

Article

Deep-Learning Model to Predict Coronary Artery Calcium Scores in Humans from Electrocardiogram Data

Changkyoung Eem ¹, Hyunki Hong ^{1,*} and Yoohun Noh ²

¹ College of Software, Chung-Ang University, Heukseok-ro 84, Dongjak-ku, Seoul 06973, Korea; ckeem@cau.ac.kr

² Famenity Co., Ltd., D1009 Indeogwon IT Valley, 40, Imi-ro, Uiwang-si, Gyeonggi-do 16006, Korea; glotesk@cau.ac.kr

* Correspondence: honghk@cau.ac.kr; Tel.: +82-2-820-5417

Received: 3 November 2020; Accepted: 4 December 2020; Published: 7 December 2020



Abstract: We introduce a deep-learning neural network model that uses electrocardiogram (ECG) data to predict coronary artery calcium scores, which can be useful for reliably detecting cardiovascular risk in patients. In our pre-processing method, each lead of the ECG is segmented into several waves with an interval, which is determined as the period from the starting point of a P-wave to the end point of a T-wave. The number of segmented waves of one lead represents the number of heartbeats of the subject per 10 s. The segmented waves of one cycle are transformed into normalized waves with an amplitude of 0–1. Owing to the use of eight-lead ECG waves, the input ECG dataset has two dimensions. We used a convolutional neural network with 16 layers and 5 fully connected layers, comprising a one-dimensional filter to examine the normalized wave of one lead, rather than a two-dimensional filter to examine the coherence among the unit waves of eight leads. The training and testing are repeated 10 times with a randomly assigned dataset (177,547 ECGs). Our network model achieves an average area under the receiver operating characteristic curve of 0.801–0.890, and the average accuracy is in the range of 72.9–80.6%.

Keywords: electrocardiogram; coronary artery calcium score; deep-learning neural network model; coronary artery disease

1. Introduction

Coronary artery disease is a cardiovascular disease that has been found to be the leading cause of death in both developed and developing countries [1]. Coronary artery calcium (CAC) scoring has been used to predict the risk of coronary heart disease [2]. Determination of the CAC score (CACS) by computed tomography (CT) is based on axial slices, with a thickness of 3 mm, without overlapping or gaps and limited to the cardiac region. Calcification is identified as areas of hyper-attenuation in the CT images by using the Agatston method [3]. The total calcium scores were calculated based on the number, area, and peak CT numbers of the calcific lesions detected. Previously, CAC scoring and its validation were performed manually.

Many methods for automatic CAC scoring are based on classical machine learning and digital signal processing techniques [4,5]. Classical machine learning methods are initially required to define specific data features and then use them as inputs to train the models. In these methods, feature engineering to transform the raw data into a suitable representation is necessary, and the used features significantly affect the performance of the classifier such as support vector machine.

Deep-learning models have overcome this limitation by extracting relevant features directly from the inputs without prior domain knowledge. Recently, deep-learning techniques have been

successfully used in many fields, including image recognition, speech recognition, and medical diagnosis applications [6–10].

Several deep-learning techniques for automatic CAC scoring and coronary artery plaque detection have been developed [11–14]. These methods mainly dealt with two tasks in CT examination: calcified region segmentation and the corresponding volume measurement.

Wang et al. introduced a neural network model with the ResNet architecture for quantization of CACS from the CT data of 530 patients [11]. First, the CT scans are converted into three-dimensional (3-D) volume data. All the voxel points with high radiodensity value are segmented as the candidate calcified regions, which are fed as input to a neural network for automated analysis. The neural network classified each suspected calcified region into five categories according to the degree of calcification.

Shadmi et al. introduced an automatic method based on fully convolutional networks (U-Net [15] and FCDenseNet) to segment coronary calcium and predict the Agaston score from 1054 non-contrast chest CTs [12]. The dataset reflects a variety of originating institutions, acquisition devices, and manufacturers. They applied a set of heuristics to predict a bounding box around the heart, which is divided into consecutive slices. After feeding all the cropped axial slices through the network, the predicted volume is assembled where each voxel intensity represents its probability of being a CAC voxel. Then, two-dimensional (2-D) candidate blobs on each axial slice are identified as coronary artery calcification via a thresholding value, and a connected components analysis is performed. The cardiovascular-disease risk is categorized into five sub-groups based on the Agaston score. It was difficult to differentiate the 1–10 category from other sub-groups because it was highly sensitive to small prediction mistakes. Specifically, the precision, recall, and F-1 scores in the category are 60.0%, 15.0%, and 24.0%, respectively.

Zreik et al. proposed a multi-task recurrent convolutional neural network (CNN) for automatic detection and classification of coronary artery plaque and stenosis in coronary CT angiography [13]. The features obtained by 3-D CNN are aggregated by a recurrent neural network that performs two simultaneous multi-class classification tasks: the type of coronary artery plaque detection and the anatomical significance of the coronary artery detection. Santini et al. introduced a CNN-based automatic calcium scoring system to identify true coronary calcifications and discard other lesions [14]. The Agatston-based risk assessment of 56 patients was calculated and compared with manual annotations provided by an expert operator, considered as the ground truth reference. This method achieved 91.1% risk categorization accuracy. However, CT images were employed, and the number of CT image datasets was limited: 45 CT volumes for training, 18 volumes for validation, and 56 volumes for testing. In contrast, our model can predict CAC scores directly using a large-scaled electrocardiogram (ECG) dataset.

Cardiac rhythm abnormalities differ from the normal rhythm of the heartbeat cycle and are known as heart arrhythmia. The heart's rhythmic irregularities can be detected using electrocardiogram (ECG). ECG is a one-dimensional (1-D) signal representing a time series, widely used as a basic diagnostic tool for the analysis of the cardiac condition of patients. ECG signals represent the recording of the bioelectrical activities of the heart. Millions of ECGs are recorded annually, with the majority being automatically analyzed, followed by an intermediate interpretation. Digital ECG programs providing diagnostic interpretation have been actively proposed [16]. However, ECG interpretation is a mixture of both subjective and objective aspects, where even experienced cardiologists or experts can disagree [17]. This significant interobserver variability makes ECG interpretation difficult; consequently, digital ECG analysis may lead to erroneous diagnosis, which may lead to unnecessary actions (such as a surgery or an operation) taken on the patient.

Several methods to diagnose patients' heart condition from ECG signals have been proposed. Sengupta et al. introduced a method to predict the presence of abnormal cardiac muscle relaxation in reference [18]. Using continuous wavelet transform, an ECG signal is converted into a normalized energy distribution in the frequency domain, which is used to calculate multiple indices, such as T-wave peak. Small changes on the surface ECG frequency spectrum, which are associated with the development of myocardial relaxation abnormalities, are magnified. Then, a random forest ensemble

classifier with a Monte Carlo cross-validation procedure is used to identify patients at risk for left ventricular diastolic dysfunction.

Recently, deep-learning networks have been widely applied to ECG data, for screening hyperkalemia [19], cardiac contractile dysfunction [20], and cardiac arrhythmias [21]. Ullah et al. introduced a 2-D CNN model for the classification of ECG signals into eight classes, including all major types of arrhythmia—from normal beat to ventricular escape beat arrhythmias [22]. By applying a short-time Fourier transform to the 1-D ECG time series signals, they obtained 2-D spectrograms that encapsulate the time and frequency information within a single matrix. The proposed CNN model works on 2-D images of ECG signals as input data, in which the augmentation method, changing the image size with operations such as cropping, is employed.

Ebrahimi et al. reviewed 75 deep-learning-based studies reported in 2017 and 2018 for arrhythmia classification using ECG signals. They concluded that CNN-based methods have shown excellent performance in classifying different types of arrhythmia [23]. Peimanker et al. presented a deep-learning model for real-time segmentation of heartbeats using ECG as inputs [24]. They combined CNN and a long short-term memory (LSTM) model to detect knowledge about the location and morphology of different segment waveforms (P-wave, QRS complex, and T-wave) in ECG records, which can be used for arrhythmia classification. QRS complex, which is a combination of the Q wave, R wave and S wave, represents ventricular depolarization. Ribeiro et al. introduced a residual neural network to recognize six ECG abnormalities with the 12-lead ECG [25].

Since ECG signals represent heart activation conditions, most studies on ECG mainly focus on cardiac arrhythmias. Sabour et al. discussed the relation of left ventricular hypertrophy and ECG abnormalities with CAC among 566 postmenopausal women selected from a population-based cohort study [26]. In their statistical-analysis-based studies, many women with ECG abnormalities reflecting subclinical ischemia have CAC. In other words, authors showed only the relevance of repolarization abnormalities (T-axis and QRS-T angle) in ECG and CAC. However, they did not propose any quantitative method to compute CAC from the patient's ECG data.

In this study, we introduce a deep-learning neural network model to predict CACS using a large scaled ECG dataset and the participant's demographic information. Here, only the participant's gender and age in the demographic information are used. To the best of our knowledge, no studies have been conducted to predict CACS directly using ECG data. Additionally, since CACS is measured using CT scans, the patient is exposed to radiation while the scan is being captured. The increased use of CT scans in a patient's life has raised concerns regarding potential cancer risks [27]. Furthermore, a CT scan is troublesome for the patient as it is time-consuming and expensive. Since a CT scan is not used for our CACS regression, the proposed method is safer, simpler, and less expensive than previous methods. In contrast, our study adopts a deep-learning model to predict the CACSs with ECG datasets.

The remainder of this paper is organized as follows. In Section 2, we describe the pre-processing step to normalize the characteristics of the ECGs and introduce our neural network architecture. Then, we explain the experimental results in Section 3 and conclude the paper in Section 4.

2. Deep-Learning Model Design

In ECG diagnostics, data from 12 ECG leads (I, II, V0–5, III, aVR, aVL, aVF) are acquired simultaneously. In a 12-lead ECG, ten electrodes are placed on the patient's limbs and chest surface. An electrode is a conductive pad that is attached to the skin and enables the recording of electrical currents. An ECG lead is a graphical description of the electrical activity of the heart, and it is created by analyzing several electrodes. In other words, each ECG lead is computed by analyzing the electrical currents detected by several electrodes [28]. ECG signals are captured over a period of time (typically, 10 s).

Specifically, leads I, II, and III compare the electrical potential differences between two electrodes. Lead I compares the electrode on the left arm (exploring electrode) with the electrode on the right arm. Lead II compares the left leg with the right arm, and lead III compares the left leg with the left arm. The spatial organization of these leads forms a triangle in the chest (Einthoven's triangle). According to

Kirchhoff's law, the sum of all currents in a closed circuit must be zero. As Einthoven's triangle can be viewed as a circuit, Kirchhoff's law can be applied to it. Einthoven's law indicates that the sum of potentials in leads I and II is equal to the potential in lead III. Additionally, aVR, aVL, and aVF leads are a linear combination of leads I, II, and III according to the Goldberger equation. Leads aVR, aVL, and aVF can be calculated using leads I, II, and III. For example, the ECG wave in lead aVF is the average of the ECG deflection in leads II and III. Therefore, leads III, aVR, aVL, and aVF do not provide any new information; however, they provide new angles to view the same information. There are six electrodes on the chest wall and thus six chest leads (V0–5). Each chest lead offers unique information that cannot be derived from other leads. Based on the ECG lead characteristics, we trained a deep-learning neural network model on eight-lead combinations of the ECGs (I, II, V0–5) among the 12-lead ECGs [28]. Figure 1a depicts the changes in the duration, amplitude, and interval of the data from the eight-lead ECG. In Figure 1a, the x and y axes represent the time (seconds) and millivoltage (mV), respectively.

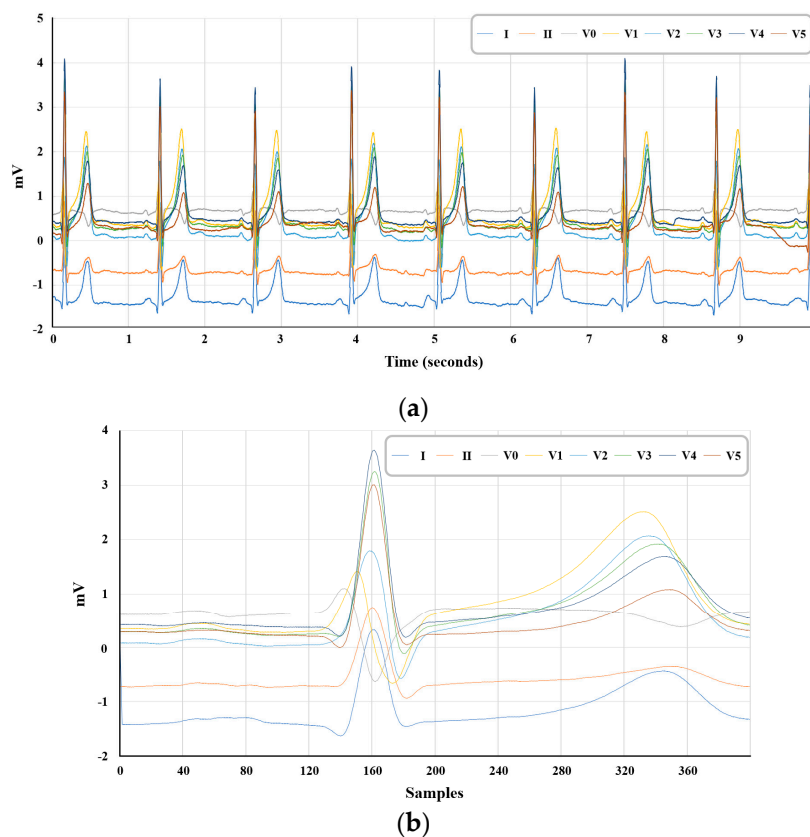


Figure 1. Eight-lead electrocardiogram (ECG) waves (I, II, V0–5): (a) eight-lead ECG waves over time (s); (b) pre-processed eight-lead ECG waves (400 samples) of one cycle.

ECG signal properties vary from person to person and depends on various factors, such as age, gender, physical conditions, and lifestyle. Specifically, the characteristics of the sub-waves (P-waves, QRS complexes, and T-waves) of the ECG, including their duration, amplitude, and R–R interval, vary according to the subject. In ECG waveforms, the first deflection (wave) is the P-wave, which represents the activation (depolarization) of the atria. Ventricular depolarization is visible as the QRS complex. The T-wave represents the repolarization of the ventricles [29]. ECG signals have very small amplitude (mV) and duration. The noise added in the capturing process usually degrades the performance of the classifier. Additionally, ECG signal detection has limitations regarding inter-patient and intra-patient variability, which means that two different beats can be of the same morphology

among different patients. Such morphological variation can be seen within the same patient as well [30]. This suggests that the characteristics of a single ECG lead can change when recording an ECG.

To improve the performance of the deep-learning model, we introduce a pre-processing step to normalize the characteristics of the ECGs. First, each lead of the ECG is segmented into several waves. The intervals of the segmented waves are determined as the period from the starting point of a P-wave to the end point of a T-wave. The number of segmented waves, H , is equal to the number of heartbeats of the subject per 10 s. Second, bidirectional weight interpolation, median pooling, and smoothing filtering are applied to the segmented waves. Here, interpolation and median pooling are implemented using library functions “scipy.interpolate.interp1d” with “cubic” argument and “skimage.measure.block_reduce” with “numpy.mean” parameter. The smoothing filtering is implemented using “numpy.convolve” function, in which a hanning widow of size 17×1 is used.

Figure 1b illustrates the pre-processed eight-lead ECG waves of one cycle, which are the eight-lead ECG waves in Figure 1a from 1–2 s. In Figure 1b, the x axis represents the sample points, and the y axis represents the voltage (mV). By choosing a sample with a maximum value between two samples, which is similar to max pooling with a 2×1 receptive field in deep learning, the number of sample points (400) of the pre-processed ECG waves is reduced to 200. Then, amplitude normalization is applied to the segmented waves. Through this pre-processing, the segmented waves of one cycle are transformed to normalized waves, which consist of 200 sample points with an amplitude of 0–1. In this work, we term the normalized wave of an interval as a unit wave for the deep-learning model.

As the unit waves of the 200 sample points are obtained from the one-lead ECG and we employ eight-lead ECGs, the input ECG dataset has two dimensions (200×8). The amplitude and interval of the unit waves are normalized, while maintaining the characteristics of the original ECGs. The unit waves of eight leads are employed as the training data. The total size of the training dataset is determined as the product of the number of times the ECGs are recorded, corresponding to the number of visits of a subject, by the heart rate, H .

The neural network model architecture is a CNN with 16 layers and 5 fully connected layers (Figure 2). We employ a 1-D filter to examine the unit wave of one lead, rather than using a 2-D filter to examine the coherence among the unit waves of eight leads. This is because, by using a 1-D filter, we can effectively derive features from the unit waves in the deep-learning model. Table 1 presents the CNN layer configuration. Our deep-learning model comprises 16 convolutional layers, some of which are followed by max pooling layers. We employ max pooling with a 2×1 receptive field and a stride of 1.

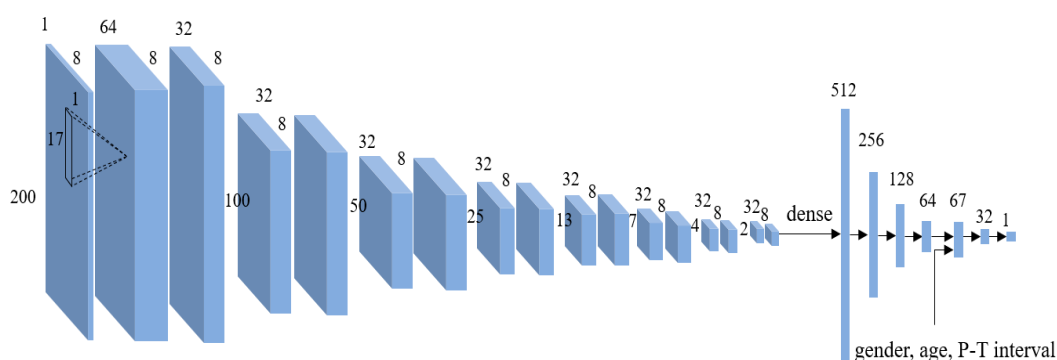


Figure 2. Proposed deep-learning model.

Table 1. Convolutional neural network (CNN) layer configuration.

Type	Patch Size/Stride	Number of Filters	Feature Map
convolution		64	200 × 8 × 64
convolution			200 × 8 × 32
max pooling + convolution			100 × 8 × 32
convolution			100 × 8 × 32
max pooling + convolution			50 × 8 × 32
convolution			50 × 8 × 32
max pooling + convolution			25 × 8 × 32
convolution			25 × 8 × 32
max pooling + convolution	17 × 1/1		13 × 8 × 32
convolution		32	13 × 8 × 32
max pooling + convolution			7 × 8 × 32
convolution			7 × 8 × 32
max pooling + convolution			4 × 8 × 32
convolution			4 × 8 × 32
max pooling + convolution			2 × 8 × 32
convolution			2 × 8 × 32

Along with the output of the CNN, the P–T interval of the segmented waves of the ECGs before pre-processing is input to the dense layer. The normalized waves obtained by the pre-processing procedure consist of 200 sample points. This means that an absolute duration of the ECG wave is lost. To further consider the participant’s ECG characteristic, the P–T interval (duration) information of the ECG wave is also input to the dense layers. Here, the participant’s gender and age in the demographic information are also considered, as shown in Figure 2. In this paper, the gender information of the male participant is coded to 0 and that of the female participant is coded to 1. The age of 100 years is scaled to 1.0. For example, the age of 45 years is scaled to 0.45. We use a rectified linear unit (ReLU) function as the activation function in the construction of our neural network architecture and use an adaptive moment estimation (Adam) optimizer for the adaptive computation of empirical gradients in the course of training process in which the gradient of the loss function with respect to the model parameters is computed following the stochastic process using a random subset of training data. Thus, the robustness of the proposed model is associated with the randomness in utilizing training data and it is naturally taken into consideration during the training phase.

As our goal is the early detection of heart disease possibility in the near future, the ECG dataset is collected from a regular medical screening at a healthcare center. Therefore, in general, the majority of participants have CACs of zero or near zero, which means the dataset has an unequal distribution. Instead of computing the sum of the squared errors between the ground truth CACs and the predicted values, a cosine similarity considering a relative magnitude (disease seriousness) of the ground truth CACs in the batch dataset is used as a cost function of our neural network. Here, both the ground truth CACs and the predicted values are represented as L2 normalized unit vectors, whose dimension number is the same as the batch size during the training stage. The L2 norm calculates the vector coordinate distance from the origin of the vector space. The degree of closeness between the ground truth CACs and the predicted values is computed with an inner product of two normalized vectors as follows:

$$\cos \theta = \frac{G \cdot P}{\|G\| \|P\|} = \frac{\sum_{i=1}^n G_i P_i}{\sqrt{\sum_{i=1}^n G_i^2} \sqrt{\sum_{i=1}^n P_i^2}}, \tag{1}$$

where G and P represent the ground truth CACs and predicted values, respectively, and G_i and P_i their corresponding vector components, n is the batch size. In our experiment, the batch size is 256, therefore, we define a 256-dimensional space as the cost domain. Our model makes the estimated

vector closer to the ground truth vector in a learning process, which means we can estimate precise CACSs in a batch learning process.

In the pre-processing, the ECG of each lead is segmented into H unit waves. Each unit wave of each lead is used as part of the training dataset; therefore, the number of training data points is increased by H . This approach implies that the training datasets are considerably larger than the number of ECG readings collected from the study participants. As a single unit wave is used per lead, the proposed model is applicable to situations in which ECG data are acquired only for a short time.

3. Experimental Results

The ECG signals are captured using an electrocardiograph (supplied by Fukuda Denshi Co. Ltd.) with a duration of 10 s. As the ECGs are recorded on the paper with the fixed grid, the electrocardiograph machines generally provide the displaying and recording sensitivity function. The sensitivity of the ECG machine by Fukuda Denshi Co. Ltd. (Tokyo, Japan) is able to set at 1/4, 1/2, 1.0, and 2.0 cm/mV. The amplitude scale of ECG is dependent on the sensitivity configuration. However, in our pre-processing procedure, the amplitude value of ECGs is normalized to 0–1, and the ECG of each lead is segmented based on the heartbeats of the subject. Here, the number of segmented waves is equal to the number of heartbeats of the subject per 10 s. That means our deep learning model did not learn the absolute amplitude values of ECGs. The relative amplitude variations within one ECG and those among eight ECGs are used in CASC regression. Therefore, the sensitivity configuration of the ECG machine does not affect the performance of our model. In our CNN model, 1-D filters (17×1) are employed.

The ECG dataset has been collected from a regular medical screening for eight years (March 2010~November 2018) at the Total Healthcare Center of Kangbuk Samsung Hospital, Seoul, South Korea. CT scan is one of major examination courses in a regular medical screening at this healthcare center. The CT images of participants have been captured, and the medical professionals measured the CACS using the participant's CT images. In our experiment, the CACS is used as the ground truth dataset. By comparing the CACS by our deep-learning neural network model with the ground truth dataset, the performance of our model can be validated quantitatively. The number of participants was 134,058. This means that our ECG dataset has been collected from a single hospital, despite large numbers of participants. Of the total number of participants, 74.24% (99,521 participants) were men and 25.76% (34,537 participants) were women. Some participants underwent the regular medical screening several times during that period. In our experiment, 177,547 ECG readings were used. The ECG characteristics of even the same participant changed over time. If the training dataset is set to a smaller percentage of the class (women), the total number of datasets decreases. Instead of balancing with the gender ratio in the dataset, our deep-learning model was trained on a large-scaled dataset. In other words, we employed a gender imbalanced dataset. The network model is trained using 142,037 ECG readings, and the test dataset consists of readings from 35,510 ECGs. The ratio between the training data and the test data is set to 8:2.

As depicted in Figure 2, the unit waves (200×8) are input into our deep-learning model. The heart rate, H , of the subjects was found to be 9–11 beats in 10 s on average, and thus the total number of datasets for training and testing increased to 1,792,919. Our model was trained for 100 epochs with a batch size of 256.

We used the CACS with the neural network model to determine whether a participant had a coronary artery disease. Traditionally, people with positive CAC, with scores in the ranges of 1–100, 100–400, and >400 , are considered to be at low, intermediate, and high risk of both ischemia and cardiovascular disease, respectively [5].

In Table 2, the proposed network model is evaluated at an operating point that is selected such that the sensitivity and specificity are equal [18]. This means that both sensitivity and specificity are the same as accuracy measure. In other words, in Table 2, sensitivity, specificity, and accuracy have the same value. This point (Q^* index) has been suggested as a possible global parameter to

summarize the test accuracy of cognitive screening instruments and as a definition for the optimal test cut-off [31]. A threshold is applied to the test datasets to evaluate the network model performance. The training and testing processes are repeated 10 times, and the obtained performances are averaged. The dataset was randomly assigned to the training and test subsets. Here, the Python NumPy library function “numpy.random.permutation” is employed to randomly permute a dataset. To compare the performances of our model under the same input (eight-lead ECG waves) condition, we implemented two network models: the deep neural network (DNN) and recurrent neural network (RNN). The first is based on the DNN model with six dense layers, in which ReLUs are used as active functions. The unit waves (200×8), by our pre-processing procedure, are flattened into 1600×1 . The number of nodes of hidden layers are 1600, 256, 128, 66, 32, and 16. The gender and age information are also input to the fourth hidden layer. The second model is based on the RNN model with two hidden layers—the one with 256 and the other with 64 states—in which the number of output states is 32. To transform ECG waves to sequence data, the unit waves (200×8) are transposed into 8×200 . Along with the output of the RNN, the gender and age information are input to the dense layers. The number of nodes of dense layers are 34, 32, and 16. In Table 2, our method and two models were compared with respect to the accuracy and area under the receiver operating characteristic curve (AUC) measures. Here, the minimum (Min), maximum (Max), and average (Avg.) of the results were presented. Our model can predict the CACS using ECG data and demographic information (gender and age). As our main goal is to generate information about the heart-disease possibility, the CACS is used to generate a clinical interpretation of the heart disease. To more clearly identify the heart-disease possibility, the cardiovascular-disease risk has been categorized into seven cases: $CACS \geq 1, 25, 50, 100, 150, 200,$ and 400. As presented in Table 2, our network model achieves an average AUC of 0.801–0.890, and the average accuracy of the proposed network model is in the range of 72.9–80.6%.

Table 2. Test dataset performance for coronary artery disease risk.

Methods	CACS	AUC			Accuracy (%)		
		Min	Max	Avg.	Min	Max	Avg.
DNN	1	0.648	0.735	0.699	61.0	67.3	64.7
	25	0.663	0.753	0.715	62.1	68.8	66.0
	50	0.674	0.766	0.727	62.8	69.8	66.8
	100	0.686	0.780	0.741	63.6	71.0	68.0
	150	0.689	0.788	0.747	64.1	71.7	68.6
	200	0.697	0.795	0.755	64.8	72.2	69.2
	400	0.735	0.838	0.796	67.5	76.1	72.6
RNN + DNN	1	0.749	0.780	0.766	68.2	71.1	69.7
	25	0.768	0.799	0.785	69.7	72.4	71.2
	50	0.782	0.813	0.799	70.8	73.5	72.3
	100	0.796	0.827	0.814	71.8	74.8	73.5
	150	0.802	0.833	0.819	72.3	75.3	74.0
	200	0.809	0.838	0.825	72.8	76.0	74.5
	400	0.848	0.874	0.863	76.3	79.0	77.8
Our Method	1	0.796	0.804	0.801	72.5	73.4	72.9
	25	0.814	0.822	0.818	74.0	74.6	74.3
	50	0.828	0.835	0.832	75.1	75.7	75.4
	100	0.841	0.847	0.845	76.1	76.8	76.4
	150	0.847	0.853	0.850	76.5	77.1	76.8
	200	0.854	0.861	0.857	77.2	77.9	77.5
	400	0.887	0.893	0.890	80.0	81.3	80.6

To validate the performance of our model, we also use metrics such as precision, recall rate, and F-1 score. Specifically, we focus on two cases ($CACS \geq 1$ and 25) in the cardiovascular-disease-risk categorization based on the CACS. The first reason is that the previous method encountered a difficulty

to differentiate the category (CACS = 1–10) from other categories [12]. In Reference [12], Shadmi et al. mentioned that it was highly sensitive to small prediction mistakes. As described in the Introduction, the precision, recall, and F-1 score in the category were 60.0%, 15.0%, and 24.0%, respectively. The second reason is that our goal is the early detection of heart-disease possibility. This means that it is important to generate a precise clinical interpretation of the coronary artery disease in the lower category. In the first case (CACS ≥ 1) and the second case (CACS ≥ 25), the percentages of the positive ECG data in the test dataset are 22.1% and 12.2%, respectively. This means that the dataset is imbalanced. In the first category, the precision, recall rate, and F-1 score of our model are 43.5%, 73.1%, and 54.6%, respectively. In the second category, the precision, recall rate, and F-1 score are 28.7%, 74.4%, and 41.1%, respectively. The experimental results indicate that our model outperformed the previous method [12] in identifying patients at risk to a coronary artery disease, with respect to the recall rate and F-1 measures.

In the quantitative evaluation of the generalization induced by our model, we employ 5-fold cross validation, wherein all the samples in the dataset are randomly split into 5 smaller sets (folds). In the first iteration, the first fold is used to test the model, and the rest are used to train the model. In the second iteration, the second fold is used as the testing set, whereas the rest serve as the training set. This process is repeated until each of the 5 folds has been used as the testing set. Figure 3 shows the performances of our model over five iterations with respect to AUC and accuracy; the performances do not change significantly during the iterations. Table 3 shows that the mean and standard deviation (std. dev.) of the Matthews correlation coefficient, accuracy, and AUC in the 5-fold cross-validation framework. The experimental results demonstrate that our model is well generalized. Figure 4a,b depict the ROC curves of the model performance when the CACS is greater than 1 and 400, respectively. The experimental results indicate that the proposed network model can effectively screen patients for the risk of coronary artery disease.

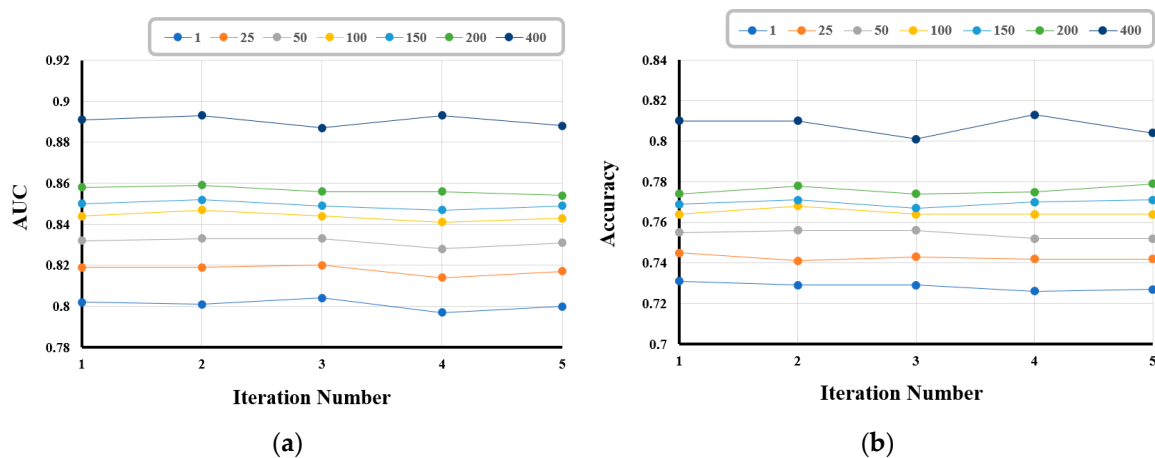


Figure 3. Performances of our model over five iterations: (a) AUC; (b) accuracy.

Table 3. Performance evaluation in 5-fold cross validation.

		Matthews Correlation Coefficient		Accuracy (%)		AUC	
		Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Training	Category 1	0.368	0.002	72.94	0.08	0.802	0.001
	Category 2	0.314	0.001	74.84	0.10	0.827	0.001
Test	Category 1	0.361	0.001	72.84	0.17	0.801	0.002
	Category 2	0.311	0.006	74.26	0.14	0.818	0.002

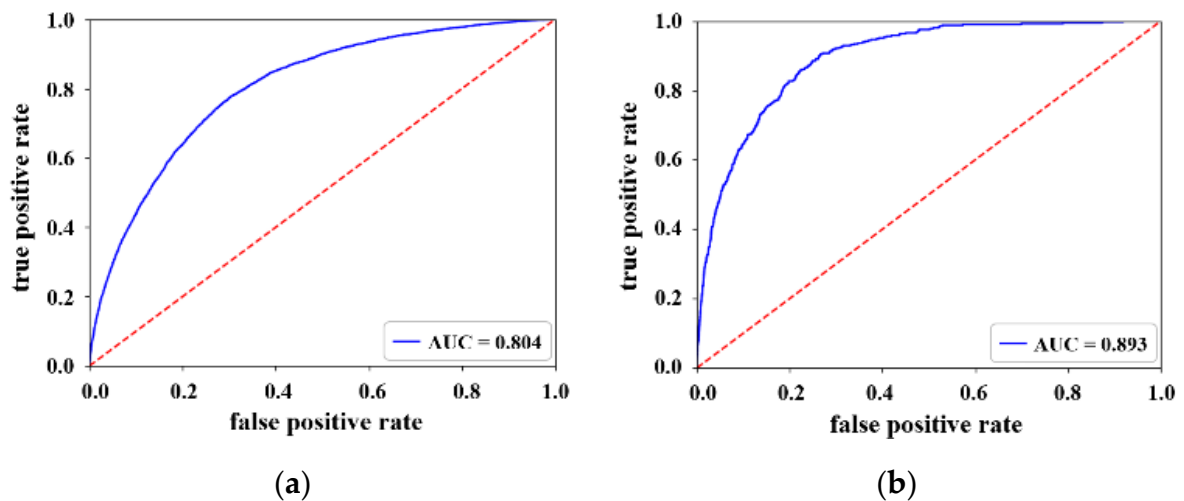


Figure 4. Classification performance for coronary artery disease risk: (a) receiver operating characteristic (ROC) curve when CACS is larger than 1; (b) ROC curve when CACS is larger than 400.

To further validate the robustness of our model, we included the experimental results according to the number of ECG leads. Table 2 presents the performance when eight-lead (I, II, and V0–5) ECG waves are used as the input to our model. In the first case, two-lead (I and II) ECG waves are input to our model. The two leads compare the electrodes of the two arms and the left leg. In the second case, we use six-lead (V0–5) ECG waves, in which the exploring electrode is on the chest overlying the heart or its vicinity. The performances obtained in the two cases are presented in Table 4. The number of ECG waves affect the performance of our model to some extent. In Table 4, even if two-lead ECG waves are used, there is no significant difference in performance, compared with that in Table 2.

Table 4. Test dataset performance according to the number of ECG leads.

Number of ECG Leads	CACS	AUC			Accuracy (%)		
		Min	Max	Avg.	Min	Max	Avg.
2	1	0.778	0.784	0.781	70.9	71.1	71.0
	25	0.796	0.801	0.798	72.3	72.5	72.4
	50	0.810	0.814	0.812	73.4	73.7	73.5
	100	0.823	0.826	0.824	74.0	74.7	74.4
	150	0.830	0.831	0.831	74.7	75.2	75.0
	200	0.837	0.838	0.837	75.3	75.7	75.5
	400	0.869	0.870	0.870	78.5	79.1	78.7
6	1	0.767	0.782	0.775	70.3	71.1	70.6
	25	0.789	0.799	0.794	72.0	72.4	72.1
	50	0.801	0.812	0.807	72.9	73.4	73.2
	100	0.813	0.824	0.820	73.9	74.6	74.2
	150	0.818	0.829	0.825	74.3	74.8	74.5
	200	0.824	0.836	0.832	74.7	75.7	75.1
	400	0.858	0.868	0.864	77.9	78.9	78.3

4. Conclusions

To obtain conventional CACS for detecting coronary heart disease risk, patients are required to have CT scans, which would involve exposure to radiation. This study introduces a deep-learning neural network model that uses ECG data to predict CACS. ECG is one of the more widely used diagnostic screening test procedures in cardiology, which is non-invasive and non-radiative. We introduce a pre-processing method to transform eight-lead ECG waves into one-cycle normalized waves. By using the normalized waves, our deep-learning neural network model can predict CACSs accurately. The neural network model is a

CNN with 16 layers and 5 fully connected layers. The experimental results indicate that our network model could be employed to screen for coronary artery diseases. Our model does not use any CT scan, but uses a deep-learning model to predict CACSs with ECG datasets, which is safer, simpler, and less expensive than CT scans. Thus, our method is believed to have widespread implications and could significantly contribute to the existing knowledge on medical diagnosis. The morphology of ECG signals is often erratic, even for one person. This means that physical states such as running, walking, and sleeping may have significant effects on ECG signals. In this paper, only the participant's gender and age in the demographic information are used. To improve the performance of our model, we will consider additional demographic information such as ethnicity, marital status, smoking, and education. Additionally, our model will be extended into wearable devices (e.g., smart watches and smart patches) to provide information regarding coronary heart condition considering the user's lifestyle.

Author Contributions: C.E. and H.H. proposed the idea of this paper; C.E., H.H., and Y.N. reviewed this paper and provided information; C.E. and H.H. conceived and designed the experiments; C.E. and H.H. performed the experiments; C.E., H.H., and Y.N. reviewed the codes in this paper and wrote this paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by CDM-based precision medical data integration platform technology development program through the Korea Evaluation Institute of Industrial Technology funded by the Ministry of Trade, Industrial and Energy (No. 20005037), and by Chung-Ang University's research grant in 2020.

Acknowledgments: The authors would like to thank Total Healthcare Center of Kangbuk Samsung Hospital for data collection.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Malakar, A.K.; Choudhury, D.; Halder, B.; Paul, P.; Uddin, A.; Chakraborty, S. A review on coronary artery disease, its risk factors, and therapeutics. *J. Cell. Physiol.* **2019**, *234*, 16812–16823. [[CrossRef](#)]
2. Polonsky, T.S.; McClelland, R.L.; Jorgensen, N.W.; Bild, D.E.; Burke, G.L.; Guerci, A.D.; Greenland, P. Coronary artery calcium score and risk classification for coronary heart disease prediction. *J. Am. Med. Assoc.* **2010**, *303*, 1610–1616. [[CrossRef](#)]
3. Agatston, A.S.; Janowitz, W.R.; Hildner, F.J.; Zusmer, N.R.; Viamonte, M.; Detrano, R. Quantification of coronary artery calcium using ultrafast computed tomography. *J. Am. Coll. Cardiol.* **1990**, *15*, 827–832. [[CrossRef](#)]
4. Zuluaga, M.A.; Magnin, I.E.; Hoyos, M.H.; Leyton, E.J.D.; Lozano, F.; Orkisz, M. Automatic detection of abnormal vascular crosssections based on density level detection and support vector machines. *J. Comput. Assist. Radiol. Surg.* **2011**, *6*, 163–174. [[CrossRef](#)]
5. Hecht, H.S. Coronary artery calcium scanning: Past, present, and future. *JACC Cardiovasc. Imaging* **2015**, *8*, 579–596. [[CrossRef](#)]
6. Lin, T.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988.
7. Chiu, C.; Sainath, T.N.; Wu, Y.; Prabhavalkar, R.; Nguyen, P.; Chen, Z.; Kannan, A.; Weiss, R.J.; Rao, K.; Gonina, E.; et al. State-of-the art speech recognition with sequence-to-sequence models. In Proceedings of the IEEE Acoustics, Speech and Signal Processing, Calgary, AB, Canada, 15–20 April 2018; pp. 4774–4778.
8. Bakator, M.; Radosav, D. Deep learning and medical diagnosis: A review of literature. *Multimodal. Technol. Interact.* **2018**, *2*, 47. [[CrossRef](#)]
9. Jiang, F.; Jiang, Y.; Zhi, H.; Dong, Y.; Li, H.; Ma, S.; Wang, Y.; Dong, Q.; Shen, H.; Wang, Y. Artificial intelligence in healthcare: Past, present and future. *Stroke Vasc. Neurol.* **2017**, *2*, 230–241. [[CrossRef](#)]
10. Schickel, B.; Tighe, P.; Bihorac, A.; Rashidi, P. Deep HER: A survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. *IEEE J. Biomed. Health Inform.* **2018**, *22*, 1589–1604. [[CrossRef](#)]

11. Wang, W.; Wang, H.; Chen, Q.; Zhou, Z.; Wang, R.; Wang, H.; Zhang, N.; Chen, Y.; Sun, Z.; Xu, L. Coronary artery calcium score quantification using a deep-learning algorithm. *Clin. Radiol.* **2020**, *75*, e11–e237. [[CrossRef](#)]
12. Shadmi, R.; Mazo, V.; Bregman-Amitai, O.; Elenkave, E. Fully-convolutional deep-learning based system for coronary calcium score prediction from non-contrast chest CT. In Proceedings of the IEEE 15th International Symposium on Biomedical Imaging, Washington, DC, USA, 4–7 April 2018; pp. 24–28.
13. Zreik, M.; Hamersvelt, R.W.; Wolterink, J.M.; Leiner, T.; Viergever, M.A.; Isgum, I. A recurrent CNN for automatic detection and classification of coronary artery plaque and stenosis in coronary CT angiography. *IEEE Trans. Med. Imaging* **2019**, *38*, 1588–1598. [[CrossRef](#)]
14. Santini, G.; Latta, D.D.; Martini, N.; Valvano, G.; Gori, A.; Ripoli, A.; Susini, C.L.; Landini, L.; Chiappino, D. An automatic deep learning approach for coronary artery calcium segmentation. In Proceedings of the European Medical and Biological Engineering Conference, Tampere, Finland, 11–15 June 2017; pp. 374–377.
15. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the 18th Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; LNCS 9351. Springer: Berlin/Heidelberg, Germany; pp. 234–241.
16. Schläpfer, J.; Wellens, H.J. Computer-interpreted electrocardiograms. *J. Am. Coll. Cardiol.* **2017**, *70*, 1183–1191. [[CrossRef](#)]
17. Mele, P.F. The ECG dilemma: Guidelines on improving interpretation. *J. Healthc. Risk Manag.* **2008**, *28*, 27–31. [[CrossRef](#)] [[PubMed](#)]
18. Sengupta, P.P.; Kulkarni, H.; Narula, J. Prediction of abnormal myocardial relaxation from signal processed surface ECG. *J. Am. Coll. Cardiol.* **2018**, *71*, 1650–1659. [[CrossRef](#)] [[PubMed](#)]
19. Galloway, C.D.; Valys, A.V.; Shreibati, J.B.; Treiman, D.L.; Petterson, F.L.; Gundotra, V.P.; Albert, D.E.; Attia, Z.I.; Carter, R.E.; Asirvatham, S.J.; et al. Development and validation of a deep-learning model to screen for hyperkalemia from the electrocardiogram. *J. Am. Med. Assoc. Cardiol.* **2019**, *4*, 428–436. [[CrossRef](#)]
20. Attia, Z.I.; Kapa, S.; Lopez-Jimenez, F.; McKie, P.M.; Ladewig, D.J.; Satam, G.; Pellikka, P.A.; Enriquez-Sarano, M.; Noseworthy, P.A.; Munger, T.M.; et al. Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. *J. Am. Med. Assoc. Cardiol.* **2019**, *25*, 70–74. [[CrossRef](#)] [[PubMed](#)]
21. Mathews, S.M.; Kambhamettu, C.; Barner, K.E. A novel application of deep learning for single-lead ECG classification. *Comput. Biol. Med.* **2018**, *99*, 53–62. [[CrossRef](#)] [[PubMed](#)]
22. Ullah, A.; Anwar, S.M.; Bila, M.; Mehmood, R.M. Classification of arrhythmia by using deep learning with 2-D ECG spectral image representation. *Remote Sens.* **2020**, *12*, 1685. [[CrossRef](#)]
23. Ebrahimi, Z.; Loni, M.; Daneshlab, M.; Gharebaghi, A. A review on deep learning methods for ECG arrhythmia classification. *Expert Syst. Appl. X* **2020**, *7*, 100033. [[CrossRef](#)]
24. Peimankar, A.; Puthusserypady, S. DENS-ECG: A deep learning approach for ECG signal delineation. *arXiv* **2020**, arXiv:2005.08689.
25. Ribeiro, A.H.; Ribeiro, M.H.; Paixão, G.M.M.; Oliveira, D.M.; Gomes, P.R.; Canazart, J.A.; Ferreira, M.P.S.; Andersson, C.R.; Macfarlane, P.W.; Meira, W., Jr.; et al. Automatic diagnosis of the 12-lead ECG using a deep neural network. *Nat. Commun.* **2020**, *11*, 1760. [[CrossRef](#)]
26. Sabour, S.; Grobbee, D.; Rutten, A.; Prokop, M.; Bartelink, M.; Schouw, Y.V.D.; Bots, M. Electrocardiogram abnormalities and coronary calcification in postmenopausal women. *J. Tehran Heart Cent.* **2010**, *5*, 19–24. [[CrossRef](#)]
27. Shi, L.; Tashiro, S. Estimation of the effects of medical diagnostic radiation exposure based on DNA damage. *J. Radiat. Res.* **2018**, *59*, ii121–ii129. [[CrossRef](#)] [[PubMed](#)]
28. Clinical ECG Interpretation, the ECG Leads: Electrodes, Limb Leads, Chest (Precordial) Leads, 12-Lead ECG (EKG). Available online: <https://ecgwaves.com/topic/ekg-ecg-leads-electrodes-systems-limb-chest-precordial/> (accessed on 17 November 2020).
29. Clinical ECG Interpretation, Cardiac Electrophysiology: Action Potential, Automaticity and Vectors. Available online: <https://ecgwaves.com/topic/cardiac-electrophysiology-ecg-action-potential-automaticity-vector/> (accessed on 17 November 2020).

30. Manisha; Dhull, S.K.; Singh, K.K. ECG beat classifiers: A journey from ANN to DNN. *Procedia Comput. Sci.* **2020**, *167*, 747–759. [[CrossRef](#)]
31. Larner, A.J. The Q* index: A useful global measure of dementia screening test accuracy? *Dement. Geriatr. Cogn. Disord. Extra* **2015**, *5*, 265–270. [[CrossRef](#)]

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).