healthcare settings. Model certainty, or the likelihood that a particular model output is accurate, could be used to improve trust of deep learning models by medical professionals, making uncertainty estimation crucial for systems targeted at healthcare. The growth of commercial wearables in healthcare further exacerbates this challenge, because along with predictive uncertainty (the model's output), the models also have to cope with aleatoric uncertainty (the data formation process). To address this, we proposed to utilize Entropy-based uncertainty filtering to remove all predictions for which the model is uncertain. In the domain of digital histopathology, Dolezal et al. Dolezal et al. (2022) previously employed the standard deviation of the distribution of predictions, generated from an ensemble of 30 models in a drop-out-enabled network, as the uncertainty metric. Our study extended this approach by introducing a novel aspect. Instead of utilizing the standard deviation, we used the probabilities obtained from the proposed deep network directly to compute the uncertainty metric (entropy). By examining our selected evaluation metrics, our empirical findings demonstrate that the proposed uncertainty estimation method significantly improves performance (Discussed further in Section 5). Entropy, a measure of randomness or impurity in a variable, is expressed as:

$$\phi(x) = -\sum_{j=1}^{K} p(y_j) * log(p(y_j))$$

Algorithm 1 Meta-training of OC-MAML

Require: S^{tr}: Set of meta-training tasks
Require: α, β: Learning Rates
Require: K, Q: Batch size for adaptation and validation
Require: c: Class Imbalance Rate (CIR)
Randomly initialize θ
while not done **do**
  Sample batch of Tasks T_i from S^{tr}: $T_i = \{T_i^{tr}, T_i^{val}\}$
  for each sampled T_i **do**
    Sample K examples B from T_i^{tr} with CIR = 0
    Initialize $\theta_i^{'} = \theta$
    for number of adaptation steps **do**
      $\theta_i^{'} = \theta_i^{'} - \alpha \nabla_{\theta_i^{'}} L_{T_i}^{tr}(f_{\theta_i^{'}})$
    Sample Q examples $B^{'}$ from T_i^{val} with CIR=50%
    Compute Loss $L_{T_i}^{val}(f_{\theta_i^{'}})$ using $B^{'}$
  Update $\theta \leftarrow \theta - \beta \nabla_\theta \sum_{T_i} L_{T_i}^{val}(f_{\theta_i^{'}})$
return meta-learned parameters θ

where K is the number of classes (binary in our case), and $p(y_j)$ is the probability that y belongs to class j. The optimal uncertainty threshold (ϕ_{opt}) below which the predictions are likely to be correct, is then determined. To determine the optimal threshold, Youden's index (J) Ruopp et al. (2008) is computed for all possible uncertainty values (ϕ_{val}) over the validation dataset, defined as:

$$J(\phi_{val}(x)) = Sensitivity(\phi_{val}(x)) + Specificity(\phi_{val}(x)) - 1$$

The optimal uncertainty threshold is then defined as the set of values that maximize the Youden Index, i.e.,

$$\phi_{opt} = \underset{\phi_{val}}{\operatorname{argmax}}\, J(\phi_{val})$$

Finally, this optimal threshold ϕ_{opt} is used by the model to filter out predictions about which it is not confident. Our instance-level confidence estimate is defined as:

$$C(x) = \begin{cases} High\ Confidence & \phi(x) < \phi_{opt} \\ Low\ Confidence & \phi(x) \geq \phi_{opt} \end{cases}$$

*3.4.3. **OC-MAML: One Class Few-Shot Classification via Meta-learning***

Conventional methods for training deep neural networks require extensive training data and are prone to overfitting when training data are limited (few examples or few-shot scenarios). Few-shot classification aims to learn a classifier that can use only a few labeled examples to distinguish the classes that unseen data belong to. The meta-learning paradigm is a promising few-shot classification approach, where the emphasis is on extracting and transferring knowledge learned from various tasks to prevent overfitting and improve generalization. Meta-learning models learn how to learn, enabling them to generalize effectively to unseen tasks with minimal data and has emerged as a promising solution for facilitating advancements in healthcare where challenges such as data scarcity and domain shifts are prevalent. Meta-learning algorithms fall into three main categories:

1. *Optimization-based algorithms:* that aim to solve an optimization problem, which can converge quickly during training.
2. *Metric-based algorithms:* that generate kernel weights by measuring the distances between samples in a latent space.
3. *Model-based approaches:* where the emphasis is on learning model parameters/architecture so that it can quickly adapt to new tasks.

While both metric- and model-based meta-learning approaches are promising, they depend on the choice of distance metrics and underlying model assumptions. If things go wrong, these selections can significantly impact performance, leading to sub-optimal results. MAML, an optimization-based model-agnostic approach, aims to learn model initialization parameters that facilitate fast adaptation to new tasks/domains with limited labeled data. To achieve this, all available tasks are divided into three disjoint sets: S^{tr} for meta-training, S^{val} for meta-validation, and S^{test} for meta-testing. Each task T_i is divided further into two disjoint sets: T_i^{tr} is used for adaptation, while T_i^{val} is for validation. Enabled by few-shot learning, the model f_θ is adapted to task T_i by taking a few gradient descent steps using a *few* data points from T_i^{tr} yielding task-specific weights $\theta_i^{'} = \theta - \alpha \nabla_\theta L_{T_i}(f_\theta)$, where α is a step size and can be fixed or meta-learned. The task-specific model f_{θ_i} is validated on T_i^{val} to compute the loss $L_{T_i}^{val}(f_{\theta_i^{'}})$ for the meta-update. Note that the loss is computed using the updated model parameters $\theta^{'}$, whereas the adaptation for each task is performed over the model parameters θ. Finally the model parameters are updated as:

$$\theta \leftarrow \theta - \beta \sum_{T_i \sim p(T)} L_{T_i}^{val}(f_{\theta_i'})$$

Where β is the meta-update size. Although MAML is a popular algorithm for learning several tasks with minimal effort, it only learns model initialization parameters that are suitable for class-balanced, few-shot classification.

One Class Classification (OCC): refers to a specific type of binary classification problem where adequate data is available for only one class. Medical problems are often considered OCC due to the fact that the occurrence of a specific disease (ill people) is typically less frequent in contrast to its absence (healthy people). Covid-19 detection aligns with the concept of OCC as being Covid-19 positive is a relatively rare event, posing challenges to the application of MAML for Covid-19 detection. Thus to overcome the inter-subject variability of vital signs for pre-symptomatic identification of Covid-19 in limited data-settings, we propose training our deep network using OC-MAML, a variant of the MAML algorithm. OC-MAML Frikha et al. (2021) adapted MAML to one-Class Few-shot classification with focus on learning model initialization parameters that can learn effectively with either no or only a few samples of minority class, and yields the same performance as doing so with a class-balanced dataset. OC-MAML achieves this by introducing a hyperparameter (c), which sets the percentage of the samples belonging to the minority class that needs to be sampled for model adaptation, e.g. setting c=0% means only majority class samples are sampled for T_i^{tr}. In contrast, for T_i^{val}, samples are class-balanced with $c = 50\%$. Algorithm 1 shows the pseudocode for OC-MAML meta-training, while for meta-testing only steps 5-8 are repeated for adapting the model to new subject.

4. Experimental Setup

4.1. Validation Protocol

To rigorously evaluate *MetaCovid* performance, we utilized 5-fold cross-validation with no overlap between training and testing segments at each fold. For evaluation, metrics that have been widely used in prior studies for the early detection of Covid-19, were utilized.

- Precision: Of all subjects that are predicted as infected, how many were really infected, expressed as $\frac{TP}{TP+FP}$
- Recall: Of all subjects that are infected, how many were accurately predicted as infected, expressed as $\frac{TP}{TP+FN}$
- Negative Predictive Value (NPV): the probability that following a negative test result, whether the individual did not truly have the specific disease, expressed as $\frac{TN}{TN+FN}$
- AUC-ROC is a probability curve that plots *TPR* and *FPR* at various thresholds and evaluates how well the model distinguishes between classes.
- F_β score is the weighted harmonic mean of precision and recall, i.e. $\frac{((1+\beta^2)\times Precision\times Recall)}{\beta^2\times Precision+Recall}$. Since identifying infected subjects is more important for containing Covid-19's spread, the F_2 score, which sets the $\beta = 2$, for evaluating MetaCovid, was utilized.

4.2. Implementation Details

Normalization: To transform all data onto the same scale, subject-wise Z-score normalization ($\frac{x-\mu}{\sigma}$) was used. Each participant's raw data and extracted physiological features were mapped on a scale of his/her standard deviation (σ) centered at his/her mean (μ) over the study period. Our results demonstrated that this normalization technique was more suitable for pre-symptomatic Covid-19 detection with its associated contraints.

Hyperparameter Tuning: To achieve optimal results, grid search was used to experiment with and tune hyperparameters for the *MetaCovid* model architecture and OC-MAML. Table 3 shows the optimal values of these hyperparameters along with their original values. For OC-MAML, disjoint sets of subjects were utilized for adaptation and validation, and each subject in the dataset was treated as one meta-task (T_i). To sample K shots for adaptation, we set c=0%, i.e, only healthy samples were used for adaptation, while for validation, $Q/2$ samples were selected from each class. During meta-training, we utilized Binary Cross Entropy Loss (BCE = $-\frac{1}{N}\sum_{i=1}^{N} y_i log(p_i) + (1-y_i)(1-log(p_i))$) for loss computation. The model's performance is assessed across five different adaptation tasks sampled from each test subject, with the reported results being an average over all adaptation episodes.

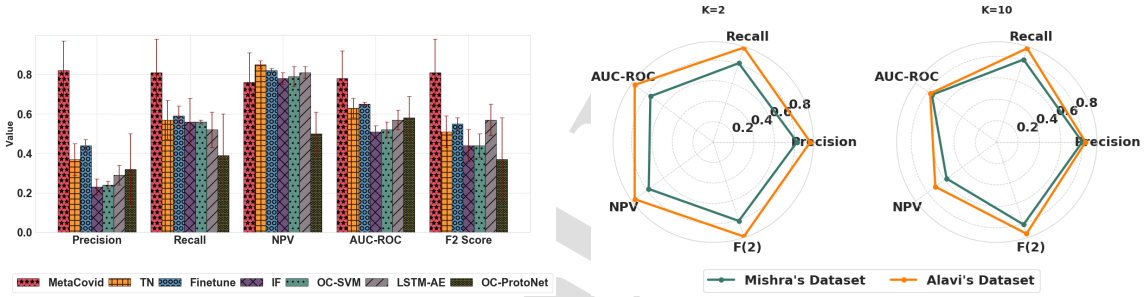
Baselines: MetaCovid was compared to six baselines selected from three categories.

1. *One Class Classification (OCC)*: Previous studies have widely explored these OCC techniques for pre-symptomatic detection of Covid-19 Cho et al. (2022) Bogu & Snyder (2021). From this category, OC-SVM, Isolation Forest (IF), and LSTM AutoEncoder-based anomaly detection were selected.
2. *Deep Embedding Networks*: Within this category, we explored deep embedding networks, which are trained on the meta-training tasks as follows: one is trained using the traditional settings for all training tasks together, followed by evaluation on test tasks. While the other was trained using the "Finetune" baseline Triantafillou et al. (2019) that utilized the support set of the given test episode to train an output layer on top of the embedding. The rationale behind using deep embedding networks as baselines was to assess their ability to address the variability of vital signs within subjects in limited data-settings.
3. *OCC via Metric-based Meta-Learning*: Lastly, one-way prototypical networks Kruspe (2019) were selected to evaluate whether OC-MAML was the most suitable choice for *MetaCovid*. One-way prototypical networks introduce a "null" class centered at zero to identify anomalies with only few examples per class.

To alleviate the data imbalance issue in deep embedding networks, a class-weighting strategy was applied during training: a value of 1 for healthy and 4 for infected. For LSTM-AE, the mean absoulte reconstruction error ($\frac{\sum_1^N |\hat{x}-x|}{N}$) over validation data was utilized as a threshold to identify anomalous examples.

Table 3. Hyperparameter Tuning

Hyperparameter	Explored Values	Optimum Value	Hyperparameter	Explored Values	Optimum Value
Model Architecture					
Recurrent Depth	1, 2, 4	1	GRU Layer Size	256, 512	256
Fully Connected Layers	1, 2, 3, 4	1 (Rhythms Network) 2 (Joint Network)	Neurons in Hidden Layer	128, 256, 512	256 (Rhythms Network) 128 (Joint Network)
Dropout	0.25, 0.5	0.25			
OC-MAML					
Inner Learning Rate (α)	0.01, 0.001, 0.0001	0.001	Outer Learning Rate (β)	0.01, 0.001, 0.0001	0.01
Meta-adaptation size (K)	2, 10	2	Validation size (Q)	2, 10	2
Gradient Steps	1, 3, 5, 10	3 (Meta-training) 1 (Meta-Testing)	Optimizer	Adam, SGD, RMS Prop	SGD with a decay of 0.0005
Epochs	1000	Early Stopping used to avoid overfitting	Batch Size	8, 16, 32	8



(a) (b)

Fig. 4. (a) *MetaCovid* vs. Baselines. The error bar shows Standard Deviation across Folds. (TN refers to Traditional Networks, IF to Isolation Forest, OC to One-Class, and AE to Autoencoders) (b) Performance of *MetaCovid* On Mishra's and Alavi's Dataset as a function of K

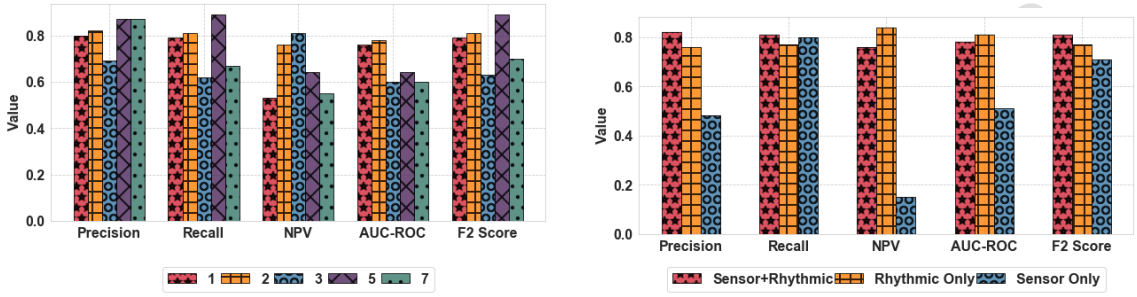
5. Evaluation and Results

5.1. Superiority of MetaCovid over Baselines

MetaCovid demonstrated exceptional performance, outperforming all baselines with a Recall of 0.81, Precision of 0.82, AUC-ROC of 0.78, and F_2 score of 0.81. Although *MetaCovid* achieved promising Recall, Precision, AUC-ROC, and F_2 score results, it did not perform as well based on NPV. Traditional Networks obtained the highest NPV of 0.85, followed by the "Finetune" Baseline (0.82) and LSTM-AutoEncoder (0.81), reflecting that *MetaCovid* equipped with its novel feature extraction, OC-MAML, and uncertainty estimation primarily focuses on identifying infected cases while achieving a desirable performance (NPV=0.76) for predicting healthy subjects. *MetaCovid* identifies 14 (61%) subjects early in the incubation period, i.e. D_{-6} to D_{-1}, 7 (30%) subjects during D_0 to D_{+2}, and had delayed prediction of 2 (9%) cases, i.e., during D_{+3} to D_{+6}, where D_0 refers to symptom onset day. The performance of *MetaCovid* was also evaluated using different values of K and Q and found that compared to K=2, using 10 shots for adaptation drops the Precision by 2%, AUC-ROC by 4%, and NPV by 24%, while Recall and F_2 score remained constant. These findings led to the conclusion that Meta-learning indeed facilitates overcoming the inter-subject vital signs variability, which the traditional machine/deep learning methods struggle with, as highlighted by the drop in the baselines' performance.

5.2. Generalizability for Other Datasets

To further examine its generalizability, *MetaCovid* was validated on a second dataset collected in Alavi et al.'s study Alavi et al. (2022), which uses finite state machines for detecting aberrant physiological and activity signals associated with early Covid-19 infection. The dataset consisted of heart rate and step signals from 2,115 participants, of whom 278 individuals reported Covid-19-positive test results, with 84 [Fitbit:49, Apple: 35] having adequate Covid-19 infection wearable data. Due to insufficient heart rate measurements from Apple Watch users (five or fewer heart rate values per day), the data of Apple users were excluded. Among 45 Fitbit users, 34 were left for further analysis after applying pre-processing



(a) (b)

Fig. 5. MetaCovid's Performance on (a) Various Window Lengths (b) Combination of Raw Sensor and Rhythmic Features

as described in section 3.2. Subsequently, the raw sensor data was segmented into overlapping 2-day windows with a 1-day offset between the start of consecutive windows. The labeled dataset was created by annotating days from D_{-6} to D_{+6} as infected, records prior to D_{-6} as healthy, and discarding all data after D_{+6}, where D_0 refers to symptom onset day. Figure 4(b) showcases the results of *MetaCovid* on Alavi's dataset and compares them to the performance on Mishra's dataset. *MetaCovid* achieved a precision of 0.95 and 0.83, recall of 0.97 and 0.92, AUC-ROC of 0.95 and 0.71, NPV of 0.95 and 0.77, and F_2 score of 0.97 and 0.8 when trained with K=2 and 10 respectively. Notably, *MetaCovid* demonstrated impressive performance with just two shots. However, post-symptomatic Covid-19 prediction contributed the most to evaluation metrics. It correctly identified 13 (38%) cases in the incubation period, 10 (29%) during D_0 to D_{+2}, delayed prediction of 7 (20%) subjects, and missed 4 (12%) cases with 2 shots, vs. 17 (50%) cases in the incubation period, 10 (29%) during D_0 to D_{+2}, delayed prediction of 3 (9%) subjects, and missed 4 (12%) cases with 10 shots.

5.3. Comparison with Previous Studies

In this section, we conduct a comparative analysis of *MetaCovid* with existing studies that aim to detect Covid-19 early using Mishra's or Alavi's datasets. We found that on Mishra's dataset, *MetaCovid* outperformed Abir et al. (2022)Bogu & Snyder (2021) in terms of both recall and F_2 score. Even though they achieved much higher precision, for highly contagious infectious diseases such as Covid-19, which has a median R_0 of 5.5 Sanche et al. (2020) (one infected person can potentially transmit the disease to 5 to 6 people), it is crucial to prioritize a lower false negative rate. Cho et al. Cho et al. (2022) and Chung et al. Chung et al. (2023) reported quite impressive results, however, their approach requires 28 days of historical data per subject for estimating baseline behavior, making it impractical to achieve the overall goal of Covid-19 identification using minimal data. In contrast, *MetaCovid* only utilizes the past 2 days to extract features that accurately predict the subject's Covid-19 status. CovidRhythm Sarwar et al. (2023) is the only study that utilizes a mere 24-hour data to predict Covid-19 one day before symptom onset and reported an AUC-ROC of 0.79, recall of 0.69, and a precision of 0.75. It is worth noting that, unlike MetaCovid, CovidRhythm uses only a subset of data to train the model. We reproduced CovidRhythm's result on the entire dataset and encountered a (40%) decline in the recall as the model struggled to handle varying vital signs patterns. On Alavi's dataset, *MetaCovid* outperformed prior work significantly regarding the recall and F_2 score [1]. To the best of our knowledge, *MetaCovid* is the first work that identifies Covid-19 early with a recall of 0.81 by leveraging only two shots per subject for training the model and requiring only data from the previous 2 days to estimate the subject's health status.

5.4. Evaluation of Effects of Different Data Window Lengths

MetaCovid aims to predict Covid-19 infection early using the heart rate and step values segmented into 2-day windows with 1-day offset between consecutive windows. In this section, we aim to determine whether using the window of 2 days achieves the highest performance or if a length exists that can achieve better/same results. Figure 5(a) provides the results for different window lengths. Upon analysis, we observed that *MetaCovid* achieved the highest performance when trained using 2 or 5 days. However, though the performance is promising regarding recall, precision, and F_2 score at the length of 5, the AUC-ROC and NPV are relatively low, suggesting that MetaCovid's ability to distinguish between healthy and infected samples is poor. We believe that utilizing heart rate and steps from the past 5 days introduced significant variations, making it challenging to learn a generalized representation of the subject's healthy behavior. We selected window length of 2 as it exhibits more stable performance in terms of AUC-ROC, NPV, and Recall. Moreover, a window length of 2 days facilitates earlier pre-symptomatic Covid-19 detection.

5.5. Efficacy of MetaCovid's Components

5.5.1. Raw Sensor Data vs. Rhythmic Features

MetaCovid utilizes raw sensor data (heart rate and step) and rhythmic features to predict Covid-19. Analysis was performed to investigate whether using raw sensor data combined with rhythmic features was the optimal approach for *MetaCovid*. Figure 5(b) illustrates the performance

[1] As vaccination may significantly alter patients' physiological signs including heart rate and step count, CovidRhythm primarily focuses on detecting Covid-19 in unvaccinated individuals. Consequently, we chose not to assess its performance on Alavi's dataset, which contains data collected from both vaccinated and unvaccinated individuals.

Table 4. *MetaCovid* vs. Previous Studies

Model	Precision	Recall	AUC-ROC	F_2 Score	Accuracy
Mishra's Dataset					
MetaCovid	0.82	0.81	0.78	0.81	0.81
PCovNet Abir et al. (2022)	0.95	0.23	-	0.27	-
Bogu et al. Bogu & Snyder (2021)	0.91	0.36	-	0.41	-
Cho et al. Cho et al. (2022)	-	-	-	-	0.85
CovidRhythm Sarwar et al. (2023)	0.75	0.69	0.79	0.7	0.83
Chung et al. Chung et al. (2023)	-	0.86	0.89	-	0.85
Alavi's Dataset					
MetaCovid	0.83	0.92	0.71	0.9	0.81
Chung et al. Chung et al. (2023)	-	0.67	0.81	-	0.79
PCovNet+ Abir et al. (2023)	0.99	0.48	-	0.53	-

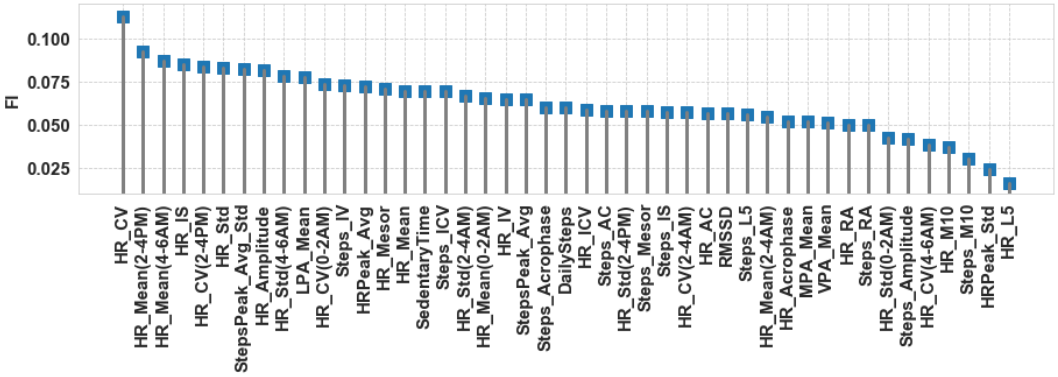


Fig. 6. Feature Importance (FI) computed using Permutation Feature Importance. The features importance presented were averaged over 5-folds. CV = Coefficients of Variation, ICV = Inter-daily Coefficient of variation, AC = Rhythm Auto-correlation, RA = Relative Amplitude, LPA = Light Intensity Physical Activity

of *MetaCovid* when trained on the combination of sensor data and rhythmic features. Our findings revealed that both could predict Covid-positive cases. However, rhythmic features also exhibited the ability to identify healthy subjects with only a few samples achieving an NPV of 0.84 in contrast to the sensor data, with an NPV of 0.15. Overall, the fusion of both raw data and rhythmic features had the most discriminative power.

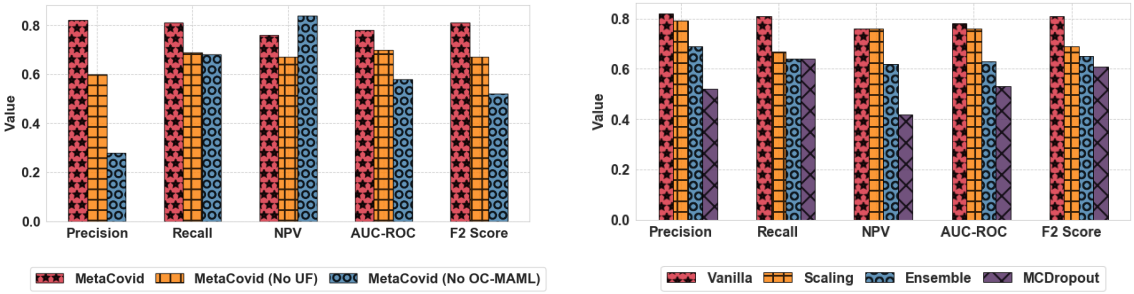
Figure 6 showcases the feature importance of rhythmic features computed using Permutation Feature Importance (PFI) Fisher et al. (2019). PFI evaluates the impact of individual features on the model's performance by measuring the change in the model's loss when a specific feature is randomly shuffled. The primary aim is to disrupt the association between the feature and the target variable. Therefore, an increase in the model loss indicates how much the model depends on that particular feature. PFI takes the trained model f_θ, feature matrix X and ground truth y, and estimates the original error, i.e., $e_{org} = \hat{L}(y, f_\theta)$, which, in our case, is represented by Binary Cross Entropy Loss (BCE). For each feature within the feature set, PFI shuffles that particular feature while keeping other features unchanged and computes the permutation error e_{perm}. The feature importance (FI) of j^{th} feature is then calculated as:

$$FI_j = e_{perm} - e_{org}$$

where FI_j refers to Feature importance of j^{th} feature, e_{perm} to BCE error with j^{th} feature shuffled, and e_{org} to the original loss. Heart rate was found to be the most affected physiological sign, as 15 of the top 20 features relate to HR disruption. Our results align with the findings of Tsai et al. Tsai et al. (2021), who reported that patients hospitalized with Covid-19 were reported to have cardiac arrhythmia. Furthermore, we also observed that abnormalities in nighttime HR were more predictive in Covid-infected patients compared to daytime HR. These findings highlight the potential of Covid-19-induced heart rate irregularities, particularly during nocturnal hours, to be leveraged for reliable, early detection of Covid-19.

5.5.2. ***OC-MAML***

Figure 7(a) shows the performance comparison of *MetaCovid* when trained with and without OC-MAML. Without OC-MAML, *MetaCovid* generates a substantial number of false positives and struggles to discriminate infected and healthy cases, as evidenced by its AUC-ROC score of 0.58, and precision of 0.28. However, the introduction of OC-MAML provided notable benefits: i) Effectively handling class imbalance, as highlighted by AUC-ROC, and ii) Addressing variability in vital signs, yielding stable recall and precision.



(a) (b)

Fig. 7. Evaluation of (a) Effectiveness of *MetaCovid's* Components. (UF refers to Uncertainty Filtering (b) Various Uncertainty Estimation Methods on *MetaCovid* Performance

5.5.3. ***Uncertainty Filtering***

Figure 7(a) is a performance comparison of *MetaCovid* with and without Uncertainty Filtering (UF). When UF is not employed, *MetaCovid* experienced a performance drop. Specifically, precision dropped by 27%, recall by 16%, NPV by 12%, AUC-ROC by 10%, and F_2 score by 17%. These results highlight the fact that *MetaCovid* exhibits more uncertainty regarding classifying healthy cases than infected ones.

Four methods for computing the uncertainty estimate were selected from the probabilistic deep learning literature and compared. These methods were selected by considering their prevalence, scalability, and practical applicability Ovadia et al. (2019). These methods are:

1. *Vanilla:* Sigmoid probability directly from the deep model.
2. *Scaling Guo et al. (2017):* Post-hoc calibration by temperature scaling the vanilla probabilities by value T.
3. *MCDropout Gal & Ghahramani (2016):* Monto-Carlo Dropout with the value p. An example was fed into the deep model N times to compute the average probability.
4. *Ensemble Lakshminarayanan et al. (2017):* Average probability generated from an ensemble of identical M networks trained independently on the same dataset with random initialization.

Figure 7(b) depicts *MetaCovid's* performance when trained using the above-mentioned approaches. It was observed that the vanilla method achieved the highest performance, whereas MCDropout exhibited the least favorable results.

5.6. Experiments with various Pre-Processing Approaches

It is crucial to preprocess the raw data appropriately to achieve optimal results. *MetaCovid* suggests normalizing the data and imputing any missing values as pre-processing methods. This section sought to identify the most suitable methods for normalizing and imputing the heart rate and step data that yield the best performance.

5.6.1. ***Normalization***

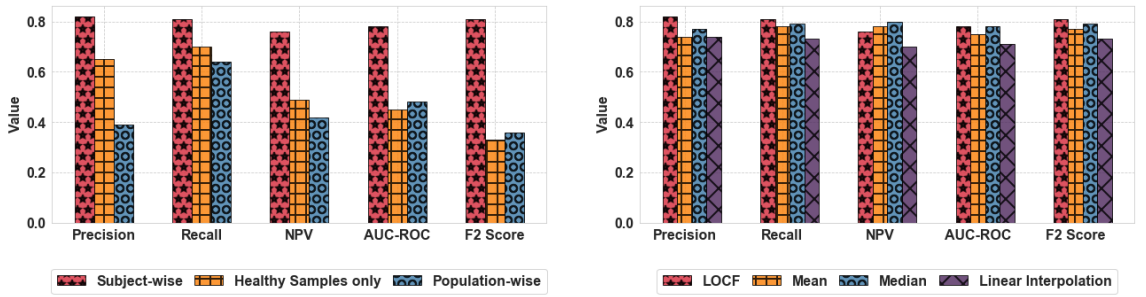
Z-score normalization has emerged as one of the most widely used methods for normalizing the physiological signs in various studies focused on early Covid-19 detection. For *MetaCovid*, we explored three variations of Z-score Normalization, i.e., i) *Subject-wise:* Scale each subject's heart rate and step data using his/her mean and standard deviation calculated over study period Romine et al. (2020), ii) *Population-wise:* Each subject's raw data is scaled using mean and standard deviation calculated over the entire population, and iii) *Healthy Samples Only:* The raw data is normalized using mean and standard deviation calculated solely from the subject's healthy samples Shu et al. (2020). Figure 8(a) illustrates the results when *MetaCovid* is trained using these variations. We found that the subject-wise normalization outperformed the other variants, emerging as the most suitable technique for normalizing the heart rate and step signals for early Covid-19 detection.

5.6.2. ***Imputation Methods***

To address the missing values in the data, we employed five imputation techniques: Last Observed Carry Forward (LOCF), Mean, Median, and Linear interpolation. LOCF replaces missing values with the most recently observed values. Mean, and Median fill missing values using day-wise mean and median, while linear interpolation estimates the missing value by connecting nearest points to the left and right in increasing order. After conducting our analysis (Figure 8(b)), LOCF was selected as the preferred method to handle missing values in the data.

6. Discussion

Main Findings: This study investigated using physiological signs (heart rate and step) collected from consumer-grade health trackers for pre-symptomatic detection of Covid-19, i.e, during D_{-6} to D_{+6}, where D_0 refers to symptom onset day. Our analyses of previous work revealed that variations in physiological signs (high intra/inter-subject variability) introduce domain shifts in the data distribution, causing traditional deep



(a) (b)

Fig. 8. Impact of Different (a) Variations of Z-score Normalization (b) Imputation Methods on MetaCovid's Performance

networks to struggle to perform well, especially with limited data. Prior studies have addressed this variability using approaches such as personalized training, estimating the baseline using extended historical data, or training on a subset of data. However, such approaches are impractical for various real-life scenarios.

Meta-learning enables *MetaCovid* to overcome vital signs variability. This is highlighted by its impressive performance, achieving a recall and F_2 score of 0.81, precision of 0.82, and AUC-ROC of 0.78 with only 2 healthy samples. Using few shots leads to accurate prediction of healthy and infected cases, however, increasing the value of K is associated with a decline in NPV. This implies that since healthy samples inherently have significant variations, the model struggles to overcome intra-subject variability as the healthy samples increase. On the other hand, increasing the value of Q in the validation set either maintains the recall or enhances it further, indicating that infected cases can be effectively identified using as few samples as 2, achieving the goal of Covid-19 screening with minimal data.

***MetaCovid* generalizes across different datasets**. Although *MetaCovid* achieved a recall of 0.97, and NPV of 0.95 with 2 shots on Alavi's dataset, most of the contribution towards evaluation metrics came from post-symptomatic identification. We hypothesize that this may be because the dataset comprises of vaccinated individuals, which can potentially weaken pre-Covid symptoms. With only two healthy samples, the model encounters difficulties in learning healthy subject patterns and distinguishing it from abnormal samples before symptom onset. Our assumption was supported by the results with K=10, where *MetaCovid* obtained a recall of 0.92, however, it identified 50% subjects as infected before symptom onset compared to 38% with K=2. Notably, with K=10, the NPV drops to 0.77, aligning with our earlier finding that increasing K leads to challenges in overcoming intra-subject vital signs variability.

Heart rate and step from the preceding 2 days can reliably identify pre-symptomatic Covid-19. While some studies have reported promising results for early detection of Covid-19, they need data from the past 28 days for estimating the individual's baseline behavior Chung et al. (2023). In contrast, *MetaCovid* only uses the heart rate and step measurements from the preceding 2 days, reducing the data requirement by 93%. Additionally, it was observed that a window length of 5, which used 5 days of historical data, yields the highest recall of 0.89. However, the poor AUC-ROC and NPV indicate that the model struggles to detect healthy subjects accurately. Our findings implied that deviations in the physiological signs from the preceding 1 or 2 days are sufficient for predicting healthy and infected samples. After 2 days, although the recall was encouraging, the overall performance dropped.

Rhythmic features had the most discriminative power for early Covid-19 identification. In contrast, raw sensor data only identified infected samples accurately, but failed to detect healthy subjects. On the other hand, rhythmic features (daily steps, night-time HR, daytime HR, MESOR, L5) achieved stable performance regarding both recall and NPV. When both sensor data and rhythmic features were employed to train *MetaCovid*, rhythmic features primarily identified healthy subjects, while both feature sets contributed equally to predicting Covid-positive cases. Among all features, heart rate parameters had the most discriminative power, revealing that the disruptions in the heart rate, specifically at night-time, were more pronounced in Covid-19 patients.

Study limitations Despite achieving encouraging results, our work had a few limitations. First, the study focused exclusively on Fitbit users that may not be representative of the general population. Second, symptom-onset dates were self-reported and may be incorrect. Third, the dataset contained a significant number of missing values, forcing us to: i) Omit data collected on several days, which also reduced the size of the dataset and potentially resulted in the loss of valuable information, and ii) Perform analyses using a 1-hour resolution. Increasing resolution to minutes could potentially capture abnormal physiological signs more accurately. Lastly, other health conditions such as influenza, chronic pain, and stress can also impact heart rate and physical activity and confound results. To strengthen our findings, additional research with a more diverse dataset that includes subjects that have these confounding conditions is required to distinguish disruptions in physiological signs caused by Covid from other health factors.

Future Work In the future, we aim to investigate the following ideas: i) Validate our results on a more extensive and diverse dataset that includes Covid-19-positive patients, healthy controls, and subjects with infections other than Covid-19 ii) Evaluation of potential enhancements of the results by applying other advanced neural networks architectures and anomaly detection techniques iii) Reproduce results using the data from other types and models of wearables iv) As one's work routine immensely impacts physical activity, investigate the impact of heart rate or resting heart rate, as observed in prior studies in order to address challenges posed by intra-subject variability.

7. Conclusion

Mobile sensing, which utilizes machine learning methods to analyze physiological data from consumer-grade wearables has opened new opportunities for passive assessment of infectious diseases including Covid-19. However, humans exhibit substantial individual differences in vital signs, posing challenges for conventional deep models to perform convincingly. To address the vital signs variability among subjects, this study introduces *MetaCovid*, which employs meta-learning for screening Covid-19 from heart rate and steps collected using Fitbit. *MetaCovid* extracts 45 novel digital biomarkers using a 2-day window, which is combined with raw sensor data and input to a deep GRU-based attention network for classification. In rigorous evaluation, *MetaCovid* outperformed all baselines, achieving a recall and F_2 score of 0.81, precision of 0.82, NPV OF 0.76, and AUC-ROC of 0.78 with only two shots per subject. *MetaCovid* generalized well on two different Covid-19 datasets, identifying 61% (14 out of 23) and 50% (17 out of 34) subjects before symptom onset. Remarkably, Covid infection can be accurately detected using only heart rate and step measurements over 2 days compared to 28 days required by state-of-art methods, reducing data requirement by 93%. We believe that the proposed meta-learning approach and insights from our study will pave the path for devising innovative interventions with limited data for containing the spread of infectious diseases in future.

References

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

Abir, F. F., Chowdhury, M. E., Tapotee, M. I., Mushtak, A., Khandakar, A., Mahmud, S., & Hasan, M. A. (2023). Pcovnet+: A cnn-vae anomaly detection framework with lstm embeddings for smartwatch-based covid-19 detection. *Engineering Applications of Artificial Intelligence*, *122*, 106130.

Alavi, A., Bogu, G. K., Wang, M., Rangan, E. S., Brooks, A. W., Wang, Q., Higgs, E., Celli, A., Mishra, T., Metwally, A. A. et al. (2022). Real-time alerting system for covid-19 and other stress events using wearable data. *Nature medicine*, *28*, 175–184.

Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, .

Banluesombatkul, N., Ouppaphan, P., Leelaarporn, P., Lakhan, P., Chaitusaney, B., Jaimchariyatam, N., Chuangsuwanich, E., Chen, W., Phan, H., Dilokthanakul, N. et al. (2020). Metasleeplearner: A pilot study on fast adaptation of bio-signals-based sleep stage classifier to new individual subject using meta-learning. *IEEE Journal of Biomedical and Health Informatics*, *25*, 1949–1963.

Biostrap (2023). Chest straps. URL: `https://shop.biostrap.com/products/cheststrap-hrm` [Online; accessed 14-September-2023].

Bogu, G. K., & Snyder, M. P. (2021). Deep learning-based detection of covid-19 using wearables data. *MedRxiv*, .

Breda, J., Springston, M., Mariakakis, A., & Patel, S. (2023). Feverphone: Accessible core-body temperature sensing for fever monitoring using commodity smartphones. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *7*, 1–23.

Cevik, M., Tate, M., Lloyd, O., Maraolo, A. E., Schafers, J., & Ho, A. (2021). Sars-cov-2, sars-cov, and mers-cov viral load dynamics, duration of viral shedding, and infectiousness: a systematic review and meta-analysis. *The lancet microbe*, *2*, e13–e22.

Cho, H. R., Kim, J. H., Yoon, H. R., Han, Y. S., Kang, T. S., Choi, H., & Lee, S. (2022). Machine learning-based optimization of pre-symptomatic covid-19 detection through smartwatch. *Scientific Reports*, *12*, 7886.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, .

Chung, H., Ko, H., Lee, H., Yon, D. K., Lee, W. H., Kim, T.-S., Kim, K. W., & Lee, J. (2023). Development and validation of a deep learning model to diagnose covid-19 using time-series heart rate values before the onset of symptoms. *Journal of Medical Virology*, .

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*, .

Dolezal, J. M., Srisuwananukorn, A., Karpeyev, D., Ramesh, S., Kochanny, S., Cody, B., Mansfield, A. S., Rakshit, S., Bansal, R., Bois, M. C. et al. (2022). Uncertainty-informed deep learning models enable high-confidence predictions for digital histopathology. *Nature communications*, *13*, 6572.

Elias, C., Sekri, A., Leblanc, P., Cucherat, M., & Vanhems, P. (2021). The incubation period of covid-19: A meta-analysis. *International Journal of Infectious Diseases*, *104*, 708–710.

Empatica (2023). E4 wristband. URL: `https://www.empatica.com/research/e4/` [Online; accessed 14-September-2023].

Finn, C., Abbeel, P., & Levine, S. (2017). Model-agnostic meta-learning for fast adaptation of deep networks. In *International conference on machine learning* (pp. 1126–1135). PMLR.

Fisher, A., Rudin, C., & Dominici, F. (2019). All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. *J. Mach. Learn. Res.*, *20*, 1–81.

Frikha, A., Krompaß, D., Köpken, H.-G., & Tresp, V. (2021). Few-shot one-class classification via meta-learning. In *Proceedings of the AAAI Conference on Artificial Intelligence* (pp. 7448–7456). volume 35.

Gal, Y., & Ghahramani, Z. (2016). Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning* (pp. 1050–1059). PMLR.

Garmin (2023). Smartwatches. URL: `https://www.garmin.com/en-US/c/wearables-smartwatches/` [Online; accessed 14-September-2023].

Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). On calibration of modern neural networks. In *International conference on machine learning* (pp. 1321–1330). PMLR.

Halberg, F., Tong, Y. L., & Johnson, E. (1967). Circadian system phase—an aspect of temporal morphology; procedures and illustrative examples. In *The cellular aspects of biorhythms* (pp. 20–48). Springer.

He, X., Lau, E. H., Wu, P., Deng, X., Wang, J., Hao, X., Lau, Y. C., Wong, J. Y., Guan, Y., Tan, X. et al. (2020). Temporal dynamics in viral shedding and transmissibility of covid-19. *Nature medicine*, *26*, 672–675.

Karjalainen, J., & Viitasalo, M. (1986). Fever and cardiac rhythm. *Archives of internal medicine*, *146*, 1169–1171.

Karlsson, L. C., Soveri, A., Lewandowsky, S., Karlsson, L., Karlsson, H., Nolvi, S., Karukivi, M., Lindfelt, M., & Antfolk, J. (2021). Fearing the disease or the vaccine: The case of covid-19. *Personality and individual differences*, *172*, 110590.

Kruspe, A. (2019). One-way prototypical networks. *arXiv preprint arXiv:1906.00820*, .

Lakshminarayanan, B., Pritzel, A., & Blundell, C. (2017). Simple and scalable predictive uncertainty estimation using deep ensembles. *Advances in neural information processing systems*, *30*.

Liono, J., Abdallah, Z. S., Qin, A. K., & Salim, F. D. (2018). Inferring transportation mode and human activity from mobile sensing in daily life. In *Proceedings of the 15th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services* (pp. 342–351).

Liu, S., Han, J., Puyal, E. L., Kontaxis, S., Sun, S., Locatelli, P., Dineley, J., Pokorny, F. B., Dalla Costa, G., Leocani, L. et al. (2022). Fitbeat: Covid-19 estimation based on wristband heart rate using a contrastive convolutional auto-encoder. *Pattern recognition*, *123*, 108403.

Liu, X., Jiang, Z., Fromm, J., Xu, X., Patel, S., & McDuff, D. (2021). Metaphys: few-shot adaptation for non-contact physiological measurement. In *Proceedings of the conference on health, inference, and learning* (pp. 154–163).

Mayer, C., Tyler, J., Fang, Y., Flora, C., Frank, E., Tewari, M., Choi, S. W., Sen, S., & Forger, D. B. (2022). Consumer-grade wearables identify changes in multiple physiological systems during covid-19 disease progression. *Cell Reports Medicine*, *3*, 100601.

Mim, T. R., Amatullah, M., Afreen, S., Yousuf, M. A., Uddin, S., Alyami, S. A., Hasan, K. F., & Moni, M. A. (2023). Gru-inc: An inception-attention based approach using gru for human activity recognition. *Expert Systems with Applications*, *216*, 119419.

Mishra, T., Wang, M., Metwally, A. A., Bogu, G. K., Brooks, A. W., Bahmani, A., Alavi, A., Celli, A., Higgs, E., Dagan-Rosenfeld, O. et al. (2020). Pre-symptomatic detection of covid-19 from smartwatch data. *Nature biomedical engineering*, *4*, 1208–1220.

Moškon, M. (2020). Cosinorpy: a python package for cosinor-based rhythmometry. *BMC bioinformatics*, *21*, 1–12.

Narin, A., Isler, Y., Ozer, M., & Perc, M. (2018). Early prediction of paroxysmal atrial fibrillation based on short-term heart rate variability. *Physica A: Statistical Mechanics and its Applications*, *509*, 56–65.

Olagoke, A. A., Olagoke, O. O., & Hughes, A. M. (2021). Intention to vaccinate against the novel 2019 coronavirus disease: The role of health locus of control and religiosity. *Journal of religion and health*, *60*, 65–80.

Oura (2023). Smart rings. URL: `https://ouraring.com/` [Online; accessed 14-September-2023].

Ovadia, Y., Fertig, E., Ren, J., Nado, Z., Sculley, D., Nowozin, S., Dillon, J., Lakshminarayanan, B., & Snoek, J. (2019). Can you trust your model's uncertainty? evaluating predictive uncertainty under dataset shift. *Advances in neural information processing systems*, *32*.

O'Brien, M. W., Kivell, M. J., Wojcik, W. R., d'Entremont, G., Kimmerly, D. S., & Fowles, J. R. (2018). Step rate thresholds associated with moderate and vigorous physical activity in adults. *International Journal of Environmental Research and Public Health*, *15*, 2454.

Physiopedia (). Vital signs. URL: `https://www.physio-pedia.com/Vital_Signs` [Online; accessed 4-September-2023].

Pulse, & HR (2022). Pulse and heart rate. URL: `https://my.clevelandclinic.org/health/diagnostics/17402-pulse--heart-rate` [Online; accessed 29-June-2023].

Romine, W. L., Schroeder, N. L., Graft, J., Yang, F., Sadeghi, R., Zabihimayvan, M., Kadariya, D., & Banerjee, T. (2020). Using machine learning to train a wearable device for measuring students' cognitive load during problem-solving activities based on electrodermal activity, body temperature, and heart rate: development of a cognitive load tracker for both personal and classroom use. *Sensors*, *20*, 4833.

Rong, X., Yang, L., Chu, H., & Fan, M. (2020). Effect of delay in diagnosis on transmission of covid-19. *Math Biosci Eng*, *17*, 2725–2740.

Ruby, D. (2023). Smartwatch statistics 2023: How many people use smartwatches? URL: `https://www.demandsage.com/smartwatch-statistics/` [Online; accessed 26-June-2023].

Ruopp, M. D., Perkins, N. J., Whitcomb, B. W., & Schisterman, E. F. (2008). Youden index and optimal cut-point estimated from observations affected by a lower limit of detection. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, *50*, 419–430.

Rykov, Y., Thach, T.-Q., Bojic, I., Christopoulos, G., Car, J. et al. (2021). Digital biomarkers for depression screening with wearable devices: cross-sectional study with machine learning modeling. *JMIR mHealth and uHealth*, *9*, e24872.

Sanche, S., Lin, Y. T., Xu, C., Romero-Severson, E., Hengartner, N., & Ke, R. (2020). High contagiousness and rapid spread of severe acute respiratory syndrome coronavirus 2. *Emerging infectious diseases*, *26*, 1470.

Sarwar, A., Agu, E. O., & Almadani, A. (2023). Covidrhythm: A deep learning model for passive prediction of covid-19 using biobehavioral rhythms derived from wearable physiological data. *IEEE Open Journal of Engineering in Medicine and Biology*, *4*, 21–30.

Shu, L., Yu, Y., Chen, W., Hua, H., Li, Q., Jin, J., & Xu, X. (2020). Wearable emotion recognition using heart rate data from a smart bracelet. *Sensors*, *20*, 718.

Triantafillou, E., Zhu, T., Dumoulin, V., Lamblin, P., Evci, U., Xu, K., Goroshin, R., Gelada, C., Swersky, K., Manzagol, P.-A. et al. (2019). Meta-dataset: A dataset of datasets for learning to learn from few examples. *arXiv preprint arXiv:1903.03096*, .

Tsai, P.-H., Lai, W.-Y., Lin, Y.-Y., Luo, Y.-H., Lin, Y.-T., Chen, H.-K., Chen, Y.-M., Lai, Y.-C., Kuo, L.-C., Chen, S.-D. et al. (2021). Clinical manifestation and disease progression in covid-19 infection. *Journal of the Chinese Medical Association*, *84*, 3–8.

Vitaterna, M. H., Takahashi, J. S., & Turek, F. W. (2001). Overview of circadian rhythms. *Alcohol research & health*, *25*, 85.

Zhang, L., Wu, H., Zhang, X., Wei, X., Hou, F., & Ma, Y. (2020). Sleep heart rate variability assists the automatic prediction of long-term cardiovascular outcomes. *Sleep medicine*, *67*, 217–224.

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: