openheart

Model for classification of heart failure severity in patients with hypertrophic cardiomyopathy using a deep neural network algorithm with a 12-lead electrocardiogram

Sanshiro Togo,[1] Yuki Sugiura,[1] Sayumi Suzuki,[2] Kazuto Ohno,[2] Keitaro Akita,[2] Kenichiro Suwa ,[2] Shin-ichi Shibata,[3] Michio Kimura,[3] Yuichiro Maekawa [2]

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/openhrt-2023-002414).

To cite: Togo S, Sugiura Y, Suzuki S, *et al*. Model for classification of heart failure severity in patients with hypertrophic cardiomyopathy using a deep neural network algorithm with a 12-lead electrocardiogram. *Open Heart* 2023;**10**:e002414. doi:10.1136/openhrt-2023-002414

Received 10 July 2023
Accepted 14 November 2023

[1]Hamamatsu University School of Medicine, Hamamatsu, Japan
[2]Division of Cardiology, Internal Medicine III, Hamamatsu University School of Medicine, Hamamatsu, Japan
[3]Department of Medical Informatics, Hamamatsu University School of Medicine, Hamamatsu, Japan

Correspondence to
Dr Yuichiro Maekawa; ymaekawa@a5.keio.jp

ABSTRACT

Objectives In hypertrophic cardiomyopathy (HCM), specific ECG abnormalities are observed. Therefore, ECG is a valuable screening tool. Although several studies have reported on estimating the risk of developing fatal arrhythmias from ECG findings, the use of ECG to identify the severity of heart failure (HF) by applying deep learning (DL) methods has not been established.

Methods We assessed whether data-driven machine-learning methods could effectively identify the severity of HF in patients with HCM. A residual neural network-based model was developed using 12-lead ECG data from 218 patients with HCM and 245 patients with non-HCM, categorised them into two (mild-to-moderate and severe) or three (mild, moderate and severe) severities of HF. These severities were defined according to the New York Heart Association functional class and levels of the N-terminal prohormone of brain natriuretic peptide. In addition, the patients were divided into groups according to Kansas City Cardiomyopathy Questionnaire (KCCQ)-12. A transfer learning method was applied to resolve the issue of the low number of target samples. The model was trained in advance using PTB-XL, which is an open ECG dataset.

Results The model trained with our dataset achieved a weighted average F1 score of 0.745 and precision of 0.750 for the mild-to-moderate class samples. Similar results were obtained for grouping based on KCCQ-12. Through data analyses using the Guided Gradient Weighted-Class Activation Map and Integrated Gradients, QRS waves were intensively highlighted among true-positive mild-to-moderate class cases, while the highlighted part was highly variable among true-positive severe class cases.

Conclusions We developed a model for classifying HF severity in patients with HCM using a deep neural network algorithm with 12-lead ECG data. Our findings suggest that applications of this DL algorithm for using 12-lead ECG data may be useful to classify the HF status in patients with HCM.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Statistical approaches and deep learning (DL) methods can predict the risk of developing fatal arrhythmias, or the occurrence of sudden cardiac death in patients with hypertrophic cardiomyopathy (HCM) using 12-lead ECG data. Previous reports on models that can predict the severity of heart failure (HF) are lacking.

WHAT THIS STUDY ADDS

⇒ We developed a model for classifying HF severity in patients with HCM using a deep neural network algorithm with 12-lead ECG data. Guided Gradient Weighted-Class Activation Map and Integrated Gradients analyses showed that it had the potential to extract effective ECG feature values to indicate HF severity.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Our findings showed that applications of a DL algorithm for using 12-lead ECG data may be useful to classify the HF status in patients with HCM.

INTRODUCTION

Hypertrophic cardiomyopathy (HCM) is a disease that requires lifelong monitoring. In certain cases, it may lead to heart failure (HF) or fatal arrhythmias.[1] In cohort studies reported worldwide, the prevalence of HCM is estimated to be approximately 1 in 500 people, without significant racial differences.[2–4] Its phenotype is diverse and is not defined solely by sarcomeric gene mutations, epigenetic modifications or environmental factors.[5] Some patients with non-obstructive HCM remain asymptomatic throughout life, while others present with HF symptoms, chest pain and palpitations as they age.[6] Asymptomatic HCM is relatively common among young people, but is associated with a higher risk of sudden cardiac death (SCD) or rapid progression of cardiac hypertrophy.[7 8] Several studies have used statistical approaches to predict the occurrence of SCD in patients with HCM.[9 10] Standard 12-lead ECG is an important screening tool for patients with HCM. In the realm of automatic ECG analysis, machine learning, like random forest, support vector machine and logistic regression models, has gained prominence alongside the conventional classification algorithm based on the Minnesota code.[11–15] Subsequently, machine learning has been replaced by deep learning (DL) in most studies. Deep neural network (DNN) models constructed with large-scale datasets have shown high classification performance that have reached expert levels. Although previous studies have used DL methods to predict SCD based on 12-lead ECG analyses, reports on models that can predict the severity of HF are lacking. Regarding the severity of HF, we used a combination of New York Heart Association (NYHA) functional class and N-terminal prohormone of brain natriuretic peptide (NT-proBNP) levels as an objective index, and a self-administered questionnaire, the Kansas City Cardiomyopathy Questionnaire-12 (KCCQ-12), as the subjective index, to construct a model and examine its usefulness. The purpose of this study was to detect the severity of HF by 12-lead ECG findings in HCM using a residual neural network (ResNet)-based model and a convolutional neural network.

METHODS

Dataset

In the present study, we analysed a consecutive dataset of patients with HCM registered at Hamamatsu University School of Medicine. The diagnosis of HCM was based on the diagnostic criteria outlined in the 2014 European Society of Cardiology and 2020 American Heart Association/American College of Cardiology 2020 guidelines.[7 8] Patients without HCM HF had history of hospitalisation for a diagnosis of HF in our university between January 2018 and December 2020. Non-HCM HF dataset was used as control. Standard 12-lead ECG data, NYHA functional class and NT-proBNP levels were examined. All ECG data were recorded at a sampling frequency of 500 Hz for a 10 s period. Samples that contained artefacts, such as pacemaker rhythms and wandering baselines, were excluded through visual confirmation by experts.

NYHA functional class and NT-proBNP levels were used to classify the severity of HF[16 17] as mild, moderate or severe. The definition of these classes was as follows: mild, 'NYHA classification I' and 'NT-proBNP <400 pg/mL'; moderate, 'NYHA classification I' and 'NT-proBNP ≥400 pg/mL' or 'NYHA classification II'; severe, 'NYHA classification III or IV' (table 1). The choice of a 400 cut-off value for NT-proBNP was based on the Japanese Circulation Society 2017/Japanese Heart Failure Society 2017 Guideline on Diagnosis and Treatment of Acute and Chronic Heart Failure-Digest Version.[18]

Of the merged HCM HF and non-HCM HF data (HCM and non-HCM HF dataset), 20% was used as the test dataset, while maintaining the ratio of each class. Of the remaining data, 80% was used as the training dataset, and 20% was used as the validation dataset (figure 1).

Table 1 Definition of the severity of heart failure (HF)

Severity of HF	
Mild	NYHA I and NT-proBNP <400
Moderate	NYHA I and NT-proBNP ≥400 or NYHA II
Severe	NYHA III or IV

NT-proBNP, N-terminal prohormone of brain natriuretic peptide; NYHA, New York Heart Association.

Transfer learning model

A DNN transfer learning model was constructed in advance using PTB-XL, an open-source ECG dataset with approximately 20 000 samples.[19] The PTB-XL dataset is a resource for structured benchmarking of ECG analysis.[20] The DNN model performed multilabel classification of ECG data into five categories: normal range, conduction abnormalities, myocardial infarction, cardiac hypertrophy and ST-T alterations. For our model, we adopted an architecture with 38 layers (11 convolutional layers) consisting of three residual blocks, using the ResNet-based DNN model, which achieved an average area under the receiver operating characteristic (ROC) curve of 0.97 for arrhythmia detection.[21] Furthermore, the ResNet-based model exhibited the best performance on the PTB-XL dataset among several models in one study.[22] Our DNN model, constructed using PTB-XL, was used as a pretrained model, and transfer learning was performed using the dataset consisting of HCM HF and non-HCM HF cases. When performing transfer learning, the number of nodes in the final output layer was changed from five to three, and the final layer activation function was changed from sigmoid to softmax, corresponding to the change in tasks from five multilabel classifications to three classifications. The residual blocks for updating weights and the learning rate (5×10^{-3} to 1×10^{-5}) were adjusted as hyperparameter tuning. Additionally, we fine-tuned the last of the three residual blocks using our ECG data. Adam was used as the optimiser. The batch size was set to 128. The

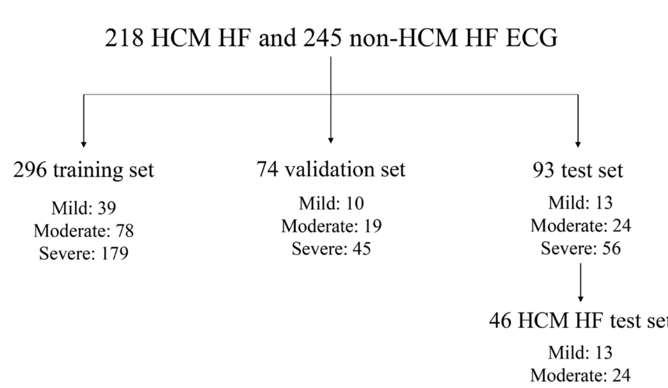

Figure 1 Set-up of experimental datasets. HCM, hypertrophic cardiomyopathy; HF, heart failure.

Table 2 Baseline characteristics of patients with hypertrophic cardiomyopathy who developed heart failure

	All (n=218)	Mild (n=62)	Moderate (n=120)	Severe (n=36)
Age, years	63.3±14.5	58.6±13.5	64.3±14.5	67.8±14.4
Female, n (%)	92 (42.2)	13 (21.0)	51 (42.5)	28 (77.8)
NYHA functional class, I/II/III/IV, n	92/90/35/1	62/0/0/0	30/90/0/0	0/0/35/1
NT-proBNP, pg/mL, median (Q1, Q3)	409.5 (127.2, 992.5)	93.0 (43.2, 191.5)	662.0 (287.8, 1292.0)	852.5 (521.0, 1897.0)
KCCQ-12, median (Q1, Q3)	87.5 (65.5, 96.9)	96.9 (90.6, 100.0)	82.3 (65.3, 94.8)	64.0 (50.2, 83.8)
Type				
HNCM, n (%)	137 (62.8)	53 (85.5)	75 (62.5)	9 (25.0)
HOCM, n (%)	81 (37.2)	9 (14.5)	45 (37.5)	27 (75.0)

Data are presented as mean±SD unless otherwise indicated.
HNCM, hypertrophic non-obstructive cardiomyopathy; HOCM, hypertrophic obstructive cardiomyopathy; KCCQ-12, Kansas City Cardiomyopathy Questionnaire-12; NT-proBNP, N-terminal prohormone of brain natriuretic peptide; NYHA, New York Heart Association.

learning rate was attenuated by a factor of 0.5 under the condition that the loss for the validation set plateaued. Only raw ECG sequence data and 60 000 potential values (500 Hz×10 s×12 leads) per case were used as input variables for the model. To improve the model's capability to predict HF, we exclusively divided HCM cases into two groups using the KCCQ-12 with a cut-off of 60, and we constructed the model using the same methodology. The KCCQ was developed to assess the quality of life of patients with HF, with scores ranging from 0 to 100, where higher scores indicate milder HF. The KCCQ score is highly correlated with the NYHA functional class. In contrast, the KCCQ-12 score (12-item instrument) is a shorter version of the original KCCQ score (23-item instrument), which has a high correlation with the original score and each domain of the original score. KCCQ-12 includes the symptom frequency, physical limitations, social limitations and quality of life domains. The KCCQ-12 is recognised for its strong correlation with the severity of HF.[23] The choice of a 60 cut-off value for KCCQ-12 is based on our prior study.[24]

Guided Gradient Weighted-Class Activation Map and Integrated Gradients

Two different saliency maps were employed to enhance the interpretation of the classification results[25]: Guided Gradient Weighted-Class Activation Map (Guided Grad-CAM) and Integrated Gradients. Guided Grad-CAM is a method that highlights and visualises features contributing to classification results by extracting gradients of weights for sample inputs and their prediction results.[26] In a prior study, the use of a Guided Grad-CAM for ECG analysis demonstrated its efficiency in detecting specific features.[27] Integrated Gradients, another saliency method, can also extract essential features from data, similar to Guided Grad-CAM.[28] Guided Grad-CAM explanations correspond to the gradient of the class scores. On the other hand, Integrated Gradients use the gradients along the straight-line path from the baseline to the input, making its process independent of class scores. We converted one-dimensional ECG sequence data into two-dimensional ECG waveforms and highlighted important waveform regions to confirm their commonality with clinical findings and to evaluate the explanatory ability of the model.

Performance of our model

To evaluate the performance of our model, we used several common performance measures, namely specificity, sensitivity (recall) and area under the ROC curve.

Precision=true positive/true positive+false positive

Recall=true positive/true positive+false negative

The F1 score is an evaluation index for a binary classification task (problem) that focuses on the trade-off relationship between precision and recall, and is the harmonic mean of the two values.

F1 score=2/(1/precision+1/recall)

The macro-average did not consider the difference in sample size between classes, whereas the weighted average is weighted according to the number of samples in each class. The ROC curve showed the true-positive rate in the test dataset calculated as true positives. Classifier performance was assessed based on the area under the ROC curve.

RESULTS

In the present study, we analysed 12-lead ECGs from 463 patients (218 with HCM HF and 245 non-HCM HF; figure 1). The characteristics of the patients with HCM HF are presented in table 2.

The classification results and the evaluation scores in severities of HF based on NYHA functional class/NT-proBNP level

The classification results of the HCM and non-HCM HF and HCM HF test datasets are shown in figures 2 and 3, respectively. For the evaluation scores of the three classes in the HCM and non-HCM HF dataset, the accuracy was 0.667 for the test dataset. The weighted-average F1 score was 0.649. The F1 score for the severe class was 0.810, whereas that for the mild class was 0.320 (table 3A).

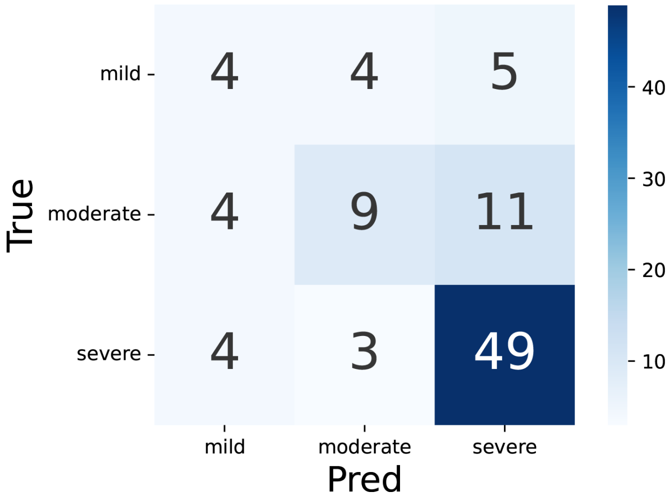

Figure 2 Confusion matrix for the classification result on HCM and non-HCM HF test dataset. The classes on the Y-axis (true) indicate the classes in which each patient was labelled. The classes on the X-axis (pred) indicate the classes in which each patient was predicted to fall by the deep neural network model. HCM, hypertrophic cardiomyopathy; HF, heart failure.

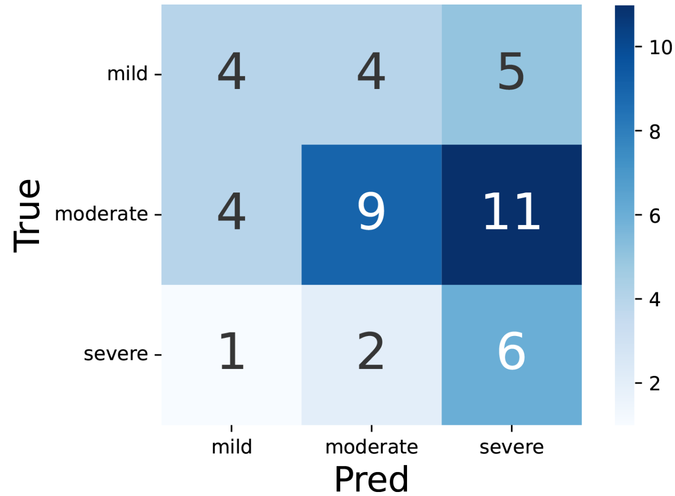

Figure 3 Confusion matrix for the classification result on HCM HF test dataset. The classes on the Y-axis (true) indicate the classes in which each patient was labelled. The classes on the X-axis (pred) indicate the classes in which each patient was predicted to fall by the deep neural network model. HCM, hypertrophic cardiomyopathy; HF, heart failure.

For the two classes of the HCM and non-HCM HF test dataset, the accuracy was 0.753 in the test dataset. The weighted-average F1 score was 0.745. The F1 score for the mild-to-moderate class was 0.646 and that for the severe class was 0.810 (table 3B). For the evaluation scores of the two classes in the HCM HF test dataset, the accuracy was 0.587. The weighted-average F1 score was 0.630. The F1 score for the mild-to-moderate class was 0.689, whereas that for the severe class was 0.387 (table 3C). The ROC curves and areas under the curve (AUC) are presented in figure 4 and online supplemental figure 1. The AUC for the three classes in the HCM and non-HCM HF test dataset was 0.80 as a micro-average, while the AUC was 0.71 for the mild class, 0.71 for the moderate class and 0.80 for the severe class (figure 4).

The evaluation scores in severities of HF based on KCCQ-12

Evaluation scores of the classification based on KCCQ-12 results on the two classes in the HCM HF test dataset showed that the final weighted-average F1 score was 0.668, which was 0.03 points higher than the same score for the NYHA/NT-proBNP two-class classification of HCM HF cases, but the accuracy was 0.630 similar to the NYHA/NT-proBNP two-class classification of HCM HF cases (table 4).

The Guided Grad-CAM and Integrated Gradients analyses

In the test dataset, the Guided Grad-CAM and Integrated Gradients analyses demonstrated a representative case with true-positive classifications in the mild-to-moderate and severe classes of HCM HF, as well as in the severe class of HCM and non-HCM HF, as shown in figure 5 and online supplemental figure 2. The results of the Guided Grad-CAM and Integrated Gradients analyses showed that the QRS complexes in the ECG were the primary features contributing to the classification of mild-to-moderate true-positive cases. In contrast, in severe true-positive cases, QRS and ST-T waves seemed to be the features that contributed to the classification.

DISCUSSION

This study demonstrated the construction of a DL-based HF severity classification model for patients with HCM using 12-lead ECG data. The mild-to-moderate class samples were classified more accurately using the proposed model. The patterns of ECG features highlighted by the Guided Grad-CAM and Integrated Gradients were more complicated and varied among the severe class samples. No previous study has proposed a DL algorithm for using 12-lead ECG data for the classification of HF status in patients with HCM.

Because the weighted-average F1 score for the HCM and non-HCM HF or HCM HF cases was reasonable, the test dataset showed that an effective DNN classification model can be built with a small dataset containing a few hundred cases by applying transfer learning of ECG analysis. The DNN model with PTB-XL qualified as the pretrained model for our study, even though the targets of the classification classes were different. A total of 60 000 potential values (500 Hz×10 s×12 leads) per sample were input into the model. Machine learning requires preprocessing to reduce the number of input variables; however, in our case, effective dimensionality reduction cannot be expected because the number of cases was small. Alternatively, summarised ECG data with measured components of ECG waves, such as heart rate, PQ interval and QRS duration, could be used instead of raw ECG sequence data. However, state-of-the-art DNN models for ECG analysis have adopted raw sequence data because it ensures that there is no loss of features, whereas summarised ECG

Table 3 (A) Classification performance for three classes of HF severity (mild, moderate and severe) in the HCM and non-HCM HF models; (B) classification performance for two classes of HF severity (mild-to-moderate and severe) in the HCM and non-HCM HF models; (C) classification performance for two classes of HF severity (mild-to-moderate and severe) in the HCM HF model

(A)	Precision	Recall	F1 score	Accuracy	Support
Mild	0.333	0.308	0.320		13
Moderate	0.563	0.375	0.450		24
Severe	0.754	0.875	0.810		56
All				0.667	93
Macro-average	0.550	0.519	0.527		93
Weighted average	0.646	0.667	0.649		93
(B)					
Mild+moderate	0.750	0.568	0.646		37
Severe	0.754	0.875	0.810		56
All				0.753	93
Macro-average	0.752	0.721	0.728		93
Weighted average	0.752	0.753	0.745		93
(C)					
Mild+moderate	0.875	0.568	0.689		37
Severe	0.273	0.667	0.387		9
All				0.587	46
Macro-average	0.574	0.617	0.538		46
Weighted average	0.757	0.587	0.63		46

HCM, hypertrophic cardiomyopathy; HF, heart failure.

data might result in loss of some features. Therefore, the method used in this study was significant because we were able to apply data from 218 cases of HCM HF to the DNN model. In addition, it is highly likely that the same method could be applied to sequence data, such as heart strain, electromyogram and electroencephalogram, for use in classifying severity in other diseases.

In this study, we applied a transfer learning method to HF cases. Transfer learning is a technique that uses a pretrained model from another study.[29] An efficient model can be built using a small dataset when a large dataset is used in the pretrained model. The number of non-HCM HF cases is overwhelmingly greater than the number of cases with HCM HF, and HCM is only one of the diseases that results in HF as a clinical condition. Therefore, we considered that patients with non-HCM HF could be used as control samples. Finally, we developed a model that could classify HF severity in patients with HCM. In addition, we classified the severity of HF based on the morphological features of ECG using the Guided Grad-CAM and Integrated Gradients.

Although many previous studies have used DNN methods targeting common cardiac diseases, such as ischaemic heart disease and arrhythmia,[30–32] few such studies have investigated HCM.[33] Asymptomatic or mildly symptomatic HCM cases are often diagnosed when 12-lead ECG findings, such as high voltage in the left precordial leads and negative T-waves, are noted during a medical examination.[34] In addition, QTc dispersion is associated with SCD,[35 36] and an association between the sum of R and S waves and SCD has also been reported.[37] These studies have analysed the measured components of ECGs.

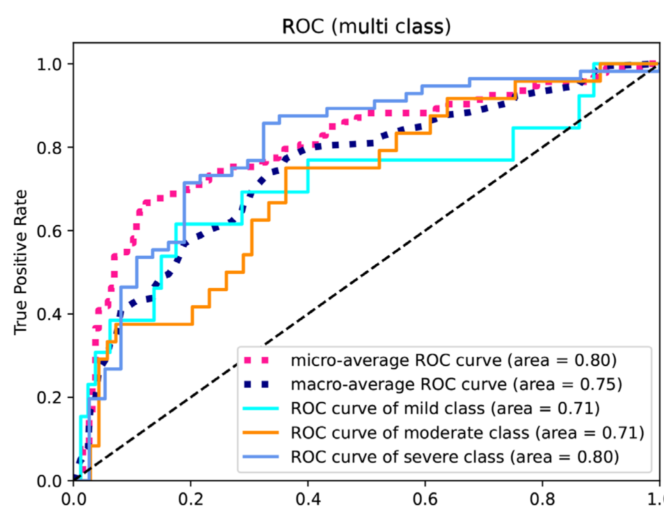

Figure 4 Receiver operating characteristic (ROC) curves and areas under the curves of the classification result for three classes in the HCM and non-HCM HF test dataset. HCM, hypertrophic cardiomyopathy; HF, heart failure.

Table 4 Classification performance for two classes based on the KCCQ-12 score (≥60 and <60) in the HCM HF model

	Precision	Recall	F1 score	Accuracy	Support
KCCQ-12 ≥60	0.885	0.622	0.730		37
KCCQ-12 <60	0.300	0.667	0.414		9
All				0.630	46
Macro-average	0.592	0.644	0.572		46
Weighted average	0.770	0.630	0.668		46

HCM, hypertrophic cardiomyopathy; HF, heart failure; KCCQ-12, Kansas City Cardiomyopathy Questionnaire-12.

In the one previous study that used raw ECG data,[33] the objective of the study was not to evaluate the severity of HF, but to screen patients.

Among the classification accuracy values for each of the three classes in the test dataset containing HF cases, those for the severe class were the highest. However, the classification accuracy values for the mild and moderate classes were lower. Adding HF cases to the dataset increased the overall number of samples; however, the severe class accounted for approximately 70% of the total, resulting in a class imbalance. Class imbalance causes a problem in that a model typically overclassifies the majority group due to its increased prior probability, leading to misclassification.[38] In such cases, data augmentation, which is a technique used to increase the data in a class with a small number of samples, or downsampling the data in a class with a large number of samples, is employed to equalise the imbalance between the classes. However, unlike for imaging data, no well-established method for ECG data augmentation exists. Downsampling of our small-scale dataset makes the analysis more unstable and reduces robustness. Therefore, we did not use these two approaches and applied class weights to the loss function and evaluation index according to the number of samples for each class. The weighted-class weighting technique has been described in detail previously.[39]

A study evaluating the diagnostic accuracy of 12-lead ECG interpretations by blinded qualified readers for HF found that the sensitivity for distinguishing patients with HF from non-HF controls was lower in patients with HF

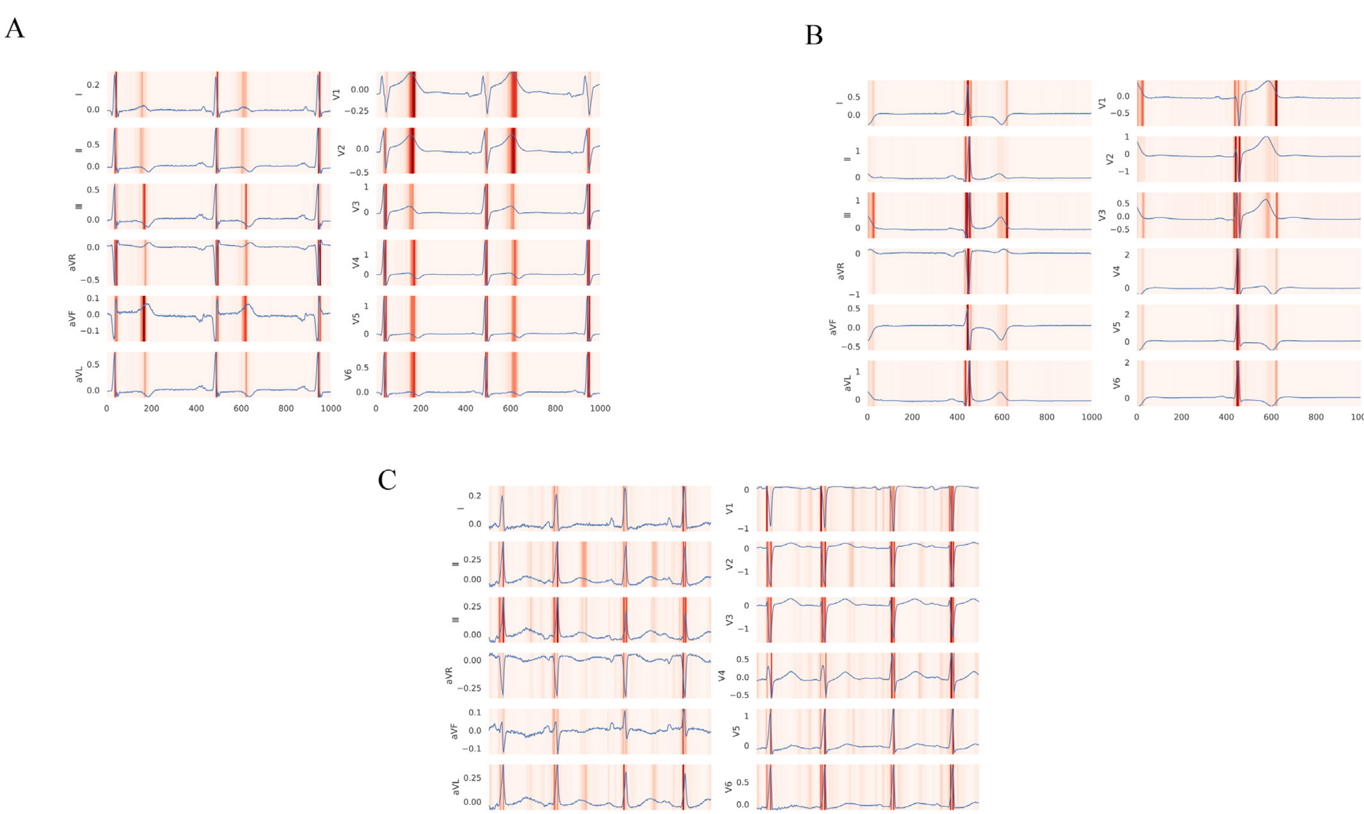

Figure 5 Twelve-lead ECG waves highlighted by Guided Gradient Weighted-Class Activation Map. (A) A case of hypertrophic cardiomyopathy who developed heart failure in the mild-to-moderate class classified as true positive. (B) A case of hypertrophic cardiomyopathy who developed heart failure in the severe class classified as true positive. (C) A case of non-hypertrophic cardiomyopathy complicated by heart failure in the severe class classified as true positive.

with preserved ejection fraction (sensitivity 65%, specificity 88%) than in those with HF with reduced ejection fraction (sensitivity 84%, specificity 88%).[40] Compared with this study, our model demonstrated similar sensitivity performance. In fact, our study showed a sensitivity of 0.875 for detecting the severe class in the two classes of HCM and non-HCM HF test datasets.

Visualisation was performed using Guided Grad-CAM and Integrated Gradients. In addition, we classified the severity of HF based on the morphological features of ECG using the Guided Grad-CAM and Integrated Gradients to assess the validity of the model's interpretation of ECGs. As QRS width and negative T waves are important findings suggestive of HCM, the appropriate feature values were generally captured and classified according to the clinical findings. Furthermore, in severe HF cases, the ECG waveform sites visualised using Guided Grad-CAM and Integrated Gradients were diverse. This reflects the diversity of the causative diseases leading to HF. From these perspectives, the Guided Grad-CAM and Integrated Gradients achieved a certain degree of interpretability of the classification results of the model. The inclusion of two classes in the pretraining, cardiac hypertrophy and ST-T alterations, which are strongly correlated with HCM HF, may have facilitated the characterisation of the ECG waveforms.

There were a few limitations to this study. First, it is necessary to consider that the robustness of the classification results may be poor because the model was constructed and evaluated using a dataset with an insufficient number of cases, even though the information on 12-lead ECG was stored in our hospital based on the Medical waveform Format Encoding Rule, which facilitated this analysis.[41] To increase the accuracy of our model, we need to validate it with a larger sample size. Second, we analysed the ECG data of the cases at one point in time; however, it may be necessary to analyse time-series ECG changes over the long term, to extract features that contribute to prognosis prediction. It is difficult to secure enough cases to apply DL to diseases with low prevalence rates, such as HCM. Third, the study population is highly heterogeneous, with a significant number of NYHA functional class I and II patients. However, this is consistent with the subjects of a previous study.[42]

CONCLUSION

We developed the classification model of HF severity in patients with HCM using DNN algorithm on data from 12-lead ECGs. The analysis using Guided Grad-CAM and Integrated Gradients showed the potential to lead to the extraction of effective ECG feature values to indicate HF severity. Our findings suggest that applications of a DL algorithm for using 12-lead ECG data may be useful to classify the HF status in patients with HCM.

Acknowledgements We thank all patients and hospital staff involved in this study.

Contributors Study design—ST, YS and YM. Data collection—ST, YS, SS, KO, S-iS, MK and YM. Data analysis—ST and YS. Data interpretation—all authors. Manuscript draft—ST, KA, KS and YM. Critical revision, editing and approval of the final manuscript—all authors. YM guarantees the overall content of the article on behalf of all authors.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests YM reports receipt of scholarship funds or donations scholarship funds from Abbott Medical Japan and BIOTRONIK JAPAN. The other authors declare no conflicts of interest.

Patient consent for publication Not required.

Ethics approval This study involves human participants. This study was performed in accordance with the Declaration of Helsinki and was approved by the Institutional Review Board of the Hamamatsu University School of Medicine (approval number: 21-043). Because this study used anonymised data in which individuals cannot be identified, the university's Ethics Review Committee has guaranteed the right to opt out by disclosing information about this study in lieu of obtaining consent.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID iDs
Kenichiro Suwa http://orcid.org/0000-0001-6934-7932
Yuichiro Maekawa http://orcid.org/0000-0002-6792-8458

REFERENCES

1 Olivotto I, Cecchi F, Poggesi C, *et al*. Patterns of disease progression in hypertrophic cardiomyopathy: an individualized approach to clinical staging. *Circ Heart Fail* 2012;5:535–46.
2 Maron BJ, Rowin EJ, Maron MS. Global burden of hypertrophic cardiomyopathy. *JACC Heart Fail* 2018;6:S2213-1779(18)30204-X:376–8.:.
3 Maron BJ, Spirito P, Roman MJ, *et al*. Prevalence of hypertrophic cardiomyopathy in a population-based sample of American Indians aged 51 to 77 years (the strong heart study). *Am J Cardiol* 2004;93:1510–4.
4 Maron BJ, Gardin JM, Flack JM, *et al*. Prevalence of hypertrophic cardiomyopathy in a general population of young adults: echocardiographic analysis of 4111 subjects in the CARDIA study. *Circulation* 1995;92:785–9.
5 Marian AJ. Hypertrophic cardiomyopathy: from genetics to treatment. *Eur J Clin Invest* 2010;40:360–9.
6 Marian AJ, Braunwald E. Hypertrophic cardiomyopathy: genetics, pathogenesis, clinical manifestations, diagnosis, and therapy. *Circ Res* 2017;121:749–70.
7 Authors/Task Force members, Elliott PM, Anastasakis A, *et al*. ESC guidelines on diagnosis and management of hypertrophic cardiomyopathy: the task force for the diagnosis and management of hypertrophic cardiomyopathy of the European society of cardiology (ESC). *Eur Heart J* 2014;35:2733–79.
8 Ommen SR, Mital S, Burke MA, *et al*. AHA/ACC guideline for the diagnosis and treatment of patients with hypertrophic cardiomyopathy: a report of the American college of cardiology/American heart association joint committee on clinical practice guidelines. *Circulation* 2020;142:e533–57.

9 O'Mahony C, Jichi F, Pavlou M, *et al*. A novel clinical risk prediction model for sudden cardiac death in hypertrophic cardiomyopathy (HCM risk-SCD). *Eur Heart J* 2014;35:2010–20.
10 Bhattacharya M, Lu D-Y, Kudchadkar SM, *et al*. Identifying ventricular arrhythmias and their predictors by applying machine learning methods to electronic health records in patients with hypertrophic cardiomyopathy (HCM-VAR-risk model). *Am J Cardiol* 2019;123:1681–9.
11 BLACKBURN H, KEYS A, SIMONSON E, *et al*. The electrocardiogram in population studies. A classification system. *Circulation* 1960;21:1160–75.
12 Macfarlane PW, Latif S. Automated serial ECG comparison based on the Minnesota code. *J Electrocardiol* 1996;29 Suppl:29–34.
13 Ganeshkumar R, YSKumaraswamy Dr. Investigating cardiac arrhythmia in ECG using random forest classification. *IJCA* 2012;37:31–4.
14 Nasiri JA, Naghibzadeh M, Yazdi HS, *et al*. ECG arrhythmia classification with support vector machines and genetic algorithm. In: *EMS 2009 - UKSim 3rd European Modelling Symposium on Computer Modelling and Simulation*. 2009.
15 Yeh YC, Wang WJ, Chiou CW. Cardiac arrhythmia diagnosis method using linear discriminant analysis on ECG signals. *Measurement* 2009;42:778–89.
16 Caraballo C, Desai NR, Mulder H, *et al*. Clinical implications of the New York heart association classification. *J Am Heart Assoc* 2019;8:e014240.
17 Januzzi JL, van Kimmenade R, Lainchbury J, *et al*. NT-proBNP testing for diagnosis and short-term prognosis in acute destabilized heart failure: an international pooled analysis of 1256 patients: the international collaborative of NT-proBNP study. *Eur Heart J* 2006;27:330–7.
18 Tsutsui H, Isobe M, Ito H, *et al*. Guideline on diagnosis and treatment of acute and chronic heart failure - digest version. *Circ J* 2019;83:2084–184.
19 Wagner P, Strodthoff N, Bousseljot R-D, *et al*. PTB-XL, a large publicly available electrocardiography dataset. *Sci Data* 2020;7:154.
20 Strodthoff N, Wagner P, Schaeffter T, *et al*. Deep learning for ECG analysis: benchmarks and insights from PTB-XL. *IEEE J Biomed Health Inform* 2021;25:1519–28.
21 Hannun AY, Rajpurkar P, Haghpanahi M, *et al*. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat Med* 2019;25:530:65–9.:.
22 Byeon YH, Pan SB, Kwak KC. Intelligent deep models based on scalograms of electrocardiogram signals for biometrics. *Sensors (Basel)* 2019;19:935.
23 Spertus JA, Jones PG. Development and validation of a short version of the Kansas City cardiomyopathy questionnaire. *Circ Cardiovasc Qual Outcomes* 2015;8:469–76.
24 Akita K, Tsuruta H, Yuasa S, *et al*. Prognostic significance of repeated brain natriuretic peptide measurements after percutaneous transluminal septal myocardial ablation in patients with drug-refractory hypertrophic obstructive cardiomyopathy. *Open Heart* 2018;5:e000786.
25 Adebayo J, Gilmer J, Muelly M, *et al*. Sanity checks for Saliency maps. 32nd Conference on Neural Information Processing Systems (NeurIPS 2018); 2018
26 Selvaraju RR, Cogswell M, Das A, *et al*. Grad-CAM: visual explanations from deep networks via gradient-based localization, Available: http://gradcam.cloudcv.org
27 van de Leur RR, Taha K, Bos MN, *et al*. Discovering and visualizing disease-specific electrocardiogram features using deep learning. Circ Arrhythm Electrophysiol 2021;14.
28 Sundararajan M, Taly A, Yan Q. Axiomatic Attribution for deep networks. Proceedings of the 34th International Conference on Machine Learning, Proceedings of Machine Learning Research; 2017:3319–28
29 Tan C, Sun F, Kong T, *et al*. A survey on deep transfer learning. In: *2018*. n.d.: 270–9.
30 Acharya UR, Fujita H, Oh SL, *et al*. Automated identification of Shockable and non-Shockable life-threatening ventricular arrhythmias using Convolutional neural network. Future Gener Comput Syst 2018;79:952–9.
31 Tan JH, Hagiwara Y, Pang W, *et al*. Application of stacked convolutional and long short-term memory network for accurate identification of CAD ECG signals. *Comput Biol Med* 2018;94:19–26.
32 Jun TJ, Nguyen HM, Kang D, *et al*. ECG arrhythmia classification using a 2-D Convolutional neural network. 2018. Available: http://arxiv.org/abs/1804.06812 [Accessed 04 Jan 2019].
33 Ko W-Y, Siontis KC, Attia ZI, *et al*. Detection of hypertrophic cardiomyopathy using a convolutional neural network-enabled electrocardiogram. *J Am Coll Cardiol* 2020;75:722–33.
34 Sugishita Y, Yamaguchi T, Ogawa T, *et al*. Variability of negative T wave in hypertrophic cardiomyopathy: possible role of beta-adrenergic function. *Acta Cardiol* 1987;42:115–33.
35 Yetman AT, Hamilton RM, Benson LN, *et al*. Long-term outcome and prognostic determinants in children with hypertrophic cardiomyopathy. *J Am Coll Cardiol* 1998;32:1943–50.
36 Ziółkowska L, Turska-Kmieć A, Petryka J, *et al*. Predictors of long-term outcome in children with hypertrophic cardiomyopathy. *Pediatr Cardiol* 2016;37:448–58.
37 Ostman-Smith I, Wettrell G, Keeton B, *et al*. Echocardiographic and electrocardiographic identification of those children with hypertrophic cardiomyopathy who should be considered at high-risk of dying suddenly. *Cardiol Young* 2005;15:632–42.
38 Johnson JM, Khoshgoftaar TM. Survey on deep learning with class imbalance. *J Big Data* 2019;6:27.
39 Pratt H, Coenen F, Broadbent DM, *et al*. Convolutional neural networks for diabetic retinopathy. *Procedia Computer Science* 2016;90:200–5.
40 Tromp J, van der Meer P, Tay WT, *et al*. Diagnostic accuracy of the electrocardiogram for heart failure with reduced or preserved ejection fraction. *J Card Fail* 2023;29:1104–6.
41 ISO/TS 22077-2:2015 health Informatics — medical Waveform format — part 2: electrocardiography. n.d. Available: https://www.iso.org/standard/64551.html
42 Ho CY, Day SM, Ashley EA, *et al*. Genotype and lifetime burden of disease in hypertrophic cardiomyopathy: insights from the sarcomeric human cardiomyopathy registry (share). *Circulation* 2018;138:1387–98.