A Deep Learning Approach to Examine Ischemic ST Changes in Ambulatory ECG Recordings

Ran Xiao, PhD[1], Yuan Xu, PhD[1], Michele M. Pelter, RN, PhD[1], David W. Mortara, PhD[1], Xiao Hu, PhD[1,2,3,4]

[1]Department of Physiological Nursing, University of California, San Francisco, CA; [2]Department of Neurological Surgery, University of California, San Francisco, CA; [3]Institute for Computational Health Sciences, University of California, San Francisco, CA; [4]Core Faculty, UCB/UCSF Graduate Group in Bioengineering, University of California, San Francisco, CA

Abstract

Patients with suspected acute coronary syndrome (ACS) are at risk of transient myocardial ischemia (TMI), which could lead to serious morbidity or even mortality. Early detection of myocardial ischemia can reduce damage to heart tissues and improve patient condition. Significant ST change in the electrocardiogram (ECG) is an important marker for detecting myocardial ischemia during the rule-out phase of potential ACS. However, current ECG monitoring software is vastly underused due to excessive false alarms. The present study aims to tackle this problem by combining a novel image-based approach with deep learning techniques to improve the detection accuracy of significant ST depression change. The obtained convolutional neural network (CNN) model yields an average area under the curve (AUC) at 89.6% from an independent testing set. At selected optimal cutoff thresholds, the proposed model yields a mean sensitivity at 84.4% while maintaining specificity at 84.9%.

1. Introduction

Patients with acute coronary syndrome (ACS) are at risk of transient myocardial ischemia (TMI), which can lead to serious medical complications. It has been found that more occurrence of myocardial infarction after admission, acute pulmonary edema and unplanned transfer from telemetry unit to the intensive care unit associated with patients with TMI compared to those without TMI[1]. As a critical step in identifying ACS, the early detection of myocardial ischemia helps reduce irreversible damage to heart tissues and prevent patient deterioration. Several deployed methods in detecting myocardial ischemia, such as coronary angiography and echocardiogram, are either invasive and/or resource demanding, or only able to access a brief time period, making them unsuitable for initial rule-out phase for ACS. On the other hand, continuous electrocardiography (ECG) provides an economical alternative and additional diagnostic values for early transient ischemia detection. It measures the electrophysiological activities of the heart in real time and noninvasively, together with other benefits including easy setup and long-term monitoring.

ECG is an important risk stratification tool in the immediate phase of ACS. ST (i.e. the isoelectric section in ECG waveform between J point and the beginning of T wave) elevation on the ECG is presented in up to 25% of ACS patients (i.e., ST elevation myocardial infarction (STEMI)), whereas the rest (non-ST elevation-ACS (NSTE-ACS) or unstable angina (UA)) show non-specific ECG changes[2,3]. This 75% of ACS patients is at risk for TMI, which can be detected with continuous ECG monitoring. However, current ECG monitoring software is underutilized due to excessive false alarms[4]. This further contributes to alarm fatigue, which is ranked as the top technology hazard in 2014 by the Emergency Care Research Institute (ECRI)[5].

In contrary to current monitoring software, expert clinicians are capable of detecting true ST changes even if the ECG is moderately contaminated (i.e., motion artifact, patient movement, etc.) and are able to differentiate between ischemic and non-ischemic changes, by examining ECG waveforms screen by screen. Therefore, representing ECG tracings as images could provide valuable discriminative features about ST change. Meanwhile, the rapid developing approach of deep learning techniques, especially the convolutional neural network (CNN), has been constantly pushing the performance boundary of image recognition by computer algorithms[6]. A well-designed CNN model has even surpassed human benchmarks in a visual recognition challenge[7]. Some pioneer studies have adopted deep learning techniques in mining ECG features to tackle challenging medical problems related to the heart. In one study, CNN was adopted to detect various types of arrhythmia in ECG[8]. In another study, CNN was utilized to learn ECG features for screening paroxysmal atrial fibrillation patients[9].

Inspired by the aforementioned studies, the present study proposes an image-based approach that transforms windowed ECG dynamics into images to train deep learning model for ST depression events. The proposed CNN model is trained from all-day ambulatory ECG recording sessions, and tested with an independent testing set to evaluate the model performance. As a preliminary study, we target to detect significant ST change as an initial step, with further discrimination between ischemic and non-ischemic ST changes as our future effort. Nevertheless, the reliable detection of ST provides accurate description of duration of ST change and its multi-lead patterns, which will be important for downstream differentiation between ischemic and non-ischemic ST episodes.

2. Methods

2.1 Data information

Data used in the present study are selected from the Long-Term ST Database (LTST database) from PhysioNet[10]. The LTST database provides 20~24-hour ambulatory 2- or 3- lead ECG recording sessions, recorded with sampling frequency at 250 Hz[11]. For each session, single-lead annotations are performed for ischemic and non-ischemic ST events, as well as noisy and unreadable segments. The annotation information is achieved semi-automatically with combination of computer software (SEMIA) and human experts[11]. Detailed information about annotation procedures in LTST database can be found in Jager et al.'s publication[11]. It is worth noting that annotation of significant ST events was based on three types of protocols by varying minimal amplitude of ST change (V_{min}) and minimal duration of the change (T_{min}). Subsequent analysis in the present study relies on annotation information from the protocol B (V_{min} = 100 μV and T_{min} = 30 s) as adopted in the original study. In this preliminary study, the first 20 sessions out of total 86 sessions in the database, which are from 20 unique patients, were selected to train our CNN model. This will ensure independent patients in the training and testing set. The following 15 sessions (see Table 1 for a full list of testing sessions), which are also from unique patients and do not contain excessive number of sudden-step ST events (<100), were selected as testing data to evaluate the performance of the proposed CNN model. The WFDB toolbox was implemented to extract ECG waveforms and annotation information from the LTST database[11,12].

2.2 Sample preparation

The continuous ECG waveform from each lead and session is firstly separated into episodes according to annotations of different events, which will later form case (significant ST changes) and control (no significant ST changes) conditions. Specifically, episodes related to significant ST change, such as ischemic ST changes and heart-rate related ST change, are grouped as the case condition, similar to previous studies[13,14]. Then the control condition consists of the remaining ECG data after further exclusion of segments related to unreadable ECG, noise, and sudden-step ST changes caused by axis shift and conduction change. For case events, time points of episode start, maxima of ST change and episode end are provided, so event episodes can be readily extracted by a reference to starting and ending time points. However, only one event time point is available for noise and sudden-step ST events, while another equally important information, the event duration, is unknown. Due to such limitation in the database, 10-second segments before and after these event time points are marked and removed from control condition. The present study aims to detect ST changes via an image-based approach, so single-lead image samples are designed to contain 10-second temporal dynamics of ECG overlaid with standard ECG grid (0.04 seconds as horizontal interval and 0.1 mV as vertical interval). Examples for both case and control conditions can be found in Figure 1.

Different sample selecting schemes are implemented for training and testing sessions, respectively. ECG episodes with ST changes are much shorter in length comparing to normal ST cases (see Table 1 for number of image samples for case and control conditions in testing sessions). To achieve balanced numbers of training samples for case and control conditions, resampling is performed based on different probability distributions for the two conditions. For control condition, downsampling is necessary to reduce the excessive number of images in order to speed up the training process. 10-second image samples are generated consecutively along the timeline with no overlapping. Then for each session, 10,000 images are selected based on uniform distribution to ensure randomness. On the other hand, oversampling is needed for case condition to obtain comparable quantity of case images to those of control condition. Based on an intuitive heuristic that samples closer to maxima of ST change hold more information about case condition, sample selection based on Gaussian distribution is proposed to augment the probability of image samples in the training set with center time points near maxima from both directions in timeline. For each event episode in the case condition, the distribution mean is set at the corresponding maxima of ST change, and standard deviation is set to be 10 seconds. This configuration ensures a large probability close to maxima of ST change, and 99% of image samples selected for training fall within 30-second radius of the maxima. With center time points selected based on the probability model, 10-second image samples are generated by combining a 5-second ECG before and after the centers. For each session, an equal number of samples is selected from each event episode, and is set to be number of samples in control (10,000) divided by the total number of episodes in that session. Samples with partial information outside episode boundaries are excluded from case condition. In total, there are 174,039 case samples and 200,000 control samples generated from the 20 training sessions. For testing sessions, testing image samples from both case and control conditions are simply selected consecutively in the timeline to preserve real-time fashion.

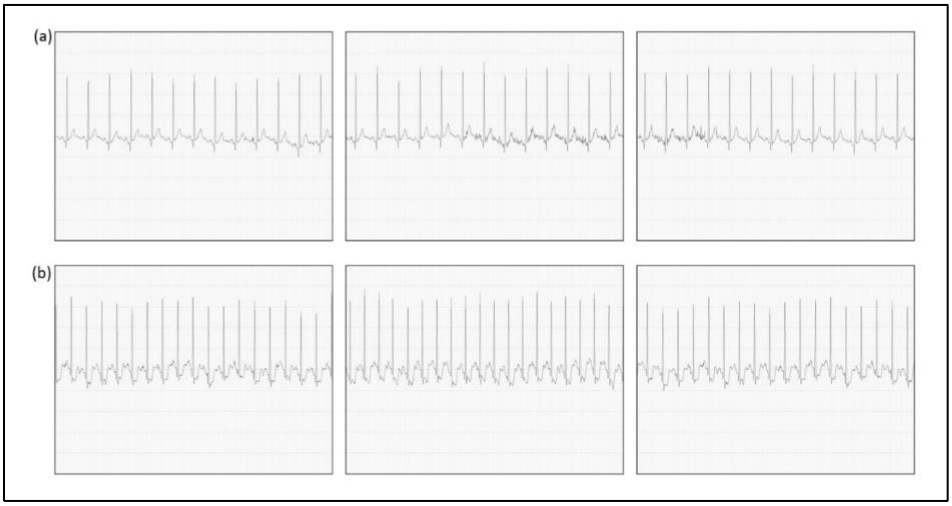

Figure 1. Exemplar image samples from control and case conditions. (a) 10-second image samples with no significant ST changes; (b) 10-second image samples with significant ST changes.

2.3 CNN model from transfer learning

The proposed CNN model takes advantage of pretrained Google Inception V3 model[15], which is trained upon millions of images and hundreds of classes from ImageNet[16], through transfer learning scheme. The idea is to retrain the existing model, which already captures rich primitive features that are common to different image applications, to classify ECG images in the present study. Figure 2 presents the schema diagram for retraining CNN model from Inception V3 via transfer learning. Layers in Inception V3 model can be separated into two parts, the feature selection part and the classification part. The feature selection part includes convolution layers, max pooling layers, dropout layers, etc., while the classification part includes fully connected layer and Softmax layer. Within transfer learning paradigm, the feature selection part of Inception V3 is kept, and transferred features are generated by passing training samples in the present study through all layers in the feature selection part. Then final layers, such as fully connected layer and Softmax layer, are trained and appended based on the class labels (i.e., case and control) in the present study to yield final classification results. The final layers are trained and updated through 50 epochs. With the retrained CNN model, classification is performed for each image sample in the testing sessions.

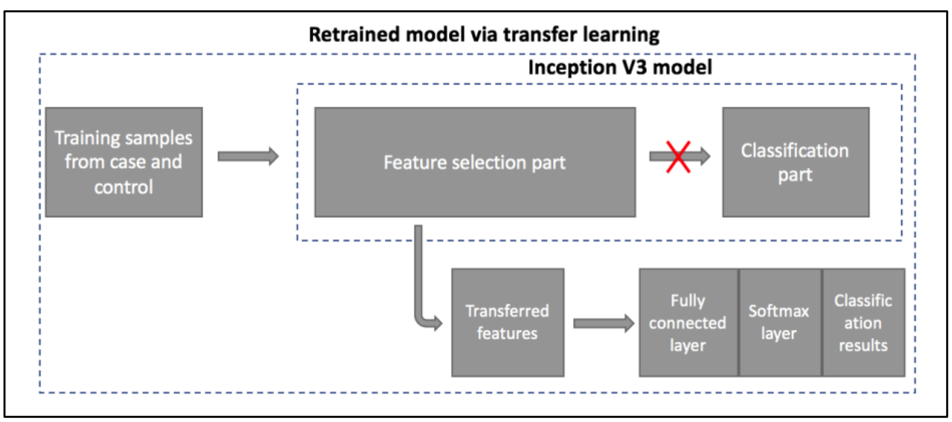

Figure 2. Schema diagram for retrained CNN model via transfer learning

2.4 Performance evaluation

To investigate the performance of proposed CNN model, common metrics for accessing binary classifiers, including the receiver operating characteristic (ROC) curve and the area under the curve (AUC), are adopted to evaluate the model performance under various discrimination thresholds. The AUCs from all testing sessions are then tested against the random guess level (50% for binary classification) using one-sample Student's t-test. For each testing session, an optimal cutoff threshold can be selected based on Youden's index[17]:

$$J = \frac{TP}{TP + FN} + \frac{TN}{TN + FP} - 1$$

where J denotes Younden's index, and TP, FP, TN and FN are counted by treating case condition as positive class. Optimal cutoff thresholds are selected where Younden's indices reach maxima, and additional metrics, including sensitivity, specificity and F1 score, are calculated at the optimal cutoff thresholds to further evaluate the classification performance.

3. Results

The retrained CNN model from Inception V3 via transfer learning uses the first 20 sessions in the LTST database (174,039 case samples and 200,000 samples) as training, and it is tested on the following 15 sessions in the database. Figure 3 presents the ROC curves achieved by the proposed model for all 15 testing sessions. The black dashed line denotes guess level, and each blue curve denotes ROC curve of one testing session. The x axis represents 1-specificity, while the y axis represents sensitivity. It shows the proposed model is able to detect ST changes from control condition above the chance level for all sessions, while some variation across different sessions in performance also presents.

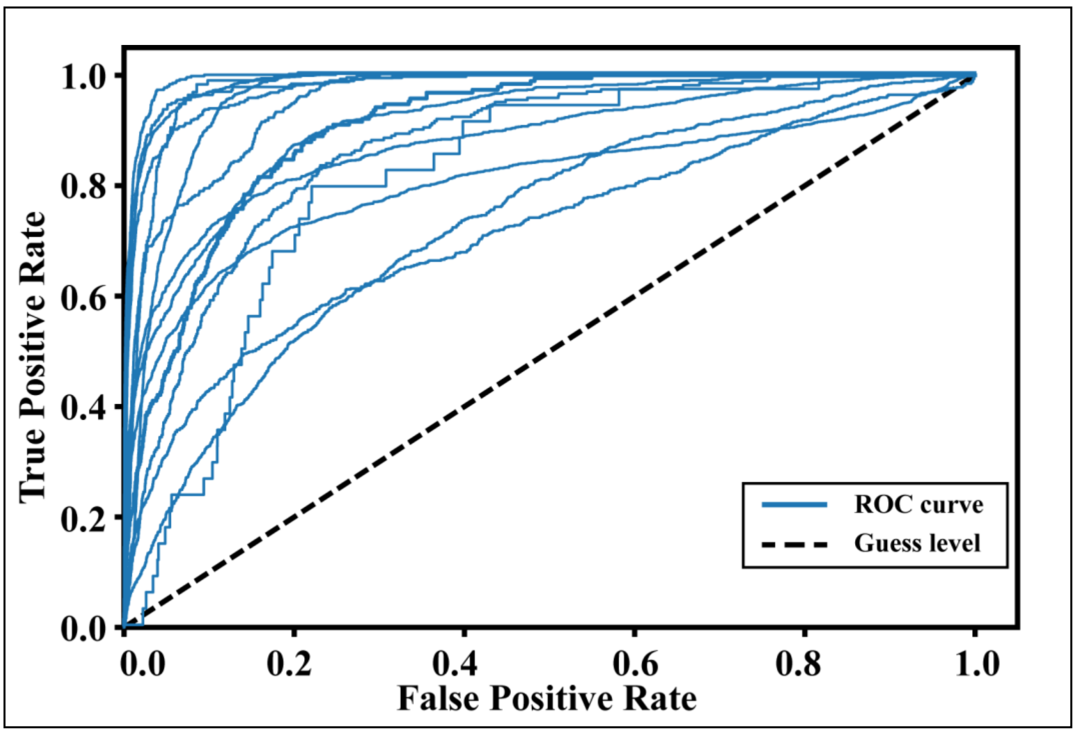

Figure 3. ROC curves of individual testing sessions. Black dashed line indicates random guess level.

Table 1 provides quantitative performance evaluation with various performance metrics from all testing sessions achieved by the proposed model. Each row presents performance information from one testing session, with last row presenting the mean and standard deviation derived from all 15 sessions. Specifically, the first column shows sessions in the LTST database selected as testing sessions. The second column (# case/control) presents number of image samples from case and control conditions for each session. It reveals highly unbalance samples between both conditions. The third to last columns present AUC in correspondence to each ROC curve in Figure 3, and sensitivity (Sen@opt), specificity (Spec@opt) and F1 score (F1@opt) at selected optimal cutoff thresholds. The AUC demonstrates above chance performance for all testing sessions, with 9 out of 15 testing sessions reaching AUC value above 90%. In average, the proposed CNN model achieves mean AUC of 89.6%± 9.3% from the 15 independent testing sessions. In addition, the Student's t-test suggests this performance is significantly higher than the random guess level ($p<0.001$). With selected optimal cutoff threshold, the proposed model achieves mean sensitivity of 84.4%± 13.9%, mean specificity of 84.9%± 8.3% and mean F1 scoreof 89.2%± 5.5%, respectively.

Table 1. List of performance information for each testing session.

Session	# case/control	AUC (%)	Sen@opt (%)	Spec@opt (%)	F1@opt (%)
s20341	480/15702	99.1	96.9	96.1	96.8
s20361	1426/15700	98.6	95.4	92.6	93.7
s20331	346/17162	98.2	94.2	94.9	96.2
s20251	141/16968	97.6	97.9	91.5	94.9
s20351	523/15893	97.5	89.7	94.2	95.3
s20231	818/16032	95.8	95.2	87.9	91.0
s20211	391/16197	95.3	96.2	79.4	86.8
s20241	1161/15653	90.8	86.9	80.1	85.0
s20371	190/15924	90.7	87.9	78.9	87.2
s20261	1183/14899	87.7	77.7	84.8	87.3
s20281	262/16796	86.5	83.2	76.9	85.7
s20321	34/17647	81.0	79.4	78.0	87.4
s20311	1803/15345	80.2	67.4	85.5	85.7
s20291	722/15724	73.0	68.1	66.2	76.2
s20301	593/16601	71.7	49.6	86.0	89.0
Mean/STD	NA	**89.6/9.3**	**84.4/13.9**	**84.9/8.3**	**89.2/5.5**

4. Discussion

The present study introduced an image-based approach in combination with a deep learning technique to monitor ischemic ST change in ambulatory ECG. A CNN model has been trained through transfer learning scheme using 24-hour ambulatory ECG recording sessions in LTST database and tested on independent sessions in the database. The proposed CNN model is able to classify testing images in real-time fashion with an average AUC at 89.6%, which is significantly higher than guessing ($p<0.001$). At selected optimal cutoff thresholds, our model achieves an average sensitivity at 84.4% at 10-second sample level. This is on par with previously reported performance using the same annotation protocol with average sensitivity at 78.10%, 78.28% and 82.13%, but achieved at whole-episode level, i.e., assigning just one class label for the whole duration of ST episode[13,14,18]. Meanwhile, our model is able to maintain a high specificity, demonstrated by the average specificity at 84.9% and F1 score at 89.2%.

To our knowledge, the present study is among one of the first studies leveraging deep learning technique for detecting ischemic ST change in ECG. The proposed approach delivers a simple training process in comparison to previous ST detection algorithms[13,14], while achieves comparable if not higher performance. It bypasses the complicate rule settings from previously used decision tree, which might increase the chance of overfitting. The image-based sample selection resembles images presented in ECG monitors for human inspection. Such approach is inspired by our previous study that presented ECG signals as images and differentiated images of good ECG quality from those of poor quality[19]. Both approaches are motivated by an insight that clinicians typically use visual pattern recognition to read ECG and identify pathological changes in ECG tracings. The combination of image-based sample selection and deep learning produces a model with both high sensitivity and specificity for monitoring significant ST change at 10-second sample level, which is crucial for early detection of myocardial ischemia. Accurate sample-level classification is also the foundation of successful detection of whole episode. Thus, our proposed approach also contributes to the effort of reducing false alarms that plague current ST monitor systems.

One limitation of present study resides in annotations for sudden-step ST change caused by axis shift and conduction change in LTST database, which are only provided with onset time. Except for the sharp change at the episode boundaries, samples associated with these events can be akin to those from ischemic ST change, and the duration varies from episode to episode. Although we have mitigated the issue by removing empirically selected 1-minute ECG centered at their onsets from control condition, residuals could still greatly limit the specificity. In addition, the large variation in number of sudden-step ST episodes across patient-independent testing sessions could shed light on the performance variation as presented in Figure 3. Thus, additional measures of event duration are needed to have these events truly accounted for during model training.

Despite the achieved performance, the present study is still preliminary and orientated as a proof of concept, with only 20 out of total 86 sessions in LTST database included in training set and 15 in testing. Future effort is needed to investigate model performance with respect to number of sessions included in the training set, as well as to validate the stability on a larger testing set. Furthermore, to expedite the training process, only 50 epochs of optimization is performed in current model, while performance improvement is expected with further optimization. Due to highly-unbalanced data in case and control, an oversampling approach based on Gaussian distribution is introduced in case condition. Our results demonstrate the efficacy of such approach in capturing representative samples in case condition. However, the selection of 10-second STD for the probability distribution is empirical. We also plan to systematically investigate the impact of its changes on the model performance in future work.

Acknowledgement

This work in partially funded by R01NHLBI128679, R18H5022860, R01GM111378, UCSF Middle-Career scientist award, and UCSF Institute of Computational Health Sciences.

References

1. Pelter MM, Loranger DL, Kozik TM, et al. Among unstable angina and non-ST elevation mi patients, transient myocardial ischemia and early invasive treatment are predictors of major in-hospital complications. J Cardiovasc Nurs. 2016;31(4):10-19.
2. Amsterdam EA, Wenger NK, Brindis RG, et al. 2014 AHA/ACC guideline for the management of patients with non-ST-elevation acute coronary syndromes: a report of the American College of Cardiology/American Heart Association task force on practice guidelines. J Am Coll Cardiol. 2014;64(24):139-228.
3. Pelter MM, Kozik TM, Loranger DL, Carey MG. A research method for detecting transient myocardial ischemia in patients with suspected acute coronary syndrome using continuous ST-segment analysis. J Vis Exp. 2012;(70):50124.
4. Patel SI, Souter MJ. Equipment-related electrocardiographic artifacts: causes, characteristics, consequences, and correction. Anesthesiology. 2008;108(1):138-48.
5. ECRI. Top 10 health technology hazard for 2014. Health devices. 2013;41:1-13.
6. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In NIPS. 2012; 1106–14.
7. He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In ICCV. 2015.
8. Rajpurkar P, Hannun AY, Haghpanahi M, Bourn C, Ng AY. Cardiologist-Level Arrhythmia Detection with Convolutional Neural Networks. arXiv:1707.01836.
9. Pourbabaee B, Roshtkhari MJ, Khorasani K. Deep convolution neural networks and learning ECG features for screening paroxysmal atrial fibrillation patients. IEEE Transactions on Systems, Man, and Cybernetics: Systems. 2017;99:1-10.

10. Goldberger AL, Amaral LAN, Glass L, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation. 2000;101(23):215-20.
11. Jager F, Taddei A, Moody GB, et al. Long-term ST database: a reference for the development and evaluation of automated ischaemia detectors and for the study of the dynamics of myocardial ischaemia. Med Biol Eng Comput. 2003;41(2):172-82.
12. Silva I, Moody G. An open-source toolbox for analysing and processing PhysioNet databases in MATLAB and Octave. J Open Res Softw. 2014; 2(1):27.
13. Dranca L, Goni A, Illarramendi A. Using decision trees for real-time ischemia detection. 19th IEEE Symposium on Computer-Based Medical Systems. 2006; 719-26.
14. Dranca L, Goni A, Illarramendi A. Real-time detection of transient cardiac ischemic episodes from ECG signals. Physiol Meas. 2009;30(9):983-98.
15. Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2015; 1-9.
16. Russakovsky O, Deng J, Su H, et al. ImageNet Large Scale Visual Recognition Challenge. Int J Comput Vis. 2015;115:211.
17. Youden WJ. Index for rating diagnostic tests. Cancer. 1950; 3(1):32-5.
18. Smrdel A, Jager F. Automated detection of transient ST-segment episodes in 24h electrocardiograms. Med Biol Eng Comput. 2004; 42:303–11.
19. Shahriari Y, Fidler R, Pelter M, Bai Y, Villaroman A, Hu X. Electrocardiogram signal quality assessment based on structural image similarity metric. IEEE Trans Biomed Eng. 2017;99:1-1.