Computer Methods and Programs in Biomedicine 219 (2022) 106762

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

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

A visually interpretable detection method combines 3-D ECG with a multi-VGG neural network for myocardial infarction identification

Rui Fang[a], Chih-Cheng Lu[a,b,c,*], Cheng-Ta Chuang[b], Wen-Han Chang[d]

[a] Graduate Institute of Manufacturing Technology, National Taipei University of Technology, Taipei 10608, Taiwan
[b] Department of Intelligent Automation Engineering, National Taipei University of Technology, Taipei 10608, Taiwan
[c] Department of Mechanical Engineering, National Taipei University of Technology, Taipei 10608, Taiwan
[d] Emergency Department, Mackay Memorial Hospital, Taipei 10449, Taiwan

ARTICLE INFO

Article history:
Received 17 September 2021
Revised 4 March 2022
Accepted 17 March 2022

Keywords:
Myocardial infarction
Electrocardiogram
3-D ECG image
Convolutional neural network
Visual interpretability

ABSTRACT

Background and Objective: The automatic recognition of myocardial infarction (MI) by artificial intelligence (AI) has been an emerging topic of academic research and an existing classification method that can recognize conventional electrocardiogram (ECG) signals with high accuracy. However, they are employed to classify one-dimensional (1-D) ECG signals rather than three-dimensional (3-D) ECG images, and it is limited to provide physicians with significant recommendations to aid in diagnosis like highlighting abnormal leads. Other studies on 3-D ECG images either did not achieve high accuracy or did not employ an inter-patient classification scheme. By proposing a multi-VGG deep neural network, this study aims to develop an automatic classification method for identifying myocardial infarction with inter-patient high accuracy and proper interpretability using 3-D ECG image and a Grad-CAM++ method.

Methods: We apply a multi-VGG deep convolutional neural network to top-view images of 3-D ECG, which are generated from simply denoised standard 12 leads ECG signals for classification. The multi-network method, which separately classifies QRS areas, ST areas, and whole heartbeats, was applied to improve classification performance. Furthermore, the Grad-CAM++ method was used to provide visually interpretable heatmaps for user's attention to improve network interpretability and assist physicians in MI diagnosis

Results: The proposed method achieved 95.65% inter-patient accuracy and exactly perfect inner-patient accuracy in the Physikalisch-Technische Bundesanstalt (PTB) diagnostic ECG database experiment. In the PTB-XL diagnostic ECG database, the proposed method achieved 97.23% inter-patient accuracy. The Grad-CAM++ experiment results also showed that the highlighted areas matched the medical diagnosis criteria for myocardial infarction.

Conclusions: Our method demonstrates that 3-D ECG images with AI classification can be efficiently employed for heart disease diagnosis with both high accuracy and visual interpretability.

1. Introduction

Myocardial infarction (MI) is a form of coronary artery disease caused by acute myocardial ischemia [1]. MI, which leads to irreversible heart damage and even death, affects approximately 16 million people worldwide each year [2]. Therefore, it is critical to developing an efficient and accurate diagnostic method for MI. Clinically, the early diagnosis of MI can be conducted using an electrocardiogram (ECG). MI can be briefly classified into ST-elevation myocardial infarction (STEMI) and non-ST elevation MI based on the ECG waveform status. Besides ST-segment elevation, MI may lead to other ECG changes, such as T-wave inversion, pathological Q wave, and ST depression [1].

The methods used in existing studies on the automatic ECG classification of myocardial infarction can be broadly classified into two categories: traditional machine learning methods and end-to-end deep convolutional neural networks (CNNs). Studies using traditional machine learning methods typically perform feature extraction combined with machine learning classification. For instance, Acharya [3] used the discrete wavelet transform (DWT) to extract features from every set of 12-lead ECG waveforms and ranked them according to the t value, followed by a k-nearest neighbor (KNN) classification, achieving 98.74% inner-patient accuracy in the V5 lead. Adam [4] used a novel DWT method combined with nonlinear feature extraction, fed features using sequen-

* Corresponding author.
E-mail address: cclu23@mail.ntut.edu.tw (C.-C. Lu).

https://doi.org/10.1016/j.cmpb.2022.106762
0169-2607/

tial forward selection technique, and then ranked them using the ReliefF method, achieving 99.27% accuracy in inner-patient-level cardiovascular disease classification, including MI. In this category, a more recent study was that Liu [5] constructed a fourth-order wavelet tensor based on the discrete wavelet packet transform and employed multilinear principal component analysis to extract features. Those features were then classified by using a bootstrap-aggregated decision tree classifier. This particular study obtained an accuracy of 99.98% at the beat level and an accuracy of 97.46% at the record level.

The performance of traditional machine learning methods significantly depends on the ability of feature extraction methods applied. Researchers have proposed more studies using such methods with the development of deep learning methods that can perform automated feature extraction. For instance, Strodthoff [6] used a CNN and residual neural network (ResNet) for ECG classification, which achieved 93.3% sensitivity and 89.7% specificity in the inter-patient scheme. Fu [7] used a novel multi-lead attention mechanism integrating with a CNN and a bi-directionally gated recurrent framework, which may achieve 96.5% inter-patient-level accuracy in the Physikalisch-Technische Bundesanstalt (PTB) database. Later Jafarian [8] proposed state of the art in this category, which achieved remarkable 99.99% MI detection performance at the inter-patient level. In this study, each lead's signal was converted into a 9×20 pseudo-time–frequency matrix in a one-dimensional convolutional layer, and then the $9 \times 20 \times 12$ matrices were inputted into a two-dimensional (2D) deep ResNet for classification. In addition, there are some additional special methods. Liu [9] used a 20th-order polynomial function to fit the ECG signal and measured them using the Akaike information criterion, achieving a 94.4% inner-patient classification accuracy.

However, all of these studies classified ECG according to the data signal itself, resulting in rather challenging information for interpretation. Furthermore, all of the studies mentioned above used the PTB database, which contains a relatively small number of patients (149 patients with MI and 52 healthy individuals). Small datasets limit the ability of these studies to validate the algorithms, whereas a more extensive dataset may be closer to real-life clinical environments. Many of these studies have also required complex denoising of ECG signals, such as DWT or notch filter, limiting their ability of clinical usage.

There are also many studies using deep neural networks to recognize 2-D ECG images, e.g. Degirmenci [10] used DNN to recognize Arrhythmic heartbeat. For the 2D ECG classification of MI, studies of Hao and Makimoto [11,12] achieved 94.73% and 81% $\pm$ 0.04% accuracy in inner-patient classification on their proprietary datasets, respectively. However, it is difficult for 2D ECGs to include 12-lead information in a single image. Thus, the classification accuracy of these studies is usually unsatisfactory, and none of them have conducted inter-patient classification experiments.

Comparatively, 3-D ECG images have been proposed to provide more information and incorporate 12-lead ECG data into one image [13]. Yet, this class of work using 3-D ECG methods for classification was relatively scarce, and there was only a recent study presented by Heo [14]. In this study, 3-D ECG images were generated, and feature extraction was conducted and classified by a KNN classifier between STEMI and normal ECGs, achieving 96.37% inner-patient classification accuracy. However, this study only focused on patients with STEMI in the PTB database and lacked inter-patient classification schemes. An inter-patient classification scheme has been considered closer to clinical applications than an inner-patient scheme as the ECG patient characteristics affect learning and classification [7]. Regarding the expression of the electrical axis activity of the 12-lead ECG, a study with a similar concept by Ozdemir [15] used a feature mapping approach that would measure the location of the features, in that case is the Gray Level Co-Occurrence Matrix, from a three-dimensional projection onto two-dimensional space, in order to form a two-dimensional color representation of the energy distribution in three-dimensional space. This approach was used to classify COVID-19 ECGs, which also accomplished the 3D spatial representation of 12-lead ECG signals.

Accordingly, the present study proposes a method that classifies generated 12-lead 3-D ECG top-view images using multi-VGG19BN. The multi-network method based on the features of these images was applied to improve the classification accuracy. In addition, this study utilized the Grad-CAM++ heatmaps of CNN for several types of MI to verify whether the regions in the image for predicting the results comply with the clinical description of this type of MI. The proposed method achieved high accuracy in the PTB and PTB-XL database inter-patient experiments. The Grad-CAM++ [16] heatmap results were also proven reasonable and can be used to aid physicians in identifying the region of interest and improving the interpretable suitability of this study.

The contributions of this study are as follows:

- This study proposes a novel method for the automatic detection of MI, which outperforms the classification accuracies obtained by the previous study using the 3-D ECG method.
- It conducts inter-patient classification experiments in the PTB and PTB-XL databases, demonstrating the generalizability of this method to MI classification at a more extensive database and clinical level.
- It presents a new method that can achieve high MI detection accuracy without complex denoising and intuitively represent the abnormal situations by a heatmap where physicians need to pay more attention.

2. Methods

The method roadmap of this study is shown in Fig. 1. First, ECG data were segmented into heartbeats based on the R-peak and normalized. Then, the heartbeat data were converted into a top view of a colored 3-D ECG and divided into three parts: ST, whole heartbeat, and QRS images for a multi-VGG19BN deep neural network (DNN) training and classification. Lastly, the classification outputs were summed to obtain the final results. Meanwhile, Grad-Cam++ was applied to the beat image network to generate the heatmap.

2.1. Data processing and 3-D display

2.1.1. Beat segmentation

In this study, we used the Pan–Tompkins method [17,18] to detect the R-peak of the ECG signal. According to the clinical criteria for MI identification, the principal ECG changes of MI are in ST waves and Q-waves, so the range of each beat was set from 0.25 s before the R-peak to 0.45 s after the R-peak to cover the whole QRS wave group and ST waves. After segmentation, each record's first and last heartbeats were removed because of possible incompleteness. As the record length and number of records per patient in the PTB database are uneven, the amount of heartbeat data for a single patient varies from 22 to 439 heartbeats, which affects the distribution of the inter-patient datasets. Therefore, 120 heartbeats from each patient were randomly sampled as experimental data, and data with less than 120 beats were randomly resampled to 120 beats. Because the number of heartbeats per patient in the PTB-XL database was well-distributed, we did not perform this process in the PTB-XL experiment.

2.1.2. Normalization

Z-score normalization, also referred to as zero-mean normalization, was conducted for every beat of these signals lead by lead,

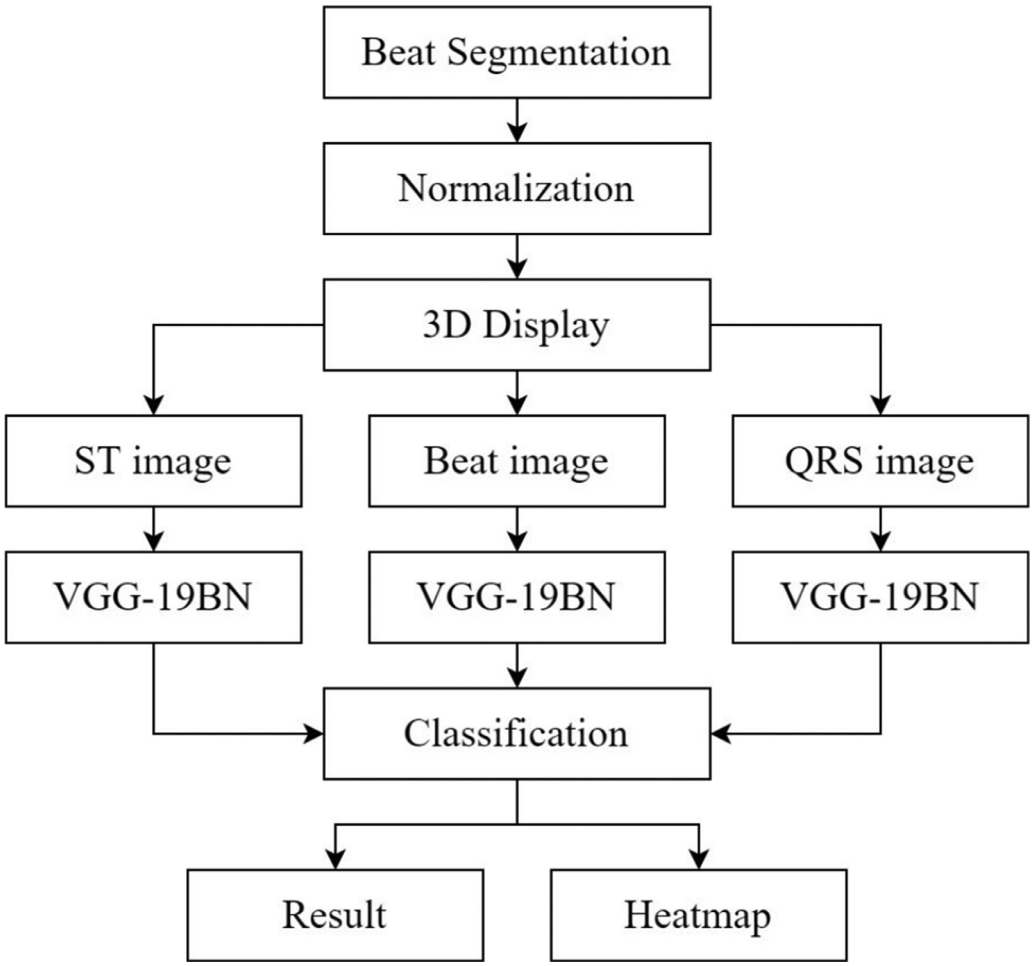

Fig. 1. Methodology roadmap of this study.

and the formula for this method is as follows:

$$x' = x - \frac{mean(p)}{std(p)} \tag{1}$$

x is the original value of the data points, x' is the normalized value, $mean(p)$ is the mean value of this set of heartbeat data, and $std(p)$ is the standard deviation of this set of heartbeat data. The Z-score normalization has been considered as a method that can reduce amplitude scaling and suppress baseline drift [19]. Note that the method provided in this study does not require any other complex denoising in pre-processing except for normalization.

2.1.3. 3-D top-view image generation

In this study, we referred to Chiang's method [13], where the filtered ECG signals are arranged by the spatial order of leads and interpolated between the leads to represent the information in the 12-lead ECG images as a 3-D ECG image. The lead arrangement is presented below, where the aVR leads are rotated by 180° to represent the information in the +30° direction.

Frontal plane: III, aVF, II, -aVR, I, aVL

Horizontal plane: V1, V2, V3, V4, V5, V6

The spatial significance of this arrangement is shown in Fig. 2, which is believed to provide more information than a 12-lead ECG image owing to the inter-lead interpolation. According to a survey among attending physicians and resident physicians [13], this presented figure provides more information, is easier to learn and read, and has the potential to replace the 12-lead ECG for diagnosis. In this study, the heartbeat time, leads, and data of the ECG signal were displayed as a top view of a 3-D colored signal. The heartbeat time and ECG leads were respectively arranged in the horizontal and vertical coordinates. The inter-lead information was interpolated by cubic spline interpolation from 700 × 12 interpolated to 700 × 1200 and further resized by the bicubic spline interpolation to 512 × 512. Then, we converted the signal strength to a colored picture according to a color bar whose range was determined by the extreme value of the signal. Finally, we generated a top-view image of the 12-lead 3-D ECG signal by this process. The 3-D ECG images from a healthy patient are shown in Fig. 3.

2.1.4. Image crop

The main features of the myocardial infarction appear in the QRS wave group and the ST wave. A part of the error in the judgment results may be because some patients with MI did not have both types of ECG abnormalities simultaneously, which affected the accuracy of the model's diagnosis. Therefore, in this study, the 3-D ECG image was further cut into the first half of the image containing QRS wave groups and the second half containing ST waves; and then rescaled to 224 × 224 pixels. Therefore, before entering the deep learning phase, three top-view image datasets of 12-lead 3-D ECGs from each heartbeat were generated, including the whole heartbeat, QRS part, and ST part. An example of these images from a MI patient is presented in Fig. 4. It also can be observed that the 3-D ECG images from a MI patient and a healthy patient (Fig. 3) have a significant difference in QRS waves and ST waves.

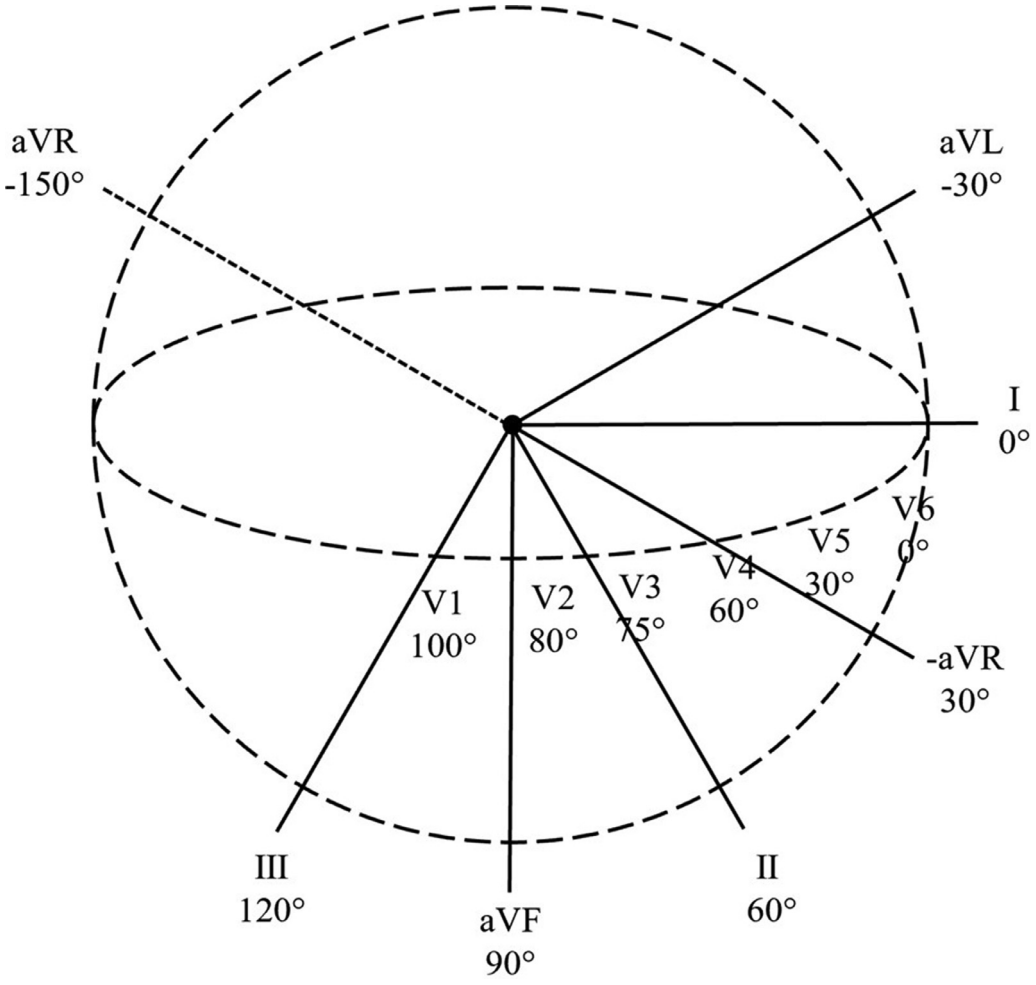

Fig. 2. Spatial denotation of 12-lead ECG sequencing.

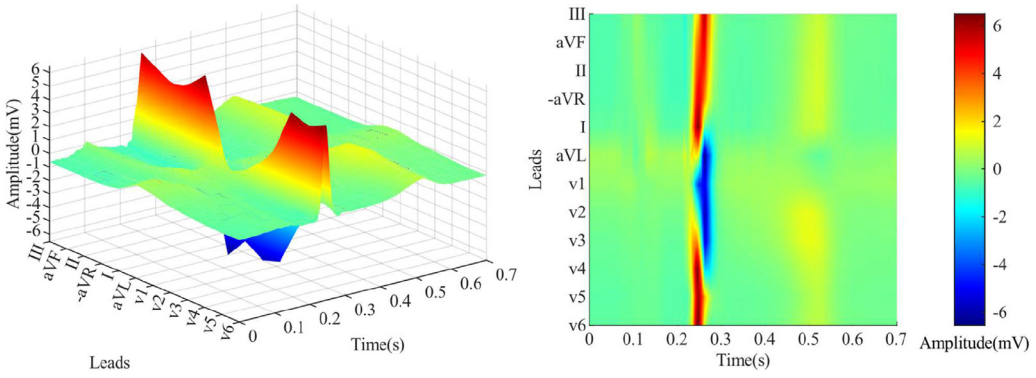

Fig. 3. A 12-lead 3-D ECG (left) and the top-view image of one beat (right) from a healthy patient.

2.2. Neural network and training

2.2.1. Neural network

For modeling, a VGG-19BN [20] neural network with batch normalization [21,22] was employed for training. VGG-19BN is a modular deep neural network consisting of 19 layers of weighted convolutional layers, which has been widely used in computer vision applications due to its significant modularity and capability of automatic feature extraction. A batch normalization layer was added after each convolutional layer. The architecture diagram of the neural network applied is shown in Fig. 5.

2.2.2. Multi-network method

According to the situation mentioned in section 2.1.4, this study proposed a multi-network method to improve classification accuracy. In this method, three separate VGG-19BN neural networks were trained by using three different image datasets, called multi-VGG-19BN. The Softmax outputs of these networks were then averaged to calculate the classification results. In Results section, we employ two different training approaches to verify the effectiveness of the multi-network method, i.e. the original VGG-19BN entitled single-VGG that uses simply the whole heartbeat signal, and

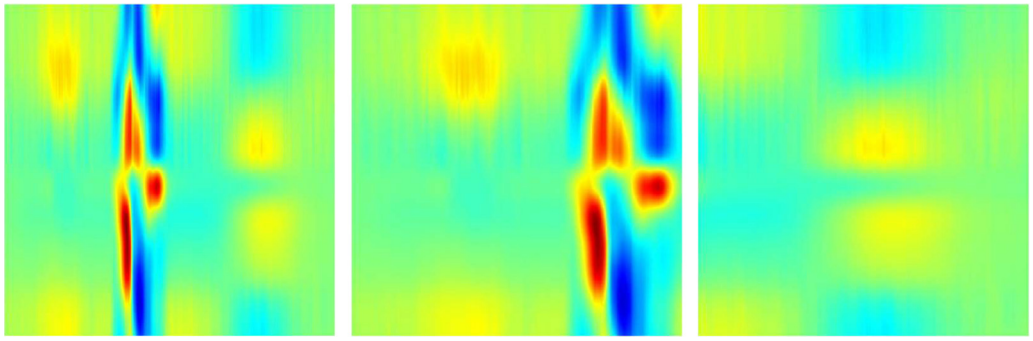

Fig. 4. A 3-D example of one beat from a MI patient: whole heartbeat, QRS part, and ST part (from left to right).

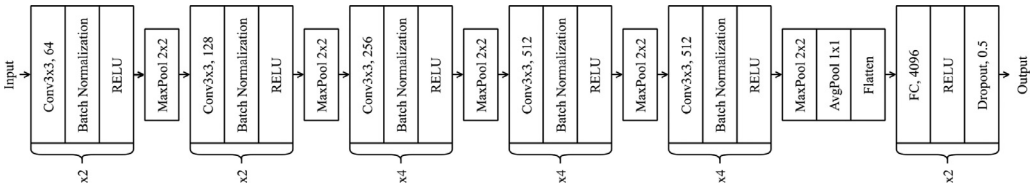

Fig. 5. Network architecture of the VGG-19BN neural network
*xN: Repeat those blocks *N* times.

the proposed experimental entitled multi-VGG that operates with the multi-network method mentioned above.

2.2.3. Assessment indicators

The following indicators were applied as benchmarks for this study: accuracy (Acc), sensitivity (Sen), specificity (Spe), receiver operating characteristic (ROC) curve, F1 score (F1S). The formulas for calculating these indicators are as follows:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

$$Sen = \frac{TP}{TP + FN} \tag{3}$$

$$Spe = \frac{TN}{TN + FP} \tag{4}$$

$$F1S = \frac{TP}{TP} + \frac{1}{2(FP + FN)} \tag{5}$$

TP: true positive; FP: false positive; TN: true negative; FN: false negative

Besides, we also used the area under ROC curve (AUROC) [23] and the Cross-Entropy Loss (CEL) as the indicators.

3. Results

3.1. Environment

A personal computer with an Intel Core I7–8700 CPU and an NVIDIA GeForce GTX 3080 GPU was used to execute the experiments. The modeling process was performed using MATLAB 2020b with the WFDB toolbox [24], and the deep-learning part was run using PyTorch 1.8.0.

3.2. Databases

In this study, we utilized the PTB database [25] and PTB-XL database [26,27] for training and testing of the model. Both databases were collected from PhysioNet [28] and contain 12-lead ECG records of healthy and MI patients. All data used are from the public ECG databases; data were disclosed with patient consent in the original study. These databases are entirely anonymous and do not contain any personal identifying marks of the subjects.

Table 1
Data information for the PTB database.

Class	Patients	Records	Beats
Healthy	52	80	6240
MI	148	368	17,760

Table 2
Data information for the PTB-XL database.

Class	Patients	Records	Beats
Healthy	6589	7184	68,322
MI	2187	2767	28,343

The PTB database contains records from 290 patients, including 148 patients with MI and 52 healthy patients. Each patient was represented with one to five records. Every record includes 12 conventional leads and 3 Frank's leads with a sampling rate of 1000 Hz and 16-bit resolution over a range of ±16.348 mV. The PTB-XL database is a new database proposed in 2020, which contains 21,837 records from 18,885 patients, including 9528 healthy records and 5486 MI records, with a recommended 10-fold split in which all records of a particular patient are assigned to the same fold. Every record includes 12 conventional leads with a sampling rate of 500 Hz and a 16-bit resolution of 1 μV/LSB. For the PTB-XL database, we only used records that were confirmed (100% likelihood) as MI or healthy. Information on the data we used and the beat segmentation results are shown in Tables 1 and 2.

3.3. Training arguments

This study used the stochastic gradient descent method [29,30] with a cyclic learning rate [31] for training. For the loss function, cross-entropy loss function was used. The images in the training set were randomly shifted by up to 20% to reduce overfitting. The specific parameters are presented in Table 3. Those hyperparameters were optimized by Bayesian hyperparameter tuning [32].

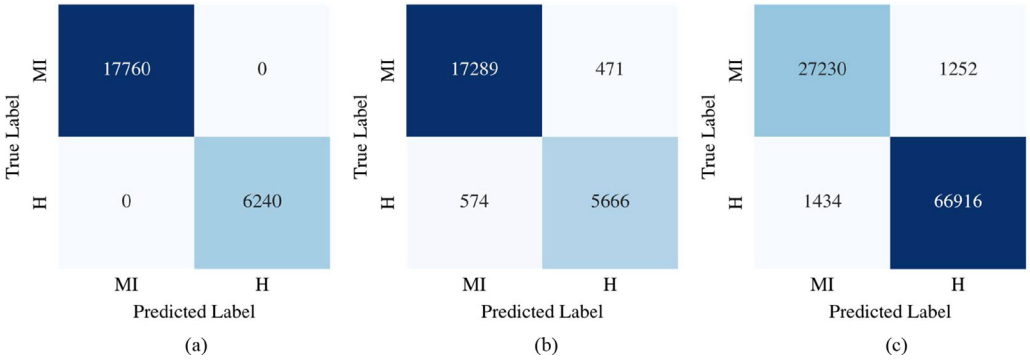

Fig. 6. Confusion matrixes. (a): PTB Inner-Patient (b): PTB Inter-Patient (c): PTB-XL Inter-Patient.

Table 3
Hyperparameter optimization.

Arguments	Value
Mini-batch size	32
Max learning rate	10^{-2}
Min learning rate	10^{-5}
Max epoch	10
Horizontal translation	20%

Table 4
Inter-patient results in the PTB database.

	Single-VGG	Multi-VGG
Accuracy	93.51%	95.65%
Sensitivity	95.88%	97.34%
Specificity	86.76%	90.80%
F1 Score	0.9563	0.9707
AUROC	0.9762	0.9861
CEL	0.1820	0.1508

3.4. PTB database

3.4.1. Inner-patient

In this experiment, whole heartbeat images from all patients in the PTB database were mixed, disordered, and then evenly partitioned into two datasets based on class. 80% of the data were divided into training sets and 20% into test sets. A K-fold cross-validation method ($K = 5$) was used. As shown in Fig. 6(a), the accuracy of this experiment reached 100% in every fold, proving that the method can achieve perfect inner-patient classification accuracy in the PTB database.

3.4.1.1. Inter-patient. In this experimental scheme, all image data were split into five folds while we ensured that the data of the same patient would only appear separately in a single fold. Similarly, a 5-fold cross-validation was used in the experiment to reduce errors. Each fold test used one of the folds as the testing set, and randomly split 90% of the remaining data as the training set whilst the rest as the validation set. Aside from the training method mentioned in Section II, we used Flooding with a flooding level of 0.07 to reduce overfitting [33]. Fig. 6(b) and Table 4 show the experiment results of the PTB database and Fig. 9(a) shows the resultant ROC curve. These results concluded that the proposed multi-VGG network method achieved better performance than the single-VGG network method in almost assessment indicators listed in Table 4. The inter-patient results are macro-average values from the testing set of each fold.

As the multi-VGG method requires three neural networks to be trained, we do not present all the training processes of three networks in this paper due to article limitation, but only the process for the whole heartbeat training as an example. As depicted in Fig. 7, the accuracy and loss curves in each fold for training and validation phases are provided. Note that due to the application of Flooding, the training result of CEL value could not be lower than 0.07. Also, the classification accuracy per patient in terms of scatter plots and histograms is shown in Fig. 8. By using 50% as the classification threshold and conducting diagnosis for each patient, one could find that there are only 6 misclassified patients out of 200 cases, resulting in 97% diagnostic accuracy.

Table 5
Inter-patient results of the PTB-XL database.

	Single-VGG	Multi-VGG
Accuracy	97.01%	97.23%
Sensitivity	95.53%	95.60%
Specificity	97.63%	97.90%
F1 Score	0.9493	0.9531
AUROC	0.9950	0.9956
CEL	0.0975	0.0810

3.5. PTB-XL database

The PTB-XL database was used as a complementary experiment to demonstrate the method's effectiveness in this study on a more extensive data set. Because the PTB database's inner-patient results had already achieved an extremely high classification accuracy, also considering the poor clinical performance of the inner-patient scheme, we have simply conducted inter-patient experiments on the PTB-XL database. The database has a 10-fold split, as recommended by the database provider, and each patient will only appear in a single fold. It is noted that 10-fold cross-validation was applied in this experiment.

The inter-patient experimental results of the PTB-XL database are presented in Table 5, Fig. 6(c), and Fig. 9(b), respectively. It's found that the proposed method achieved 97.23% in accuracy, 95.60% in sensitivity, and 97.90% in specificity with the PTB-XL database. It can be seen in Fig. 9 that the difference in classification results between different folds is not as significant as the PTB database, proving that the classification ability of the proposed method can be improved with an enormous dataset.

Furthermore, the performance variation between the single-VGG and multi-VGG network methods with the PTB-XL database seemed insignificant, only a 0.22% increase in accuracy, as the expansion of the database had probably already made the single-VGG network method reach an adequate accuracy for recognition.

3.6. Grad-CAM++ verification and assistance in identification

Grad-CAM++ is a visual explanation method commonly used in deep neural networks to compute the critical regions of an image

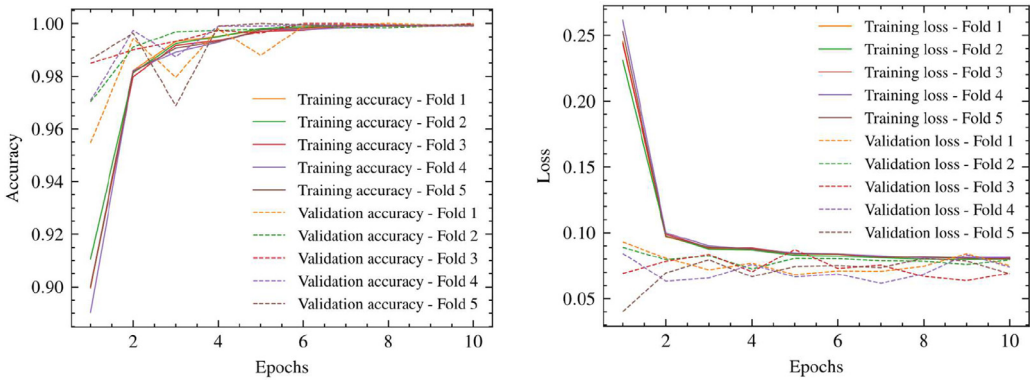

Fig. 7. Accuracy and loss curves of PTB inter-patient experiments with the single-network method (whole heartbeat).

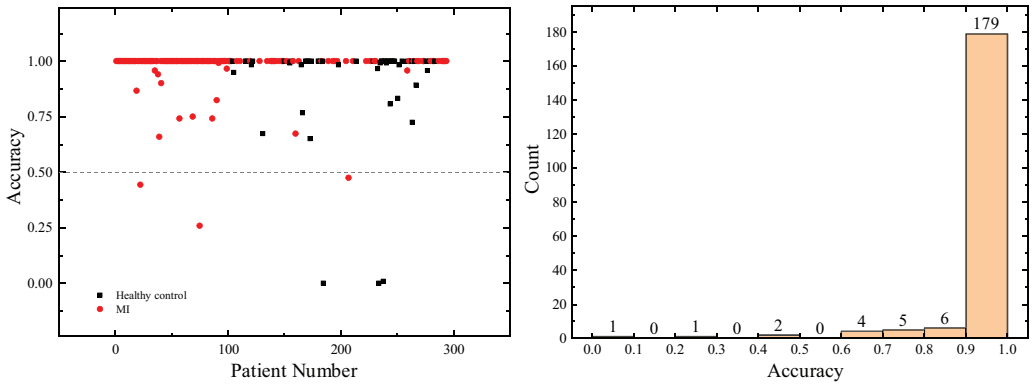

Fig. 8. Scatter plots and histograms of accuracy per patient.

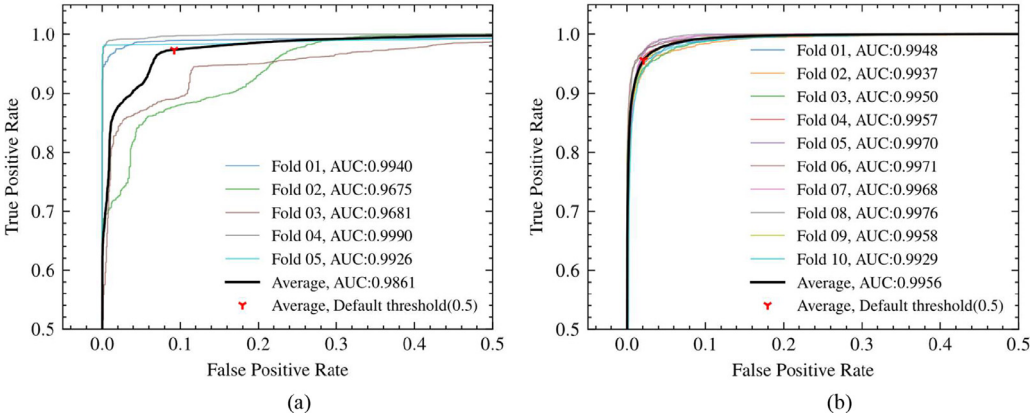

Fig. 9. ROC curves of inter-patient experiments of (a) PTB database; (b): PTB-XL database.

that are used for the neural network by using a weighted combination of the positive partial derivatives of the last convolutional layer [16,34]. Grad-CAM++ was used in this study to generate heatmaps in terms of the color image to validate the classification targets of the network and provide the information of the referred abnormal region, i.e., the leads, wave groups and beat time for further purpose in diagnosis.

In the Grad-CAM++ experiment, a network trained on the PTB-XL database only using whole heartbeats was considered the subject. The model used in this experiment is trained by the first fold in 10-fold cross-validation experiments, where folds 1∼9 as the training set and fold 10 as the testing set. All images mentioned below were obtained from the testing set. Fig. 10 shows the Grad-CAM++ results of the 3-D ECG image from a patient with anterior MI (AMI). The heatmap representation is referred to as the color bar of a 3-D ECG image, and the region in brown or red color essentially retains a higher weighted value for the neural network. Grad-CAM++ indicates that the neural network determination results from QRS abnormalities and ST elevation in the V1∼V4 leads, which is consistent with the clinical ECG diagnosis of the patients with AMI, and the ECG abnormalities in V1–V4 leads [35]. Similarly Fig. 11 illustrates the Grad-CAM++ results of the 3-D ECG image from a patient with inferior MI (IMI), where the principal foci of the neural network are the ST waves in the aVF, II, and aVL leads and QRST waves in the V3∼V5 lead, complied with the characteristics of a right ventricular IMI [36]. Fig. 12 exhibits the average diagrams of Grad-CAM++ heatmaps for the three groups of all MI, AMI and IMI patients. Compared with all the MI categories, the network had more interest in frontal plane leads while classifying patients with IMI. Conversely, the network paid more attention to

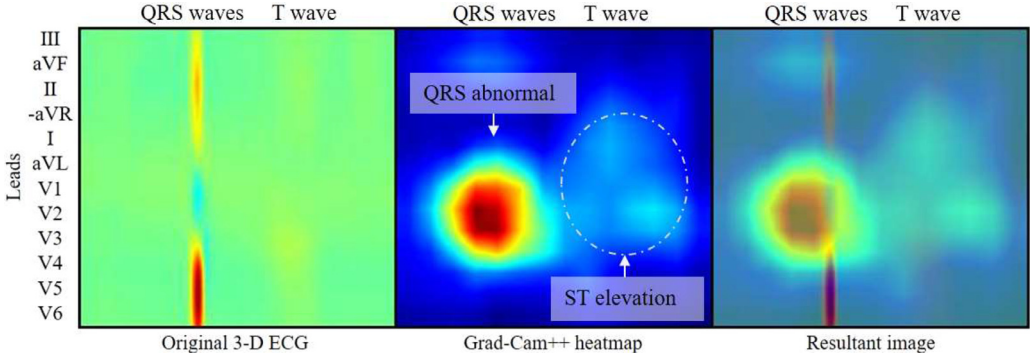

Fig. 10. Grad-CAM++ results on a patient with AMI.

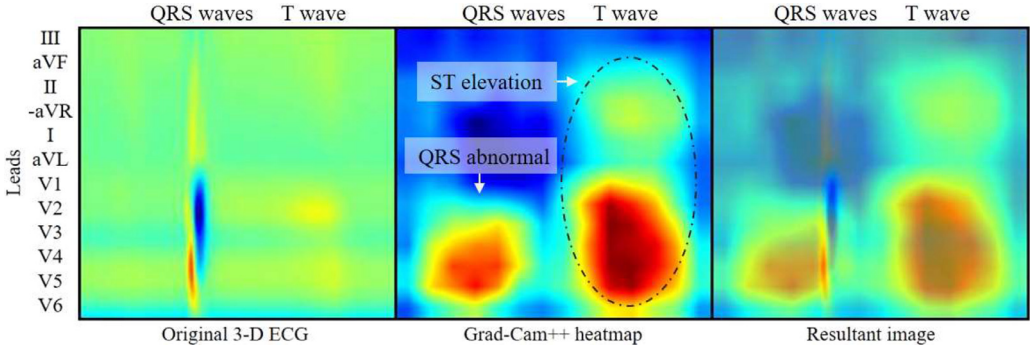

Fig. 11. Grad-CAM++ results on a patient with IMI.

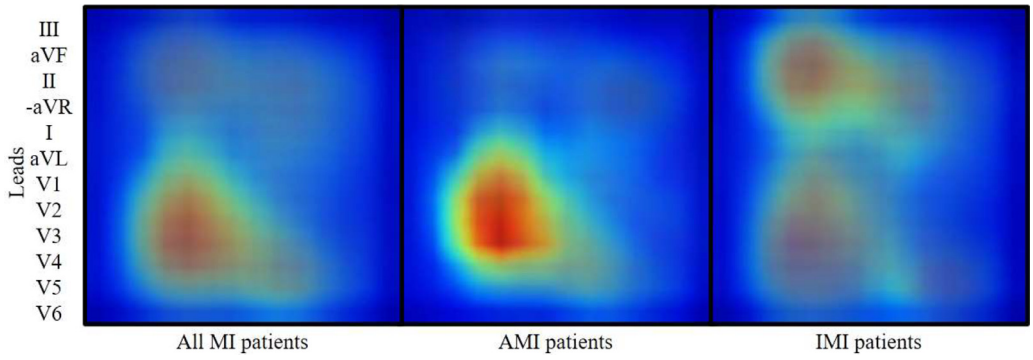

Fig. 12. Average diagrams of Grad-CAM++ heatmaps for all MI, AMI and IMI patients.

the QRS wave groups in the V1∼V4 leads while the classified patients were with AMI. Thus, these results are in accordance with the medical characteristics of the MI diagnosis categories.

4. Discussion

In this section, we compared our method with several common neural network structures to verify the capability of the proposed 3-D ECG and multi-network method. An ablation study was also conducted to further demonstrate the effectiveness of the multi-network structure proposed in this study. Finally, we compared the results with recent studies on automatic identification of myocardial infarction and discussed the limitations and future perspective of the proposed method.

4.1. Network comparison and the ablation study

In the network comparison experiments, an inter-patient scheme of the PTB database was employed, and the image of the whole heartbeat was used for training and testing. The results in Table 4 and Table 5 show that good classification accuracy can be achieved by using the previous standard neural network, proving that the images generated by this method are suitable for neural network classification. The experiment results suggest that the proposed method exhibited higher classification accuracy in the PTB-XL database than the PTB database, which proved that this method may have better performance with a more extensive database. However, the improved accuracy of the PTB-XL database was not very substantial, possibly because the single network has already achieved a sufficient level of classification accuracy owing to the increased data size. Furthermore, the comparison among several neural networks are shown in Table 6; for example, the classification accuracy using ResNet18 or VGG-19BN was both higher than that (= 93.08%) of MFB-RNN proposed by Liu et al. [39]. Also, the other assessment indicators of the proposed multi-network method were proven higher than those of published neural network structures except specificity.

On the ablation study, we conducted several experiments as follows: S-VGG uses a single VGG-19BN to train on the whole im-

Table 6
A comparison of neural networks.

Network	Acc (%)	Sens (%)	Spec (%)	F1S	AUROC
ResNet18 [37]	93.21	95.64	86.32	0.9542	0.9672
ResNet50 [37]	91.89	91.52	**92.91**	0.9434	0.9685
AlexNet [38]	91.12	92.55	87.06	0.9391	0.9578
VGG16 [19]	91.57	95.70	79.90	0.9438	0.9666
VGG19BN [22]	93.51	95.88	86.76	0.9563	0.9762
As-proposed	**95.65**	**97.34**	90.80	**0.9707**	**0.9861**

Table 7
Results of the ablation study.

network	acc (%)	sens (%)	spec (%)	f1s	auroc
s-vgg	93.51	95.88	86.76	0.9563	0.9762
m-vgg-1	95.51	**97.74**	89.18	0.9699	0.9859
m-vgg-2	94.81	96.48	90.07	0.9650	0.9809
m-vgg-3	94.40	97.19	86.39	0.9624	0.9837
m-vgg	**95.65**	97.34	**90.80**	**0.9707**	**0.9861**

age of heartbeats; M-VGG-1 employs two neural networks which are trained on the QRS and ST images, respectively; M-VGG-2 utilizes two neural networks which are trained on the whole image of heartbeats and the QRS segment image, respectively, and M-VGG-3 contains two neural networks using the images of the whole heartbeats and the QRS segment for training.

The experimental evaluation of ablation results are shown in Table 7, indicating that M-VGG multi-network approach is able to improve performance in nearly all assessment indicators. The comparison between S-VGG and M-VGGs shows that the multi-network method effectively improves the classification ability, and exhibits that using more than one neural network for classification can reduce the error rate. M-VGG-1 network, which did not use the whole heartbeat image, apparently obtained better classification accuracy than M-VGG-2 and M-VGG-3 networks, indicating that splitting QRS and ST segments can effectively improve the network's ability to identify myocardial infarction. In addition, due to the lower sensitivity of M-VGG-2 and S-VGG networks, the separate training on the ST segment can also increase the sensitivity of MI classification and reduce the false negative rate, which is also consistent with the clinical criteria for the ST-segment elevation in the myocardial infarction case. Comparatively, identifying the signal of the QRS segment will reduce the false positive rate, implying that the heartbeat characteristics of healthy individuals are more significant in the QRS segment.

4.2. MI identification comparison with recent studies

Table 8 summarizes the recent findings in various fields of MI recognition in terms of classification accuracy. Heo [14] proposed the only study on 3-D ECG image classification using feature extraction and a KNN classifier for categorizing STEMI; nonetheless, it only employed the inner-patient paradigm. Compared to this study, our presented work demonstrates an absolute inner-patient classification accuracy (100%), broader classification targets of MI disease, more comprehensive inter-patient experiments, and a successful application to the vaster database PTB-XL.

This study achieved a nearly perfect accuracy for inner-patient classification though this type of scheme is probably not suitable for clinical practice, and it outperformed most of the previous studies [5,7,14,39,41], which may be attributed to the fact that the 3-D ECG images contain more information than traditional 2-D ECG images. Compared with the earlier studies of MI identification using the inter-patient scheme [5,7,8,39,40,42], our method achieved similar accuracy, and the MI sensitivities were better than those of the investigations except the study [8]. Considering that clinical diagnoses have much less tolerant of false negatives than

Table 8
Summary of recent studies in automatic MI detection.

Author	Year	Database	Beats	Based on	Method	Inter-patient (%)	Inner-patient (%)
Liu et al. [5].	2019	PTB	MI: 47,215 H: 10,342	ECG Signal	TQWT + PCA + Tree-bagger	Acc: 97.46** Sen: 99.09** Spe: 90.26**	Acc: 99.98 Sen: 100 Spe: 99.9
Fu et al. [7].	2020	PTB	MI: 632,940 H: 127,188	ECG Signal	Multi Lead Attention + CNN	Acc: 96.5 Sen: 97.1 Spe: 93.34	Acc: 99.93 Sen: 99.99 Spe: 99.63
Liu et al. [39].	2020	PTB	MI: 53,712 H: 10,638	ECG Signal	RNN + LSTM	Acc: 93.08 Sen: 94.42 Spe: 86.29	Acc: 99.9 Sen: 99.97 Spe: 99.54
Prabhakararao et al. [40].	2020	PTB	MI: 8029 H: 4427 non-MI: 1532	ECG Signal and Clinical Features	Attentive RNN	Acc: 98.3 Sen: 97.6 Spe: 99.2	
Alghamdi et al. [41].	2020	PTB	MI: 80,364 H: 21,092	2-D ECG Image	VGG-MI		Acc: 99.22 Sen: 99.15 Spe: 99.49
Han et al. [42].	2020	PTB	MI: 28,213 H: 5373	ECG Signal	ML-ResNet	Acc: 95.49 Sen: 94.85 Spe: 97.37	Acc: 99.92 Sen: 99.98 Spe: 99.77
Jafarian et al. [8].	2020	PTB	MI: 4082 H: 1886	*pseudo*-time–frequency matrix	Residual DNN	Acc: 99.99 Sen: 100 Spe: 100	
Heo et al. [14].	2020	PTB (STEMI*)	MI: 281,328 H: 28,944	3-D ECG Image	Feature Extraction + KNN		Acc: 96.37 Sen: 99.56 Spe: 72.44
Proposed method in this study	2021	PTB	MI: 18,350 H: 6060	3-D ECG Image	DNN + multi-network	Acc: 95.65 Sen: 97.34 Spe: 90.80	Acc: 100 Sen: 100 Spe: 100
		PTB-XL	MI: 28,343 H: 68,322			Acc: 97.23 Sen: 95.60 Spe: 97.90	–

*: In this study, classification was only between healthy beats and STEMI beats, rather than all MI data from the PTB database.
**: Record level: the PTB database contains several records from the same patient, so this training method wasn't completely patient-independent.
Notation: Acc = accuracy; Sen = sensitivity; Spe = specificity.

false positives, high sensitivity is more crucial for a method primarily used to assist physicians in MI identification. It is noteworthy that studies whose inter-patient accuracy is higher than our result [7,40] actually used the attention mechanism, which mimics human attention mechanism and is believed to be effective in improving the accuracy of various tasks. The attention mechanism can also be applied to the 3-D ECG image classification in the future. The study from Jafarian [8] achieved perfect results in an inter-patient scheme, but its DNN classification method was based on *pseudo*-time–frequency matrix generated by one-dimensional convolutional layers from 12-lead ECG signals, instead of the 1-D signals or 2-D images, which means that this method is still challenging to generate naturally interpretable Grad-CAM++ images for disease identification. Our proposed work with the Grad-CAM++ method can provide a clear recommendation that has good visual interpretability, and can intuitively and accurately label leads and waves of interest in 3-D ECG images., which are highly compatible with the current medical diagnostic criteria for MI, as described in section 3.6.

4.3. Limitation and outlook

Nonetheless, there exists two major challenges in the proposed study. First, its identification performance seems not particularly significant when compared with other methods because this approach used the common VGG neural network as the sub-network to validate the method. In addition, the experiments we conducted have shown that the accuracy of MI localization with the proposed method is relatively low at about 50%. As far as we know, it is still challenging at present to perform MI location just using ECG data. Han's recent study [42] on ML–ResNet network presented only less than 60% accuracy in MI localization. In the future, we may develop a more dedicated neural network for 3-D ECG images to advance classification performance or add some complex network structures like attention mechanisms. Besides, further efforts can be made on the investigation of early MI identification based on the AI-assisted ECG image analysis.

5. Conclusion

This study proposes to utilize 3-D ECG images with a Multi-VGG DNN for MI diagnosis, achieving 95.65% in accuracy, 97.34% in sensitivity, and 90.80% in specificity for the PTB database with an inter-patient scheme and near-perfect classification in an inner-patient scheme. The proposed method can also provide interpretable diagnosis recommendations via the Grad-Cam++ method, which is helpful for physicians to improve their confidence in AI diagnosis. Considering the current black-box state of most neural networks, higher interpretability of classified images is undoubtedly necessary for the widespread use of deep neural networks in clinical practice. The experimental results reveal that the proposed method can effectively classify 3-D ECG images with high accuracy without the need for complex denoising. The multi-network approach is found effective on the recognition accuracy in the PTB database. In addition, this work outperforms a previous 3-D ECG study and achieves similar classification accuracy with higher sensitivity compared to other previous studies using ECG signal data directly for classification. Moreover, the method can provide visually interpretable heatmaps to annotate 3-D ECG images and observe the network interpretation that meets the medical diagnostic criteria for MI. The presented work offers a promising balance between classification accuracy and interpretability.

Author contributions

Each author has participated in the study and the manuscript, and all authors have read and approved the manuscript.

Conflicts of interest

The authors declare no conflict of interest in this study.

Acknowledgments

This work was financially supported by the Ministry of Science and Technology, Taiwan, R.O.C. under grants of MOST 107–2221-E-027–121 and MOST 109–2221-E-027–040-MY2.

References

[1] K. Thygesen, et al., Fourth universal definition of myocardial infarction, Eur. Heart J. 40 (3) (2018) 237–269 2019.
[2] T. Vos, et al., Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015, Lancet 388 (10053) (2016) 1545–1602, doi:10.1016/s0140-6736(16)31678-6.
[3] U.R. Acharya, et al., Automated identification of shockable and non-shockable life-threatening ventricular arrhythmias using convolutional neural network, Future Gener. Comput. Syst. 79 (2018) 952–959 2018/02/01/, doi:10.1016/j.future.2017.08.039.
[4] M. Adam, et al., Automated characterization of cardiovascular diseases using relative wavelet nonlinear features extracted from ECG signals, Comput. Methods Programs Biomed. 161 (2018) 133–143 /07/01/2018, doi:10.1016/j.cmpb.2018.04.018.
[5] J. Liu, C. Zhang, Y. Zhu, T. Ristaniemi, T. Parviainen, F. Cong, Automated detection and localization system of myocardial infarction in single-beat ECG using Dual-Q TQWT and wavelet packet tensor decomposition, Comput. Methods Programs Biomed. 184 (2020), doi:10.1016/j.cmpb.2019.105120.
[6] N. Strodthoff, C. Strodthoff, Detecting and interpreting myocardial infarction using fully convolutional neural networks, Physiol. Meas. 40 (1) (2019) 015001.
[7] L. Fu, B. Lu, B. Nie, Z. Peng, H. Liu, X. Pi, Hybrid network with attention mechanism for detection and location of myocardial infarction based on 12-lead electrocardiogram signals, Sensors (Switzerland) 20 (4) (2020), doi:10.3390/s20041020.
[8] K. Jafarian, V. Vahdat, S. Salehi, M. Mobin, Automating detection and localization of myocardial infarction using shallow and end-to-end deep neural networks, Appl. Soft Comput. J. 93 (2020) 106383 106383, doi:10.1016/j.asoc.2020.106383.
[9] B. Liu, et al., A novel electrocardiogram parameterization algorithm and its application in myocardial infarction detection, Comput. Biol. Med. 61 (2015) 178–184 /06/01/2015, doi:10.1016/j.compbiomed.2014.08.010.
[10] M. Degirmenci, M. Ozdemir, E. Izci, A. Akan, Arrhythmic heartbeat classification using 2d convolutional neural networks, IRBM (2021).
[11] P. Hao, X. Gao, Z. Li, J. Zhang, F. Wu, C. Bai, Multi-branch fusion network for Myocardial infarction screening from 12-lead ECG images, Comput. Methods Programs Biomed. 184 (2020) 105286.
[12] H. Makimoto, et al., Performance of a convolutional neural network derived from an ECG database in recognizing myocardial infarction, Sci. Rep. 10 (1) (2020) 1–9.
[13] H.K. Chiang, C.-.W. Chu, G.-.Y. Chen, C.-.D. Kuo, A new 3-D display method for 12-lead ECG, IEEE Trans. Biomed. Eng. 48 (10) (2001) 1195–1202 [Online]. Available https://ieeexplore.ieee.org/document/951523/ .
[14] J. Heo, J.J. Lee, S. Kwon, B. Kim, S.O. Hwang, Y.R. Yoon, A novel method for detecting ST segment elevation myocardial infarction on a 12-lead electrocardiogram with a three-dimensional display, Biomed. Signal Process. Control 56 (2020) 101700 101700, doi:10.1016/j.bspc.2019.101700.
[15] M.A. Ozdemir, G.D. Ozdemir, O. Guren, Classification of COVID-19 electrocardiograms by using hexaxial feature mapping and deep learning, BMC Med. Inform. Decis. Mak. 21 (1) (2021) 1–20.
[16] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad–cam: visual explanations from deep networks via gradient-based localization, in: Proceedings of the IEEE international conference on computer vision, 2017, pp. 618–626.
[17] H. Sedghamiz, Matlab implementation of Pan Tompkins ECG QRS detector, Code Available at the File Exchange Site of MathWorks (2014).
[18] J. Pan, W.J. Tompkins, A real-time QRS detection algorithm, IEEE Trans. Biomed. Eng. (3) (1985) 230–236.
[19] S.K. Pandey, R.R. Janghel, Automatic detection of arrhythmia from imbalanced ECG database using CNN model with SMOTE, Australas. Phys. Eng. Sci. Med. 42 (4) (2019) 1129–1139.
[20] K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, arXiv pre-print server (2015) 2015-04-10doi:arxiv:1409.1556.

[21] S. Ioffe, C. Szegedy, Batch normalization: accelerating deep network training by reducing internal covariate shift, in: International conference on machine learning, 2015, pp. 448–456. PMLR.
[22] M. Simon, E. Rodner, J. Denzler, Imagenet pre-trained models with batch normalization, arXiv preprint (2016) arXiv:1612.01452.
[23] A.P. Bradley, The use of the area under the ROC curve in the evaluation of machine learning algorithms, Pattern Recognit. 30 (7) (1997) 1145–1159.
[24] I. Silva, G.B. Moody, An open-source toolbox for analysing and processing physionet databases in matlab and octave, J. Open Res. Softw. 2 (1) (2014).
[25] R. Bousseljot, D. Kreiseler, A. Schnabel, Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das Internet, Biomed. Eng./Biomedizinische Technik 40 (s1) (1995) 317–318.
[26] P. Wagner, et al., PTB-XL, a large publicly available electrocardiography dataset, Sci. Data 7 (1) (2020) 1–15.
[27] P. Wagner, Strodthoff, N., Bousseljot, R., Samek, W., & Schaeffter, T., "PTB-XL, a large publicly available electrocardiography dataset (version 1.0.1).". [Online]. Available: https://doi.org/10.13026/x4td-x982.
[28] A.L. Goldberger, et al., PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals, Circulation 101 (23) (2000) e215–e220.
[29] L. Bottou, Large-scale machine learning with stochastic gradient descent, in: Proceedings of COMPSTAT'2010, 2010, pp. 177–186. Springer.
[30] I. Sutskever, J. Martens, G. Dahl, G. Hinton, On the importance of initialization and momentum in deep learning, in: International conference on machine learning, 2013, pp. 1139–1147. PMLR.
[31] J. Wang, J. Yang, W. Wu, Convergence of cyclic and almost-cyclic learning with momentum for feedforward neural networks, IEEE Trans. Neural Netw. 22 (8) (2011) 1297–1306.
[32] J. Wu, X.-.Y. Chen, H. Zhang, L.-.D. Xiong, H. Lei, S.-.H. Deng, Hyperparameter optimization for machine learning models based on Bayesian optimization, J. Electr. Sci. Technol. 17 (1) (2019) 26–40.
[33] T. Ishida, I. Yamane, T. Sakai, G. Niu, M. Sugiyama, Do We Need Zero Training Loss After Achieving Zero Training Error? arXiv preprint (2020) arXiv:⟨13:italic⟩2002.08709⟨/13:italic⟩.
[34] A. Chattopadhay, A. Sarkar, P. Howlader, V.N. Balasubramanian, Grad-cam++: generalized gradient-based visual explanations for deep convolutional networks, in: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), 2018, pp. 839–847. IEEE.
[35] E. Bozbeyoğlu, et al., The established electrocardiographic classification of anterior wall myocardial infarction misguides clinicians in terms of infarct location, extent and prognosis, Ann. Noninvasive Electrocardiol. 24 (3) (2019) e12628, doi:10.1111/anec.12628.
[36] H.S. Roshdy, I.I. El-Dosouky, M.H. Soliman, High-risk inferior myocardial infarction: can speckle tracking predict proximal right coronary lesions? Clin. Cardiol. 41 (1) (2018) 104–110.
[37] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.
[38] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Adv. Neural Inf. Process. Syst. 25 (2012) 1097–1105.
[39] W. Liu, F. Wang, Q. Huang, S. Chang, H. Wang, J. He, MFB-CBRNN: a Hybrid Network for MI Detection Using 12-Lead ECGs, IEEE J. Biomed. Health Inform. 24 (2) (2020) 503–514, doi:10.1109/JBHI.2019.2910082.
[40] E. Prabhakararao, S. Dandapat, Attentive rnn-based network to fuse 12-lead ecg and clinical features for improved myocardial infarction diagnosis, IEEE Signal Process. Lett. 27 (2020) 2029–2033, doi:10.1109/LSP.2020.3036314.
[41] A. Alghamdi, et al., Detection of myocardial infarction based on novel deep transfer learning methods for urban healthcare in smart cities, Multimed. Tools Appl. (2020), doi:10.1007/s11042-020-08769-x.
[42] C. Han, L. Shi, ML–ResNet: a novel network to detect and locate myocardial infarction using 12 leads ECG, Comput. Methods Programs Biomed. 185 (2020) October, doi:10.1016/j.cmpb.2019.105138.