

## Article

# Classification of Prehospital Electrocardiograms Performed in Ambulances According to Severity Using a Deep Learning Neural Network

Ryo Oikawa <sup>1</sup>, Akio Doi <sup>2,\*</sup> , Tomonori Itoh <sup>3,4</sup>, Toshiaki Sakai <sup>5</sup> and Osamu Nishiyama <sup>5</sup>

<sup>1</sup> Graduate School of Software and Information Science, Iwate Prefectural University, 152-52 Sugo, Takizawa 020-0693, Iwate, Japan; g231t006@s.iwate-pu.ac.jp

<sup>2</sup> Faculty of Software and Information Science, Iwate Prefectural University, 152-52 Sugo, Takizawa 020-0693, Iwate, Japan

<sup>3</sup> Department of Internal Medicine, Division of Cardiology, Iwate Medical University, 1-1-1 Idaidori, Shiwa 028-3695, Iwate, Japan; tomoitoh@iwate-med.ac.jp

<sup>4</sup> Department of Medical Education, Division of Community Medicine, Iwate Medical University, 1-1-1 Idaidori, Shiwa 028-3695, Iwate, Japan

<sup>5</sup> Iwate Prefectural Ninohe Hospital, 38-2 Okawarage, Ninohe 028-6193, Iwate, Japan; tsakai201189@gmail.com (T.S.); yamari@mopera.net (O.N.)

\* Correspondence: doi@iwate-pu.ac.jp

**Abstract:** Prehospital electrocardiogram (PH-ECG) transmission is an important technology for reducing door-to-balloon time, but the decision to transmit often depends on the discretion of emergency medical technicians (EMTs). Additionally, studies based on real-world data remain insufficient. This study reports a machine learning-based method for classifying the severity of PH-ECG images and explores its feasibility. PH-ECG data were compiled from 120 patients between September 2017 and September 2020. The model we created from these data was the first classification model for PH-ECG images using data from a Japanese study population and showed a weighted F1-score of 0.85 and an Area Under the Curve (AUC) of 0.93. This result can be interpreted as having an excellent balance of sensitivity and specificity. The Cohen's Kappa coefficient between AI's inferences and the correct labels created by two cardiologists was 0.68 ( $p < 0.05$ ), which is considered "substantial" according to the guidelines presented by Landis and Koch. In this study, although we were not able to remove noise caused by patient movement or electrode detachment, the results indicate that image-based abnormality detection from PH-ECGs is feasible and effective, particularly in regions like Japan where ECG data are often stored and transmitted as images. In addition, in our region, paramedics follow a multi-step process to decide whether to transmit an ECG, which takes time for the first screening. However, if the ECG is transmitted when either the paramedics or the deep learning model detects an abnormality, it is expected to reduce reading time and door-to-balloon time, as well as decrease false negatives.

**Keywords:** PH-ECG; ECG classification; myocardial infarction; emergency medical services; deep learning



**Citation:** Oikawa, R.; Doi, A.; Itoh, T.; Sakai, T.; Nishiyama, O. Classification of Prehospital Electrocardiograms Performed in Ambulances According to Severity Using a Deep Learning Neural Network. *Emerg. Care Med.* **2024**, *1*, 280–298. <https://doi.org/10.3390/ecm1030029>

Academic Editor: George Notas

Received: 10 April 2024

Revised: 9 June 2024

Accepted: 13 August 2024

Published: 2 September 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

A prehospital electrocardiogram (PH-ECG) is an electrocardiogram performed by paramedics on patients suspected of having a myocardial infarction or other conditions. Several countries and regions have also implemented systems that transmit data from the field to the medical facility, allowing a cardiologist to evaluate the ECG remotely. This optimizes the time until the patient arrives and enables more prompt initiation of treatment. This is particularly useful for patients requiring immediate medical attention, such as those with ST-elevation myocardial infarction (STEMI) [1]. Several studies have also shown that PH-ECGs improve door-to-balloon time and in-hospital mortality. This is because very early ECGs can detect important signs of acute coronary syndrome (ACS) [1–4] that may

disappear before arrival at the emergency department (ED) [5,6]. Due to these advantages, the use and transmission of prehospital ECGs are recommended in various guidelines [7,8]. A 2016 statistical survey in Japan [9] found that 85.3% of PH-ECGs were evaluated by paramedics, with only 2.9% of institutions performing physician evaluations. One reason why PH-ECG transmission systems are not widely used is the high burden placed on physicians. In our medical area, cardiologists are currently constantly carrying tablet terminals to respond to patients even outside of working hours [10,11]. It is also difficult for physicians to evaluate all ECGs at the first contact stage, and even in areas where PH-ECG transmission systems have been introduced, the EMTs decide whether to transmit or not. The criteria depend on the region and the emergency medical service (EMS) team, and if the decision is discretionary, the sensitivity and specificity depend on the knowledge and experience of the EMTs [12]. Using machine learning models to assist screening during the PH-ECG transmission phase may help reduce the burden on EMTs and physicians and contribute to the widespread use of PH-ECG transmission systems.

A major challenge in anomaly detection from PH-ECGs is the lack of data accumulation. This is especially problematic with machine learning, which requires a lot of data. A number of machine learning-based approaches for detecting anomalies from ECG data have been reported [13–15] and offer superior generalizability compared to traditional, rule-based approaches using Minnesota codes, etc. Few studies have constructed their own clinical datasets [16], and specific public datasets (MIT-BIH [17], Physikalisch-Technische Bundesanstalt (PTB) [18], etc.) are widely used, but in-hospital ECGs such as these usually do not include important signs that can be observed in very early ECGs. Although the gold standard for automatic ECG analysis is the signal format, in clinical practice, only digital images of ECGs displayed by ECG Viewer may be stored, leaving room to investigate their potential use in PH-ECGs, where data are scarce.

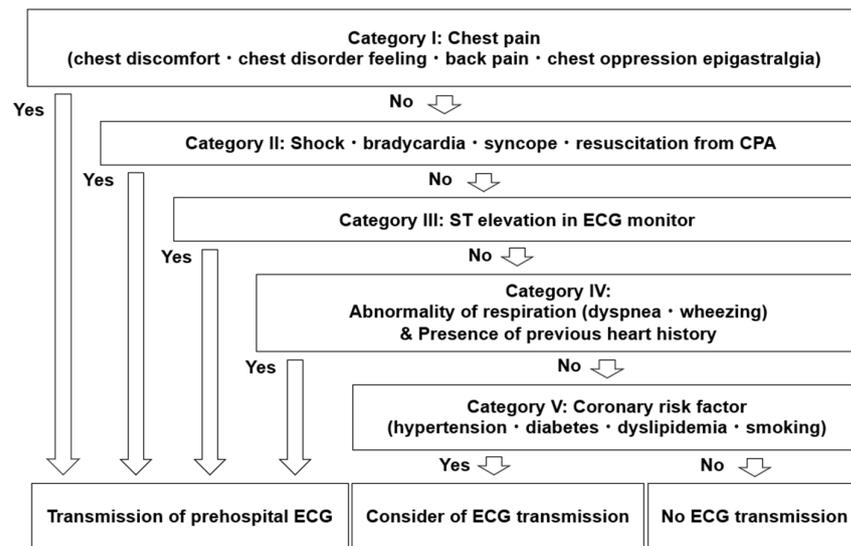
The purpose of this study was to evaluate the predictive power of image-based abnormality detection from PH-ECGs. The results presented indicate that deep learning models may be able to assist in screening PH-ECGs at the transmission stage.

## 2. Materials and Methods

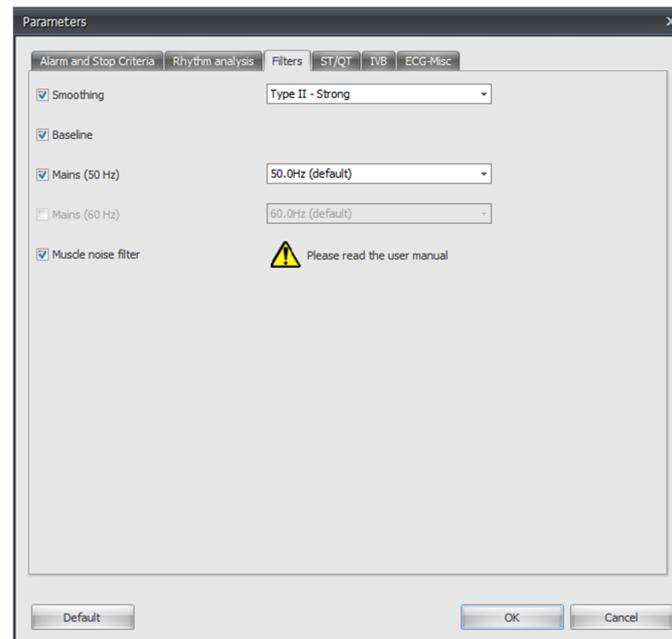
### 2.1. Data Collection and Preprocessing

Patients for whom ECGs were transmitted to Iwate Prefectural Ninohe Hospital between September 2017 and September 2020 were enrolled. Figure 1 shows the prehospital ECG transmission criteria and flowchart. The number of patients included in this study was 120, with a mean age of  $77 \pm 14.5$  years. Fifty percent of the patients were male. PH-ECGs were performed and transmitted by EMTs from five fire stations in the Ninohe Medical Area in northern Iwate Prefecture. The medical area has a population of 50,000, spans an area equivalent to half the size of Tokyo's metropolitan area, and includes Ninohe City, two towns, and one village (Ichinohe Town, Karumai Town, and Kudo Village). PH-ECGs were acquired from a PC-based electrocardiograph (EC-12RS; Labtech, Debrecen, Hungary) and transmitted by the "Fuji no Kuni" wireless 12-lead ECG transmission system (Good Care, Osaka, Japan) [19]. The 12-lead ECGs were placed according to the standard 12-lead ECG placement guidelines. Data were converted to JPEG format image files by the "MFER Image Converter" image conversion application (MFER Committee, Japan) before being sent.

The ECG signal suffers from various artifacts, and the inclusion of these noise signals reduces diagnostic accuracy [20]. Specifically, eliminating Baseline Wander (BW) and Power Line Interference (PLI) is crucial for an accurate cardiac disease diagnosis [21,22]. The primary causes of BW include patient motion and respiration [23], while PLI primarily stems from electromagnetic interference originating from the AC power supply. In our process, we employed the EC-12RS electrocardiograph's noise reduction feature to counteract these disturbances. The noise reduction settings are shown in Figure 2.



**Figure 1.** PH-ECG transmission criteria and flowchart. Adapted with permission from [11], Elsevier, 2018. Ideally, all electrocardiograms should be transmitted and analyzed by physicians. However, in regions with a limited number of doctors, unlike in large cities, such a flowchart is effective for the efficient detection of STEMI and the reduction in DTBT.



**Figure 2.** Noise reduction settings for the EC-12RS used. Smoothing of waveforms, BW removal, PLI removal (50 Hz), and EMG signal filter are enabled.

Out of the 120 patients, 21 cases were found unsuitable for analysis and were excluded, leaving 99 cases for the final data analysis. The exclusion criteria are detailed in Table 1. The class breakdown of the 99 subjects is shown in Table 2. In the analysis of ECGs from images, the calibration height can impact the results. To simplify this issue, the calibration of the input images in this study was standardized at 10 mm/mV, leading to the exclusion of 18 cases with varying heights. Furthermore, two ECGs that were printed on paper and then rescanned were excluded as these situations were not anticipated in this study. Lastly, we excluded one ECG that was excessively noisy and could not be assessed by a cardiologist. Thus, all reasons for exclusion were due to the quality of the recordings, and ECG findings did not influence the exclusion process. For the PH-ECG

data of each patient, the cardiologists performed a three-level classification of severity: “normal,” “mild/moderate,” or “severe. A single label was created through cross-checking by two cardiologists. This labeling was primarily performed on a per-lead basis. For data where this was not possible, labeling was conducted in parts divided into three segments per lead. The classification criteria are shown in Table 3. The definitions of ECG findings in this study are based on the ACS guidelines of the Japanese Circulation Society [24] and adhere to the strictly defined and widely accepted standards, such as the guidelines for the Fourth Universal Definition of Myocardial Infarction. ECG findings that are not included in the current dataset are not listed in Table 3 (e.g., Wide QRS complex tachycardia). During this labeling, the cardiologists had access to the patient’s final diagnosis. To make the image size suitable for network input, we first extracted 12 single-lead ECG images from one 12-lead ECG. Each of these single-lead ECGs was then further divided into three segments. Thus, from one 12-lead ECG, we obtained a total of 36 single-lead ECG images. Each image was thinned and binarized. Of the 3564 final data images, 2439 showed normal waveforms and 1125 had abnormal (mild/moderate or severe) waveforms. The breakdown is shown in Table 4. In the experiments described in the next section, 80% of the data were used as training data and 20% as test data. Additionally, 20% of the training data were used as validation data and cross-validated five times. A flow diagram of the dataset is shown in Figure 3.

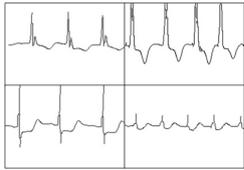
**Table 1.** Number of cases excluded and reasons.

Reason	Number of Cases Excluded
Calibration differing from 10 mm/mV	18
ECGs printed and rescanned	2
Excessive noise-preventing analysis	1

**Table 2.** Subject classification.

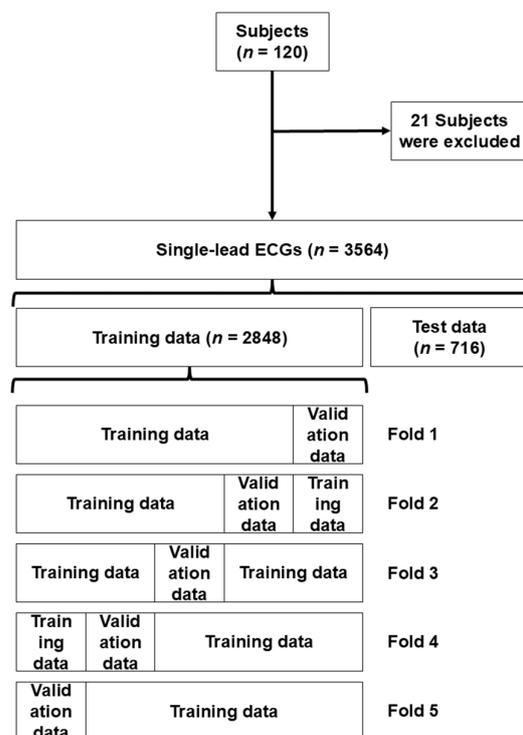
Class	Subjects <i>n</i>
Normal	14
Mild/moderate	34
Severe	51

**Table 3.** Definition of severity. The “Examples” column consists of preprocessed data that were randomly selected from each severity.

Severity	Abnormal Findings	Examples
Mild	Sinus tachycardia, left axis deviation, low voltage, counterclockwise rotation, flat T wave, premature atrial contraction (PAC), premature ventricular contraction (PVC), mild ST depression of 0.5 mm, incomplete right bundle branch block (IRBBB), negative T waves in V1-2, first-degree atrioventricular block (AVB), and sinus bradycardia.	
Moderate	Right bundle branch block (RBBB), negative T wave, atrial fibrillation, right ventricular hypertrophy, QT prolongation, ST depression of 1 mm, QS pattern in V1-3, and paroxysmal supraventricular tachycardia (PSVT).	
Severe	Q wave formation, left bundle branch block, pacemaker rhythm, ST elevation, reduced R wave amplitude, Q waves in II, III, and aVF leads, complete right bundle branch block (CRBBB) with negative T waves in precordial leads, second-degree AVB, and third-degree AVB.	

**Table 4.** Waveform classification.

Class	Waveforms n
Normal	2439
Mild/moderate	489
Severe	636



**Figure 3.** A flow diagram of the dataset. The training data are divided into 5 folds for cross-validation. This technique evaluates the performance of a model by dividing the data into multiple subsets. The model is trained on some subsets and tested on the remaining ones. This process is repeated several times, and the results are averaged to ensure the model's robustness and to mitigate overfitting.

Data augmentation was not implemented because preliminary experiments showed no significant effect. In particular, affine transformations resulted in a decrease in accuracy, likely due to the loss of positional information in the waveforms, making it more difficult to detect abnormalities.

## 2.2. Neural Network

The conceptual figures of this study and the model structure are shown in Figures 4 and 5. EfficientNetB0 [25] was used as the basis for the severity classification model from PH-ECG. Unless otherwise noted, hyperparameters and the network architecture conform to the original EfficientNetB0. All input images were resized to  $224 \times 224$  pixels using a linear interpolation algorithm. The mini-batch size was set to 512 and the epochs to 100. The Adam optimizer [26] was selected as the optimization method. The Adam optimizer has four hyperparameters, all of which are considered important [27]. The search was therefore conducted using Bayesian optimization, a method of searching for the optimal solution of a target function from an unknown space using a prior distribution of Gaussian processes. Four search iterations were conducted for the following hyperparameters:  $\beta_1$  and  $\beta_2$ , which represent the exponential decay rates used for moment estimation;  $\epsilon$ , an offset to prevent zero division; and  $\alpha$ , the initial learning rate. The ranges for these values were set as 0.8 to 0.99, 0.9 to 0.999,  $1 \times 10^{-9}$  to  $1 \times 10^{-7}$ , and  $1 \times 10^{-6}$  to  $1 \times 10^{-2}$ , respectively. After 50 iterations of search, the resulting estimated optimal solutions were 0.97167, 0.97262,

$2.9343 \times 10^{-8}$ , and 0.0010413, respectively. However, over the 50 searches, no solution exceeded the results obtained when the Adam optimizer parameters were set to the default values of  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ ,  $\epsilon = 1 \times 10^{-8}$ , and  $\alpha = 0.001$ . Model depth, width, and input resolution are also hyperparameters that affect learning, but were fixed to the initial values of EfficientNetB0 and were not explored. This is because these values are correlated and width and depth are already appropriately adjusted according to the input resolution at the baseline of Tan et al. [25].

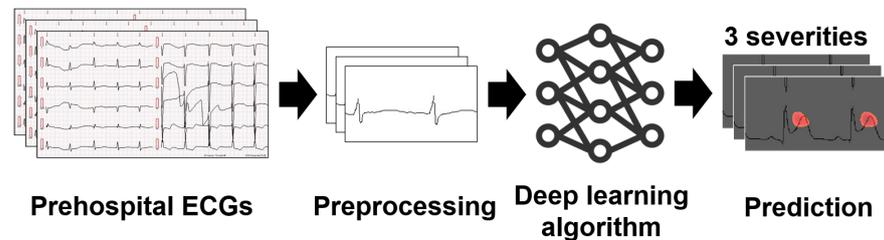


Figure 4. Conceptual figure of this study.

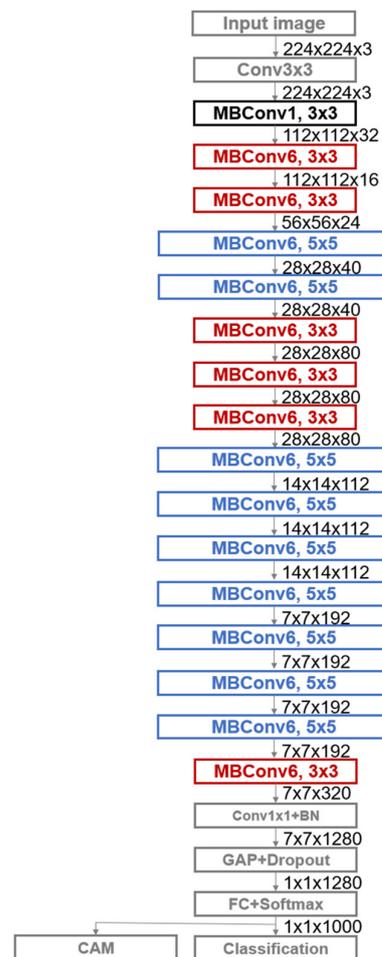
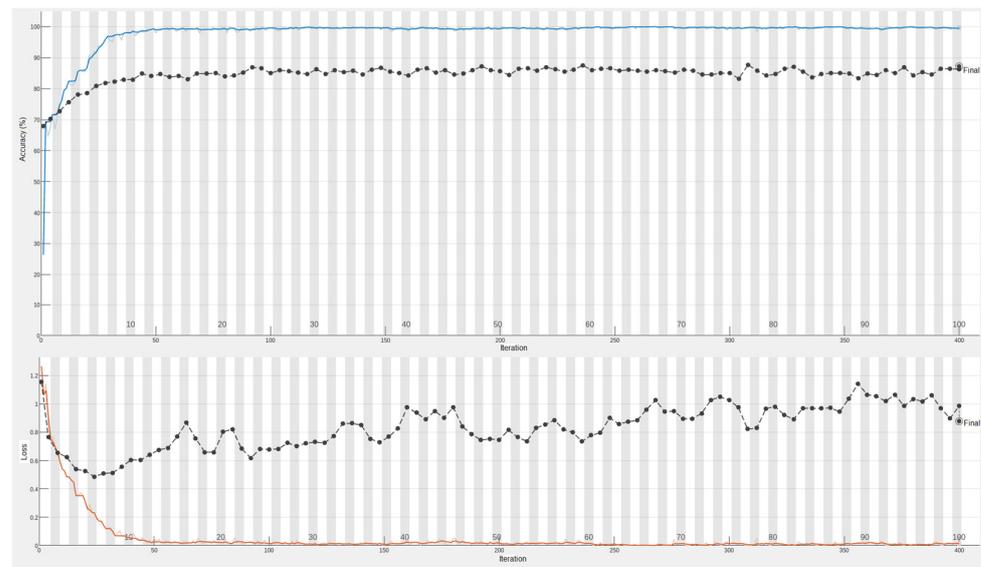


Figure 5. Adapted with permission from [25], arXiv, 2019. The structure of the convolutional neural network used is based on EfficientNet. The MBConv block refers to the Inverted Residual Blocks with an added Squeeze-and-Excitation module.

Input ECG images were feature-extracted by a convolution layer responsible for local feature extraction of the image and a pooling layer that summarizes the features for each locality. The batch normalization layer eliminated differences in distribution between

layers while maintaining sample distribution characteristics. Global average pooling (GAP) calculated the average value in the image space direction for each channel of the feature map, and the average value was used as the value for each feature map. Dropout (0.25) was performed just before the fully connected layer to suppress overfitting of the model. The final output layer was activated using a SoftMax function that provides probabilities. To account for class imbalances in the dataset, we gave the loss function for each class an inverse class frequency weight.

As a result of training, validation loss was minimized at Epoch 7, and validation accuracy continued to level off from around Epoch 15, as shown in Figure 6. Even with dropout, the loss function tended to overfit. This may be due to the paucity of data.



**Figure 6.** Learning curve for the best fold. Blue represents accuracy; red represents loss, and black represents validation.

### 2.3. Visualization of Features Identified by Neural Networks

To visualize what the proposed model sees, a class activation map (CAM) [28] of the final convolution layer was generated. A heat map was outputted showing the areas the model focused on. If the index of the fully connected layer is  $k$ , the weight connecting the GAP layer and output layer is  $\omega_k^c$ , and the output corresponding to the  $(x, y)$  coordinates of the  $k$ th feature map is  $f_k(x, y)$ . The output  $S_c$  for class  $c$  is calculated as follows:

$$S_c = \sum_k \omega_k^c \sum_{x, y} f_k(x, y) = \sum_{x, y} \sum_k \omega_k^c f_k(x, y)$$

The equation was generalized to eliminate model constraints (Grad-CAM).

## 3. Results

Four classifiers were created using the proposed network, and ECG images from the test dataset were classified. A summary of each classifier is shown in Table 5. The class breakdown of the test data is shown in Table 6. Accuracy, kappa coefficient, F1-score, and weighted F1-score resulting from inference on the test data are shown in Table 7 and Figure 7. The definition of accuracy is given in Equation (1) [29].

$$Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + True\ Negative + False\ Positive + False\ Negative} \quad (1)$$

**Table 5.** Overview of each model.

Models	Bayesian Optimization	Classes
1	No	Normal/mild or moderate/severe
2	Yes	Normal/mild or moderate/severe
3	No	Normal/STEMI
4	No	Normal/mild or moderate/severe

**Table 6.** Number of test data by class.

Class	n
Normal	489
Mild/moderate	99
Severe	128

**Table 7.** Prediction performance comparison.

	Fold	Accuracy	Kappa	F1-Score	Weighted F1-Score
Model 1	1	0.86	0.69	0.78	0.85
	2	0.85	0.67	0.76	0.84
	3	0.85	0.67	0.78	0.85
	4	0.86	0.70	0.79	0.86
	5	0.84	0.67	0.77	0.84
	Avg.	0.85	0.68	0.78	0.85
Model 2	1	0.85	0.67	0.77	0.84
	2	0.85	0.69	0.79	0.85
	3	0.86	0.70	0.80	0.86
	4	0.85	0.68	0.77	0.85
	5	0.85	0.68	0.78	0.84
	Avg.	0.85	0.68	0.78	0.85
Model 3	1	0.90	0.71	0.85	0.90
	2	0.92	0.78	0.89	0.92
	3	0.91	0.73	0.86	0.91
	4	0.91	0.72	0.86	0.91
	5	0.91	0.71	0.86	0.90
	Avg.	0.91	0.73	0.86	0.91
Model 4	1	0.84	0.66	0.76	0.84
	2	0.82	0.63	0.73	0.82
	3	0.83	0.64	0.75	0.83
	4	0.84	0.66	0.77	0.84
	5	0.82	0.63	0.75	0.82
	Avg.	0.83	0.65	0.75	0.83

Precision and recall are shown in Table 8. The confusion matrices are shown in Table 9. The ROC curves and AUCs for each model are shown in Figures 8 and 9.

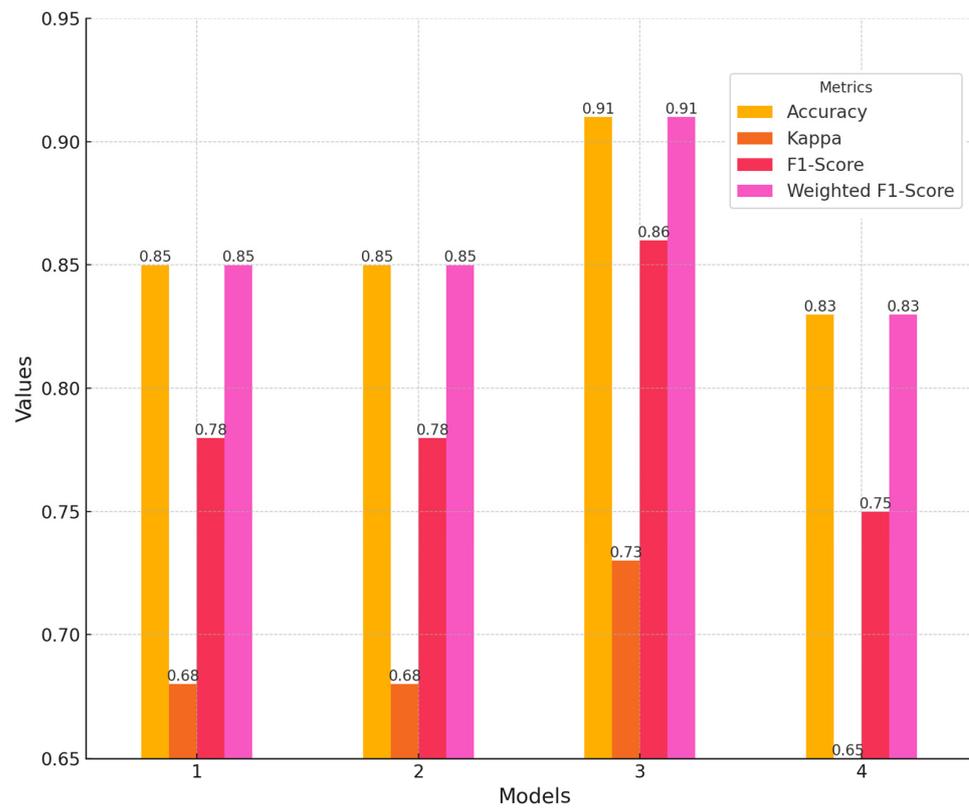


Figure 7. Prediction performance comparison.

Table 8. Precision and recall.

Classes	Precision	Recall
<b>Model 1</b>		
Normal	0.88	0.92
Mild/moderate	0.73	0.59
Severe	0.79	0.77
Avg.	0.80	0.80
Weighted avg.	0.85	0.85
<b>Model 2</b>		
Normal	0.89	0.92
Mild/moderate	0.79	0.62
Severe	0.79	0.77
Avg.	0.80	0.77
Weighted avg.	0.85	0.85
<b>Model 3</b>		
Normal	0.92	0.97
STEMI	0.88	0.71
Avg.	0.90	0.84
Weighted avg.	0.91	0.91
<b>Model 4</b>		
Normal	0.89	0.89
Mild/moderate	0.72	0.56
Severe	0.69	0.81
Avg.	0.77	0.75
Weighted avg.	0.83	0.83

**Table 9.** Confusion matrix.

<b>Model 1</b>		<b>Confusion Matrix</b>			
Fold 1		Predicted			
Actual	Normal	452	Mild/moderate	33	29
	Mild/moderate	15		60	6
	Severe	22		6	93
Fold 2		Predicted			
Actual	Normal	445	Mild/moderate	32	19
	Mild/moderate	18		63	6
	Severe	26		4	103
Fold 3		Predicted			
Actual	Normal	455	Mild/moderate	34	26
	Mild/moderate	14		64	6
	Severe	20		1	96
Fold 4		Predicted			
Actual	Normal	457	Mild/moderate	43	21
	Mild/moderate	14		53	7
	Severe	18		3	100
Fold 5		Predicted			
Actual	Normal	439	Mild/moderate	30	21
	Mild/moderate	22		65	7
	Severe	28		4	100
<b>Model 2</b>		<b>Confusion matrix</b>			
Fold 1		Predicted			
Actual	Normal	495	Mild/moderate	25	23
	Mild/moderate	21		86	8
	Severe	15		4	117
Fold 2		Predicted			
Actual	Normal	486	Mild/moderate	34	26
	Mild/moderate	28		77	6
	Severe	17		4	116
Fold 3		Predicted			
Actual	Normal	499	Mild/moderate	27	33
	Mild/moderate	25		83	1
	Severe	7		5	114
Fold 4		Predicted			
Actual	Normal	504	Mild/moderate	34	33
	Mild/moderate	23		79	5
	Severe	4		2	110

**Table 9.** *Cont.*

Fold 5		Predicted			
		Normal	Mild/moderate	Severe	
	Actual	Normal	489	29	38
		Mild/moderate	32	83	9
		Severe	10	3	101
<b>Model 3</b>		<b>Confusion matrix</b>			
Fold 1		Predicted			
		Normal	STEMI		
	Actual	Normal	283	26	
		STEMI	11	61	
Fold 2		Predicted			
		Normal	STEMI		
	Actual	Normal	282	17	
		STEMI	12	70	
Fold 3		Predicted			
		Normal	STEMI		
	Actual	Normal	290	29	
		STEMI	4	58	
Fold 4		Predicted			
		Normal	STEMI		
	Actual	Normal	282	23	
		STEMI	12	64	
Fold 5		Predicted			
		Normal	STEMI		
	Actual	Normal	288	29	
		STEMI	6	58	
<b>Model 4</b>		<b>Confusion matrix</b>			
Fold 1		Predicted			
		Normal	Mild/moderate	Severe	
	Actual	Normal	445	39	18
		Mild/moderate	12	52	5
		Severe	32	8	105
Fold 2		Predicted			
		Normal	Mild/moderate	Severe	
	Actual	Normal	435	30	17
		Mild/moderate	16	49	6
		Severe	38	20	105
Fold 3		Predicted			
		Normal	Mild/moderate	Severe	
	Actual	Normal	443	38	24
		Mild/moderate	17	52	3
		Severe	29	9	101
Fold 4		Predicted			
		Normal	Mild/moderate	Severe	
	Actual	Normal	439	36	18
		Mild/moderate	11	57	6
		Severe	39	6	104

Table 9. Cont.

Fold 5		Predicted					
		Normal	419	Mild/moderate	25	Severe	19
Actual	Normal						
	Mild/moderate	23		65		7	
	Severe	47		9		102	

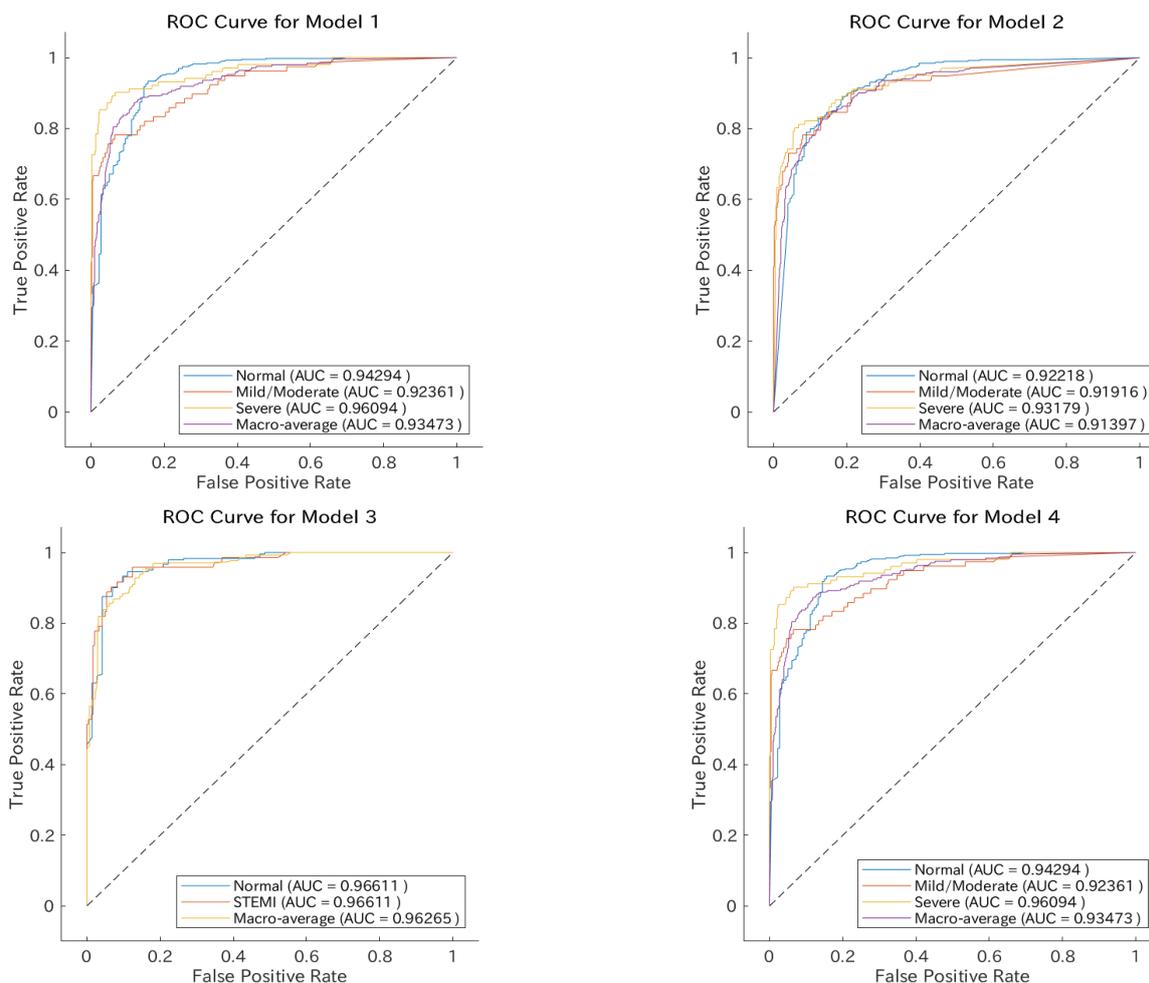
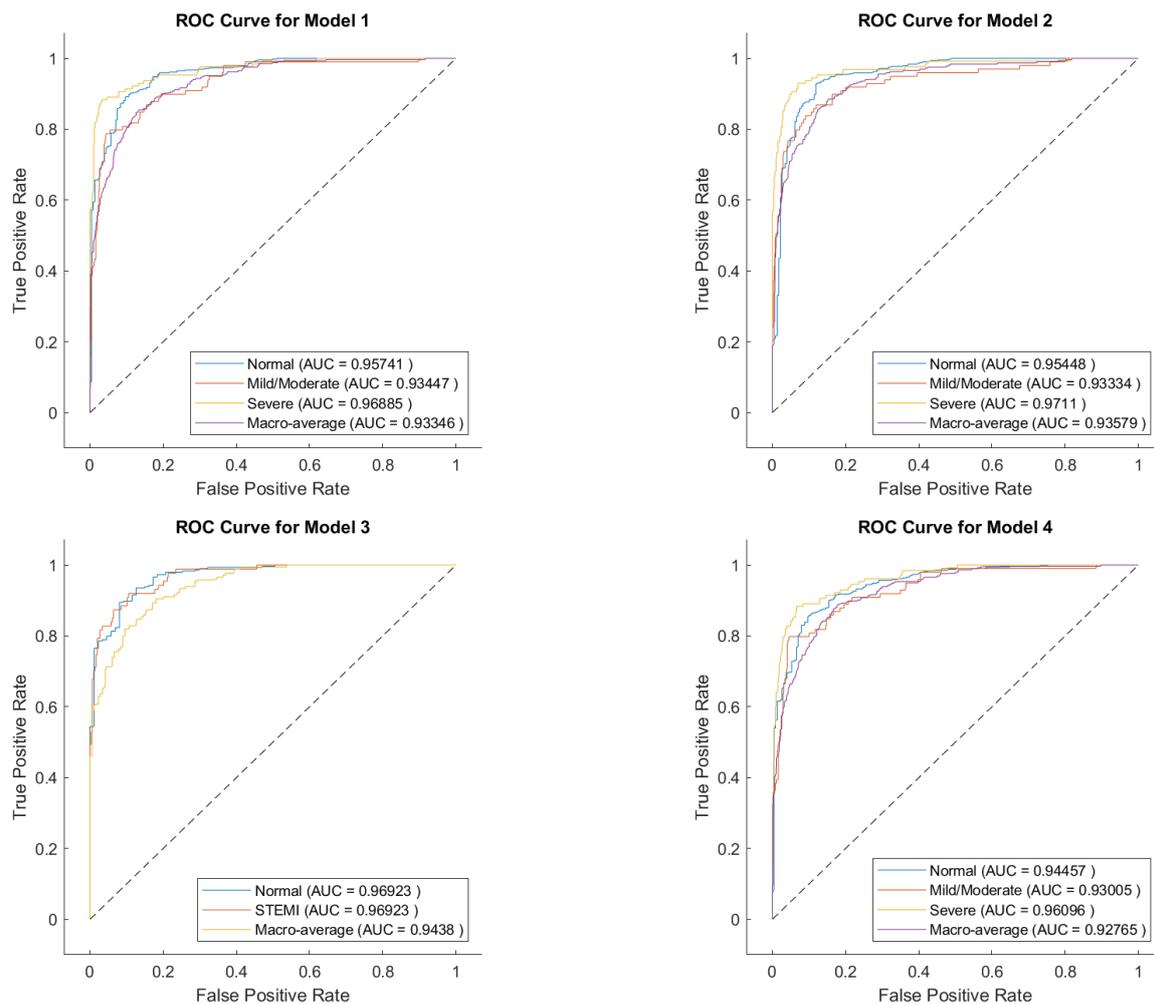


Figure 8. ROC curves at the time of validation for each model. Models 1, 2, and 4: blue line—ROC curve for the normal class; orange line—ROC curve for the mild/moderate class; yellow line—ROC curve for the severe class; and purple line—macro average. Model 3: blue line—ROC curve for the normal class; orange line—ROC curve for the STEMI class; and yellow line—macro average.

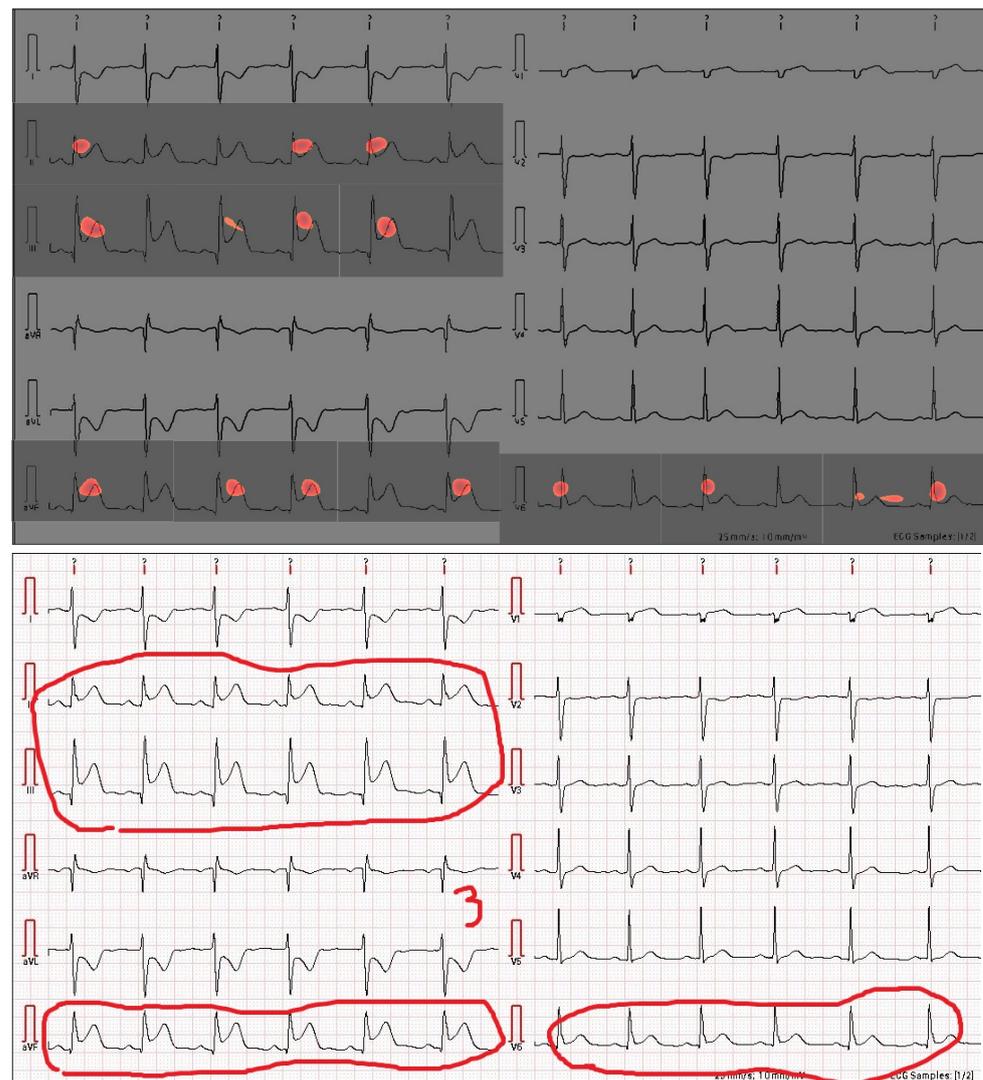
Model 1 recorded a weighted F1-score of 0.85 and an AUC of 0.93. The kappa coefficient was 0.68. The kappa coefficient is a statistic that expresses the degree of agreement between observations of a phenomenon by two different observers. This was within the range of 0.61–0.80, which is considered “substantial” under the guidelines provided by Landis and Koch [30]. Model 2 explored the Adam optimizer hyperparameters based on Bayesian optimization, but did not achieve greater accuracy than Model 1 on any of the evaluation measures.



**Figure 9.** ROC curves at the time of testing for each model. Models 1, 2, and 4: blue line—ROC curve for the normal class; orange line—ROC curve for the mild/moderate class; yellow line—ROC curve for the severe class; purple line—Macro average. Model 3: blue line—ROC curve for the normal class; orange line—ROC curve for the STEMI class; and yellow line—macro average.

Because it is important not to miss STEMI, which represents the most lethal form of acute coronary syndrome (ACS), we created Model 3, a binary normal/STEMI classification, using normal and STEMI cases in the dataset. Therefore, Model 3 is unable to perform a three-class classification of severity. Then, based on the output of Model 1, Model 4 was created to classify test data determined to be STEMI in Model 3 as severe. Compared to Model 1, Model 4 decreased precision for the severe class by 0.10 but improved recall by 0.04. In general, there is a trade-off between recall and precision, and since recall is more important than precision in the diagnosis of the severe class, Model 4, which compensates for the shortcomings of Model 1, has room for adoption. The lower F1-scores for atypical syndromes in Models 1, 2, and 4 compared to Model 3 may be attributed to an insufficient number of data points for certain abnormalities. Our study had limitations, and apart from STEMI, we were unable to access the exact number of abnormalities, which may have resulted in data imbalances.

The results obtained by applying CAM to Model 1, which had the highest AUC, are shown in Figure 10.

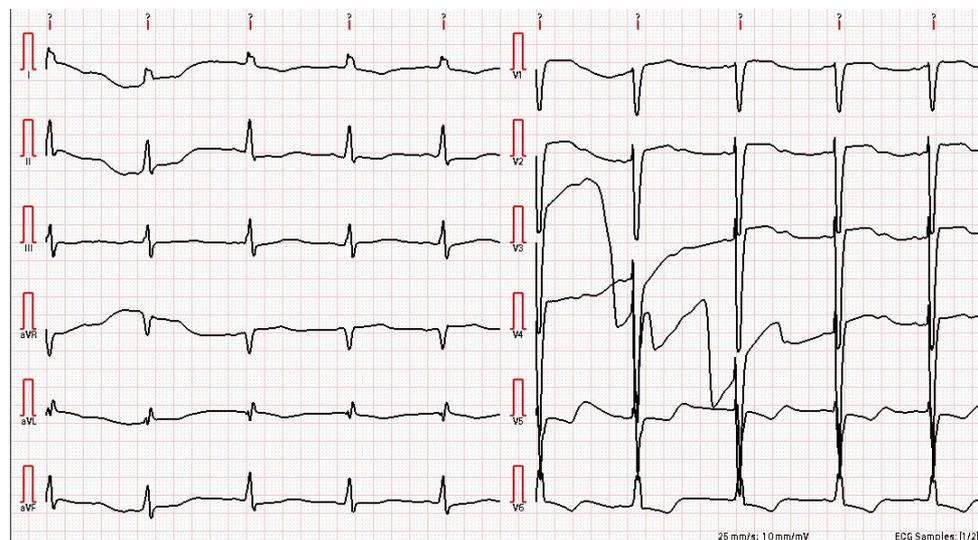


**Figure 10.** Prediction using a class activation map (CAM) applied to a case of myocardial infarction, and the correct labels. In the prediction, severe cases are colored red. In the correct labels, the physician has circled areas of abnormality with red lines. In this case, the focus is on the ST segment. Although an elevated ST segment is not definitive of myocardial infarction [31], it is suggested that this visualization may assist the EMTs in determining the PH-ECG transmission.

#### 4. Discussion

The severity classification of ECGs from ambulances using real-world data shows various limitations. Compared to ECGs performed in the hospital, the accumulated data were small. For example, the STEMI rate was only 12.5% (447/3564). The classification of severity was experimental in order to allow for some lump-sum weighting correction of such data biases, and there is room for improvement in how to select abnormal findings (i.e., the “Severe” class) that should be preferentially detected. Additionally, the ECG data from our study population were transmitted and stored as images rather than as actual waveforms. The ability to remove noise depends on the hardware, and in our case, it was difficult to remove noise that may have been caused by patient body movement or electrode dropout (Figure 11). Models thus needed to be constructed with the noise that could not be removed from ECG data. In Japan, prehospital 12-lead ECG is not routinely performed by emergency medical service personnel at first medical contact sites [32], making a large dataset difficult to construct. The performance of the classification model created in this study thus tended to show low recall for classes with small datasets. In particular, the

amount of data required for feature training is considered to be large for cases that change over time, such as STEMI, and we speculate that classification accuracy can be markedly improved by increasing the amount of training data.



**Figure 11.** Cases with noise. V3 and V4 inductions show abnormalities, most likely due to ambulance vibrations or patient body movements, which may have caused electrode misalignment or dropout.

We next discuss the trade-offs between precision and recall. Model 4, which is a composite model of Model 1 and Model 3 and prioritizes not missing STEMI cases, had the lowest precision and AUC but the highest recall. Considering the importance of minimizing false negatives in medical situations, Model 4 can be considered superior to Models 1 and 2.

Diagnosis using PH-ECG images is worthwhile because ECGs performed in the field by emergency personnel often show earlier ECG changes in cardiac disease and different characteristics than in-hospital ECGs. A small number of machine learning-based prediction algorithms using PH-ECG transmission have been reported (Table 10). However, most of these are waveform-based algorithms, making them difficult to apply to image data. Al-Zaiti et al. predicted ACS for PH-ECG signals in 1244 Americans and showed an AUC of 0.82 with a combination of logistic regression, gradient boosting machine, and artificial neural network [33]. Chen et al. performed STEMI prediction on 2907 PH-ECG signals obtained from ambulances in central Taiwan and showed an AUC of 0.997 using a combination of a 1D-convolutional neural network and long short-term memory [34]. This method is highly effective when waveform data are accessible, but it cannot be applied to our current data or to many hospitals in our region. Also, based on the differences in reference values of ECGs between races shown by Simonson [35], we can infer that the use of race-specific models is preferable for AI analyses of ECGs, and only one study appears to have been conducted on Japanese subjects [36]. Takeda et al. [36] used 17 features including vital signs, three-lead ECG monitoring, and symptoms from 555 individuals obtained in an urban area of Japan to predict diagnoses and subcategories of ACS using a support vector machine, showing an AUC of 0.864.

To the best of our knowledge, no examples of severity classification of PH-ECG images from ambulances have been provided for Japanese populations. While the acquisition of signals directly from instruments is efficient for processing ECGs in a computer, PH-ECG data are scarce, and it can be difficult to collect large amounts of data in signal form alone. In addition, the ECG standards used vary among ECG device manufacturers of ECG devices, and many data formats do not provide interoperability [37]. The advantage of image-based analysis over digital signal-based analysis is the versatility since the method can be applied relatively easily, even to viewers from different manufacturers, as long as the recording speed and calibration are consistent.

**Table 10.** Comparison with previous studies.

	Target	Method	Race	Training Data	Test Data	Input	AUC (95% CI)
Ours	Severity classification	DL	Japanese	2848	716	1-lead PH-ECG images	0.933 [0.915–0.951]
	Prediction of STEMI	DL	Japanese	1563	381	1-lead PH-ECG images	0.943 [0.920–0.966]
Al-Zaiti et al. [33] (2020)	Prediction of ACS	ML	American	745	499	12-lead PH-ECG signals	0.82 [0.77–0.86]
	Prediction of NSTEMI-ACS	ML	American	745	499	12-lead PH-ECG signals	0.78 [0.73–0.84]
Chen, K.-W. et al. [34] (2022)	Prediction of STEMI	DL	(Data acquired in Taiwan)	2907	362	12-lead PH-ECG signals	0.997
M. Takeda et al. [36] (2022)	Prediction of ACS	ML	Japanese	555	61	Vital signs 3-lead ECG monitoring 43 symptoms	0.839 [0.734–0.931]
	Prediction of AMI	ML	Japanese	555	61	Vital signs 3-lead ECG monitoring 17 symptoms	0.850 [0.817–0.882]
	Prediction of STEMI	ML	Japanese	555	61	Vital signs 3-lead ECG monitoring 17 symptoms	0.862 [0.831–0.894]

Interpretability is an issue when using the developed models in medical practice [38]. Since the days when rule-based analysis was mainstream, reports have found that automated analyses of ECGs by computer should be used as an adjunct [39,40], and the difficulty in explaining model results from machine learning-based analyses, which have a more black box nature than rule-based analyses, has hindered confidence. Various reports have described CAM feature visualization in hospital ECG signals, with promising results [41], but further research is needed to determine if the same holds true for PH-ECG images.

## 5. Limitations

There are several constraints we have not yet discussed. Due to restrictions on sensitive information and data accessibility of the study subjects, detailed patient demographics and comorbidities could not be ascertained. Analysis at this stage also requires data acquired in a uniform environment, including the same ECG device and its settings. Future research needs to validate our findings in diverse real-world settings. Additionally, because of the limited sample size, further studies are necessary to establish the reliability and clinical utility of our method.

## 6. Conclusions

This study suggests that image-based anomaly detection from PH-ECGs is feasible and effective, particularly in regions like Japan where ECG data are often stored and transmitted as images.

In our region, EMTs decide whether to transmit ECGs according to the flow shown in Figure 1. If we modify this flow so that ECGs are transmitted whenever either the EMTs or the deep learning models detect an anomaly, it is expected to reduce reading time and door-to-balloon time, as well as decrease false negatives.

However, despite these promising results, we faced significant limitations related to noise removal from ECG data. Noise is a critical factor in accurately assessing heart problems, and our model's performance is influenced by the quality of the input data. Our model theoretically accounts for noise caused by patient movement or electrode detach-

ment, but unlike waveform-based methods, it has difficulty removing noise embedded in the images.

In conclusion, while deep learning models hold promise for clinical application in PH-ECG screening, further research is needed to verify the impact of noise and ensure the model's reliability and clinical utility.

**Author Contributions:** Conceptualization, A.D., T.I., T.S. and O.N.; data curation, R.O., A.D., T.I., T.S. and O.N.; formal analysis, R.O., A.D., T.I., T.S. and O.N.; funding acquisition, A.D. and T.I.; investigation, R.O., A.D., T.I., T.S. and O.N.; methodology, R.O. and A.D.; project administration, A.D. and T.I.; resources, R.O., A.D., T.I., T.S. and O.N.; software, R.O. and A.D.; supervision, A.D.; validation, R.O. and A.D.; visualization, R.O. and A.D.; writing—original draft preparation, R.O., A.D., T.I., T.S. and O.N.; writing—review and editing, R.O. and A.D. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by a Grant-in-Aid for Scientific Research (C) from the Japan Society for the Promotion of Science (grant no. JP20K08142) as well as by the “FY2022 Research and Development Subsidy Program” of the Japan Keirin Autorace (JKA) Foundation (grant no. 259) and the “FY2020 Iwate Strategic Research and Development Promotion Program” of Iwate Prefecture (grant no. 3). We would like to express our gratitude to these organizations.

**Institutional Review Board Statement:** This study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Iwate Medical University Ethics Committee (MH2020-133, 13 November 2020).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in this study.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

- Martinoni, A.; De Servi, S.; Boschetti, E.; Zanini, R.; Palmerini, T.; Politi, A.; Musumeci, G.; Belli, G.; De Paolis, M.; Etori, F.; et al. Importance and limits of pre-hospital electrocardiogram in patients with ST elevation myocardial infarction undergoing percutaneous coronary angioplasty. *Eur. J. Cardiovasc. Prev. Rehabil.* **2011**, *18*, 526–532. [[CrossRef](#)] [[PubMed](#)]
- Nam, J.; Caners, K.; Bowen, J.M.; Welsford, M.; O'Reilly, D. Systematic review and meta-analysis of the benefits of out-of-hospital 12-lead ECG and advance notification in ST-segment elevation myocardial infarction patients. *Ann. Emerg. Med.* **2014**, *64*, 176–186.e9. [[CrossRef](#)] [[PubMed](#)]
- Brunetti, N.D.; Di Pietro, G.; Aquilino, A.; I Bruno, A.; Dellegrottaglie, G.; Di Giuseppe, G.; Lopriore, C.; De Gennaro, L.; Lanzone, S.; Caldarola, P.; et al. Pre-hospital electrocardiogram triage with tele-cardiology support is associated with shorter time-to-balloon and higher rates of timely reperfusion even in rural areas: Data from the Bari- Barletta/Andria/Trani public emergency medical service 118 registry on primary angioplasty in ST-elevation myocardial infarction. *Eur. Heart J. Acute Cardiovasc. Care* **2014**, *3*, 204–213. [[CrossRef](#)] [[PubMed](#)]
- Quinn, T.; Johnsen, S.; Gale, C.P.; Snooks, H.; McLean, S.; Woollard, M.; Weston, C.; On behalf of the Myocardial Ischaemia National Audit Project (MINAP) Steering Group. Effects of prehospital 12-lead ECG on processes of care and mortality in acute coronary syndrome: A linked cohort study from the Myocardial Ischaemia National Audit Project. *Heart* **2014**, *100*, 944–950. [[CrossRef](#)]
- Ownbey, M.; Suffoletto, B.; Frisch, A.; Guyette, F.X.; Martin-Gill, C. Prevalence and interventional outcomes of patients with resolution of ST-segment elevation between prehospital and in-hospital ECG. *Prehospital Emerg. Care* **2014**, *18*, 174–179. [[CrossRef](#)]
- Bouza, Z.; Faramand, Z.; Martin-Gill, C.; Sereika, S.M.; Callaway, C.W.; Saba, S.; Gregg, R.; Badilini, F.; Sejdic, E.; Al-Zaiti, S.S. Incorporation of Serial 12-Lead Electrocardiogram with Machine Learning to Augment the Out-of-Hospital Diagnosis of Non-ST Elevation Acute Coronary Syndrome. *Ann. Emerg. Med.* **2023**, *81*, 57–69. [[CrossRef](#)]
- JRC Resuscitation Guidelines 2015*; Resuscitation Council: Tokyo, Japan, 2016. (In Japanese)
- O'Gara, P.T.; Kushner, F.G.; Ascheim, D.D.; Casey, D.E.; Chung, M.K.; De Lemos, J.A.; Ettinger, S.M.; Fang, J.C.; Fesmire, F.M.; Franklin, B.A. 2013 ACCF/AHA guideline for the management of ST-elevation myocardial infarction: A report of the American College of Cardiology Foundation/American Heart Association task force on practice guidelines. *J. Am. Coll. Cardiol.* **2013**, *61*, e78–e140. [[CrossRef](#)]
- Otani, H.; Tanaka, H.; Maki, A.; Takyu, H.; Harikae, K.; Ueta, H.; Sone, E.; Sagisaka, R. The utilization of pre-hospital 12 lead electrocardiogram by emergency life-saving technicians and its education. *J. Jpn. Soc. Emerg. Med.* **2017**, *20*, 703–711. (In Japanese)
- Ogita, M.; Suwa, S.; Ebina, H.; Nakao, K.; Ozaki, Y.; Kimura, K.; Ako, J.; Noguchi, T.; Yasuda, S.; Fujimoto, K.; et al. Off-hours presentation does not affect in-hospital mortality of Japanese patients with acute myocardial infarction: J-MINUET substudy. *J. Cardiol.* **2017**, *70*, 553–558. [[CrossRef](#)]

11. Sakai, T.; Nishiyama, O.; Onodera, M.; Matsuda, S.; Wakisawa, S.; Nakamura, M.; Morino, Y.; Itoh, T. Predictive ability and efficacy for shortening door-to-balloon time of a new prehospital electrocardiogram-transmission flow chart in patients with ST-elevation myocardial infarction—Results of the CASSIOPEIA study. *J. Cardiol.* **2018**, *72*, 335–342. [CrossRef]
12. Feldman, J.A.; Brinsfield, K.; Bernard, S.; White, D.; Maciejko, T. Real-time paramedic compared with blinded physician identification of ST-segment elevation myocardial infarction: Results of an observational study. *Am. J. Emerg. Med.* **2005**, *23*, 443–448. [CrossRef] [PubMed]
13. Acharya, U.R.; Fujita, H.; Oh, S.L.; Hagiwara, Y.; Tan, J.H.; Adam, M. Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals. *Inf. Sci.* **2017**, *415*, 190–198. [CrossRef]
14. Wu, C.-C.; Hsu, W.-D.; Islam, M.; Poly, T.N.; Yang, H.-C.; Nguyen, P.-A.; Wang, Y.-C.; Li, Y.-C. An artificial intelligence approach to early predict non-ST-elevation myocardial infarction patients with chest pain. *Comput. Methods Programs Biomed.* **2019**, *173*, 109–117. [CrossRef]
15. Ahmed, A.A.; Ali, W.; Abdullah, T.A.A.; Malebary, S.J. Classifying Cardiac Arrhythmia from ECG Signal Using 1D CNN Deep Learning Model. *Mathematics* **2023**, *11*, 562. [CrossRef]
16. Jambukia, S.H.; Dabhi, V.K.; Prajapati, H.B. Classification of ECG signals using machine learning techniques: A survey. In Proceedings of the International Conference on Advances in Computer Engineering and Applications, Ghaziabad, India, 19–20 March 2015; pp. 714–721.
17. Moody, G.B.; Mark, R.G. The impact of the MIT-BIH arrhythmia database. *IEEE Eng. Med. Biol. Mag.* **2001**, *20*, 45–50. [CrossRef] [PubMed]
18. Boussejot, R.; Kreiseler, D.; Schnabel, A. Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das Internet. *Biomed. Eng./Biomed. Tech.* **1995**, *40* (Suppl. S1), 317–318. [CrossRef]
19. Fuji-no-kuni' Product Introduction Website. Available online: <https://www.goodcare.jp/product/fujinokuni/> (accessed on 13 March 2023).
20. Alickovic, E.; Subasi, A. Effect of Multiscale PCA De-noising in ECG Beat Classification for Diagnosis of Cardiovascular Diseases. *Circuits, Syst. Signal Process.* **2014**, *34*, 513–533. [CrossRef]
21. Sharma, R.R.; Pachori, R.B. Baseline wander and power line interference removal from ECG signals using eigenvalue decomposition. *Biomed. Signal Process. Control.* **2018**, *45*, 33–49. [CrossRef]
22. Van Alste, J.A.; Schilder, T.S. Removal of base-line wander and power-line interference from the ECG by an efficient FIR filter with a reduced number of taps. *IEEE Trans. Biomed. Eng.* **1985**, *32*, 1052–1060. [CrossRef]
23. Ji, T.; Lu, Z.; Wu, Q.; Ji, Z. Baseline normalisation of ECG signals using empirical mode decomposition and mathematical morphology. *Electron. Lett.* **2008**, *44*, 82–84. [CrossRef]
24. Kimura, K.; Kimura, T.; Ishihara, M.; Nakagawa, Y.; Nakao, K.; Miyauchi, K.; Sakamoto, T.; Tsujita, K.; Hagiwara, N.; Miyazaki, S.; et al. JCS 2018 Guideline on Diagnosis and Treatment of Acute Coronary Syndrome. *Circ. J.* **2019**, *83*, 1085–1196. [CrossRef]
25. Tan, Q.L.M. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In Proceedings of the 36th International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; Volume 97, pp. 6105–6114.
26. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.
27. Choi, D.; Shallue, C.J.; Nado, Z.; Lee, J.; Maddison, C.J.; Dahl, G.E. On Empirical Comparisons of Optimizers for Deep Learning. *arXiv* **2019**, arXiv:1910.05446.
28. Zhou, B.; Khosla, A.; Lapedriza, A.; Oliva, A.; Torralba, A. Learning deep features for discriminative localization. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2921–2929.
29. Cleverdon, C.; Mills, J.; Keen, M. *Aslib Cranfield Research Project—Factors Determining the Performance of Indexing Systems*; Cranfield: Bedford, UK, 1966; Volume 1.
30. Landis, J.R.; Koch, G.G. The measurement of observer agreement for categorical data. *Biometrics* **1977**, *33*, 159–174. [CrossRef]
31. Hygesen, K.; Alpert, J.S.; Jaffe, A.S.; Chaitman, B.R.; Bax, J.J.; Morrow, D.A.; White, H.D. Fourth universal definition of myocardial infarction. *Rev. Esp. Cardiol.* **2019**, *14*, 72.
32. Mori, H.; Maeda, A.; Akashi, Y.; Ako, J.; Ikari, Y.; Ebina, T.; Tamura, K.; Namiki, A.; Fukui, K.; Michishita, I.; et al. The impact of pre-hospital 12-lead electrocardiogram and first contact by cardiologist in patients with ST-elevation myocardial infarction in Kanagawa, Japan. *J. Cardiol.* **2021**, *78*, 183–192. [CrossRef]
33. Al-Zaiti, S.; Besomi, L.; Bouzid, Z.; Faramand, Z.; Frisch, S.; Martin-Gill, C.; Gregg, R.; Saba, S.; Callaway, C.; Sejdić, E. Machine learning-based prediction of acute coronary syndrome using only the pre-hospital 12-lead electrocardiogram. *Nat. Commun.* **2020**, *11*, 3966. [CrossRef]
34. Chen, K.W.; Wang, Y.C.; Liu, M.H.; Tsai, B.Y.; Wu, M.Y.; Hsieh, P.H.; Wei, J.T.; Shih, E.S.; Shiao, Y.T.; Hwang, M.J. Artificial intelligence-assisted remote detection of ST-elevation myocardial infarction using a mini-12-lead electrocardiogram device in prehospital ambulance care. *Front. Cardiovasc. Med.* **2022**, *9*, 1001982.
35. Simonson, E. Differentiation between Normal and Abnormal in Electrocardiography. *Acad. Med.* **1962**, *37*, 161.
36. Takeda, M.; Oami, T.; Hayashi, Y.; Shimada, T.; Hattori, N.; Tateishi, K.; Miura, R.E.; Yamao, Y.; Abe, R.; Kobayashi, Y.; et al. Prehospital diagnostic algorithm for acute coronary syndrome using machine learning: A prospective observational study. *Sci. Rep.* **2022**, *12*, 14593. [CrossRef] [PubMed]

37. Cuevas-González, D.; García-Vázquez, J.P.; Bravo-Zanoguera, M.; López-Avitia, R.; Reyna, M.A.; Zermeño-Campos, N.A.; González-Ramírez, M.L. ECG Standards and Formats for Interoperability between mHealth and Healthcare Information Systems: A Scoping Review. *Int. J. Environ. Res. Public Health* **2022**, *19*, 11941. [[CrossRef](#)] [[PubMed](#)]
38. Schläpfer, J.; Wellens, H.J. Computer-Interpreted Electrocardiograms: Benefits and Limitations. *J. Am. Coll. Cardiol.* **2017**, *70*, 1183–1192. [[CrossRef](#)] [[PubMed](#)]
39. De Champlain, F.; Boothroyd, L.J.; Vadeboncoeur, A.; Huynh, T.; Nguyen, V.; Eisenberg, M.J.; Joseph, L.; Boivin, J.F.; Segal, E. Computerized interpretation of the prehospital electrocardiogram: Predictive value for ST segment elevation myocardial infarction and impact on on-scene time. *Can. J. Emerg. Med.* **2014**, *16*, 94–105. [[CrossRef](#)] [[PubMed](#)]
40. Bhalla, M.C.; Mencl, F.; Gist, M.A.; Wilber, S.; Zalewski, J. Prehospital electrocardiographic computer identification of ST-segment elevation myocardial infarction. *Prehospital Emerg. Care* **2012**, *17*, 211–216. [[CrossRef](#)] [[PubMed](#)]
41. Mehta, S.; Fernandez, F.; Villagrán, C.; Niklitschek, S.; Ávila, J.; Botelho, R.; Frauenfelder, A.; Vieira, D.; Ceschim, M.; Merchant, S.; et al. Applicability of novel, class activation maps (CAM) in the development of artificial intelligence-guided, single and 12-lead ECG to detect ST-elevation myocardial infarction. *J. Am. Coll. Cardiol.* **2020**, *75*, 3474. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.