



ELSEVIER

Contents lists available at ScienceDirect

Pattern Recognition

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

Fitbeat: COVID-19 estimation based on wristband heart rate using a contrastive convolutional auto-encoder



Shuo Liu^{a,*}, Jing Han^{a,b}, Estela Laporta Puyal^{c,d}, Spyridon Kontaxis^{c,d}, Shaoxiong Sun^e, Patrick Locatelli^f, Judith Dineley^a, Florian B. Pokorny^{a,g}, Gloria Dalla Costa^h, Letizia Leocani^h, Ana Isabel Guerreroⁱ, Carlos Nosⁱ, Ana Zabalzaⁱ, Per Soelberg Sørensen^j, Mathias Buron^j, Melinda Magyari^j, Yatharth Ranjan^e, Zulqarnain Rashid^e, Pauline Conde^e, Callum Stewart^e, Amos A Folarin^{e,k}, Richard JB Dobson^{e,k}, Raquel Bailón^{c,d}, Srinivasan Vairavan^l, Nicholas Cummins^{a,e}, Vaibhav A Narayan^l, Matthew Hotopf^{m,n}, Giancarlo Comi^o, Björn Schuller^{a,p}, RADAR-CNS Consortium^q

^a EIHW – Chair of Embedded Intelligence for Health Care and Wellbeing, University of Augsburg, Augsburg, Germany

^b Department of Computer Science and Technology, University of Cambridge, Cambridge, United Kingdom

^c BSICoS Group, Aragón Institute of Engineering Research (I3A), IIS Aragón, University of Zaragoza, Zaragoza, Spain

^d CIBER of Bioengineering, Biomaterials and Nanomedicine (CIBER-BNN), Madrid, Spain

^e The Department of Biostatistics and Health Informatics, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK

^f Department of Engineering and Applied Science, University of Bergamo, Bergamo, Italy

^g Division of Phoniatrics, Medical University of Graz, Graz, Austria

^h Università Vita Salute San Raffaele and Experimental Neurophysiology Unit, Institute of Experimental Neurology, Scientific Institute Hospital San Raffaele, Milan, Italy

ⁱ Multiple Sclerosis Centre of Catalonia (Cemcat), Department of Neurology/Neuroimmunology, Hospital Universitari Vall d'Hèbron, Universitat Autònoma de Barcelona, Barcelona, Spain

^j Danish Multiple Sclerosis Centre, Department of Neurology, Copenhagen University Hospital Rigshospitalet, Copenhagen, Denmark

^k Institute of Health Informatics, University College London, London, United Kingdom

^l Janssen Research and Development LLC, Titusville, NJ, USA

^m The Department of Psychological Medicine, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom

ⁿ South London and Maudsley National Health Service Foundation Trust, London, United Kingdom

^o Università Vita Salute San Raffaele, Casa di Cura Privata del Policlinico, Milan, Italy

^p GLAM – Group on Language, Audio, & Music, Imperial College London, London, United Kingdom

^q The RADAR-CNS Consortium, London, United Kingdom

ARTICLE INFO

Article history:

Received 5 April 2021

Revised 30 August 2021

Accepted 24 October 2021

Available online 26 October 2021

Keywords:

COVID-19

Respiratory tract infection

Anomaly detection

Contrastive learning

Convolutional auto-encoder

ABSTRACT

This study proposes a contrastive convolutional auto-encoder (contrastive CAE), a combined architecture of an auto-encoder and contrastive loss, to identify individuals with suspected COVID-19 infection using heart-rate data from participants with multiple sclerosis (MS) in the ongoing RADAR-CNS mHealth research project. Heart-rate data was remotely collected using a Fitbit wristband. COVID-19 infection was either confirmed through a positive swab test, or inferred through a self-reported set of recognised symptoms of the virus. The contrastive CAE outperforms a conventional convolutional neural network (CNN), a long short-term memory (LSTM) model, and a convolutional auto-encoder without contrastive loss (CAE). On a test set of 19 participants with MS with reported symptoms of COVID-19, each one paired with a participant with MS with no COVID-19 symptoms, the contrastive CAE achieves an unweighted average recall of 95.3%, a sensitivity of 100% and a specificity of 90.6%, an area under the receiver operating characteristic curve (AUC-ROC) of 0.944, indicating a maximum successful detection of symptoms in the given heart rate measurement period, whilst at the same time keeping a low false alarm rate.

© 2021 Elsevier Ltd. All rights reserved.

* Corresponding author.

E-mail address: shuo.liu@informatik.uni-augsburg.com (S. Liu).

URL: <http://www.radar-cns.org> (R.-C. Consortium)

<https://doi.org/10.1016/j.patcog.2021.108403>

0031-3203/© 2021 Elsevier Ltd. All rights reserved.

1. Introduction

Remote passive monitoring of physiological and behavioural characteristics using smartphones and wearable devices can be used to rapidly collect a variety of data in huge volumes with minimal effort from the wearer. Such data has the potential to improve our understanding of the interplay between a variety of health conditions at individual and population level, if rigorously collected and validated [1]. Passive data collection is typically implemented with a high temporal resolution [2]. Wearable fitness trackers, for example, estimate parameters such as heart rate up to every second and up to 24 hours a day. Monitoring individuals with a range of health states, lifestyles, and demographic variables in combination with data artefacts and missing data leads to high variability, while multiple data streams, from heart rate and physical activity to GPS-based location, can be collected. Therefore, studies using wearables and smartphones in this way exhibit several vs of big data: velocity, volume, variability and variety. As such, advanced analysis methodologies such as deep learning can potentially make a significant contribution [3], particularly in the context of infectious diseases, such as COVID-19, the disease caused by the novel corona virus (SARS-CoV-2). Specific applications include individual screening and population-level monitoring that minimise contact with infected individuals [4,5].

Since the outbreak of the COVID-19 pandemic in 2020, several deep learning methodologies have been applied to computed tomography (CT) scans [6] and 2D X-ray images [7] to detect COVID-19. These methods require specific clinical equipment and the patient must attend a clinical facility. Consequently, it cannot achieve early, automatic detection when COVID-19 symptoms first appear. In contrast, heart rate can be measured remotely and non-intrusively using wearable devices.

Heart rate is a biomarker of particular value in such applications. Patterns in heart rate fluctuations over time have been found to provide clinically relevant information about the integrity of the physiological system generating these dynamics. Previous studies have not only revealed an altered heart rate variability in a number of medical conditions [8], but also demonstrated that the degree of short-term heart rate alteration correlates with illness severity. Analysis of the autonomic regulation of heart rate has also been discussed as a promising approach for detecting infections earlier than conventional clinical methods and making prognoses [9].

Wearables such as Fitbit fitness trackers¹ provide indirect measurements of the heart rate through pulse rate estimates made using photoplethysmography (PPG). In the ongoing DETECT² study [5], researchers are focusing on monitoring outbreaks of viral infections including COVID-19 based on the resting heart rate collected in this way [10]. Other similar ongoing endeavours include the German project *Corona-Datenspende*³, which has a cohort of over 500 000 volunteers, and the TemPredict study in the US⁴.

Applied to such data sets, deep learning has the potential to automatically identify individuals with COVID-19 purely on the basis of data passively acquired by means of wearable devices [5,11,12]. To the best of our knowledge, the present study is the first to compare deep learning approaches in predicting the presence or absence of COVID-19-like symptoms using Fitbit-measured heart rate data. We aim to exploit state-of-the-art methods to represent the

problem by feature maps, including convolutional neural networks (CNNs) and a convolutional auto-encoder (CAE) [13].

Considering the deficiency of class information in training a standard CAE, in some previous works, the class information was applied to latent attribute layers, leading to the supervised auto-encoder introduced in [14]. Cross-entropy losses are used to minimise the difference between predicted labels from latent attributes and true labels. This approach provides a certain preservation of the reconstructed feature map, taking the cross-entropy loss as a regularisation method. The reconstruction error and cross entropy loss are jointly optimised. However, the optimisation of the joint loss requires a proper combination factor in order to balance the optimisation on reconstruction error and prediction error. Since the two types of errors originate from different stages of the auto-encoder model, leading to their different scale level, the difficulty lies in seeking a good combination scale.

To circumvent this problem, we consider the task at hand as analogous to anomaly detection [15] and propose a self-supervised training strategy by means of fitting the reconstruction error into the format of contrastive loss [16] instead of conventional loss like root mean square error (RMSE). In this way, contrastive loss can be employed directly on the reconstruction error for positive and negative input pairs. The method also enables validation of whether the model has learnt discriminative latent attributes for different classes in the auto-encoder framework.

We investigate the effectiveness of the proposed technique, comparing its performance to a CAE without contrastive loss, in addition to other standard deep learning methods including a multi-layer perceptron (MLP), a long short-term memory (LSTM) neural network, and a conventional CNN [13].

2. Related work

Recent work has investigated data streams that could potentially be used to detect COVID-19 and can be easily captured using smart devices and wearable equipment [17,18], including recordings of coughing and breathing [19] and speech signals [20,21]. Un et al. [22] proposed a machine learning-derived index reflecting overall health status of the patients with mild COVID-19, using the data captured from wearable biosensors. Hirten and colleagues [23] performed an evaluation of heart rate variability (HRV) collected by a wearable device to identify and predict COVID-19 and its related symptoms. Radin and colleagues [4] analysed the resting heart rate alongside with sleep duration data in over 47 000 individuals to improve model predictions of influenza rates in five US states. Quer et al. [5] and Mishra et al. [12] have shown the potential of using heart rate, sleep duration, and activity data, retrieved from smart wearable devices for COVID-19 recognition. Natarajan and colleagues [11] used a CNN to predict illness on a given day using Fitbit data from 1 181 individuals, reporting an area under the receiver operating characteristics curve (AUC-ROC) of 0.77 ± 0.03 .

This paper describes a deep learning approach applied to Fitbit measurement of heart rate to predict the presence or absence of COVID-19-like symptoms. We explore the suitability of using a CAE with contrastive loss, expecting to learn feature representations by contrasting symptomatic and asymptomatic samples. Contrastive learning has already been applied to detect COVID-19 from CT scans or X-ray images as in [24] and [25]. By using contrastive learning for a CAE, we aim to incorporate the class information into its reconstruction error to assist the model in achieving more differentiable latent attributes, and reaching sufficient distance between the reconstruction errors of symptomatic and asymptomatic samples.

¹ <https://www.fitbit.com/> [as of 03 August 2021].

² <http://detectstudy.org/> [as of 03 August 2021].

³ <http://corona-datenspende.de/science/en/> [as of 03 August 2021].

⁴ <http://osher.ucsf.edu/research/current-research-studies/tempredict> [as of 03 August 2021].

Table 1
Gender-, age-, and site-related distribution of participants per data subset.

		Pre-training	Positive participants for testing	Health control for testing
Genders	Female	14	5	5
	Male	35	14	14
Locations	Italy	18	7	7
	Spain	19	6	6
	Denmark	12	6	6
Ages	≤ 30	1	2	2
	30 - 39	10	3	4
	40 - 49	12	6	5
	50 - 59	19	6	6
	60 - 69	6	1	1
	≥ 70	1	-	-

3. Data collection

The data used in this work was collected as part of the IMI2 RADAR-CNS programme⁵, which is currently being conducted at multiple clinical sites in several European countries. Participant recruitment and data collection in the RADAR-MS study started in June 2018. As of March 1, 2020, 499 participants had been enrolled and 403 (81%) remained in the study [26].

Heart rate data was collected continuously 24-hours-a-day/7-days-a-week using a Fitbit wristband combined with participants' own Android smartphones where available, or a provided Motorola G5, G6, or G7. Fitbit Charge 2 or Charge 3 devices were provided to participants, who were asked to wear the device on their non-dominant hand. Meanwhile, an app-based questionnaire was distributed to all active participants on March 25, 2020 and again on April 8, 2020. By April 15, 2020, at least one of the questionnaires was completed by 399 participants (99%).

We used two definitions to determine the prevalence of COVID-19 in participants [26]: In the first, referred to as CD1, participants experience fever or anosmia/ageusia in combination with any other COVID-19 symptoms including respiratory symptoms, tiredness and gastrointestinal symptoms, or respiratory symptoms plus two other COVID-19 symptoms. In the second definition, CD2, participants experience fever plus any other COVID-19 symptoms, or respiratory symptoms plus anosmia/ageusia. Laboratory-confirmed cases are included in both case definitions [26].

We considered Fitbit heart rate measurements made between 21 February and 20 May 2020, from 87 participants in Denmark, Italy and Spain, with an age range from 23 to 73 years (mean = 46.5 ± 10.5 standard deviation). Sixty eight of these MS participants (30 female, 38 male) reported symptoms characteristic of COVID-19. However, in 49, symptoms did not meet CD1 or CD2 criteria. Heart rate data from these 49 participants was used for model pre-training (pre-training set). For testing, we applied leave one subject out (LOSO) cross-validation (CV) [27] on the data of the 19 MS participants, whose symptoms meet CD1 or CD2 criteria. Each of these 19 symptomatic participants was paired with a COVID-19-like symptom-free control participant with MS matched for site and gender and being at a similar age (cross-validation set). Table 1 summarises the numbers of participants per data subset as a function of the independent variables gender, age, and location.

Heart rate data of the participants were assigned into temporal segments, defining a 14-day interval extending from 7 days preceding symptom onset to 7 days following symptom onset in which we sought to identify infection-related variations in heart rate. The interval mainly covers the duration of the COVID-19 incubation pe-

riod [28], and minimises the anomalous effects of day-to-day variations in activity, such as those observed between weekdays and weekends.

Fig. 1 demonstrates the segmentation and subsequent data pre-processing procedure for the heart rate data of a participant with reported COVID-19-like symptoms. A heart rate segment over 14 days centred at 00:00 at the day of reported symptom onset, i. e., 7 consecutive days before the day of reported symptom onset plus 7 consecutive days starting with the day of reported symptom onset (red box on top of Fig. 1) is referred to as *symptomatic segment*. In contrast, an *asymptomatic segment* stands for any 14-days interval of consecutive heart rate data again starting at 0 o'clock that is at least 7 days distant from a symptomatic segment (green box in top of Fig. 1).

Asymptomatic segments were created by shifting a 14-days window in full day steps over periods at least 7 days distant from the boundaries of a symptomatic segment. With the chosen 7-days distance of asymptomatic segments from symptomatic segments we presume, that (i) a participant might not have already been infected 14 or more days prior to the onset of symptoms, and (ii) participants might have recovered from illness 14 days after the onset of symptoms at the latest. From the 49 participants of the pre-training set, totally 49 symptomatic segments and 1 470 asymptomatic segments are extracted. Since the number of available symptomatic and asymptomatic segments is highly imbalanced, we replicate the symptomatic segments to the number of asymptomatic segments to guide the detection model weighted in favour of the minority class. For the LOSO CV procedure, 19 symptomatic and 570 asymptomatic segments are acquired from participants with reported symptoms, and 1 140 asymptomatic segments from the control participants. An overview of available symptomatic and asymptomatic segments is given in Table 2.

Instantaneous heart rate estimates were derived from the PPG signal and ideally uploaded every five seconds (blue curve in the middle of Fig. 1). The mean of 5 min is taken to smooth heart rate measurements while still tracking slow short-term changes in the heart rate. Moreover, this approach alleviates the effect of different sampling rates of heart rate measures in Fitbit data, and missing estimates observed in real living conditions, both of which making it very difficult to study other features than mean heart rate in 5-minutes intervals. Furthermore, 5 min represents the time interval usually recommended for short-term heart rate variability analysis, assuming a constant mean heart rate (HRV).

Missing data over full 5-minutes intervals was filled with the median value of the overall 14-days segment, which guarantees for more robustness against outliers as compared to the mean value. Finally, we have a single heart rate value every 5 min. Despite the high completeness of heart rate segments (see Table 2), the missing data can be potentially spread over an entire heart rate segment. A too small short-term duration leads to more empty mean values, whereas a too large short-term duration results in the information loss of the variations in heart rate segments. The resulting smoothed heart rate trajectory is considered appropriate to detect global heart rate patterns associated with COVID-19-like symptoms (red curve in middle of Fig. 1).

We then transform the averaged heart rate data of each segment into a feature map, i. e., an image of size 24×168 pixels (bottom of Fig. 1), in which each pixel represents a 5-minutes heart rate sampling point. Thereby, each column encodes a heart rate trajectory of 2 h (24×5 min), resulting in a covered interval of overall 14 days by 168 columns (168×2 h). In our experiments, we verify that this set-up of the feature map is effective as the input of our deep learning models, leading to promising detection results.

⁵ <https://www.radar-cns.org/>

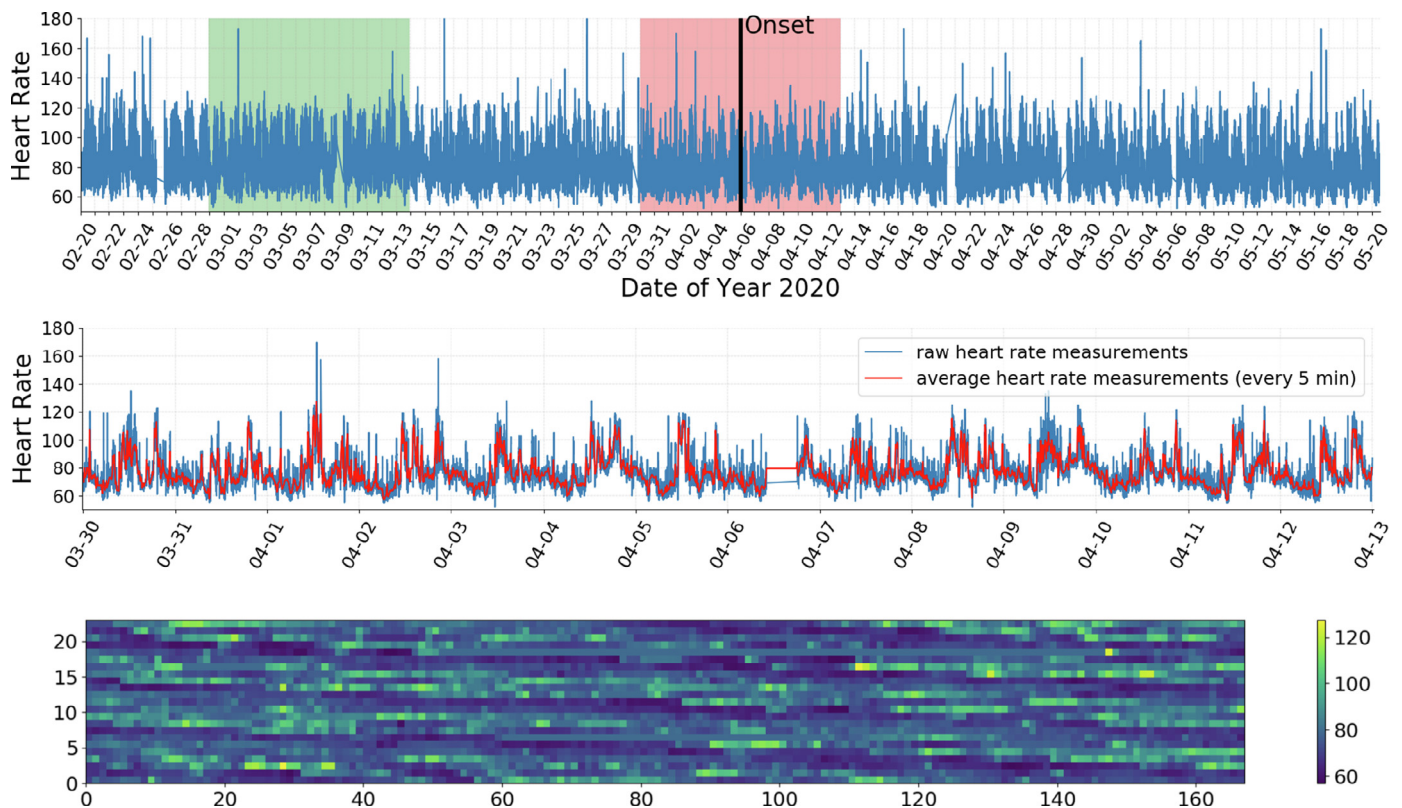


Fig. 1. Segmentation and pre-processing of heart rate data of a participant with reported COVID-19-like symptoms. **Top:** Heart rate data recorded 24-hours-a-day/7-days-a-week from 21 February to 20 May 2020 (total 90 days). *Onset* (black vertical bar) indicates 0 o'clock at the reported symptom onset date. Red rectangle – 7 days heart rate data before and after symptom onset representing a symptomatic segment; green rectangle – asymptomatic segment. **Middle:** Symptomatic segment. Blue curve – unprocessed heart rate trajectory of the red rectangle above; red curve – heart rate trajectory averaged over 5-minutes intervals. **Bottom:** Representation of the symptomatic segment as 24×168 sized image of 5-minutes heart rate data related pixels. Each column represents an interval of 2 h, the 168 columns sum up to 14 days.

Table 2

Available symptomatic and asymptomatic segments per data subset. Data completeness [%] of respective heart rate segments is given in parentheses (mean + std).

# (%)	Pre-training	Positive participants for testing	Health control for testing
Symptomatic	49 (98.7 ± 0.3)	19 (97.6 ± 0.2)	-
Asymptomatic	1470 (98.1 ± 0.4)	570 (97.4 ± 0.2)	1140 (99.2 ± 0.5)

4. Methodology

An approach to learn representations from a feature map is to use a CAE [29], which contains an encoder to learn latent attributes of the original input, and a decoder for reconstructing the original input from the learnt latent attribute. The dimensionality of the latent attributes is designed as a bottleneck imposed in the architecture. It hence can be seen as a compressed knowledge representation of the input. To reproduce the original input at the output of the decoder, the reconstruction error is minimised when optimising an auto-encoder network. To incorporate the class information during the optimisation, we apply contrastive loss [16] to the CAE reconstruction error in order to guide it to learn sufficiently discriminative latent attributes for different classes.

4.1. Architecture of CAE

The encoder part of our CAE is a stack of convolutional layers, an example of 4 layers is illustrated in Fig. 2. Following each convolutional layer, batch normalisation is used and a parametric rectified linear unit (PReLU) performs as the activation function. Max-pooling is then used to process the activations to reduce the spatial size of the feature maps. The encoder part is therefore a

sequential cascade of convolutional layer – batch normalisation – PReLU – max pooling. Given N heart rate segments, their features $[x_1, x_2, \dots, x_i, \dots, x_N]$ are created as introduced in Section 3. The encoder $f^{\text{enc}}(\cdot)$ processes a feature x_i , and its flattened output is linearly projected to latent attributes

$$h = f^{\text{enc}}(x_i). \quad (1)$$

The decoder presents an inverse processing of the encoder. For each decoder layer, the feature map mainly passes through transposed convolution and transposed max-pooling, also known as deconvolution and de-pooling. Batch normalisation is employed in between, followed by PReLU as the activation function. The decoder $f^{\text{dec}}(\cdot)$ outputs the reconstructed feature map

$$\hat{x}_i = f^{\text{dec}}(h). \quad (2)$$

The specifications of our CAE are given in Table 3. In experiments, we consider different numbers of convolutional layers in the CAE. The last encoder layer determines the dimensionality of the flatten layer, we hence adjust the length of its following fully-connected layer to optimise the CAE performance.

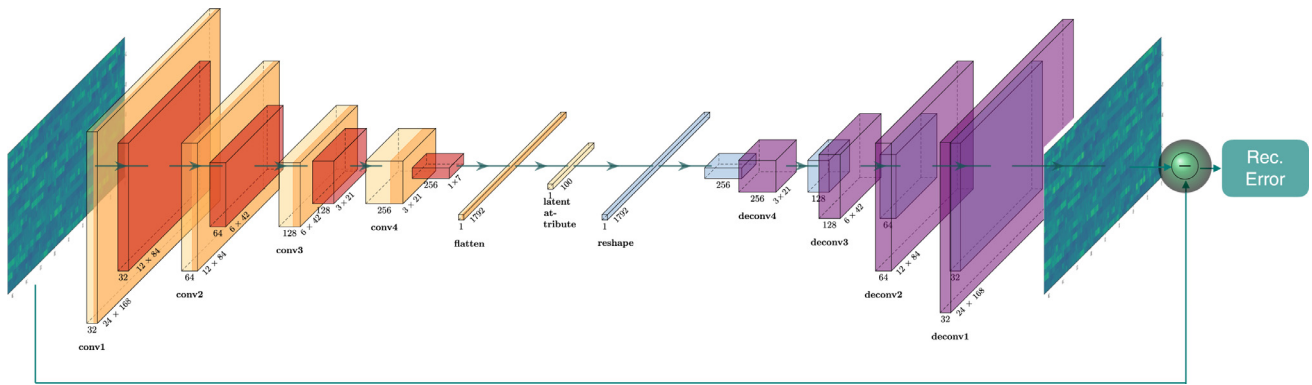


Fig. 2. The convolutional auto-encoder (CAE) architecture with 4 encoder layers and 4 decoder layers as an example. An encoder layer is a sequence of **convolution – batch-normalisation – PReLU – max-pooling**. A decoder layer is a sequence of **transposed convolution – batch-normalisation – PReLU – transposed max-pooling**. The distance between the original and reconstructed image represents the reconstruction error.

Table 3

Specifications of our CAE models. Each convolution and pooling layer, as well as de-convolution and de-pooling layer contains its own kernel size, stride, padding size, and number of channels. * = dimensionality depends on the total number of layers, ** = dimensionality of latent attributes. fc abbreviates fully-connected layer.

	Blocks	Kernel	Stride	Padding	# Channels
Encoder	conv1	(5,5)	(1,1)	(2,2)	32
	pool1	(2,2)	(2,2)	-	32
	conv2	(5,5)	(1,1)	(2,2)	64
	pool2	(2,2)	(2,2)	-	64
	conv3	(5,5)	(1,1)	(2,2)	128
	pool3	(2,2)	(2,2)	-	128
	conv4	(5,5)	(1,1)	(2,2)	256
	pool4	(3,3)	(3,3)	-	256
	conv5	(3,3)	(1,1)	(1,1)	512
	conv6	(3,3)	(1,1)	(1,1)	1024
	flatten		*		
	fc		**		
	fc		**		
Decoder	deconv6	(3,3)	(1,1)	(1,1)	512
	deconv5	(3,3)	(1,1)	(1,1)	256
	deconv4	(3,3)	(1,1)	(1,1)	128
	depool3	(3,3)	(3,3)	-	128
	deconv4	(5,5)	(1,1)	(2,2)	64
	depool4	(2,2)	(2,2)	-	64
	deconv5	(5,5)	(1,1)	(2,2)	32
	depool5	(2,2)	(2,2)	-	32
	deconv6	(5,5)	(1,1)	(2,2)	1
	depool6	(2,2)	(2,2)	-	1

An auto-encoder is typically optimised by minimising the reconstruction error, such as the root mean squared error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_i^N |x_i - \hat{x}_i|^2}. \quad (3)$$

The difficulty in finding good latent attributes lies in setting it to a proper dimensionality. Too long latent attributes may contain redundancies for easier reconstructing the original input, but fall short of concentrating on learning the salient discriminative features for different classes. Meanwhile, shorter latent attributes can have less or limited representation capability. Besides, the optimisation of an auto-encoder considers no class information, and hence the learnt latent attributes are not well oriented to be discriminative for different classes. Specifically, for our classification task, the auto-encoder may tend to learn the latent attributes that can better reconstruct the original feature map, while ignoring some salient attributes that indicate the difference between symptomatic and asymptomatic segments.

4.2. Contrastive loss

To incorporate class information – symptomatic and asymptomatic – into the optimisation of CAE, we fit the reconstruction error of the two classes into contrastive loss [16]. As analogues to anomaly detection, we expect the CAE to output a low reconstruction error for asymptomatic segments, and a high reconstruction error for symptomatic segments. Therefore, the loss function for our contrastive CAE can be seen as

$$\text{Loss} = \sqrt{\frac{1}{N} \sum_i^N |x_i^n - \hat{x}_i^n|^2} + (\mathbf{m} - \sqrt{\frac{1}{N} \sum_i^N |x_i^p - \hat{x}_i^p|^2}), \quad (4)$$

where the superscripts p and n are used to distinguish positive (symptomatic) and negative (asymptomatic) samples. Ideally, the reconstruction error for a negative pair, i. e., an original and a reconstructed feature map for an asymptomatic segment, is expected to be 0, indicating a successful reconstruction of the original input at the decoder. In contrast, the reconstruction error for a positive input pair, i. e., an original and a reconstructed feature map for a symptomatic segment, is expected to be the margin value \mathbf{m} . Therefore, the difference in classes leads to different reconstruction errors from our CAE.

5. Experiments & results

We conducted a series of experiments to test the model presented in Section 4. The contrastive CAE was pre-trained with the heart rate segments of 49 participants that reported COVID-19-like symptoms, but did not meet the CD1 or CD2 criterion. We then applied LOSO CV to the heart rate segments of the 19 individuals who meet CD1 or CD2, and their corresponding symptom-free control group.

The performance is mainly compared to a CNN of the same architecture of our CAE encoder, and a CAE that is optimised using RMSE loss. Models of different layers are tested using mean unweighted average recall (UAR, chance-level is 50%), sensitivity, and specificity, the area under receiver operating characteristic curve (AUC-ROC), and Matthews correlation coefficient (MCC) as the evaluation metrics throughout the experiments. We consider latent attributes of different lengths, namely 50, 100, 300, 500, and 1 000. For each length, a two-layers MLP is separately optimised to project the learnt latent attributes to classes – symptomatic or asymptomatic. Further, the contrastive CAE provides the possibility to directly perform a classification based on its reconstruction error using classic machine learning techniques, for instance, logistic regression.

Table 4

Evaluation results for the binary COVID-19 yes/no (based on the symptom CD1/CD2 definitions above) classification [%] of the baseline methods and contrastive CAE models with a different number of (#) layers. For the contrastive CAE, classification is performed based on reconstruction error using logistic regression.

	# Layers	UAR	Sensitivity	Specificity	AUC-ROC	MCC
MLP (1D)		61.0	63.2	58.8	0.542	0.046
LSTM (1D)		67.3	73.7	61.0	0.577	0.074
LSTM (2D)		72.8	73.7	71.9	0.685	0.105
CNN (2D)		76.0	78.9	73.1	0.705	0.122
	1	58.8	70.2	47.4	0.508	0.044
	2	83.0	84.2	81.9	0.769	0.176
Contrastive CAE	3	90.6	100.0	81.3	0.878	0.213
	4	95.3	100.0	90.6	0.944	0.310
	5	93.9	100.0	87.7	0.931	0.270
	6	90.9	100.0	81.9	0.883	0.217

The models' parameters are optimised using an Adam optimiser. The learning rate decays from 0.03 to about 0.0001 with a decay factor of 0.33 after every 50 epochs. We keep using a batch size of 32 for all experiments. The hyper-parameters are selected after careful fine-tuning to assure stable and fast convergence of our models.

5.1. Contrastive CAE vs CNN vs MLP & LSTM

We first compare our proposed contrastive CAE with MLP and LSTM neural networks [13] directly applied to the one-dimensional 5-min average heart rate segments without formatting it into feature maps (noted as "1D" in Table 4). LSTM and CNN are then tested with the two-dimensional formatted feature maps (noted as "2D" in Table 4) for fair comparisons.

The MLP is found to be best with 4 layers (1000 – 250 – 50 – 20 hidden units of each layer), and its performance is shown in Table 4. The LSTM model performs best when using 64 hidden units in its recurrent cell for 1D segments, and 128 hidden units for 2D feature maps. The CNN model achieves its best performance with 3 convolutional layers, demonstrating significant improvements over the baseline models applied to 1D segments according to paired *t*-tests at significance level $\alpha = 0.05$. For the 2D feature maps, the CNN outperforms the LSTM neural network in general.

Our proposed contrastive CAE with 4 encoder and 4 decoder layers performs best reaching a considerable performance improvement over other methods. For this, we apply logistic regression to the reconstruction error of the test set, and achieve a UAR of 95.3%, a sensitivity of 100.0%, a specificity of 90.6%, an AUC-ROC of 0.944, and the MCC of 0.310. Across all LOSO CV folds, the best result yields significant improvements over the CNN approaches in paired *t*-tests ($p < 0.05$).

5.2. Contrastive CAE vs conventional CAE

We next compare our proposed CAE using contrastive loss to a conventional CAE using RMSE. We first explore the improvements in learning discriminative latent attributes, and then investigate the approach of applying classification directly on reconstruction errors of contrastive CAE.

For each different dimension of latent attributes, a two-layers MLP classifier is separately tuned to project the learnt latent attributes to classes. The conventional CAE reaches its optimum UAR, specificity, and MCC when using the latent attributes of the size of 50, and optimum sensitivity and AUC-ROC when using the latent attributes of the size of 500, as given in Table 5. Its best performance indicates its limited capability in learning discriminative latent attributes between symptomatic and asymptomatic segments. As it considers no class information when learning latent

Table 5

Comparison of results [%] between convolutional auto-encoders (CAEs) with 4 encoder and 4 decoder layers trained with RMSE loss vs contrastive loss. Classification is performed based on the latent attributes. # Attr: dimensionality of latent attributes.

	# Attr	UAR	Sensitivity	Specificity	AUC-ROC	MCC
	50	66.6	57.9	75.4	0.545	0.080
CAE	100	58.5	47.4	69.5	0.465	0.038
	300	63.4	63.2	63.7	0.527	0.058
	500	65.8	68.4	63.2	0.591	0.068
	1000	55.3	47.4	63.2	0.448	0.023
	50	92.0	100.0	83.9	0.904	0.233
Contrastive CAE	100	92.2	100.0	84.3	0.907	0.236
	300	90.9	100.0	81.9	0.890	0.217
	500	90.9	94.7	87.1	0.881	0.247
	1000	71.9	68.4	75.4	0.597	0.105

attributes, it leaves the classification difficulty to the MLP classifiers. The conventional CAE performs even worse than the CNN model, further stressing the need of involving the class information in training a more efficient CAE.

For the binary classification task, the classes' difference can be implicitly modelled in the contrastive loss as in Eq. (4) for training the CAE, since the positive and negative reconstruction error are guided to produce a margin between each other in a discriminative manner. Hence, the contrastive CAE is capable of learning latent attributes that represent salient features to distinguish between symptomatic and asymptomatic segments. In our experiments, the contrastive CAE with an attribute dimensionality of 100 achieves its best result in terms of UAR, sensitivity and AUC-ROC, and when the dimensionality increases to 500, the proposed contrastive CAE achieves the best specificity result, and a MCC of 0.247 which considerably outperforms the conventional CAE.

Applying classification directly on the reconstruction errors, rather than the learnt latent attributes, is a more efficient way to use the contrastive CAE for our binary decision task. The decision threshold between the reconstruction errors of symptomatic and asymptomatic classes is determined using logistic regression on the training part for each cross-validation round. A 14-days heart rate segment is decided for as COVID-19 symptomatic (CD1/CD2 criterion) if the reconstruction error is above the decision boundary.

The best performance, shown in Table 6, is achieved with the attributes' length equalling 100, achieving an UAR of 95.3%, a sensitivity of 100.0%, a specificity of 90.6%, an AUC-ROC of 0.944, and a MCC of 0.310. Generally, the contrastive CAE performs stably over different attribute dimensionalities, reducing the difficulty in setting its proper dimensionality. An extreme case is to combine the encoder and decoder by removing the latent attributes layer. The

Table 6

Classification results [%] of the contrastive CAE with 4 encoder and 4 decoder layers based on the reconstruction error (rec. error) using logistic regression. # Attr: dimensionality of latent attributes. The last row indicates removing the latent attributes layer.

	# Attr.	UAR	Sensitivity	Specificity	AUC-ROC	MCC
Contrastive CAE (rec. error)	50	93.9	100.0	87.7	0.927	0.270
	100	95.3	100.0	90.6	0.944	0.310
	300	91.5	100.0	83.0	0.890	0.226
	500	92.4	100.0	84.8	0.895	0.240
	1000	94.4	100.0	88.9	0.936	0.284
-	93.3	100.0	86.6	0.923	0.258	

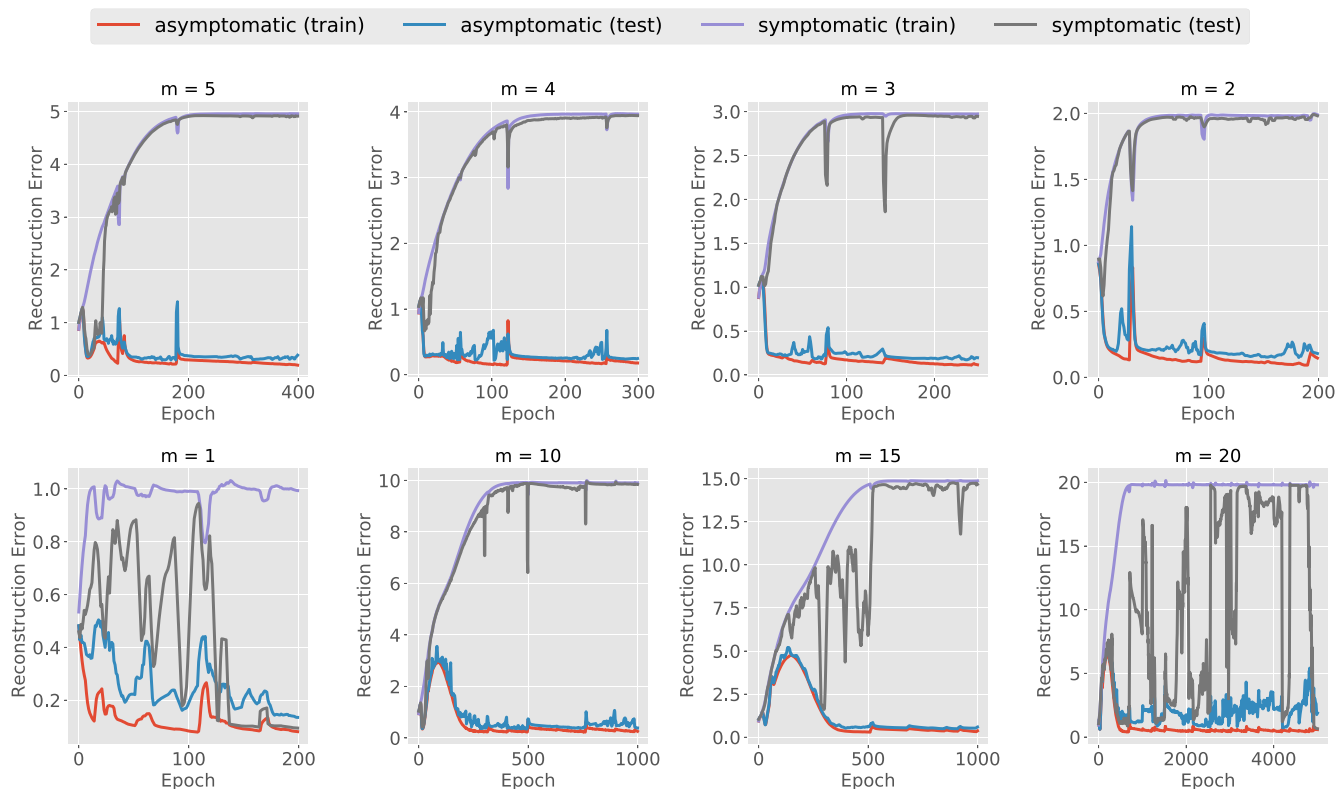


Fig. 3. Training and testing curves illustrated by the reconstruction errors when using different margin sizes.

performance, however, maintains stable as given in the last row of Table 6.

5.3. Effect of margin size

Margin size represents the expected distance between the reconstruction errors of positive and negative samples. Ideally, the reconstruction error of a positive input pair is expected to be 0, and that of a negative input pair to the margin m according to Eq. (4). In practice, during the optimisation of the contrastive CAE, the reconstruction errors can fluctuate around the expected output in some range. Therefore, setting a too small margin may lead to an insufficient fluctuating region. For example, when setting $m = 1$, the model cannot converge according to training and testing curves depicted in Fig. 3. Increasing the margin to above 2, the model can successfully converge after enough training epochs, by creating the margin between the reconstruction errors of symptomatic and asymptomatic segments. However, an unfit large margin (like $m = 15$) can lead to strong oscillation before the positive reconstruction error reaches its expected margin value. Even a larger margin size can result in convergence failure. Besides, a proper margin should provide enough space for setting the deci-

sion threshold between the reconstruction errors of symptomatic and asymptomatic segments. The impact of the margin size on classification results can be seen in Table 7.

An interesting phenomenon can be observed for the successful training cases, especially when m is set to 10 or 15. At the begin of the training phase, the reconstruction errors of positive and negative samples vary in the same direction, until a turning point from where the two reconstruction errors diverge and then approach to their individual expected output. One can understand the training procedure according to Eq. (4). The optimisation of the contrastive CAE starts with reconstructing the input of the encoder at the output of the decoder. Then it makes a concession to the creation of margin between the positive and negative reconstruction errors, leading to their parallel increase for several epochs. Finally, it compromises feature reconstruction and margin creation, resulting in the divergence of the two reconstruction errors.

5.4. Necessity of pre-training

Previous work has demonstrated the generalisation effect when applying pre-training in some representation learning techniques, such as auto-encoders [30]. In this section, we compare the use of

Table 7
Classification results [%] of the contrastive CAE with 4 encoder and 4 decoder layers based on the reconstruction error (rec. error) using different margin sizes.

	(m)argin	UAR	Sensitivity	Specificity	AUC-ROC	MCC
Contrastive CAE (rec. error)	2	78.9	84.2	73.6	0.753	0.136
	3	91.4	100.0	82.8	0.905	0.224
	4	94.1	100.0	88.2	0.920	0.275
	5	95.3	100.0	90.6	0.944	0.310
	10	90.5	94.7	86.2	0.861	0.238
	15	90.9	94.7	87.0	0.861	0.247

Table 8
Classification results [%] of the contrastive CAE with 4 encoder and 4 decoder layers based on the reconstruction error (rec. error) using different numbers of (#) participants for pre-training.

	# Participants	UAR	Sensitivity	Specificity	AUC-ROC	MCC
Contrastive CAE (rec. error)	49	95.3	100.0	90.6	0.944	0.310
	40	95.9	100.0	91.7	0.950	0.329
	30	95.2	100.0	90.3	0.940	0.305
	20	82.3	84.2	80.3	0.823	0.167
	10	79.8	84.2	75.4	0.737	0.143
	0	76.4	78.9	73.8	0.696	0.124

Table 9
Test results [%] for shifting the sliding window by days.

	# Days	UAR	Sensitivity	Specificity	AUC-ROC	MCC
Contrastive CAE	-3	57.4	52.6	62.2	0.420	0.032
	-2	64.7	68.4	61.0	0.558	0.063
	-1	95.6	100.0	91.2	0.946	0.320
	0	95.3	100.0	90.6	0.944	0.310
	1	95.4	100.0	90.8	0.945	0.313
	2	96.1	100.0	92.1	0.957	0.337
	3	94.9	100.0	89.9	0.949	0.298
	4	87.4	94.7	80.2	0.823	0.193
	5	61.5	68.4	54.6	0.517	0.048

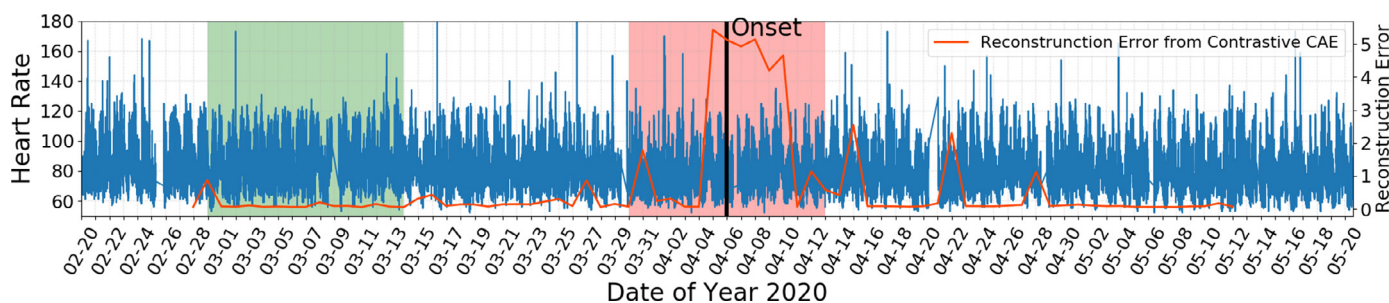


Fig. 4. Reconstruction errors for continuous binary COVID-19 yes/no classification on 14-days heart rate windows of an exemplary individual (the same as in Fig. 1, top).

different numbers of participants for pre-training the contrastive CAE.

For each considered number of participants for pre-training, we apply random selection of participants. We then pre-train our model using the selected participants while keeping the LOSO CV procedure unchanged for evaluation. We run the selection procedure five times and the average testing results are given in Table 8. The model turns out to be effective when using at least 30 participants for pre-training. As the number of participants drops below 20, the classification performance declines, revealing the necessity of supplying the model with enough pre-training data to reach its optimal performance.

5.5. Shifting of symptomatic segments

Throughout all previous experiments, we keep assuming that the participant-reported onset date is identical to the real symptom onset. The contrastive CAE performs effectively on the symptomatic segments that are centred at the reported symptom onset.

To explore the possibility to even make binary COVID-19 yes/no decisions based on segments with a decentralised reported symptom onset, we shift the window for sliding over the symptomatic segments to earlier and later days, while still containing the onset date. The asymptomatic segments are kept unchanged as in the previous experiments.

The experimental results, as seen in Table 9, reveal that the model works well for the heart rate segments that are shifted one day forward or three days backward. However, segments that further deviate from the original symptomatic segments, i. e., shifting the sliding window to two more previous days or four days later, results in decreased classification performance. Potentially, participants may have noticed the onset of their symptoms, but only reported this days later, resulting in an inaccurate reported date. Further, segments shifted up to a few days later (maximally three days in our experiments) have higher certainty that symptoms are indeed contained. Therefore, in both cases, our model achieves stable performance. Also, there might be some participants whose symptoms started earlier, and eased soon. In this case, segments that

are shifted many days later, may exclude the true symptomatic episode, leading to a low classification performance.

Fig. 4 illustrates using our contrastive CAE to continuously classify COVID-19 yes/no based on CD1/CD2 symptom presence based on the example given on the top of Fig. 1. The estimated reconstruction errors indicate that the onset detection of the COVID-19-like symptoms can be estimated in their earlier stage up to several days later.

6. Conclusions

We proposed a contrastive CAE to make a machine-learning-based COVID-19 yes/no decision based on symptom presence defined by two criteria (CD1/CD2) given 14-days heart rate measurements from a Fitbit wristband. The models were pre-trained based on the heart rate data of 49 participants with MS who reported having COVID-19-like symptoms. The models were then tested on data of 19 MS participants whose reported symptoms met the criteria of CD1 or CD2, by means of LOSO CV. In this process, each of the 19 symptomatic MS participants was paired with a site-, gender-, and age-matched symptom-free MS participant. Experimental results indicate that our proposed approach, incorporating class information into optimising the CAE with contrastive loss, achieved considerable improvements over the conventional CNN, CAE and other typical deep learning models in terms of performance, evaluated as UAR, sensitivity, specificity, AUC-ROC, and MCC. We tested the proposed model with different numbers of layers, and different dimensions of latent attributes. The need of using enough data for pre-training was verified by having achieved a reliable performance. In addition, adjusting the margin size within a proper range was shown to be crucial to stable convergence and classification performance.

Although the results have been obtained using heart rate estimates provided by Fitbit, they are expected to be generalisable to any other device providing accurate heart rate measurements, and better results could be obtained if the recorded PPG signal is accessible. The efficacy of contrastive CAE demonstrated in this work provides the basis for further research. As representing a general binary classification method, we expect its widespread adoption, especially for the prediction of diseases other than COVID-19. In a departure from conventional unsupervised learning methods for anomaly detection with auto-encoders, we present a self-supervised learning approach by supplying a training target that adapts the model to the objective of anomaly detection during its optimisation. The benefit of this target-oriented optimisation strategy should not stay reserved to our COVID-19 yes/no scenario. Since the proposed method introduces an additional parameter, i. e., margin size, the challenge lies in setting a proper margin size for new scenarios. Besides, the proposed model requires an appropriate amount of data for pre-training, which hampers its adoption, e. g., to the detection of rare diseases. In the short term, our proposed contrastive CAE will be extended to multi-class paradigms in order to fit for a wider range of applications.

Since the set-up of our experiments was chosen to detect whether or not the COVID-19-like symptoms appeared during a period of recorded heart rate data, the models show limitations in a causal set-up, i. e., when trying to predict potential symptoms before they are present. To this end, future work shall try to answer the question of how many days in advance we can reliably predict the potential imminent onset of COVID-19-like symptoms. As the acquisition of data in the RADAR-CNS programme is still ongoing, the improvement of our proposed binary COVID-19 yes/no (based on the symptom CD1/CD2 definitions above) classification model based on a broader data foundation is expected. Further to that, other windows of time should be analysed. Overall, we are optimistic that an applicable decision can be made as to COVID-19

presence based on the symptoms defined herein based on machine learning analysis of consumer-type heart rate measurement.

Declaration of Competing Interest

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.

Acknowledgement

This project has received funding from the Innovative Medicines Initiative⁶ 2 Joint Undertaking under grant agreement No 115902. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA. This communication reflects the views of the RADAR-CNS consortium and neither IMI nor the European Union and EFPIA are liable for any use that may be made of the information contained herein.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.patcog.2021.108403](https://doi.org/10.1016/j.patcog.2021.108403).

References

- [1] L. Piwek, D.A. Ellis, S. Andrews, A. Joinson, The rise of consumer health wearables: promises and barriers, *PLoS Med.* 13 (2) (2018) 9pages.
- [2] Y. Ranjan, Z. Rashid, C. Stewart, P. Conde, M. Begale, D. Verbeeck, S. Boettcher, T. Hyve, R. Dobson, A. Folarin, T.R.-C. Consortium, RADAR-base: open source mobile health platform for collecting, monitoring, and analyzing data using sensors, wearables, and mobile devices, *JMIR Mhealth Uhealth* 7 (8) (2019) 13pages.
- [3] M. Mohammadi, A. Al-Fuqaha, S. Sorour, M. Guizani, Deep Learning for IoT big data and streaming analytics: a survey, *IEEE Commun. Surv. Tutor.* 20 (4) (2018) 2923–2960.
- [4] J. Radin, N. Wineinger, E. Topol, S. Steinhubl, Harnessing wearable device data to improve state-level real-time surveillance of influenza-like illness in the USA: population-based study, *Lancet Digital Health* 2 (2) (2020) 85–93.
- [5] G. Quer, J. Radin, M. Gadaleta, K. Baca-Motes, L. Ariniello, E. Ramos, V. Kheterpal, E. Topol, S. Steinhubl, Wearable sensor data and self-reported symptoms for COVID-19 detection, *Nat. Med.* 2 (2020) 1–5.
- [6] J. Chen, L. Wu, J. Zhang, L. Zhang, D. Gong, Y. Zhao, Q. Chen, S. Huang, M. Yang, X. Yang, et al., Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography, *Sci. Rep.* 10 (1) (2020) 1–11.
- [7] Z. Wang, Y. Xiao, Y. Li, J. Zhang, F. Lu, M. Hou, X. Liu, Automatically discriminating and localizing COVID-19 from community-acquired pneumonia on chest X-rays, *Pattern Recognit.* 110 (2021) 107613.
- [8] S.N. Karmali, A. Sciusco, S.M. May, G.L. Ackland, Heart rate variability in critical care medicine: a systematic review, *Intensive Care Med. Exp.* 5 (1) (2017) 1–15.
- [9] S. Ahmad, A. Tejuja, K.D. Newman, R. Zarychanski, A.J.E. Seely, Clinical review: a review and analysis of heart rate variability and the diagnosis and prognosis of infection, *Crit. Care* 13 (232) (2009) 1–7.
- [10] J.M. Radin, G. Quer, E. Ramos, K. Baca-Motes, M. Gadaleta, E.J. Topol, S.R. Steinhubl, Assessment of prolonged physiological and behavioral changes associated with COVID-19 infection, *JAMA Netw. J.* 4 (7) (2021) 4pages.
- [11] A. Natarajan, H.-W. Su, C. Heneghan, Assessment of physiological signs associated with COVID-19 measured using wearable devices, *Digital Med.* 3 (156) (2020) 1–8.
- [12] T. Mishra, M. Wang, A.A. Metwally, G.K. Bogu, A.W. Brooks, A. Bahmani, A. Alavi, A. Celli, E. Higgs, O. Dagan-Rosenfeld, B. Fay, S. Kirkpatrick, R. Kellogg, M. Gibson, T. Wang, E. Hunting, P. Mamic, A. Gany, B. Rolnik, A.B. Ganz, X. Li, M.P. Snyder, Pre-symptomatic detection of COVID-19 from smartwatch data, *Nat. Biomed. Eng.* 4 (12) (2020) 1208–1220.
- [13] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, *Nature* 521 (7553) (2015) 436–444.
- [14] L. Le, A. Patterson, M. White, Supervised autoencoders: Improving generalization performance with unsupervised regularizers, in: *Proceedings of the Conference on Neural Information Processing Systems*, 2018, pp. 107–117. Montreal, Canada
- [15] G. Pang, C. Shen, L. Cao, A.V.D. Hengel, Deep learning for anomaly detection: a review, *ACM Comput. Surv.* 54 (2) (2021) 1–38.
- [16] P. Khosla, P. Teterwak, C. Wang, A. Sarna, Y. Tian, P. Isola, A. Maschinot, C. Liu, D. Krishnan, Supervised contrastive learning, in: *Proceedings of the Conference on Neural Information Processing Systems*, 2020, p. 13. Vancouver, Canada

¹ <https://www.imi.europa.eu/>

- [17] B.L. Smarr, K. Aschbacher, S.M. Fisher, A. Chowdhary, S. Dilchert, K. Pulton, A. Rao, F.M. Hecht, A.E. Mason, Feasibility of continuous fever monitoring using wearable devices, *Sci. Rep.* 10 (1) (2020) 1–11.
- [18] A. Shapiro, N. Marinsek, I. Clay, B. Bradshaw, E. Ramirez, J. Min, A. Trister, Y. Wang, T. Althoff, L. Foschini, Characterizing COVID-19 and influenza illnesses in the real world via person-generated health data, *Patterns* 2 (1) (2021) 100188.
- [19] C. Brown, J. Chauhan, A. Grammenos, J. Han, A. Hasthanasombat, D. Spathis, T. Xia, P. Cicuta, C. Mascolo, Exploring automatic diagnosis of COVID-19 from crowdsourced respiratory sound data, in: *Proceedings of the SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2020, pp. 3474–3484. New York, NY, USA.
- [20] K. Qian, M. Schmitt, H. Zheng, T. Koike, J. Han, J. Liu, W. Ji, J. Duan, M. Song, Z. Yang, Z. Ren, S. Liu, Z. Zhang, Y. Yamamoto, B.W. Schuller, Computer audition for fighting the SARS-CoV-2 corona crisis introducing the multi-task speech corpus for COVID-19, *IEEE Internet Things J.* (2021) 12pages.
- [21] T.K. Dash, S. Mishra, G. Panda, S.C. Satapathy, Detection of COVID-19 from speech signal using bio-inspired based cepstral features, *Pattern Recognit.* 117 (2021) 107999.
- [22] K.-C. Un, C.-K. Wong, Y.-M. Lau, J.C.-Y. Lee, F.C.-C. Tam, W.-H. Lai, Y.-M. Lau, H. Chen, S. Wibowo, X. Zhang, et al., Observational study on wearable biosensors and machine learning-based remote monitoring of COVID-19 patients, *Sci. Rep.* 11 (1) (2021) 1–9.
- [23] R.P. Hirten, M. Danieletto, L. Tomalin, K.H. Choi, M. Zweig, E. Golden, S. Kaur, D. Helmus, A. Biello, R. Pyzik, A. Charney, R. Miotto, B.S. Glicksberg, M. Levin, I. Nabeel, J. Aberg, D. Reich, D. Charney, E.P. Bottinger, L. Keefer, M. Suarez-Farinas, G.N. Nadkarni, Z.A. Fayad, Use of physiological data from a wearable device to identify SARS-CoV-2 infection and symptoms and predict COVID-19 diagnosis: observational study, *J. Med. Internet Res.* 23 (2) (2021) e26107.
- [24] M. Shorfuzzaman, M.S. Hossain, MetaCOVID: a siamese neural network framework with contrastive loss for n-shot diagnosis of COVID-19 patients, *Pattern Recognit.* 113 (2021) 107700. This issue
- [25] X. Chen, L. Yao, T. Zhou, J. Dong, Y. Zhang, Momentum contrastive learning for few-shot COVID-19 diagnosis from chest CT images, *Pattern Recognit.* 113 (2021) 107826. This issue
- [26] G. Dalla Costa, L. Leocani, X. Montalban, A.I. Guerrero, P.S. Sørensen, M. Magyari, R. Dobson, N. Cummins, V. Narayan, M. Hotopf, G. Comi, T.R.-C. consortium, Real-time assessment of COVID-19 prevalence among multiple sclerosis patients: a multicenter European study, *Neurol. Sci.* 41 (7) (2020) 1647–1650.
- [27] T.-T. Wong, Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation, *Pattern Recognit.* 48 (9) (2015) 2839–2846.
- [28] S. Lauer, K. Grantz, Q. Bi, F. Jones, Q. Zheng, H. Meredith, A. Azman, N. Reich, J. Lessler, The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application, *Ann. Intern. Med.* 172 (9) (2020) 577–582.
- [29] P. Vincent, H. Larochele, I. Lajoie, Y. Bengio, P.-A. Manzagol, Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion, *J. Mach. Learn. Res.* 11 (2010) 33713408.
- [30] Y. Bengio, P. Lamblin, D. Popovici, H. Larochelle, et al., Greedy layer-wise training of deep networks, *Proc. Adv. Neural Inf. Process. Syst.* 19 (2006) 153–160.

Shuo Liu received his M.Sc in Technical University of Darmstadt, Germany, in 2017. He worked as a researcher in Sivantos group for hearing aids solutions. He is currently a Ph.D candidate at the Chair of Embedded Intelligence for Health Care and Wellbeing, University of Augsburg, Germany. His research focuses are deep learning for digital health, affective computing, and audio processing.

Jing Han is a postdoctoral researcher at the Department of Computer Science and Technology at the University of Cambridge, UK. She received her doctoral degree in Computer Science from the University of Augsburg, Germany, in 2020. Her research interests are deep learning methods for human-centric multimodal affective computing and health care.

Estela Laporta Puyal obtained the B.Sc. degree in Telecommunications Engineering and the M.Sc. in the same field, both at University of Zaragoza. She joined BSICoS group in 2019.

Patrick Locatelli graduated in Computer Engineering in 2013 and received the Ph.D. degree in Engineering and Applied Sciences in 2019 at University of Bergamo. There, he worked on the RADAR-CNS project as a post-doctoral research fellow until December 2020. Today, Patrick works at Bosch Sensortec as an IMU Sensor Expert.

Florian Pokorny received his doctoral degree at the Technical University of Munich, Germany. As a postdoctoral researcher at the Chair of Embedded Intelligence for Health Care and Wellbeing, University of Augsburg, Germany, and the Division of Phoniatrics, Medical University of Graz, Austria, he works in the field of digital health.

Carlos Nos is a neurologist working at the Multiple Sclerosis Center of Catalonia, Vall d'Hebron Barcelona Hospital Campus. His main research interest is Therapeutics and he has been involved in more than 50 international Clinical Trials over the last 25 years.

Matthew Hotopf CBE FMedSci is a psychiatrist and epidemiologist who researches the relationships between physical and mental health. He is principal investigator of the Innovative Medicines Initiative's Remote Measurement in Disease and Relapse in CNS conditions (RADAR-CNS) which explores the use of mobile technologies to track and predict health states.

Björn Schuller, FIEEE, FISCA, FBSC (h-Index 87, 36k+ citations) is Professor of AI at Imperial College London/UK, Professor and Chair of Embedded Intelligence for Health Care and Wellbeing at the University of Augsburg/Germany, co-founding CEO of audEERING, and Field Chief Editor of *Frontiers in Digital Health* amongst other Affiliations.