Feasibility of the deep learning method for estimating the ventilatory threshold with electrocardiography data

Kotaro Miura1,2,3, Shinichi Goto1,3, Yoshinori Katsumata1,2,5, Hidehiko Ikura1,2, Yasuyuki Shiraishi1,2, Kazuki Sato2 and Keiichi Fukuda1

Regular aerobic physical activity is of utmost importance in maintaining a good health status and preventing cardiovascular diseases (CVDs). Although cardiopulmonary exercise testing (CPX) is an essential examination for noninvasive estimation of ventilatory threshold (VT), defined as the clinically equivalent to aerobic exercise, its evaluation requires an expensive respiratory gas analyzer and expertise. To address these inconveniences, this study investigated the feasibility of a deep learning (DL) algorithm with single-lead electrocardiography (ECG) for estimating the aerobic exercise threshold. Two hundred sixty consecutive patients with CVDs who underwent CPX were analyzed. Single-lead ECG data were stored as time-series voltage data with a sampling rate of 1000 Hz. The data of preprocessed ECG and time point at VT calculated by respiratory gas analyzer were used to train a neural network. The trained model was applied on an independent test cohort, and the DL threshold (DLT; a time of VT estimated through the DL algorithm) was calculated. We compared the correlation between oxygen uptake of the VT (VT–VO2) and the DLT (DLT–VO2). Our DL model showed that the DLT–VO2 was confirmed to be significantly correlated with the VT–VO2 (r = 0.875; P < 0.001), and the mean difference was nonsignificant (−0.05 ml/kg/min; P > 0.05), which displayed strong agreements between the VT and the DLT. The DL algorithm using single-lead ECG data enabled accurate estimation of VT in patients with CVDs. The DL algorithm may be a novel way for estimating aerobic exercise threshold.

npj Digital Medicine (2020) 3:141; https://doi.org/10.1038/s41746-020-00348-6

INTRODUCTION
Adequate regular physical activity is paramount to maintaining good health1,2 and preventing cardiovascular diseases (CVDs)3,4. In contrast, an unexpected high-intensity exercise could be a cause of deterioration or hospitalization in patients with CVDs5,6; current clinical practice guidelines and expert statements recommend aerobic exercise for patients with CVDs7,8. Although cardiopulmonary exercise testing (CPX) is an essential examination for noninvasively detecting the ventilatory threshold (VT), that is, defined as clinically equivalent to aerobic exercise, its assessment requires an expensive respiratory gas analyzer and expertise. In this context, expansion of the exercise therapy with a simple, versatile methodology to facilitate its introduction and persistency is warranted to improve clinical outcomes of patients with CVDs.

Advancement of high-performance computer technology and deep learning (DL) technology has enabled generation of models that accurately predict outcomes, detect diseases, and automatically classify or quantitate measurements from various modalities including electrophysiological and imaging data9 (e.g., electrocardiography [ECG]10, echocardiography11, computed tomography12, single-photon emission computed tomography13, and magnetic resonance imaging14) in cardiovascular medicine. In addition, DL aids in the interpretation of clinically important findings from imaging data to support clinical judgment by physicians15. However, there is limited evidence for estimating the threshold of aerobic exercise with DL algorithms combined with neural networks. Herein, we aimed to investigate the feasibility of a DL algorithm with single-lead ECG during incremental exercise for estimating the aerobic exercise threshold in patients with CVDs.

RESULTS
Patients’ selection
From April 2014 to May 2019, 404 patients underwent CPX in Keio University Hospital (Fig. 1). Among patients who were eligible for screening, we extracted 327 patients who had CVDs (chronic heart failure, coronary artery disease, pulmonary hypertension, or arrhythmias). The exclusion criteria of this study were patients with missing data, patients whose CPXs were terminated at the physician’s discretion before peak exercise load, and patients with pacing rhythm due to implantation of a pacemaker or a cardiac resynchronization therapy device. Two hundred sixty individuals who had CVD with eligible data were included in the final analysis.

Patient characteristics
Tables 1 and 2 show the baseline characteristics and respiratory gas data during CPX of the train and test cohorts. Overall, the patients were predominantly men (73.1%), with an average age of 58.9 ± 14.6 years, and mean body mass index was 23.5 ± 3.9 kg/m². Comorbidities included hypertension (46.5%), diabetes mellitus (18.1%), and dyslipidemia (40.4%). Medical history of CVD included chronic heart failure (46.5%), coronary artery disease (38.4%), pulmonary hypertension (15.8%), and arrhythmias (14.2%). The details of each medical history are provided in Supplementary Table 1. Twenty-five (9.6%) patients had atrial fibrillation (AF) during CPX. There were no significant differences in patients’ backgrounds among the two cohorts (the train cohort and the test cohort). In respiratory gas analysis data, there was significant difference of the oxygen uptake at VT (VT–VO2) in the train and test cohort (14 ± 4.5 vs. 15.7 ± 5.8 mL/kg/min; P = 0.041, Table 2).

1Department of Cardiology, Keio University School of Medicine, Tokyo, Japan. 2Institute for Integrated Sports Medicine, Keio University School of Medicine, Tokyo, Japan. 3These authors contributed equally: Kotaro Miura, Shinichi Goto. 5Email: goodcentury21@gmail.com
Evaluation of the DL algorithm to estimate the threshold of aerobic exercise

The VO₂ estimated from a single-lead ECG using DL algorithm (DL threshold, DLT–VO₂) was compared with the VT–VO₂ manually detected from a respiratory gas analyser during CPX. The relationship between the VT–VO₂ and the DLT–VO₂ of the train cohorts showed a satisfactory result (derivation cohort: \( r = 0.873 \), \( P < 0.001 \), and validation cohort: \( r = 0.749 \), \( P < 0.001 \), Fig. 2a, b). Further, the relationship between the VT–VO₂ and the DLT–VO₂ of the test cohort indicated that our DL algorithm was a clinically effective tool for estimating the threshold (\( r = 0.875 \), \( P < 0.001 \), Fig. 2c). The Bland–Altman plot revealed that the mean difference between the VT–VO₂ and the DLT–VO₂ of all three cohorts (derivation cohort: \(-0.14 \text{ mL/kg/min} \), validation cohort: \(-0.38 \text{ mL/kg/min} \), and test cohort: \(-0.05 \text{ mL/kg/min} \)) was nonsignificant (\( P > 0.05 \), Fig. 2b, d, f). Therefore, these findings demonstrated that there was no bias between the mean values, which displayed strong agreements between the VT and the DLT.

### DISCUSSION

In the present study, we found that the DL algorithm constructed with neural networks from single-lead ECG data during exercise enabled estimation of the VT in patients with CVDs. This study is unique in that we focused on the fact that electrical activity of the heart is dynamic during exercise and that the changes derived from hidden big data might resemble the feature of the VT.

DL is bringing a paradigm shift to healthcare, which is powered by increasing availability of healthcare data and rapid progress of analytic techniques. In recent years, the application of DL to cardiovascular medicine has advanced rapidly. Especially since the learning method called neural network has spread, we have made remarkable progress in precision. In the area of cardiovascular medicine, there are many applications of DL using neural networks; however, few reports are available in the area of cardiac rehabilitation. Myers et al.\(^{16}\) reported that deep neural networks could have application in the context of CPX for detecting cardiovascular outcome in patients with heart failure. Hearn et al.\(^{17}\) also described that harnessing neural networks with CPX data improved the prognostication of outcomes compared with the conventional prognostic method in patients with heart failure. These two studies implemented DL for the detection of
prognosis. The strong point of our study is that our DL algorithm could recognize hidden features from single-lead ECG and estimate the VT, which was a clinically essential parameter for determining the level of exercise.

Our network structure combining one-dimensional convolution (1D-conv) and long–short term memory (LSTM) efficiently learned the complex time-series pattern of voltage in an ECG record and effectively estimated the time point of VT without the need for a respiratory gas analyzer. This combination of 1D-conv and LSTM has been shown to be effective in learning 12-lead ECG in our previous investigation\(^1\). The two-dimensional convolution network is useful in extracting information from a still image\(^2\). This is done by abstracting local information within the special axis of the images. The 1D-conv network does the similar abstraction but on only a single axis; in our case, it was a time axis. The LSTM is an improved version of the recurrent neural network, which is effective in learning time-series data. However, this network expands the steps of back propagation as the length of the input data increases. This causes the problem of vanishing gradient and makes the training computationally expensive. The combination of 1D-conv and LSTM may have been powerful for dealing with long time-series data by allowing abstraction of complex ECG patterns by 1D-conv and reducing the complexity of the data that the LSTM should learn. However, in order to deal with a larger number of voltages recorded in a single dataset, our study modified the network to form a more complex structure but was based on the same principal of the previously reported combination. We applied an ECG data length of 30 s to run our network structure combining 1D-conv and LSTM. We did not compare the model performance using other ECG length. A systematic research of the best ECG length may have further improved the model. Our study again shows that the combination of 1D-conv and LSTM is a powerful tool for dealing with time-series data of voltage recordings from ECG.

![Fig. 2](image_url)
Cardiac rehabilitation defined by multidisciplinary professionals plays an important role in the disease management program for patients with CVDs, leading to improved exercise tolerance and quality of life, and reduced hospitalization. Nevertheless, the application of cardiac rehabilitation for patients with CVDs is extremely low, especially in the outpatient setting. Some factors contribute to such situations: complexity of CPX, time conflicts, and patients’ disinterest and uncertainty about the management of aerobic exercise in daily life. Alternative methods are needed to facilitate the estimation of aerobic exercise thresholds and expand exercise therapy to the outpatient setting. We have previously demonstrated that a real-time evaluation of the HR variability (HRV) with single-lead ECG during CPX could be helpful for detecting the aerobic exercise threshold. The study targeted patients who had myocardial infarction and sinus rhythm on ECG but not arrhythmia, such as AF, because if the HRV analysis is not applicable in patients with irregular R–R intervals on ECG. In contrast, the method combined with the DL algorithm in the present study could be expanded to a wide range of CVD patients, including those with AF during CPX. Further, if the algorithm of this study is mounted on wearable devices that can record ECGs, it can improve the persistence of cardiac rehabilitation programs in outpatients and relocate their bases from hospitals to other institutions, such as commercial fitness clubs or even patients’ homes.

Our findings should be interpreted in light of the following limitations. Firstly, the current analysis was performed in a single university hospital in Japan. The selection of patients who underwent CPX in the hospital may be biased. Further validation analyses using external datasets are necessary to establish the validity of our DL model. Second, variables of neural networks contained age, sex, and exercise time in addition to ECG data. Previous studies have suggested a correlation between VT, and age, sex, and peak VO₂. We also tested model excluding ECG data by training the same DL architecture with dummy ECG (all recorded voltage for each 1 ms. The neural network was constructed and trained using the Keras framework with TensorFlow as backend. The neural network was trained using the back-propagation supervised training algorithm. The loss function of binary cross entropy was minimized using the RMSprop optimizer. The network was trained for 60 epochs, and the model that performed best with the validation cohort was selected as the final model (Fig. 5). The performance of the final model was tested only once on the test dataset to confirm that the model was not over-fitted.

**METHODS**

**Exercise testing protocol**

The patients performed the test in the upright position on an electronically braked ergometer (Strength Ergo 8; Mitsubishi Electric Engineering Company, Tokyo, Japan). At first, the patients rested for 2 min on the ergometer until their heart rate (HR) and respiratory condition slowed down. Following a 2-min rest (rest phase), the patients performed a 2-min warm-up pedaling at 0 W (warm-up phase). The intensity was increased with a RAMP protocol ergometer (10–15 W/min), depending on the exercise capacities of each patient (exercise phase). The patients exercised with a progressive intensity until they could no longer maintain the pedaling rate (volitional exhaustion). After the exercise tests were terminated, the patients were instructed to stop pedaling and to stay on the ergometer for 3 min (recovery phase). The blood pressure was measured every minute with an indirect automatic manometer. Single-lead and 12-lead electrocardiograms were continuously recorded during whole test from the beginning of the rest phase to the end of the recovery phase.

**Respiratory gas analysis during CPX**

The expired gas flows were measured using a breath-by-breath automated system (Vmax; Nihon Kohden, Tokyo, Japan). The respiratory gas exchange, including the ventilation, VO₂, and carbon dioxide production, were monitored continuously and measured using a 30-s average. This system was subjected to a 3-way calibration process, involving a flow volume sensor, gas analyzer, and delay time calibration. VT was determined conventionally using the procedures described by Gaskill et al. (i.e., the ventilatory equivalent, excess carbon dioxide, and modified V-slope methods). The peak VO₂ was calculated as the average oxygen consumption during the last 30 s of exercise. The ventilation/carbon dioxide (ventilator efficiency) slope (VE-VCO₂ slope) was based on data from the onset of exercise to the respiratory compensation point, and it was obtained by performing a linear regression analysis of the data acquired throughout the entire period of exercise. The respiration quotients at VT and peak exercise were measured.

**Electrocardiographic sampling, preprocessing data, and construction of the DL model**

Among 260 cardiovascular patients, 97 (37.3%), 72 (27.7%), and 91 (35.0%) patients were randomly assigned to derivation, validation, and test cohorts, respectively. A combination of the derivation and validation cohorts were grouped with the train cohort (Fig. 1). The overall process of this study is illustrated in Fig. 3. The single-lead ECG data were stored as measurements of time-series voltage with a sampling rate of 1000 Hz by the LRR-03 (Crosswell, Yokohama, Japan). The conversion of ECG data to literature was previously published with slight modification. The ECG data from the beginning to the end of exercise phase (the ECG data painted yellow in Fig. 3) were extracted, and divided into multiple sections of 30 s. Each section was labeled independently as before VT (0) and including or after VT (1). Each labeled section of the 30 s-series of ECG data along with the patients’ demographic data (age and sex) and exercise time (duration) was independently fed into the network for training. The network structure of the DL model is shown in Fig. 4. We constructed the structure with a combination of one-dimensional (1D) convolution and LSTM to deal with the time-series data points in single-lead ECG data, converted to a one-dimensional matrix containing the recorded voltage for each 1 ms. The neural network was constructed and trained using the Keras framework (https://keras.io) with Tensorflow as backend. The neural network was trained using the back-propagation supervised training algorithm. The loss function of binary cross entropy was minimized using the RMSprop optimizer (https://www.coursera.org/learn/neural-networks/home/welcome). The network was trained for 60 epochs, and the model that performed best with the validation cohort was selected as the final model (Fig. 5). The performance of the final model was tested only once on the test dataset to confirm that the model was not over-fitted.

**DL threshold**

The DLT was defined as an initial time zone of 30 s-series including the VT estimated by the DL model. We validated the relationships between VT-VO₂ and DLT-VO₂ as the derivation, validation, and test cohorts to confirm that the DLT was the good threshold for clinically estimating the VT (Fig. 3).
The results are represented as the mean ± standard deviation for continuous variables and as a percentage for categorical variables, as appropriate. The relationships between the studied methods of the VT and the DLT as derivation, validation, and test cohorts were investigated by using the Pearson correlation coefficient test. The Bland–Altman technique was applied to verify the similarities between the different methods (VT and DLT). This comparison was a graphical representation of the difference between the methods and the average of these methods. In addition, we stratified patients into groups based on disease history (chronic heart failure, coronary artery disease, and pulmonary hypertension), left ventricular ejection fraction, rhythm of AF during CPX, and prescription of a β blocker, and estimated the relationship between the subgroups in the test cohort. Correlation coefficients between the full

Statistical analyses
The results are represented as the mean ± standard deviation for continuous variables and as a percentage for categorical variables, as appropriate. The relationships between the studied methods of the VT and the DLT as derivation, validation, and test cohorts were investigated by using the Pearson correlation coefficient test. The Bland–Altman technique was applied to verify the similarities between the different methods (VT and DLT). This comparison was a graphical representation of the difference between the methods and the average of these methods. In addition, we stratified patients into groups based on disease history (chronic heart failure, coronary artery disease, and pulmonary hypertension), left ventricular ejection fraction, rhythm of AF during CPX, and prescription of a β blocker, and estimated the relationship between the subgroups in the test cohort. Correlation coefficients between the full
model and the model without ECG were compared using the correlation coefficient difference test. All probability values were two-tailed, and $P$ values $<0.05$ were considered to be statistically significant. All statistical analyses were performed with SPSS version 25.0 software (IBM Corp., Armonk, NY).

Ethics and registration
The study protocol was approved by the Institutional Review Board of Keio University School of Medicine (permission number: 2014023) and conducted in accordance with the Declaration of Helsinki. All patients provided written informed consent.

Reporting summary
Further information on research design is available in the Nature Research Reporting Summary linked to this article.

DATA AVAILABILITY
The data that support the findings of this study are available from the corresponding author upon reasonable request.

CODE AVAILABILITY
The codes that support the findings of this study are available from the corresponding author upon reasonable request.

**Fig. 5** Training and testing of the model. Schematic illustration of the process of training and testing the model. The model was trained with data from the derivation cohort, and the performance of each model was calculated using data from the validation dataset on the end of each epoch. The final model was chosen as the model that performed best for the 60 epochs in the validation cohort. The performance of the final model was calculated only once using data from the testing dataset.
ACKNOWLEDGEMENTS
This research was supported by the financial support from the Vehicle Racing Commemorative Foundation to Shinichi Goto, and a Grant-in-Aid for Clinical Research from the Japanese Circulation Society to Yasuyuki Shiraishi (2019). The authors thank C. Yoshida and K. Takeuchi for their technical assistance (Keio University School of Medicine).

AUTHOR CONTRIBUTIONS
K.M., S.G., and Y.K. designed the study; K.M. and S.G. analyzed the data. K.M. and S.G. wrote the first draft of the paper which was reviewed, modified, and approved by all other authors. All the authors vouch for the accuracy and completeness of the data reported and for the fidelity of the study to the protocol.

COMPETING INTERESTS
Dr. Katsumata has a financial relationship with Suzuken Memorial Foundation. Dr. Shiraishi is affiliated with an endowed department by Nippon Shinyaku CO., Ltd, and received a research grant from the SECOM Science and Technology Foundation and an honorarium from Otsuka Pharmaceutical Co., Ltd.

ADDITIONAL INFORMATION
Supplementary information is available for this paper at https://doi.org/10.1038/s41746-020-00348-6.

Correspondence and requests for materials should be addressed to Y.K.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.