

Deep Learning-Based Electrocardiogram Signal Noise Detection and Screening Model

Dukyong Yoon^{1,2}, Hong Seok Lim³, Kyoungwon Jung⁴, Tae Young Kim¹, Sukhoon Lee⁵

¹Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, Korea

²Department of Biomedical Sciences, Graduate School of Medicine, Ajou University, Suwon, Korea

³Department of Cardiology, Ajou University School of Medicine, Suwon, Korea

⁴Division of Trauma Surgery, Department of Surgery, Ajou University School of Medicine, Suwon, Korea

⁵Department of Software Convergence Engineering, College of Convergence Engineering, Kunsan National University, Gunsan, Korea

Objectives: Biosignal data captured by patient monitoring systems could provide key evidence for detecting or predicting critical clinical events; however, noise in these data hinders their use. Because deep learning algorithms can extract features without human annotation, this study hypothesized that they could be used to screen unacceptable electrocardiograms (ECGs) that include noise. To test that, a deep learning-based model for unacceptable ECG screening was developed, and its screening results were compared with the interpretations of a medical expert. **Methods:** To develop and apply the screening model, we used a biosignal database comprising 165,142,920 ECG II (10-second lead II electrocardiogram) data gathered between August 31, 2016 and September 30, 2018 from a trauma intensive-care unit. Then, 2,700 and 300 ECGs (ratio of 9:1) were reviewed by a medical expert and used for 9-fold cross-validation (training and validation) and test datasets. A convolutional neural network-based model for unacceptable ECG screening was developed based on the training and validation datasets. The model exhibiting the lowest cross-validation loss was subsequently selected as the final model. Its performance was evaluated through comparison with a test dataset. **Results:** When the screening results of the proposed model were compared to the test dataset, the area under the receiver operating characteristic curve and the F1-score of the model were 0.93 and 0.80 (sensitivity = 0.88, specificity = 0.89, positive predictive value = 0.74, and negative predictive value = 0.96). **Conclusions:** The deep learning-based model developed in this study is capable of detecting and screening unacceptable ECGs efficiently.

Keywords: Electrocardiography, Noise, Deep Learning, Signal Detection Analysis, Physiologic Monitoring

Submitted: May 14, 2019

Revised: 1st, June 14, 2019; 2nd, July 8, 2019

Accepted: July 16, 2019

Corresponding Author

Dukyong Yoon

Department of Biomedical Informatics, Ajou University School of Medicine, 206 World cup-ro, Yeongtong-gu, Suwon 16499, Korea.
Tel: +82-31-219-4476, E-mail: d.yoon.ajou@gmail.com (<https://orcid.org/0000-0003-1635-8376>)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

© 2019 The Korean Society of Medical Informatics

1. Introduction

Electrocardiograms (ECGs) are commonly performed cardiology tests that record the electrical activity of the heart over a period using electrodes [1]. These electrodes detect small electrical changes caused by depolarization and repolarization in the electrophysiological pattern of the heart muscle during each heartbeat [1]. ECGs are widely used for various purposes, including measuring heart rate consistency, size, and location; identifying damage to the heart; and observing the effects of devices, such as pacemakers or heart-regulating medications. The information obtained via ECGs can also be used for medical diagnosis. ECGs represent the best method

for measuring and diagnosing abnormal heart rhythms [2]—an especially useful trait when applied to the measurement of damaged conduction tissues that transmit electrical signals [3]. ECGs can be used to detect damage to specific portions of the myocardium during myocardial infarctions [4]. Furthermore, digitally gathered and stored ECG data can be used for automatic ECG signal analysis [5].

ECG signals, however, are frequently interrupted by various types of noise and artifacts (Figure 1A–1C). Previous works have classified these into common artifact types [6–8], such as baseline wandering (BA), muscle artifacts (MA), and powerline interference (PLI). Subject movements or respiratory activities cause BA, which manifests as slowly wandering baselines primarily related to random body movements. ECGs with MAs are contaminated with muscular contraction artifacts. PLIs, caused by electrical power leakage or improper equipment grounding, are indicated by varying ECG amplitudes and indistinct isoelectric baselines. Because such noise or artifacts may lead to disturbances of further automatic ECG signal analysis, their detection and elimination is of great importance, as this could prevent ECG noise-related misclassifications or misdiagnosis.

Previous studies have attempted to de-noise ECG signals using a wide range of approaches, including wavelet transformation [9–11], weighted averages [12,13], adaptive filtering [14], independent component analysis [15], and empirical mode decomposition (EMD) [16–19]. However, existing methods have several noise-removal limitations [20]. For example, EMD-based approaches may filter out p- and t-waves. Adaptive filters, proposed by Rahman et al. [21], can apply filters such as the signed regression algorithm and normalized least-mean square, but they encounter difficulties obtaining noise-signal references from a typical ECG signal acquisition.

Recent ECG-related research has required a substantially different approach to noise because the scale of data collection has become very large. The Electrocardiogram Vigilance with Electronic data Warehouse II (ECG-VIEW II) released 979,273 ECG results from 461,178 patients, with plans to add all 12-lead data in the next version [22,23]. More recent attempts have, additionally, been made to acquire ECG measurements from patient monitoring equipment in intensive care units (ICUs) [24,25]. The MIMIC III (Medical Information Mart for Intensive Care) dataset contains 22,317 waveform records (ECG, arterial blood pressure [ABP] waveforms, fingertip photoplethysmogram [PPG] signals, and respiration). The biosignal repository used in this study collected 23,187,206 waveform records (like MIMIC III,

all kinds of waveforms captured in the ICU were included) from over 8,000 patients with an observational period of approximately 517 patient-years as of October 2018.

A different approach to noise was thus believed to be required. First, because there were sufficient data, researchers did not need to use de-noised data, which might still contain incorrect information (some real-world signals captured noise exclusively without any ECG information, as shown in Figure 1C). Second, a deep learning algorithm is needed. Deep learning algorithms possess multiple advantages. They do not require feature extraction processes performed by domain experts; the abovementioned biosignal repositories collected ECG alongside many other biosignal data types (such as respiration, ABP, PPG, and central venous pressure [CVP]), and because each type of waveform possesses unique characteristics, it requires a customized algorithm. Because a feature extraction process is unnecessary, it is

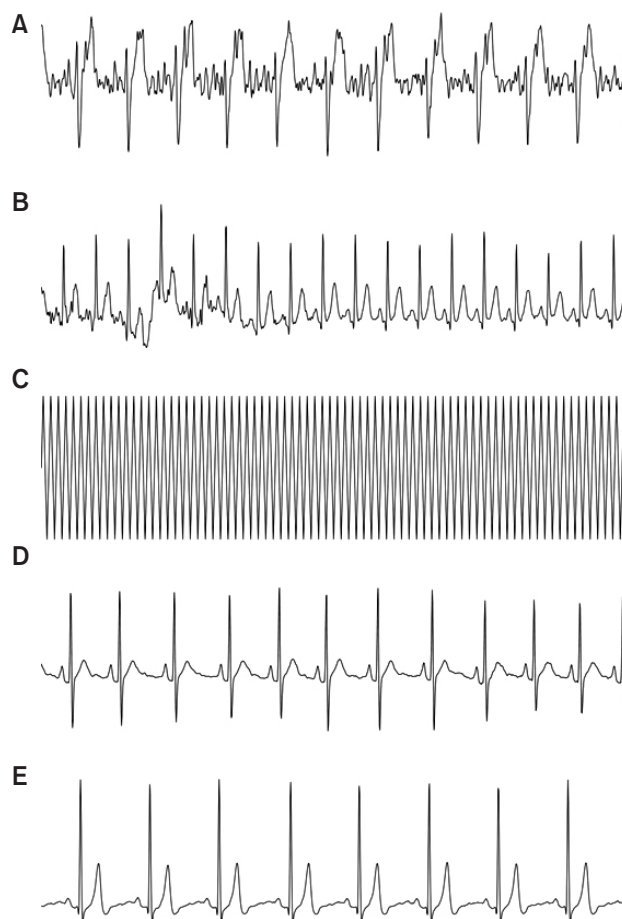


Figure 1. Acceptable and non-acceptable electrocardiogram (ECG) examples. Non-acceptable ECGs have global (A) or local (B) noise, or signals from ECG equipment rather than patients (C). (D) and (E) are examples of acceptable ECGs.

easier to develop a deep learning-based algorithm to screen unacceptable signals than to develop an algorithm for each signal.

In this context, a new unacceptable ECG (ECG with noise) detection and screening deep learning-based model for further automatic ECG signal analysis was developed in this study. In the development process, we minimized the manual review effort of the medical expert by pre-screening ECG data using non-experts.

II. Methods

Informed consent was waived in this study by the Ajou

University Hospital Institutional Review Board (No. AJIRB-MED-MDB-16-155). Only de-identified data were used and analyzed, retrospectively.

1. Data Source

The data used for this study were obtained from a biosignal repository constructed by us through our previous research [25] (Figure 2). From September 1, 2016 to September 30, 2018, the biosignal data collected from the trauma ICU of Ajou University Hospital comprised a total of 2,767,845 PPG, measuring peripheral oxygen saturation (SpO₂); 2,752,382 ECG lead II; 2,597,692 respiration data; 1,864,864 ABP; and 754,240 CVP (each indicating 10-minute data files). Because

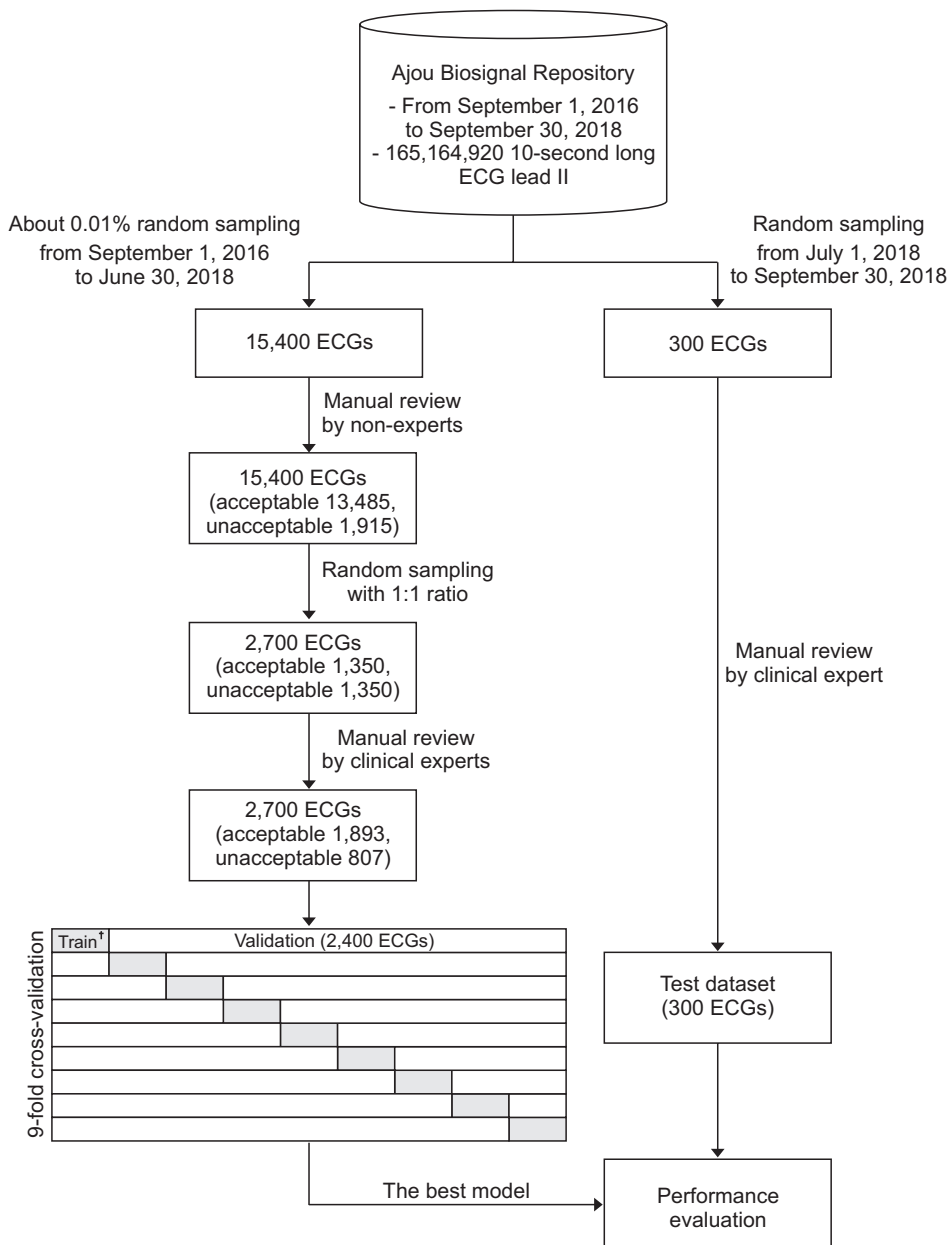


Figure 2. Study process flow. From 15,400 electrocardiograms (ECGs) reviewed by non-experts, 2,400 and 300 ECGs were confirmed by a medical expert and used as training and validation dataset. Three hundred ECGs independently gathered from different periods were reviewed by the expert and used to evaluate the performance of the developed deep learning-based model.

the data were collected from the trauma ICU, most patients were admitted because of physical trauma rather than any disease the patients had. Approximately 52.1% of patients in the trauma ICU experienced surgery to repair physical damage to their bodies. The average \pm standard deviation of the age of the patients was 50.7 ± 20.5 . Males constituted 73.0% of the patients.

2. Unacceptable ECG (ECG with Noise) Definition

The ECG waveforms were classified into two types: acceptable or unacceptable. An acceptable waveform was defined as a waveform in the normal category that is able to be used for further analysis. Unacceptable waveforms included the following subtypes: (1) BA, waveforms with variations in the measured voltage baseline; (2) MA, partial noise caused by patient movements or body instability; (3) PLI, noise generated across the entire waveform owing to close contact between the measurement sensor and voltage interference; (4) unacceptable (other reasons), waveforms not categorized as normal waveforms because of other causes; and (5) unclear, waveforms in which the preceding type-judgements were inappropriate. All waveforms that were not acceptable were classified as unacceptable (ECG with noise).

3. Labeling Tool Development

A web-based tool was developed to label the two types of ECG waveforms defined above. This facilitated rapid evaluation and efficient management in labeling the results of each ECG signal (Figure 3). The tool displayed a 10-second ECG result, allowing the evaluator to select one of the two types: acceptable or unacceptable. Using these tools, randomly selected ECGs were reviewed by two non-experts and a medical expert. A short pre-training (approximately 10 min-

utes) session was conducted for each non-expert evaluator. Labeling results were manually reviewed and corrected by a medical expert before they were used for model training (see Section II-4 and Figure 2).

4. Training, Validation, and Test Datasets

The datasets for model development training and validation were initially reviewed by two non-experts. If one evaluator classified an ECG as unacceptable, the ECG was considered unacceptable even if the other disagreed. From 15,400 labeled ECGs, 13,485 (87.6%) were classified as acceptable by the non-experts. Among these pre-screened ECGs, 2,700 were randomly sampled ECGs with a 50:50 balance ratio (acceptable:unacceptable) to adjust the imbalance between them in the real-world dataset. The sampled ECGs were confirmed via manual review by a medical expert. These ECGs were randomly divided into a training dataset (2,400 ECGs) and a validation dataset (300 ECGs) for 9-fold cross-validation (Figure 2). The test dataset (300 ECGs) was randomly sampled from ECGs gathered from various periods and data from patients whose data were included in the training or validation dataset were excluded (Figure 3). The data were also evaluated as acceptable or unacceptable by a medical expert using the tool described above. Finally, the dataset for 9-fold cross-validation (training and validation) and test datasets were generated in a ratio of 9:1.

5. Deep Learning Model Development

1) Model development

Waveform data representing 10-second ECGs were used as input because they were labeled for this timespan. For the original ECGs, there were 2,500 data points because they were measured at 250 Hz. Sampling was performed at a 50%

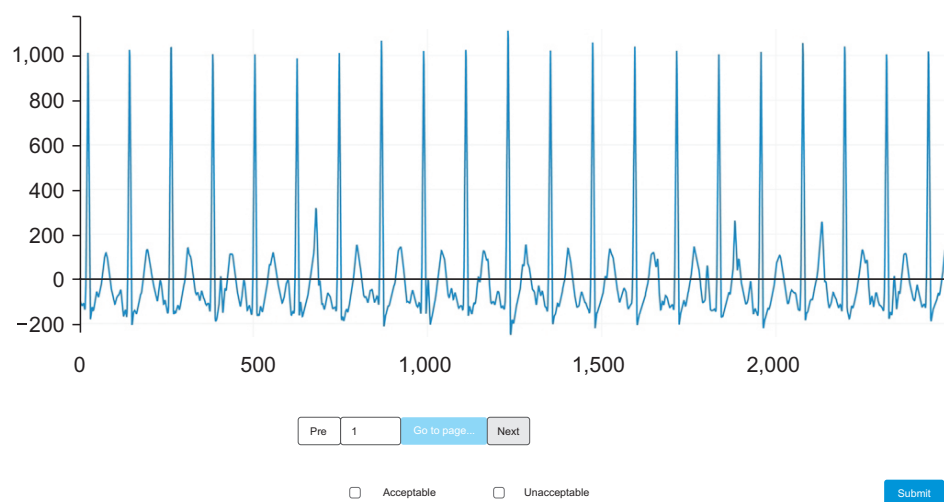


Figure 3. Web-based labeling tool. The tool presented 10-second electrocardiogram waveforms and asked evaluators if they were acceptable or unacceptable. All answers were recorded and managed in the backend database.

ratio to reduce input size, and the input data was ultimately defined as a 1×1250 vector. In addition, the frequency domain information of the ECG signals was also taken into consideration. The ECG signal energy in each frequency band was extracted using fast Fourier transform (FFT), and the structure of the transformed data was the same as that of a 1×1250 vector (real part of the double-sided FFT results). The input data were normalized via min-max normalization.

Because convolutional neural network (CNN) models have not previously been used for unacceptable ECG signal de-

tection, there are no references on an optimal architecture. Therefore, we tested four different internal architectures, shown in Figure 4 and Table 1, moving from simple to more complex networks. In model architecture #1, which is the simplest approach, the network consists of a single convolutional layer with 64 feature maps for the time domain and frequency domain data. Furthermore, a single fully connected layer with 64 neurons is used to combine the time domain and frequency domain information. The number of convolutional layers, feature maps, fully connected layers,

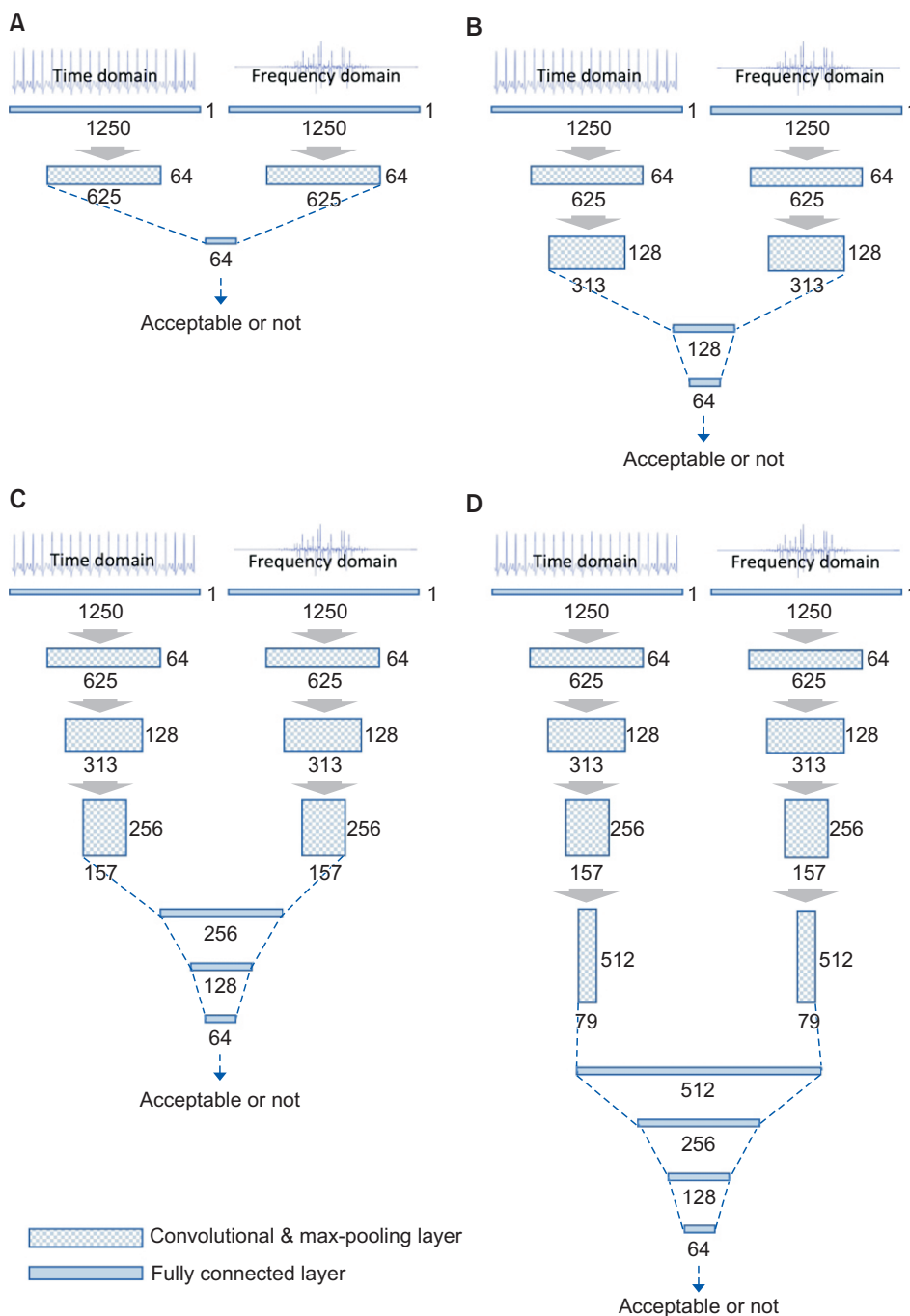


Figure 4. Architecture of the four convolutional neural network models: (A) model 1, (B) model 2, (C) model 3, and (D) model 4. After time domain and frequency domain data were abstracted by dependent convolutional layers, they were combined in the ensuing fully connected layers.

Table 1. Description of the convolutional neural network architecture

	Model 1	Model 2	Model 3	Model 4
Number of convolutional layers	1	2	3	4
Number of feature maps in each convolutional layer				
Convolutional layer 1	64	64	64	64
Convolutional layer 2	-	128	128	128
Convolutional layer 3	-	-	256	256
Convolutional layer 4	-	-	-	512
Size of kernel in convolutional layers ^a	16 / 48 / 96	16 / 48 / 96	16 / 48 / 96	16 / 48 / 96
Stride of kernel in convolutional layers	1	1	1	1
Size of kernel in max-pooling layers	20	20	20	20
Stride of kernel in max-pooling layers	2	2	2	2
Number of fully connected layers	1	2	3	4
Number of neurons in each fully connected layer				
Fully connected layer 1	64	128	256	512
Fully connected layer 2	-	64	128	256
Fully connected layer 3	-	-	64	128
Fully connected layer 4	-	-	-	64
Learning rate	1.00E-04	1.00E-04	1.00E-04	1.00E-04
Decay rate for learning rate	0.8	0.8	0.8	0.8
Minimum learning rate	1.00E-06	1.00E-06	1.00E-06	1.00E-06
Keep probability for dropout layers	0.8	0.8	0.8	0.8

^aIt is tested three different sizes (16, 48, and 96) in all the models.

and neurons gradually increase in model architectures #2, #3, and #4. The ReLU activation function was used in all the fully connected layers, and a final layer that outputs the binary classification result (i.e., acceptable or not) by using the softmax activation was also present. A threshold of 0.5 was used for the probabilities from the softmax layer to classify unacceptable (= 1) and acceptable (= 0). To define the optimum size of the kernel in the convolutional layers, we also tested three different sizes (16, 48, and 96) in all the models.

2) Model optimization

A cross-entropy loss function was selected as the cost function. The adaptive moment estimation (Adam) optimizer (learning rate = 0.0001, decay = 0.8, minimum learning rate = 0.000001) was used to train the model. We repeated the training process up to 100 epochs with a batch size of 200 (i.e., up to 600 iterations). During the iterations, the results that showed the lowest loss in the validation set was selected for the finally trained model.

To evaluate the robustness of the results, we conducted 9-fold cross-validation with variation of the training and validation datasets. The model that exhibited the lowest loss

in cross-validation was selected as the final model, and its performance was evaluated.

6. Performance Evaluation

Model performance was evaluated via comparison with the test dataset of 300 ECG signals labeled by a medical expert. The performance evaluation considered unacceptable screening as a positive value and calculated the sensitivity, specificity, positive and negative predictive values (PPV and NPV), F1-score, and the area under the receiver operating characteristic (AUROC) curve. Sensitivity (true positive rate) refers to the ability of the model to correctly detect unacceptable ECGs not meeting the gold standard. Specificity (true negative rate) indicates the ability of the model to correctly reject acceptable ECGs identified as accurate. PPV and NPV represent the proportion of true to modeled positive and negative results, respectively. The F1-score is the harmonic average of the sensitivity and PPV.

7. Cutoff for Classifying Unacceptable ECGs

To set an appropriate cutoff to screen out as many unacceptable ECGs as possible, even at the expense of some accept-

able ECGs, we conducted sensitivity analysis on the cutoff and determined the optimal cutoff for our study.

8. Software Tools and Computational Environment

MS-SQL 2017 was used for data management, and Python (version 3.6.1) and the TensorFlow library (version 1.2.1) were used to develop the data preprocessing and deep learning models. The machine used for model development had one Intel Xeon CPU E5-1650 v4 (3.6 GHz), 128 GB RAM, and one GeForce GTX 1080 graphics card.

III. Results

Among the model architectures, model architecture #3 exhibited the best average cross-validation accuracy and loss (Figure 5 and Supplementary Table S1). With increasing model architecture complexity, accuracy increased, and loss decreased. With model architecture #4, however, neither the accuracy nor the loss improved. During evaluation of the kernel size in the convolutional layers, a smaller size showed better performance in all models. Therefore, model architecture #3 with a kernel size of 16 was selected as the final model architecture.

When the final model showing the lowest loss in cross-validation among models with model architecture #3 was applied to the gold standard dataset (test dataset), it gave a result of 0.93 AUROC (Figure 6). In the sensitivity analysis to set the cutoff to classify unacceptable ECGs, when 0.05 was used as the cutoff value, high sensitivity was achieved with reasonable performance in the other indexes (Table 2).

When 0.05 was used as the cutoff value, 88% of the unacceptable ECGs were detected, and 11% of the acceptable data

were incorrectly evaluated as unacceptable (Table 3). The signals defined as unacceptable by the algorithm were 74% unacceptable and 26% misclassified acceptable waveforms. In contrast, only 9 of 76 unacceptable ECGs in the test dataset were misclassified as acceptable, and 96% of the signals evaluated as acceptable by the model were acceptable.

The time required to evaluate the 300 gold-standard dataset signals was 0.48 seconds, taking an average of 0.58 ms for each ECG.

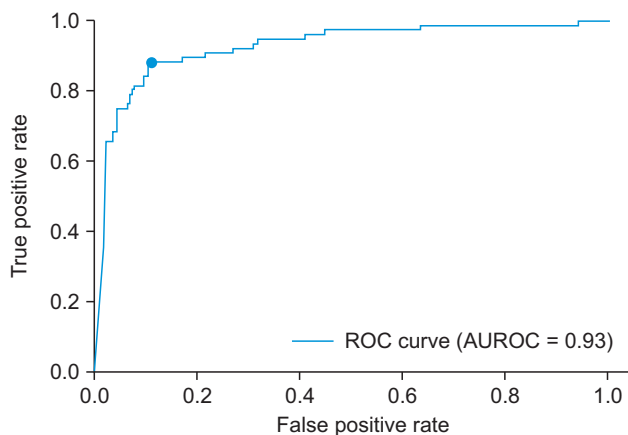


Figure 6. Receiver operating characteristics (ROC). When the final model was applied to the test dataset, the area under the ROC curve (AUROC) reached 0.93. The blue point represents the results when 0.05 was used as the threshold for the probabilities output by the convolutional neural network. Depending on the user's purpose (for example, higher sensitivity), the threshold can be adjusted.

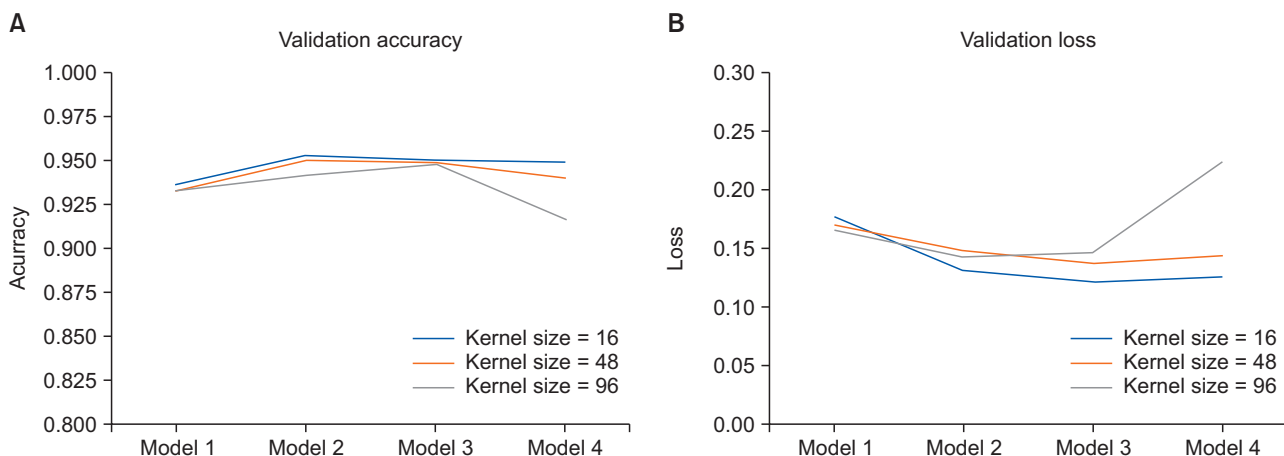


Figure 5. Classification accuracy (A) and loss (B) of the four convolutional neural network models with various kernel sizes. The more complex model architecture with smaller kernel size exhibited better classification accuracy and less loss on the validation dataset.

Table 2. Result of sensitivity analysis to set optimal cutoff value

	Cutoff values				
	0.005	0.05	0.5	0.95	0.995
Sensitivity	0.91	0.88	0.76	0.70	0.66
Specificity	0.76	0.89	0.93	0.96	0.97
PPV	0.57	0.74	0.79	0.85	0.89
NPV	0.96	0.96	0.92	0.90	0.89
F1-score	0.70	0.80	0.78	0.77	0.76

The performance when the finally selected cutoff value was used is shown in bold.

PPV: positive predictive value, NPV: negative predictive value.

Table 3. Confusion matrix when cutoff value of 0.05 was applied to the finally selected convolutional neural network model

		Classification of our model	
		Unacceptable	Acceptable
Test dataset	Unacceptable	67	9
	Acceptable	24	200

IV. Discussion

This study developed a deep learning model for screening unacceptable ECG waveforms. The developed model was able to identify most (88%) of the unacceptable ECGs detected by a clinical expert. Because the time required to analyze a single 10-second waveform is only approximately 0.58 ms, this model can be used in real-time and for large volume ECG analysis.

Previous studies have primarily attempted to de-noise ECG signals [26], for example, using principal component analysis (PCA) to abstract original signal data into a few eigenvectors with low noise-levels. A discrete wavelet transformation (DWT) concentrated on true signals, which possessed larger coefficients than noise data; however, threshold definition was critical. Other methods, such as wavelet Wiener filtering and pilot estimation, have also been used.

The background to the problem addressed in this study is different from that of existing attempts. As described earlier, there are now sufficient data to establish the groundwork for a future algorithm to achieve exclusively accurate data, without input noise. This is because when noisy input is entered into a learning model, its performance declines. This study therefore aimed to leave only clean data as much as possible, even at the expense of some acceptable data. Additionally, a significant portion of noise in real-world ECG data is generated not by alterations from specific effects, but data rendered unacceptable because they were of a type of system-

generated abnormal signal, as shown in Figure 1C. These data points cannot be de-noised, as they are exclusively noise without any true ECG signal information.

Our approach, which minimizes intervention of domain experts by conducting non-experts' pre-screening, would be also viable for application to other biosignals. Because the enrollment of medical expertise is also accompanied by high costs, it was necessary to test whether accurate noise evaluations could be developed while reducing the effort required of medical expertise. Conventional de-noising or quality assessment methods have been designed considering the features of specific waveforms; however, they require much effort from domain experts, which could be a barrier to the development of a model that can be applied to a wide variety of biosignals.

Our results in CNN model optimization suggest that deeper but smaller convolutional filter (kernel) sizes provide better performance. This finding was also observed in another domain, image recognition. The VGG-16 model, which won first place in the ImageNet Challenge 2014, improved the performance by increasing network depth with very small convolutional filters [27].

In analysis of ECG signal, we chose 1D CNNs rather than the 2D CNN with the spectrogram of models because we assumed that noise or unacceptable ECG signal is independent from time. For the same reason, we did not use recurrent neural network (RNN)-based models (long short-term memory models, gated recurrent unit, etc.) and only focused on the morphological characteristics of signal.

In addition, to evaluating the advantages of using both time and frequency domain input rather than time domain data only or frequency domain data only, we conducted ad hoc analysis by using time domain data only or frequency domain data only in model architecture #3, which showed the best performance in our study. Based on the results, we confirmed that accuracy and loss were better in 9-fold cross-

validation when both data were used together (in mean \pm standard deviation; accuracy = 0.95 ± 0.02 , loss = 0.12 ± 0.03) than when only time domain data (accuracy = 0.93 ± 0.02 , loss = 0.22 ± 0.06) or only frequency domain data was used (accuracy = 0.94 ± 0.02 , loss = 0.19 ± 0.05).

Some specific waveforms due to pathology status may have been screened as unacceptable. In some cases, signal modification caused by cardiovascular system diseases might be quite similar to those affected by noise or artifacts (for example, atrial fibrillation). Therefore, we conducted ad hoc analysis by applying the proposed model to the results of portable ECGs that include interpretation by the ECG machine [23]. When our model was applied to the randomly selected 10,000 ECG lead II data, 3,337 ECGs were classified as unacceptable. Further, we could observe the tendency that waveforms of certain types of arrhythmia are likely to be classified as unacceptable (Table 4). However, there is a possibility that situations of measuring ECGs for arrhythmic patients or the statuses of arrhythmic patients were less stable and led to more unacceptable ECGs. Therefore, this model would be proper for filtering noise data for the preparation of a noise-free training dataset. If this model is used to filter noise signal before the application of certain arrhythmia detection models, sufficient input data must be prepared considering that the filtering rate could be high in pathologic

situations.

The proposed model may be used in two ways: planted in monitoring devices, or to process centrally collected data. As this model runs very quickly on an Intel Xeon CPU E5-1650 v4 (3.6 GHz) with 128 GB RAM and one GeForce GTX 1080 graphic card computing environment, time constraints would remain insignificant in the latter application because central systems would possess sufficient computing power. In the near future, monitoring devices themselves will analyze signals and provide warnings of danger. Because prediction models are highly complex, the proposed unacceptable ECG detection model needed to be as computationally simple as possible. Various approaches attempted previously in this regard [28] require further research.

This study also encountered the following limitations. First, the test dataset used as the gold standard was generated by only one expert. As previously mentioned, this study did not aim to diagnose diseases requiring a high level of domain knowledge; thus, the gap between different experts should not be significant. However, there was no supplemental evaluation to correct for potential mistakes. Second, the specific category of noise was not evaluated. Noise could be classified into five categories: BA, MA, PLI, unacceptable, and unclear. However, data collection in an actual clinical environment resulted in a majority of normal waveform points and a small

Table 4. Distribution of ECG types in ECGs classified as acceptable or unacceptable by our model

Type of ECGs	Acceptable	Unacceptable	<i>p</i> -value ^a
Total ECGs	6,663 (66.6)	3,337 (33.4)	
Normal sinus rhythm	4,641 (69.5)	2,033 (30.5)	<0.001
Atrial arrhythmia			
Sinus bradycardia	1,040 (71.0)	424 (29.0)	<0.001
Premature atrial complexes	72 (48.6)	76 (51.4)	<0.001
Supraventricular tachycardia	10 (55.6)	8 (44.4)	0.456
Atrial flutter	25 (55.6)	20 (44.4)	0.158
Atrial fibrillation	188 (50.0)	188 (50.0)	<0.001
Junctional arrhythmia	53 (52.0)	49 (48.0)	0.002
Ventricular arrhythmia			
Premature ventricular complexes	140 (45.0)	171 (55.0)	<0.001
Heart blocks			
1st degree	236 (64.0)	133 (36.0)	0.311
2nd degree	3 (33.3)	6 (66.7)	0.068
Bundle	251 (51.5)	236 (48.5)	<0.001

Values are presented as number (%).

ECG: electrocardiograms.

^aChi-square test or Fisher exact test against the expectation of unacceptable in total ECGs.

volume of noise data. Therefore, the data were insufficient for the deep learning model to learn to classify all defined categories. Furthermore, the noise categories were integrated without distinction because distinguishing between noise causes in actual applications was inconsequential (as this study aimed to eliminate all unacceptable signals, regardless of their cause). All waveforms evaluated as non-acceptable were thus also integrated as unacceptable ECGs. Finally, the PPV was not quite as high as 0.74, meaning that many of the unacceptable ECGs identified by the proposed algorithms were truly acceptable ECGs. However, as mentioned above, this study aimed to increase sensitivity, and some normal waveform loss was acceptable because these data were sufficiently represented. Instead, the proposed algorithm was able to ensure high sensitivity and successfully screened 88% of unacceptable ECGs.

In conclusion, this study developed a model capable of efficiently detecting unacceptable ECGs. The developed unacceptable ECG detection model is expected to provide a first step for future automated large-scale ECG analyses.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

Acknowledgments

This research was supported by grants from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI) funded by the Ministry of Health & Welfare, Republic of Korea (No. HG18C0067, Government-wide R&D Fund project for infectious disease research). This work was also supported by the faculty research fund of Ajou University School of Medicine.

ORCID

Dukyong Yoon (<http://orcid.org/0000-0003-1635-8376>)
 Hong Seok Lim (<http://orcid.org/0000-0002-3127-2071>)
 Kyoungwon Jung (<http://orcid.org/0000-0001-7895-0362>)
 Tae Young Kim (<http://orcid.org/0000-0002-2591-0129>)
 Sukhoon Lee (<http://orcid.org/0000-0002-3390-5602>)

Supplementary Materials

Supplementary materials can be found via <http://doi.org/10.4258/hir.2019.25.3.201>.

<http://doi.org/10.4258/hir.2019.25.3.201>. Table S1. Performance of the model in cross-validation.

References

1. Kumar A. ECG-simplified. [place unknown]: LifeHugger; 2010.
2. Kaplan NM. Systemic hypertension therapy. In: Braunwald E, editor. Braunwald's heart disease: a textbook of cardiovascular medicine. Philadelphia (PA): Saunders; 1997.
3. Van Mieghem C, Sabbe M, Knockaert D. The clinical value of the ECG in noncardiac conditions. *Chest* 2004;125(4):1561-76.
4. American Health Association. Part 8: Stabilization of the patient with acute coronary syndromes. *Circulation* 2005;112(24_Suppl):IV.89-IV.110.
5. Li H, Yuan D, Wang Y, Cui D, Cao L. Arrhythmia classification based on multi-domain feature extraction for an ECG recognition system. *Sensors (Basel)* 2016;16(10):E1744.
6. Rodrigues J, Belo D, Gamboa H. Noise detection on ECG based on agglomerative clustering of morphological features. *Comput Biol Med* 2017;87:322-34.
7. Sivaraks H, Ratanamahatana CA. Robust and accurate anomaly detection in ECG artifacts using time series motif discovery. *Comput Math Methods Med* 2015;2015:453214.
8. Satija U, Ramkumar B, Manikandan MS. Automated ECG noise detection and classification system for unsupervised healthcare monitoring. *IEEE J Biomed Health Inform* 2018;22(3):722-32.
9. Ercelebi E. Electrocardiogram signals de-noising using lifting-based discrete wavelet transform. *Comput Biol Med* 2004;34(6):479-93.
10. Ho CY, Ling BW, Wong TP, Chan AY, Tam PK. Fuzzy multiwavelet denoising on ECG signal. *Electron Lett* 2003;39(16):1163-4.
11. Tikkanen PE. Nonlinear wavelet and wavelet packet denoising of electrocardiogram signal. *Biol Cybern* 1999;80(4):259-67.
12. Iravanian S, Tung L. A novel algorithm for cardiac bio-signal filtering based on filtered residue method. *IEEE Trans Biomed Eng* 2002;49(11):1310-7.
13. Leski JM. Robust weighted averaging. *IEEE Trans Biomed Eng* 2002;49(8):796-804.
14. Almenar V, Albiol A. A new adaptive scheme for ECG enhancement. *Signal Process* 1999;75(3):253-63.

15. Barros AK, Mansour A, Ohnishi N. Removing artifacts from electrocardiographic signals using independent components analysis. *Neurocomputing* 1998;22(1-3):173-86.
16. Blanco-Velasco M, Weng B, Barner KE. ECG signal denoising and baseline wander correction based on the empirical mode decomposition. *Comput Biol Med* 2008;38(1):1-13.
17. Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc Math Phys Eng Sci* 1998;454(1971):903-95.
18. Kabir MA, Shahnaz C. Denoising of ECG signals based on noise reduction algorithms in EMD and wavelet domains. *Biomed Signal Process Control* 2012;7(5):481-9.
19. Karagiannis A, Constantinou P. Noise-assisted data processing with empirical mode decomposition in biomedical signals. *IEEE Trans Inf Technol Biomed* 2011;15(1):11-8.
20. Xiong P, Wang H, Liu M, Zhou S, Hou Z, Liu X. ECG signal enhancement based on improved denoising auto-encoder. *Eng Appl Artif Intell* 2016;52:194-202.
21. Rahman MZ, Shaik RA, Reddy DR. Efficient and simplified adaptive noise cancelers for ECG sensor based remote health monitoring. *IEEE Sens J* 2012;12(3):566-73.
22. Kim YG, Shin D, Park MY, Lee S, Jeon MS, Yoon D, et al. ECG-VIEW II, a freely accessible electrocardiogram database. *PLoS One* 2017;12(4):e0176222.
23. Chung D, Choi J, Jang JH, Kim TY, Byun J, Park H, et al. Construction of an electrocardiogram database including 12 lead waveforms. *Healthc Inform Res* 2018;24(3):242-6.
24. Johnson AE, Pollard TJ, Shen L, Lehman LW, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. *Sci Data* 2016;3:160035.
25. Yoon D, Lee S, Kim TY, Ko J, Chung WY, Park RW. System for collecting biosignal data from multiple patient monitoring systems. *Healthc Inform Res* 2017;23(4):333-7.
26. Reddy CK, Aggarwal CC. *Healthcare data analytics*. Boca Raton (FL): CRC Press; 2015.
27. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [Internet]. Ithaca (NY): arXiv.org; 2014 [cited at 2019 Jul 10]. Available from: <https://arxiv.org/pdf/1409.1556.pdf>.
28. Huynh LN, Lee Y, Balan RK. Deepmon: mobile GPU-based deep learning framework for continuous vision applications. *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*; 2007 Jun 19-23; Niagara Falls, NY. p. 82-95