

RESEARCH ARTICLE

Open Access



Estimating critical values from electrocardiogram using a deep ordinal convolutional neural network

Guodong Wei^{1†}, Xinxin Di^{2†}, Wenrui Zhang³, Shijia Geng¹, Deyun Zhang¹, Kai Wang¹, Zhaoji Fu^{1,4} and Shenda Hong^{5,6*} 

Abstract

Background: Critical values are commonly used in clinical laboratory tests to define health-related conditions of varying degrees. Knowing the values, people can quickly become aware of health risks, and the health professionals can take immediate actions and save lives.

Methods: In this paper, we propose a method that extends the concept of critical value to one of the most commonly used physiological signals in the clinical environment—Electrocardiogram (ECG). We first construct a mapping from common ECG diagnostic conclusions to critical values. After that, we build a 61-layer deep convolutional neural network named CardioV, which is characterized by an ordinal classifier.

Results: We conduct experiments on a large public ECG dataset, and demonstrate that CardioV achieves a mean absolute error of 0.4984 and a ROC-AUC score of 0.8735. In addition, we find that the model performs better for extreme critical values and the younger age group, while gender does not affect the performance. The ablation study confirms that the ordinal classification mechanism suits for estimating the critical values which contain ranking information. Moreover, model interpretation techniques help us discover that CardioV focuses on the characteristic ECG locations during the critical value estimation process.

Conclusions: As an ordinal classifier, CardioV performs well in estimating ECG critical values that can help people quickly identify different heart conditions. We obtain ROC-AUC scores above 0.8 for all four critical value categories, and find that the extreme values (0 (no risk) and 3 (high risk)) have better model performance than the other two (1 (low risk) and 2 (medium risk)). Results also show that gender does not affect the performance, and the older age group has worse performance than the younger age group. In addition, visualization techniques reveal that the model pays more attention to characteristic ECG locations.

Keywords: Critical value, Deep neural network, Ordinal classification, Electrocardiogram

Background

Critical values, which are also known as panic values, tell “when to panic” over abnormal health-related conditions [1]. With a few numerical critical values that summarize and simplify the complex medical circumstances, health professionals can provide timely and effective responses. Also, people without any medical background can easily assess the health problems based on the values. Critical

[†]Guodong Wei and Xinxin Di contributed equally to the article

*Correspondence: hongshenda@pku.edu.cn

⁶ Institute of Medical Technology, Health Science Center of Peking University, Beijing 100191, China

Full list of author information is available at the end of the article



values are always defined with decision boundaries estimated from laboratory tests or vital signs [2–5]. For example, if a routine blood test shows a person's serum potassium level is less than 3.0 mmol/L, it is a serious warning that the subject is at risk for hypokalemia and will need hospitalization, otherwise he or she might suffer from severe ventricular arrhythmia due to digoxin toxicity. Moreover, mortality risk scores in the Intensive Care Unit (ICU), such as APACHE III score [6] or SAPS II score [7], can also be viewed as an extended concept of critical values.

The above mentioned critical values require clinical laboratory tests or medical monitoring devices, which limits their usage in everyday life. In this paper, we aim to extend the concept of critical value to Electrocardiogram (ECG or EKG), which is one of the most commonly used non-invasive diagnostic or health management tools for heart-related problems. Compared with laboratory tests or simple vital signs, it is more difficult to build a critical value estimator for physiological signals, such as ECGs. Due to the high sampling frequencies, complex patterns, and long trends, traditional machine learning methods might not learn effectively from these complex signals.

Recently, deep neural networks (or deep learning methods) have achieved state-of-the-art performances in many areas such as speech recognition, computer vision, and natural language processing [8]. They also show great potentials on cardiovascular management [9, 10], disease detection [11–22], and biometric human identification [23, 24], and many other ECG analysis tasks [25–29]. However, there are no deep learning models designed for ECG critical value estimation so far. Unlike the above mentioned research, critical value estimation is neither a regression task nor a pure classification task. It is actually an ordinal classification task [30], which outputs categories that involve certain order relationships. As a result, the existing regression or classification models cannot be used directly to solve the ECG critical value estimation task.

In this paper, we present a method to estimate ECG critical values based deep learning techniques. We first propose a mapping from common ECG diagnostic conclusions (ECG statements) to ECG critical values. Then, we build an automatic critical value estimation model named CardioV, which is a 61-layer deep neural network based on neural architecture search and other advanced techniques in general Artificial Intelligence (AI) research areas [31–38]. Since the critical values have orders, we define the problem as a novel ordinal classification multi-task problem [30]. In order to obtain the probabilities of “no risk (critical value 0)”, “low risk (critical value 1)”, “medium risk (critical value 2)”, and “high risk (critical value 3)”, we train the model to learn the probabilities of

being greater than “no risk”, “low risk” or “medium risk”, then convert them into probabilities of the four ordered critical values. We conduct experiments on a large public ECG dataset named PTB-XL [39]. The mean absolute error on the test set is 0.4984, and the average ROC-AUC score is 0.8735. Results also show that the agreement of model–cardiologist is comparable with that of cardiologist–cardiologist. In addition, our ablation study reveals that CardioV is better than the baseline deep learning models. With ECG critical values, people can easily assess their heart conditions and be aware of the critical situations.

Methods

Dataset

We use PTB-XL¹ [39] from PhysioNet [40] to build and evaluate our method. The PTB-XL ECG dataset is a large publicly available ECG dataset, which contains 21,837 10-s clinical 12-lead ECG recordings from 18,885 patients (52% male and 48% female), ranging in age from 0 to 95. The waveform files are stored in WaveForm DataBase (WFDB) format with 16 bit precision at a resolution of 1 μ V/LSB and a sampling frequency of 500 Hz. A downsampled version of the waveform data with a sampling frequency of 100 Hz is released for the convenience of users. We use the 500 Hz ECG data in our experiments and preprocess the raw data using a bandpass filter of 0.5–50 Hz. The raw ECG data are annotated by up to two cardiologists, who assign potentially multiple ECG statements to each record [39]. There are 71 different ECG statements within five categories: Normal ECG (9528 records), Myocardial Infarction (5486 records), ST/T Change (5250 records), Conduction Disturbance (4907 records), and Hypertrophy (2655 records). The details of the 71 ECG statements can be found in our critical value mapping table (Table 2). The overall statistics is shown in Table 1.

Mapping from ECG statements to critical values

We start with ECG statements conforming to the Standard Communication Protocol for Computer-assisted Electrocardiography (SCP-ECG) standard, which covers diagnostic, form, and rhythm statements. Based on the SCP-ECG standard and the 2017 Chinese expert consensus [41], we create the mapping between critical values and ECG statements as shown in Table 2. The resulting ECG critical values have four levels, which are 0 (No Risk), 1 (Low Risk), 2 (Medium Risk), and 3 (High Risk). Their ordinal relationships are as follow:

¹ <https://physionet.org/content/ptb-xl/1.0.1/>.

Table 1 Overall statistics of PTB-XL dataset

Item	Statistics
# Records	21,837
Normal ECG	9528
Myocardial infarction	5486
ST/T change	5250
Conduction disturbance	4907
Hypertrophy	2655
Duration	10 s
Total patients	18,885
Gender	
Male	11,379 (52%)
Female	10,458 (48%)
Age	
Mean, (min, max)	60, (2, 95)
< 65, ≥ 65	12,331, 9417

$$0(\text{no risk}) < 1(\text{low risk}) < 2(\text{medium risk}) < 3(\text{high risk}) \tag{1}$$

Deep neural network for modeling ECG

Deep learning methods especially convolutional neural networks (CNNs) have achieved state-of-the-art performances in ECG modeling [28]. We design our ECG classification CNN model with the neural architecture space searching technique that is adapted to find the

best models for image classification [31]. The resulting network contains 61 layers which includes 7 stages of convolutional blocks connected with shortcut residual connection [32, 33], one global average pooling layer, and one fully connected dense layer. Each block consists of one convolutional layer with kernel size 1 (Conv1), one aggregated convolutional layer [34] with kernel size 16 and 16 groups (ConvK), and another convolutional layer with kernel size 1 (Conv1). Before each convolution layer, we apply batch normalization (BN) [35], Swish activation [36], and dropout (DO) [37]. We also introduce the channel-wise attention mechanism (SE block) [38] to improve the model performance. The first block of each stage downsample its input by a factor of 2, and the corresponding shortcut connections downsample the identity input using a max pooling operation by a factor of 2. The detailed model architecture is shown in Table 3.

Formally, we use $X \in \mathbb{R}^{d \times n}$ to represent input ECG data, where n is the length of ECG, d is the number of leads which is 12 in our case. We also use \mathcal{F} to represent our deep neural network. The predicted logits $z \in \mathbb{R}^c$ can then be represented as:

$$z = \mathcal{F}(X). \tag{2}$$

Training via ordinal classification

For the ECG critical value estimation task, the common idea would be to build a deep model to implement a

Table 2 The mapping of ECG statements to critical values

Critical Values	Labels
0, No risk (9148 Records)	Normal ECG (NORM), non-specific T-wave changes (NT_), sinus rhythm (SR)
1, Low Risk (5322 Records)	Non-diagnostic T abnormalities (NDT), left anterior fascicular block (LAFB), incomplete right bundle branch block (IRBBB), non-specific ST changes (NST_), complete right bundle branch block (CRBBB), non-specific ischemic (ISC_), normal functioning artificial pacemaker (PACE), first degree AV block (1AVB), left atrial overload/enlargement (LAO/LAE), ischemic in lateral leads (ISCLA), incomplete left bundle branch block (ILBBB), right atrial overload/enlargement (RAO/RAE), left posterior fascicular block (LPFB), right ventricular hypertrophy (RVH), septal hypertrophy (SEHYP), sinus tachycardia (STACH), atrial premature complex (PAC), abnormal QRS (ABQRS), high QRS voltage (HVOLT), inverted T-waves (INVT), low amplitude T-waves (LOWT), prolonged PR interval (LPR), low QRS voltages in the frontal and horizontal leads (LVOLT), sinus arrhythmia (SARRH), sinus bradycardia (SBRAD), non-specific ST depression (STD_), non-specific ST elevation (STE_), T-wave abnormality (TAB_)
2, Medium Risk (3241 Records)	Left ventricular hypertrophy (LVH), complete left bundle branch block (CLBBB), ischemic in anterolateral leads (ISCAL), non-specific intraventricular conduction disturbance (block) (IVCD), ventricular premature complex (PVC), subendocardial injury in anteroseptal leads (INJAS), ischemic in anteroseptal leads (ISCAS), ischemic in inferior leads (ISCIN), ischemic in inferolateral leads (ISCIL), atrial flutter (AFLT), atrial fibrillation (AF), electrolytic disturbance or drug (former EDIS) (EL), paroxysmal supraventricular tachycardia (PSVT), digitalis-effect (DIG), ischemic in anterior leads (ISCAN), second degree AV block (2AVB), subendocardial injury in inferolateral leads (INJIL), subendocardial injury in lateral leads (INJLA), subendocardial injury in inferior leads (INJIN), bigeminal pattern (unknown origin, SV or Ventricular) (BIGU), premature complex(es) (PRC(S)), Q waves present (QWAVE), supraventricular arrhythmia (SVARR), supraventricular tachycardia (SVTAC), trigeminal pattern (unknown origin, SV or Ventricular) (TRIGU), voltage criteria (QRS) for left ventricular hypertrophy (VCLVH)
3, High Risk (4126 Records)	Inferior myocardial infarction (IMI), anteroseptal myocardial infarction (ASMI), inferolateral myocardial infarction (ILMI), anterior myocardial infarction (AMI), anterolateral myocardial infarction (ALMI), long QT-interval (LNGQT), Wolf-Parkinson-White syndrome (WPW), lateral myocardial infarction (LMI), inferoposterolateral myocardial infarction (IPLMI), subendocardial injury in anterolateral leads (INJAL), inferoposterior myocardial infarction (IPMI), posterior myocardial infarction (PMI), third degree AV block (3AVB), ST-T changes compatible with ventricular aneurysm (ANEUR)

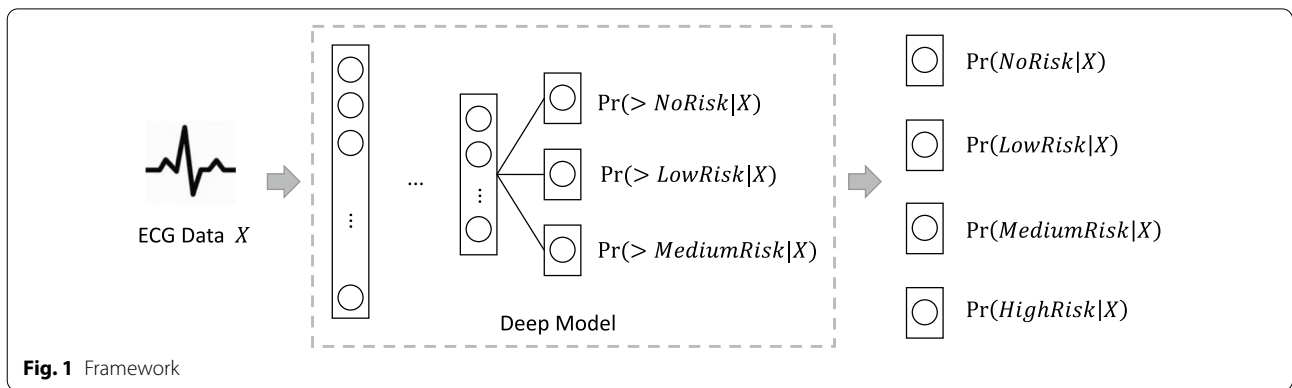


Fig. 1 Framework

Table 3 Model architecture

Stage	Layers	Output size
Input		(* , 12, 5000)
Stage 1	(Conv1, ConvK, Conv1) × 2	(* , 64, 2500)
Stage 2	(Conv1, ConvK, Conv1) × 2	(* , 160, 1250)
Stage 3	(Conv1, ConvK, Conv1) × 2	(* , 160, 625)
Stage 4	(Conv1, ConvK, Conv1) × 3	(* , 400, 312)
Stage 5	(Conv1, ConvK, Conv1) × 3	(* , 400, 156)
Stage 6	(Conv1, ConvK, Conv1) × 4	(* , 1024, 78)
Stage 7	(Conv1, ConvK, Conv1) × 4	(* , 1024, 39)
Pooling	Global Average	(* , 1024)
Prediction	Dense	(* , 3)

ECG length $n = 5000$, number of leads $d = 12$, and output dimension $c = 3$. The first dimension * represents number of samples in a batch

Table 4 Computing referenced probabilities from critical values

Critical value	$Pr(> NoRisk X)$	$Pr(> LowRisk X)$	$Pr(> MediumRisk X)$
0 (no risk)	0	0	0
1 (low risk)	1	0	0
2 (medium risk)	1	1	0
3 (high risk)	1	1	1

classification task. That is, given predicted logits $z \in \mathbb{R}^c$, for classification task we first apply softmax on z to get probabilities $p \in [0, 1]^c$, then optimize deep neural network \mathcal{F} via cross-entropy loss. However, the classification task only distinguishes different classes, which does not model the ordinal relationship shown in Eq. 1.

To solve this problem, we define the task as an ordinal classification task [30] rather than a simple classification task. The framework of our method is shown in Fig. 1. Ordinal classification task has the ability to learn from the ordinal relationship of classes. Intuitively, in our setting, the ordinal classification task can be regarded as a multi-task classification problem with the following three tasks:

- Task 1: whether the ECG critical value is higher than no risk? The probability is denoted as $Pr(> NoRisk|X)$;
- Task 2: whether the ECG critical value is higher than low risk? The probability is denoted as $Pr(> LowRisk|X)$;
- Task 3: whether the ECG critical value is higher than medium risk? The probability is denoted as $Pr(> MediumRisk|X)$.

Formally, given predicted logits $z \in \mathbb{R}^c$, for ordinal classification task we first apply sigmoid on z to get probabilities $p \in [0, 1]^c$ (Eq. 3), then optimize objective L of deep neural network \mathcal{F} via multi-task binary cross entropy (BCE) loss (Eqs. 4, 5). The label y is computed based on Table 4. In addition, c is set to 3, $p[0] = Pr(> NoRisk|X)$, $p[1] = Pr(> LowRisk|X)$, and $p[2] = Pr(> MediumRisk|X)$.

$$p = \text{sigmoid}(z) \tag{3}$$

$$L = \frac{1}{c} \sum_{i=1}^c BCE(p[i], y[i]) \tag{4}$$

$$BCE(p, y) = y \cdot \log p + (1 - y) \cdot \log(1 - p) \tag{5}$$

Then, we transform $Pr(> NoRisk|X)$, $Pr(> LowRisk|X)$ and $Pr(> MediumRisk|X)$ into probability of each critical value: $Pr(NoRisk|X)$, $Pr(LowRisk|X)$, $Pr(MediumRisk|X)$ and $Pr(HighRisk|X)$ based on Eq. 6. The final output is the class with the highest probability.

$$\begin{aligned} Pr(NoRisk|X) &= 1 - Pr(> NoRisk|X) \\ Pr(LowRisk|X) &= Pr(NoRisk|X) - Pr(> LowRisk|X) \\ Pr(MediumRisk|X) &= Pr(LowRisk|X) - Pr(> MediumRisk|X) \\ Pr(HighRisk|X) &= Pr(> MediumRisk|X). \end{aligned} \tag{6}$$

Implementation details

We split the entire dataset by subject and obtain a training set with 17,741 samples (80% subjects), a validation set with 2193 samples (10% subjects), and a test set with 2203 samples (10% subjects). The model is built and trained with the PyTorch Python package. We choose Adam [42] optimizer with back-propagation, and add weight normalization to avoid overfitting. The batch size is set to be 256 samples, and the original learning rate is set to be 0.001. When the validation performance stops improving, we reduce the learning rate by a factor of 0.3. Compared with conventional classification, it is more difficult to train the ordinal multi-task classification. To solve the problem, we first train the model with conventional cross-entropy loss, and then conduct a finetuning after replacing the objective to ordinal loss. The results are reported on the test set.

Evaluations

Our evaluation measurements include mean absolute error (MAE), receiver operating characteristic (ROC) curve of each class, area under the ROC curve (ROC-AUC, or just AUC) of each class, and the average value of AUC scores. The ROC curve is first computed based on the predicted probability and ground truth of each label directly without a predefined threshold, then defined as the curve of the true positive rate versus the false positive rate at various thresholds ranging from zero to one. We also ask the cardiologists to revise wrong predicted cases and calculate the agreement of model-cardiologist and the agreement of cardiologist-cardiologist. Moreover, we analyze $MAE = 3$ cases, which represent serious errors in the model, one by one. Finally, in order to explain the model, we use the Grad-CAM [43] method to obtain the corresponding heat maps for the layers of interest. Through the heat maps, we can find the positions of the signal that the model is concerned about in the corresponding layers.

Results

Classification results

The results of the ROC curves for each class are shown in Fig. 2. We can see that all four classes achieve higher than 0.8 ROC-AUC scores. We also observe that 0 (No Risk) and 3 (High Risk) are higher than the other two. The reason might be that the intermediate values (1 and 2) are more difficult to predict than the extreme values (0 and 3).

We then evaluate the model performance on different patient subtypes. We further divide the test set into subgroups by gender (male, female), and by age (age < 65, age \geq 65), and show the results in Table 5. In terms of

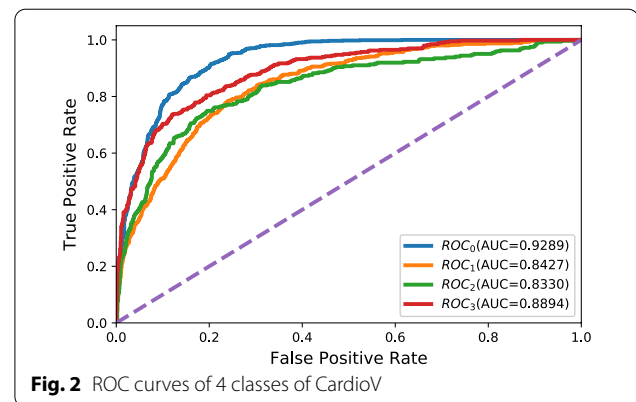


Fig. 2 ROC curves of 4 classes of CardioV

genders, the male and female groups have close performances on all evaluations, which indicates that the model is fair towards different genders. For ages, the age < 65 group is much better than the age \geq 65 group. The reason might be that elders have age-related issues which could affect the heart but are difficult to be identified with ECG.

Moreover, we compare the agreement of model-cardiologist and the agreement of cardiologist-cardiologist. We first extract wrong predicted cases ($MAE \geq 1$), and then ask an individual cardiologist to revise these cases. After that, we analyze these results by comparing original labels (annotated by other cardiologists), model predictions, and revised labels. The total number of incorrectly predicted cases is 674. After revising, the cardiologist agrees with the original labels in 259 samples, agrees with model predictions in 361 samples, and disagrees with both in 54 samples. In this case, we can see that the agreement between model and cardiologist (model-cardiologist) is $361/674 = 53.56\%$, which is higher than the agreement between cardiologist and cardiologist (cardiologist-cardiologist) $259/674 = 38.43\%$. The disagreement of ECG diagnosis among cardiologists has already been discovered in previous research [25]. The result suggests that our method has at least comparable performance with cardiologists.

Ablation study

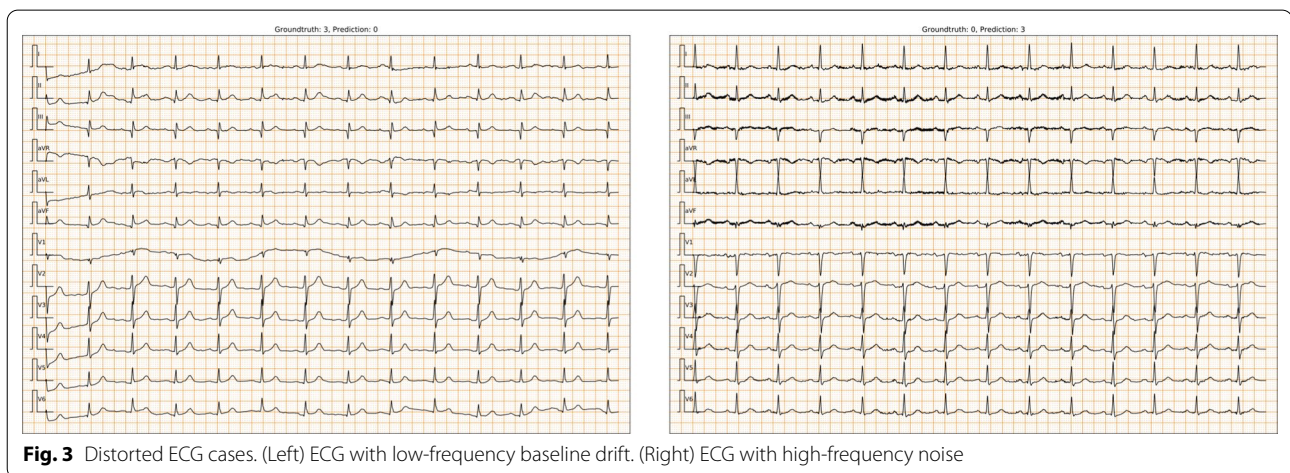
We compare CardioV with two ablation study baselines: classification and regression. We implement classification with the same model architecture, and replace our ordinal classification objective with the four-class cross-entropy objective. We also implement regression with the same model architecture, but optimize a mean squared error (MSE) objective to predict the numerical critical values. From the results shown in Table 6, we see that CardioV performs better than both classification and regression, which suggests that ordinal classification is a

Table 5 Results of different subtypes

Subtype	MAE	AUC_0	AUC_1	AUC_2	AUC_3	AUC
Full	0.4934	0.9289	0.8421	0.8330	0.8894	0.8735
Male	0.4686	0.9304	0.8549	0.8254	0.9001	0.8777
Female	0.5197	0.9222	0.8298	0.8381	0.8754	0.8676
Age < 65	0.3882	0.8971	0.8725	0.8132	0.8975	0.8701
Age \geq 65	0.6108	0.9324	0.7974	0.8214	0.8634	0.8537

Table 6 Results of different methods

Method	MAE	AUC_0	AUC_1	AUC_2	AUC_3	AUC
CardioV	0.4934	0.9289	0.8421	0.8330	0.8894	0.8735
Classification	0.5247	0.9259	0.8233	0.8360	0.8838	0.8673
Regression	0.5811	–	–	–	–	–

**Fig. 3** Distorted ECG cases. (Left) ECG with low-frequency baseline drift. (Right) ECG with high-frequency noise

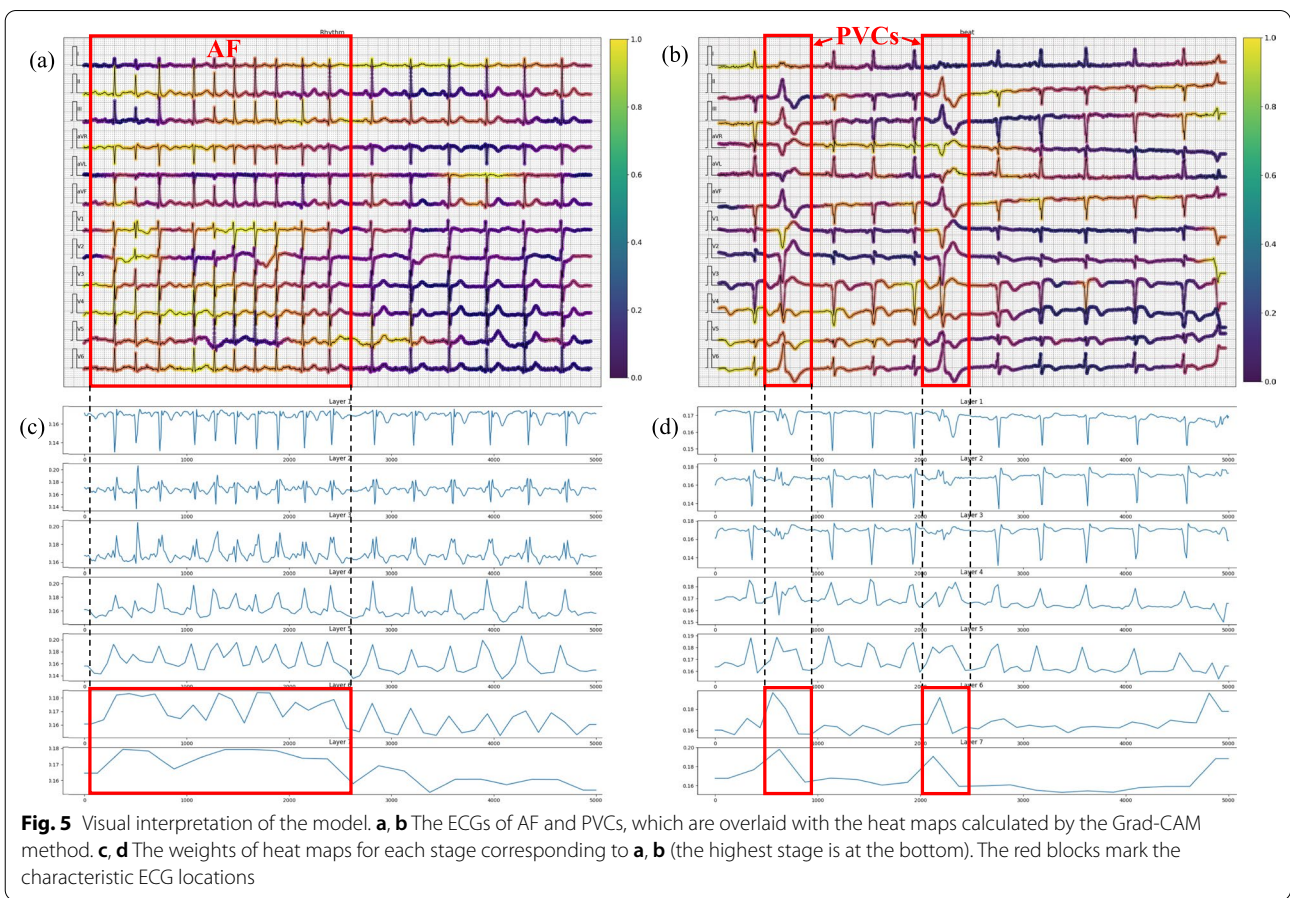
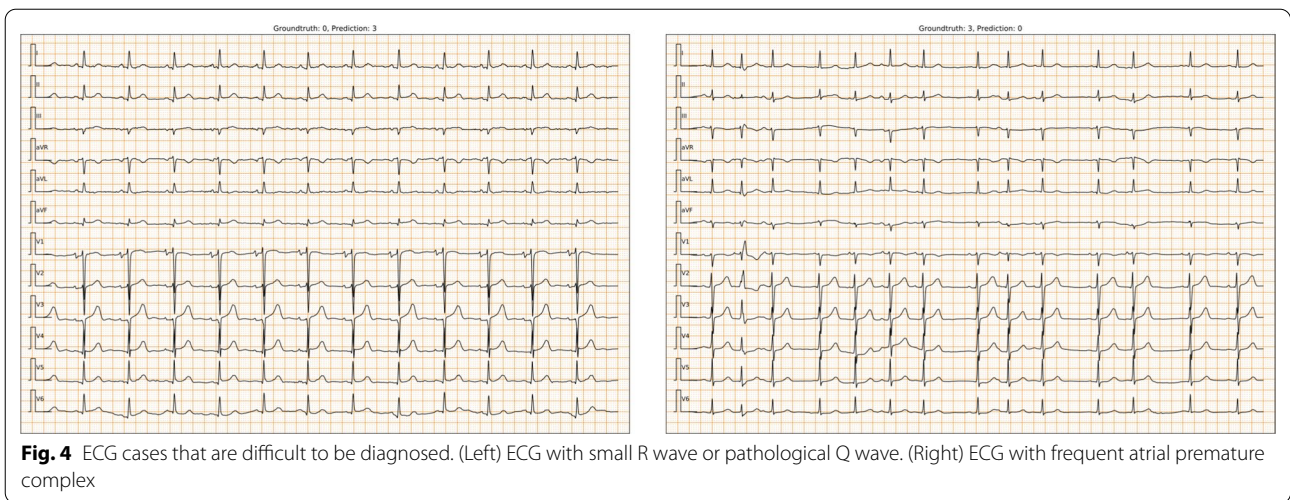
good choice when dealing with classification tasks with definite ordinal relationships among categories.

Case study

Finally, we examine the $MAE = 3$ cases which might lead to serious consequences in the real-world applications. The total number of $MAE = 3$ wrong predicted cases is 15. Among these cases, we find that 9 have distortions with low-frequency baseline drift (Fig. 3 Left) or high-frequency noise (Fig. 3 Right). The other cases are themselves difficult to be identified. For example, Fig. 4 (Left) shows the tiny R wave or pathological Q wave, which mainly exists in ECGs of people who have old inferior wall myocardial infarction (old IMI), but could also appear in ECGs of healthy people. Figure 4 (Right) shows

an ECG with frequent atrial premature complex, which might be recognized as sinus arrhythmia.

In addition, we apply gradient-weighted class activation mapping (Grad-CAM) to obtain the heat maps of the last convolutional layers for each stage to interpret the model. The highlighted areas represent the locations that the model focus on. To visualize them, we select two representative types of ECGs, which are ECGs of “rhythm-type” AF (characterized by an irregularly irregular rhythm) and ECGs of “beat-type” PVCs (characterized by wide QRS complexes). Figure 5a, b show the selected ECGs of AF and PVCs, which are overlaid with heat maps of the last convolution layer calculated by the Grad-CAM method. We see that most of the characteristic locations are brighter than



other areas. To further explore the model’s hierarchical focus locations, we combine the heap map weights of all 12 leads for the last layer of each stage, and plot the weights from top to bottom in layer arrangement order (see Fig. 5c, d). The results show that the higher layers pay more attention to the characteristic ECG locations.

Discussion

Critical value is a concept that is easy to understand, and even people without medical background can use it to identify different health conditions. The traditional critical values are associated with laboratory tests results and simple vital signs. However, in many cases, we need more

complex signals to accurately assess the health conditions. As a physiological signal that can be collected easily and quickly, ECG is a good candidate to be mapped into critical values. Compared with laboratory tests and vital signs, ECG signals provide much more high-frequency information about the heart conditions, which also means it is difficult for traditional machine learning models to learn the features. In this paper, we propose a deep ordinal convolutional neural network named CardioV to automatically estimate the ECG critical value categories.

From the experimental results we find that the two extreme critical values (0 and 3) have better model performance than the two middle ones (1 and 2). The extreme critical values include normal and high risk conditions, which might be easier to predict than the other two, since the healthy state and severe condition could have more easily identifiable ECG characteristics. On the other hand, older people may have complex ECG signatures because their hearts may be affected by age-related diseases, so the model performs poorly in the older age group compared to the younger age group.

Since the main objective of this work is to demonstrate the feasibility of assessing critical values with ECGs, there are still several limitations. First, it does not combine other information of the patients, such as the blood routine test results. Second, without considering additional stratification rules, the same ECG recording can reflect a variety of heart diseases and can be subdivided into different critical grades. In the end, no specific suggestion of actions are associated with each critical value.

In the future, we plan to collect more data to enhance our model, and build a hierarchy of critical value estimator to support tiered medical services. Moreover, we would like to extend similar ideas to other physiological data, such as photoplethysmogram (PPG), electroencephalogram (EEG), and electromyogram (EMG), so people can easily understand these complex signals and take quick actions in life-threatening situations.

Conclusion

In our study, we propose CardioV, an ordinal classifier, to estimate ECG critical value categories that can help people quickly identify different heart health conditions. Test results show that the model performs well in all four critical value categories. Furthermore, we observe three phenomena: extreme values (0 and 3) have better model performance than the other two; gender does not affect the performance; the older age group has worse performance than the younger age group. We also find that the agreement of model-cardiologist is comparable with that of

cardiologist-cardiologist. The ablation study reveals that CardioV outperforms baseline deep learning models and validates that ordinal classification is suitable for identifying categories with ranking information. In addition, we interpret our model through activation visualization techniques, and discover that the model pays more attention to characteristic ECG locations, whether in “rhythm-type” or “beat-type” arrhythmia.

Acknowledgements

We would like to thank the editors and reviewers for their time and efforts.

Author contributions

GW, XD and SH conceptualized the study idea. GW and XD conducted the data curation. GW, WZ, and SG performed the formal analysis. SH acquired the funding. GW and SH applied the methodology. SH did the project administration. DZ and KW conducted the validation. GW finished the visualization. GW, WZ and SH made the original draft writing. All the authors did the review and editing, and all authors have read and approved the manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (No. 62102008).

Availability of data and materials

The PTB-XL ECG dataset used in this study is available on the PhysioNet website <https://physionet.org/content/ptb-xl/1.0.1/>. The code of CardioV can be downloaded from the github repository: <https://github.com/hsd1503/CardioV>. We also deploy our trained model and create an online application: https://www.heartvoice.com.cn/criticalvalue/index_en. With the application, users can test the critical value analysis function using their own data or the provided sample data.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹HeartVoice Medical Technology, Hefei 230027, China. ²Department of Electrocardiogram, The First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, Anhui 230001, China. ³Department of Mathematics, National University of Singapore, Singapore 119077, Singapore. ⁴School of Management, University of Science and Technology of China, Hefei 230026, China. ⁵National Institute of Health Data Science, Peking University, Beijing 100191, China. ⁶Institute of Medical Technology, Health Science Center of Peking University, Beijing 100191, China.

Received: 6 June 2022 Accepted: 28 October 2022

Published online: 16 November 2022

References

- Lundberg G. When to panic over abnormal values. *MLO Med Lab Obs.* 1972;4(1):47–54.
- Kuperman GJ, Boyle D, Jha A, Rittenberg E, Ma'Luf N, Tanasijevic MJ, Teich JM, Winkelman J, Bates DW. How promptly are inpatients treated for critical laboratory results? *J Am Med Inform Assoc.* 1998;5(1):112–9.
- Kuperman GJ, Teich JM, Tanasijevic MJ, Ma'Luf N, Rittenberg E, Jha A, Fiskio J, Winkelman J, Bates DW. Improving response to critical

- laboratory results with automation: results of a randomized controlled trial. *J Am Med Inform Assoc.* 1999;6(6):512–22.
4. Tillman J, Barth J. A survey of laboratory 'critical (alert) limits' in the UK. *Ann Clin Biochem.* 2003;40(2):181–4.
 5. Dighe AS, Jones JB, Parham S, Lewandrowski KB. Survey of critical value reporting and reduction of false-positive critical value results. *Arch Pathol Lab Med.* 2008;132(10):1666–71.
 6. Knaus WA, Wagner DP, Draper EA, Zimmerman JE, Bergner M, Bastos PG, Sirio CA, Murphy DJ, Lotring T, Damiano A. The apache iii prognostic system: risk prediction of hospital mortality for critically ill hospitalized adults. *Chest.* 1991;100(6):1619–36.
 7. Le Gall J-R, Lemeshow S, Saulnier F. A new simplified acute physiology score (saps ii) based on a European/North American multicenter study. *JAMA.* 1993;270(24):2957–63.
 8. LeCun Y, Bengio Y, Hinton GE. Deep learning. *Nature.* 2015;521(7553):436–44. <https://doi.org/10.1038/nature14539>.
 9. Siontis KC, Noseworthy PA, Attia ZI, Friedman PA. Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. *Nat Rev Cardiol.* 2021;18(7):465–78.
 10. Fu Z, Hong S, Zhang R, Du S. Artificial-intelligence-enhanced mobile system for cardiovascular health management. *Sensors.* 2021;21(3):773.
 11. Acharya UR, Fujita H, Lih OS, Hagiwara Y, Tan JH, Adam M. Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network. *Inf Sci.* 2017;405:81–90.
 12. Acharya UR, Fujita H, Lih OS, Adam M, Tan JH, Chua CK. Automated detection of coronary artery disease using different durations of ECG segments with convolutional neural network. *Knowl Based Syst.* 2017;132:62–71.
 13. Acharya UR, Fujita H, Oh SL, Hagiwara Y, Tan JH, Adam M. Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals. *Inf Sci.* 2017;415:190–8.
 14. Attia ZI, Noseworthy PA, Lopez-Jimenez F, Asirvatham SJ, Deshmukh AJ, Gersh BJ, Carter RE, Yao X, Rabinstein AA, Erickson BJ. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. *The Lancet.* 2019;394(10201):861–7.
 15. Erdenebayar U, Kim YJ, Park J-U, Joo EY, Lee K-J. Deep learning approaches for automatic detection of sleep apnea events from an electrocardiogram. *Comput Methods Programs Biomed.* 2019;180:105001.
 16. Raghunath S, Cerna AEU, Jing L, Stough J, Hartzel DN, Leader JB, Kirchner HL, Stumpe MC, Hafez A, Nemani A, et al. Prediction of mortality from 12-lead electrocardiogram voltage data using a deep neural network. *Nat Med.* 2020;2020:1–6.
 17. Ribeiro AH, Ribeiro MH, Paixão GM, Oliveira DM, Gomes PR, Canazart JA, Ferreira MP, Andersson CR, Macfarlane PW, Wagner M Jr. Automatic diagnosis of the 12-lead ECG using a deep neural network. *Nat Commun.* 2020;11(1):1–9.
 18. Hong S, Zhou Y, Wu M, Shang J, Wang Q, Li H, Xie J. Combining deep neural networks and engineered features for cardiac arrhythmia detection from ECG recordings. *Physiol Meas.* 2019;40(5):054009.
 19. Hong S, Xiao C, Ma T, Li H, Sun J. Mina: multilevel knowledge-guided attention for modeling electrocardiography signals. In: *Proceedings of the 28th international joint conference on artificial intelligence, AAAI Press.* 2019. p. 5888–94.
 20. Zhou Y, Hong S, Shang J, Wu M, Wang Q, Li H, Xie J. K-margin-based residual-convolution-recurrent neural network for atrial fibrillation detection. In: *IJCAI 2019.*
 21. Hong S, Xu Y, Khare A, Priambada S, Maher K, Aljiffry A, Sun J, Tumanov A. Holmes: health online model ensemble serving for deep learning models in intensive care units. In: *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery and data mining.* 2020. p. 1614–24.
 22. Hong S, Zhang W, Sun C, Zhou Y, Li H. Practical lessons on 12-lead ECG classification: meta-analysis of methods from physionet/computing in cardiology challenge 2020. *Front Physiol.* 2022;2022:2505.
 23. Labati RD, Muñoz E, Piuri V, Sassi R, Scotti F. Deep-ECG: convolutional neural networks for ECG biometric recognition. *Pattern Recogn Lett.* 2019;126:78–85.
 24. Hong S, Wang C, Fu Z. Cardiod: learning to identification from electrocardiogram data. *Neurocomputing.* 2020;412:11–8.
 25. Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP, Ng AY. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat Med.* 2019;25(1):65.
 26. Sinnecker D. A deep neural network trained to interpret results from electrocardiograms: better than physicians? *Lancet Dig Health.* 2020;2(7):332–3.
 27. Elul Y, Rosenberg AA, Schuster A, Bronstein AM, Yaniv Y. Meeting the unmet needs of clinicians from AI systems showcased for cardiology with deep-learning-based ECG analysis. *Proc Natl Acad Sci.* 2021;118(24):e2020620118.
 28. Hong S, Zhou Y, Shang J, Xiao C, Sun J. Opportunities and challenges of deep learning methods for electrocardiogram data: a systematic review. *Comput Biol Med.* 2020;122: 103801.
 29. Somani S, Russak AJ, Richter F, Zhao S, Vaid A, Chaudhry F, De Freitas JK, Naik N, Miotto R, Nadkarni GN, Narula J, Argulian E, Glicksberg BS. Deep learning and the electrocardiogram: review of the current state-of-the-art. *EP Europace.* 2021;23(8):1179–91. <https://doi.org/10.1093/europace/eaab377>.
 30. Frank E, Hall M. A simple approach to ordinal classification. In: *European conference on machine learning, Springer.* 2001. p. 145–56.
 31. Radosavovic I, Kosaraju RP, Girshick R, He K, Dollár P. Designing network design spaces. 2020.
 32. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2016. p. 770–8.
 33. He K, Zhang X, Ren S, Sun J. Identity mappings in deep residual networks. In: *European conference on computer vision, Springer.* 2016. p. 630–45.
 34. Xie S, Girshick R, Dollár P, Tu Z, He K. Aggregated residual transformations for deep neural networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2017. p. 1492–500.
 35. Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. 2015. arXiv preprint [arXiv:1502.03167](https://arxiv.org/abs/1502.03167).
 36. Ramachandran P, Zoph B, Le QV. Searching for activation functions. 2017. arXiv preprint [arXiv:1710.05941](https://arxiv.org/abs/1710.05941).
 37. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res.* 2014;15(1):1929–58.
 38. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2018. p. 7132–41.
 39. Wagner P, Strodthoff N, Boussejot R-D, Kreiseler D, Lunze FI, Samek W, Schaeffter T. Ptb-xl, a large publicly available electrocardiography dataset. *Sci Data.* 2020;7(1):1–15.
 40. Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE. PhysioBank, physioToolkit, and physionet: components of a new research resource for complex physiologic signals. *Circulation.* 2000;101(23):215–20.
 41. Chinese Electrocardiographic Society, CVEWG 2017 consensus of Chinese experts on ECG critical value (in chinese). *J Clin Electrocardiol.* 2017;026(006):401–02.
 42. Kingma DP, Ba J. Adam: a method for stochastic optimization. 2014. arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980).
 43. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-cam: Visual explanations from deep networks via gradient-based localization. In: *Proceedings of the IEEE international conference on computer vision (ICCV).* 2017.

Publisher's Note

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