Computers in Biology and Medicine 105 (2019) 92–101

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Automated beat-wise arrhythmia diagnosis using modified U-net on extended electrocardiographic recordings with heterogeneous arrhythmia types

Shu Lih Oh[a], Eddie Y.K. Ng[b], Ru San Tan[c], U. Rajendra Acharya[a,d,e,*]

[a] Department of Electronics and Computer Engineering, Ngee Ann Polytechnic, Singapore
[b] School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore
[c] National Heart Centre Singapore, Singapore
[d] Department of Biomedical Engineering, School of Science and Technology, Singapore University of Social Sciences, Singapore
[e] School of Medicine, Faculty of Health and Medical Sciences, Taylor's University, 47500, Subang Jaya, Malaysia

ARTICLE INFO

Keywords:
ECG
Deep learning
U-net
Autoencoder
Arrhythmia

ABSTRACT

Abnormality of the cardiac conduction system can induce arrhythmia — abnormal heart rhythm — that can frequently lead to other cardiac diseases and complications, and are sometimes life-threatening. These conduction system perturbations can manifest as morphological changes on the surface electrocardiographic (ECG) signal. Assessment of these morphological changes can be challenging and time-consuming, as ECG signal features are often low in amplitude and subtle. The main aim of this study is to develop an automated computer aided diagnostic (CAD) system that can expedite the process of arrhythmia diagnosis, as an aid to clinicians to provide appropriate and timely intervention to patients. We propose an autoencoder of ECG signals that can diagnose normal sinus beats, atrial premature beats (APB), premature ventricular contractions (PVC), left bundle branch block (LBBB) and right bundle branch block (RBBB). Apart from the first, the rest are morphological beat-to-beat elements that characterize and constitute complex arrhythmia. The novelty of this work lies in how we modified the U-net model to perform beat-wise analysis on heterogeneously segmented ECGs of variable lengths derived from the MIT-BIH arrhythmia database. The proposed system has demonstrated self-learning ability in generating class activations maps, and these generated maps faithfully reflect the cardiac conditions in each ECG cardiac cycle. It has attained a high classification accuracy of 97.32% in diagnosing cardiac conditions, and 99.3% for R peak detection using a ten-fold cross validation strategy. Our developed model can help physicians to screen ECG accurately, potentially resulting in timely intervention of patients with arrhythmia.

1. Introduction

Arrhythmia — abnormal heart rhythm — is a group of conditions in which the heartbeat is too slow or too fast, and can be either regular or irregular. It is caused by changes in the cardiac conduction system due to inheritable or acquired pathologies, or is the result of aging. Indeed, the incidence of arrhythmia increases with global population aging [1,2]. As one ages, the cardiovascular system undergoes structural and functional changes brought about by increased apoptosis of cardiomyocytes, fibrofatty degeneration of heart muscle [3,4] and conductive tissue. The disruption of the cardiac conduction system induces arrhythmia. Some arrhythmias induce adverse cardiac remodeling and heart failure (tachycardia-induced cardiomyopathy), or can result in cardiovascular complications like thromboembolic stroke (from atrial fibrillation) [5]. Less commonly, arrhythmia can be life threatening due to compromise of mechanical cardiac output, for example during ventricular tachycardia, complete heart block, etc.

The current medical routine for arrhythmia screening involves manual examination of ECGs by skillful clinicians. This process is often laborious and taxing. The ECGs taken in clinical settings are usually insufficient for the doctors to diagnose the activity of the heart comprehensively. Therefore, diagnosis of suspected arrhythmias typically requires patients to wear a small recorder over their chest for continuous monitoring of the heart's functioning during daily activities [6]. Data collected from such devices often last over a day or two. The interpretation of ECG recordings is based on identification of the abnormal heart beats, which is a manually intensive process. The heart beats are manually classified into the most common types: normal sinus

* Corresponding author. Department of Electronics and Computer Engineering, Ngee Ann Polytechnic, Singapore, 599489.
E-mail address: aru@np.edu.sg (U.R. Acharya).

https://doi.org/10.1016/j.compbiomed.2018.12.012
Received 14 November 2018; Received in revised form 18 December 2018; Accepted 18 December 2018

beats, atrial premature beats (APB), premature ventricular contraction (PVC), left bundle branch block (LBBB) and right bundle branch block (RBBB) [7–11]. Apart from the first, the rest are morphological beat-to-beat elements that characterize and constitute complex arrhythmia.

Many authors (shown in Table 5)have employed conventional machine learning methods using some form of pre-processing, including handcrafted features extraction and selection for automated arrhythmia detection [12–23]. Engineering a handcrafted feature is no easy task: even with good domain knowledge, repeated trial-and-error is obligatory. Furthermore, machine learning methods tend to have more false positive results [24]. Therefore, the deep learning method is used in this work to alleviate these problems, so as to achieve a better diagnosis performance, by feeding the ECG signals without preprocessing.

In recent years, deep learning has overtaken classical machine learning techniques. Deep learning algorithms are increasing being employed in healthcare for complex tasks, such as segmentation of brain images [25] and clinical decision-making [26]. For arrhythmia detection, several deep learning models have shown promise [27–31]. The convolutional neural network (CNN) and long short-term memory (LSTM) network are the two most commonly-used algorithms for ECG arrhythmia classification. CNN is robust to noise, and is thus able to extract useful predictors even when the data are noisy [32]. This property is fully realized in a deep hierarchical structure, where the learned features tend to become more abstract as the network gets deeper. Acharya et al. [33] studied CNN for diagnosis of myocardial infarction on ECG, and found only a marginal decrease in classification accuracy for noisy compared to non-noisy ECG signals. The rationale for using LSTM networks in ECG analysis lies in its ability to interpret the idea of time. They can hence learn complex temporal dynamics within time-varying data, which explains their utility in natural language translation [34], speech analysis [35–38] and handwriting recognition applications [38–40]. While both CNN and LSTM networks can adequately distinguish morphological differences in ECG signals, they are less adept at detecting subtle differences that are finely resolved. Importantly, prediction models are based on ECG signal readouts with single arrhythmic conditions [27,29–31,33,41,42,59,60]. However, in real-life, ECG signals are of variable length and may contain mixed arrhythmic conditions. The question of whether deep learning can accurately classify variable length ECGs has not been investigated. We thus aimed to develop a deep learning autoencoder model for beat-wise analysis of ECG signals of variable lengths, for arrhythmia classification.

A deep autoencoder operates by encoding the original data into a lower dimension through a series of compression steps. The model then learns to decode and express the data as output. Since locality information of the data is preserved during compression, restoring the compressed data back to its original form is easily achievable [43]. In the domain of ECG studies, Yildirim et al. [44] exploited this property to model a compression system for ECG signals. Other studies utilized the model for signal preprocessing and noise reduction [45,46]. The autoencoder is the most frequently used model in image segmentation and pixel-wise classification [47–50]. The U-net is a convolutional autoencoder developed with skipped connection to perform high precision cellular segmentation on microscopic images [48]. The skipped connections are added to the network to recover the spatial information that are lost during compressions. As a result, the U-net is able to produce localized outputs of higher resolution when compared to the conventional autoencoder. Hitherto, the autoencoder has not been explored for temporal and beat-wise classification of ECG signals.

2. Data description

We studied 48 ECG recordings obtained from 47 subjects, all sampled at 360 Hz, derived from the MIT-BIH arrhythmia dataset from PhysioNet [51]. Twenty-three recordings were of the 24-h ambulatory ECG type, collected at Boston's Beth Israel Hospital, and twenty-five were specially selected to include those arrhythmias of clinical significance. Each ECG record was scrutinized carefully, and beat-by-beat diagnostic classification was rendered at the beginning of each beat, i.e., the R peak of the ECG signal, by cardiologists through mutual consensus, where necessary. For this study, we analyzed only modified limb lead II ECG signals, which were acquired via torso electrodes. Being a bipolar lead, modified limb lead II exhibited similar measured potential as Einthoven (standard) limb lead II [52].

Table 1
Number of ECG segments and their corresponding conditions.

Condition(s) present in the segment	No of Segments
Normal	21253
Normal, APB	445
Normal, APB, PVC	25
Normal, PVC	4156
Normal, RBBB	27
APB	411
APB, PVC	1
APB, PVC, RBBB	2
APB, LBBB	2
APB, RBBB	340
PVC	342
PVC, LBBB	267
PVC, LBBB, RBBB	2
PVC, RBBB	87
LBBB	2444
LBBB, RBBB	1
RBBB	2327
Total	32132

3. Heterogeneous segmentation

The data used for training the U-net were preprocessed into standardized 1000-sample length segments that may or may not contain mixed arrhythmia beats. The ECG signals were segmented into 99 samples before the first annotated R peak and 160 samples after the last identified R peak. When the subsequent beat was not of interest, or when the length of segment exceeded 1000 samples, the process was terminated with the preceding R peak identified as the last R peak of the segment. Zeros were padded for the segments having fewer than 1000 samples. Table 1 shows the number of segmented ECG signals with the corresponding R-peak annotated conditions within.

Without constraining all the beats within individual segments to a single particular arrhythmia condition, the segments could contain multiple arrhythmia conditions. Heterogeneous segmentation of the ECG records allowed more data to be analyzed, and rendered the training data more diversified and complex. Table 2 depicts the total number of beats corresponding to the R-peak annotated conditions found among all 32132 segmented signals, that was available for training and testing the U-net.

Deep learning models often take a lot of time to train, a process that can be accelerated by normalization of the intrinsic data value variation that occur naturally. Normalization compresses the values of the original data by scaling the values to a smaller range, thereby helping to

Table 2
Total number of ECG beats by respective condition.

Types	No of ECG beats
Normal	71337
APB	2123
PVC	6194
LBBB	7890
RBBB	7123
Total	94667

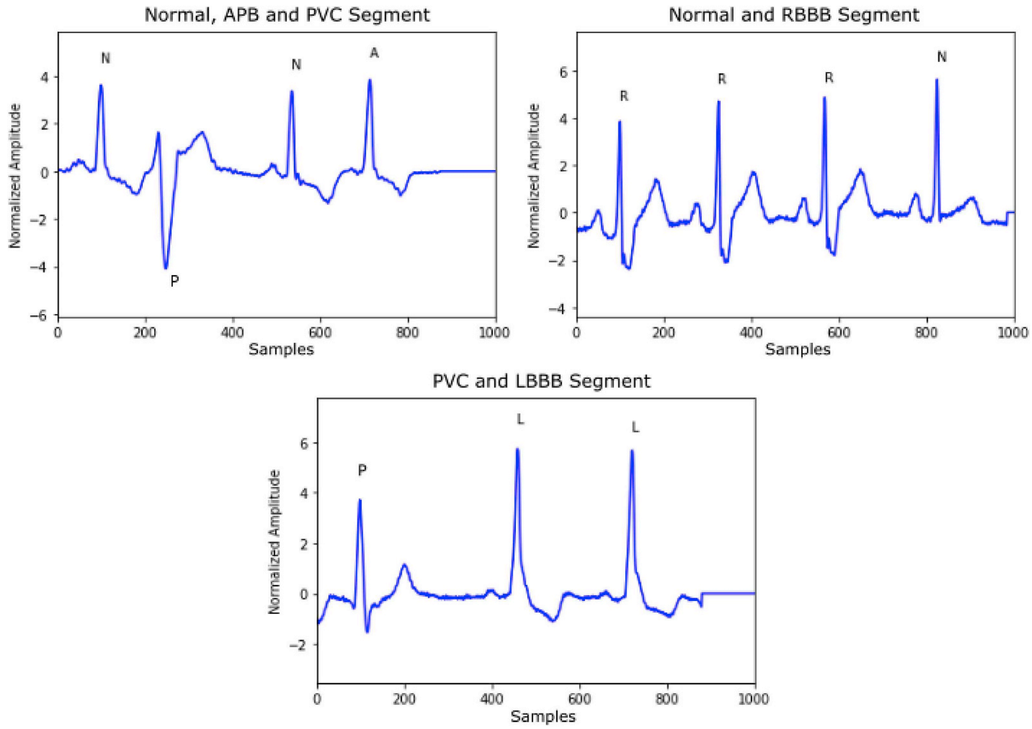

Fig. 1. Normalized ECG signals that were heterogeneously segmented with annotated R peaks (N = Normal, L = LBBB, R = RBBB, A = APB, P = PVC).

standardize the values as well as improving the backpropagation process by speeding up the convergence rate. In this study, Z-score normalization was used in amplitude scaling of the ECGs [53] Fig. 1 depicts examples of normalized ECG segments used in training the U-net model.

4. Modified U-net model

ECGs are intrinsically heterogenous — often containing combinations of beats, conditions and sequence patterns. We developed a U-net model that could handle the heterogeneity and extract localized information from the ECG segment for beat-wise analysis. In this project, the U-net was modified with multiple classification heads: one head for R peak detection; the others, for mapping out the conditions. Fig. 2 illustrates the modified U-net architecture; Table 3 shows the sequential workflow of the newly proposed model.

The bulk of the proposed U-net comprised 1D convolution operations of kernel size 3×1. Unlike the conventional U-net model which uses valid convolution [48], we applied the same convolution operation to obtain output feature maps of the same size. Cropping was therefore not required for concatenation of feature maps. This model has three compression stages. With each compression, the feature maps were halved and the number of feature maps doubled. In the expansive path, high resolution features were copied directly from the contracting path and then combined with the upsampled features for subsequent convolutions, effectively transmitting the encoded context information to the subsequent layer. At layer 25, 5×1 kernels were used for convolving the R peak prediction branch while 3×1 kernels were used in the confidence map branch. Finally, 1×1 convolution was used in the last layer for reducing the number of feature maps to the corresponding training targets [54]. Since the model allowed samples other than those located at the R peaks to converge freely to any class, a global average pooling layer was added to the model during training to prevent the confidence maps from converging to a non-existent class within the segment.

5. Training targets

In this project, the proposed U-net has three training targets. The first training targets were for peak prediction. Peak prediction training targets were generated by converting the segment annotations to a binary vector, where the annotated R peaks were set to 1, while the other samples of the segment were set to 0. The second training targets localized the conditions. A 5×1000 array was created for each segment. Depending on the annotated conditions, the corresponding class row of the R peak was set to 1, while the other rows were set to 0. Columns with no annotated conditions were set to -1. These columns were ignored during the training procedure and no loss would be backpropagated — this enabled the outputs to converge to any condition without restriction. Lastly, the training targets for class presence were acquired to prevent the confidence map from converging to a class that was non-existent within the segment. The class presence targets was a binary encoded class vector whereby conditions present within the entire segment were set to 1 and classes that were not present in the segment were set to 0.

6. Global pooling

Unlike conventional pooling operations, the filter size for global average pooling operations — first described by Lin et al. [54] — is defined the same as the size of input. Hence, the features dimensionality of a global pooled feature map is vastly reduced by outputting only a single element. In this project, global average pooling was applied in the final layer of the U-net model for generation of class activation maps (CAM). The application was done by replacing the dense network structure with global average pooling operations to generate a single class corresponding feature element for classification. Instead of directly vectorizing the feature maps and feeding them into fully connected layers for class prediction, each feature map was averaged and softmaxed. This conformed the final layer of the model to learn the correspondences between feature maps and their respective categories. As a result, visualization of class specific confidence maps was easily

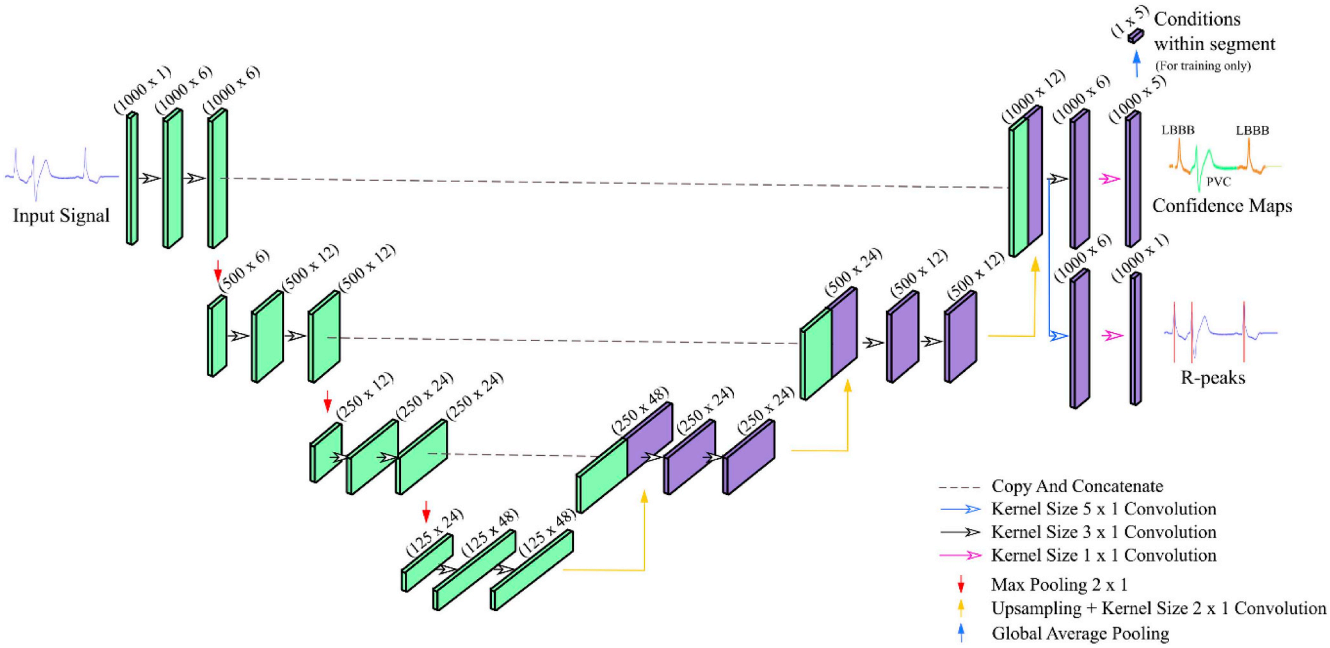

Fig. 2. Modified U-net architecture.

Table 3
Overview of the proposed U-net model.

Layers	Types	Activation function	Output Shapes	Size of kernel	No. of Filters	Stride	No. of trainable parameters
0	Input	–	1000 × 1	–	–	–	0
1	1D Same convolution	ReLU	1000 × 6	3 × 1	6	1	24
2	1D Same convolution	ReLU	1000 × 6	3 × 1	6	1	114
3	1D Max-pooling	–	500 × 6	2 × 1	6	2	0
4	1D Same convolution	ReLU	500 × 12	3 × 1	12	1	228
5	1D Same convolution	ReLU	500 × 12	3 × 1	12	1	444
6	1D Max-pooling	–	250 × 12	2 × 1	12	2	0
7	1D Same convolution	ReLU	250 × 24	3 × 1	24	1	888
8	1D Same convolution	ReLU	250 × 24	3 × 1	24	1	1752
9	1D Max-pooling	–	125 × 24	2 × 1	24	2	0
10	1D Same convolution	ReLU	125 × 48	3 × 1	48	1	3504
11	1D Same convolution	ReLU	125 × 48	3 × 1	48	1	6960
12	Upsampling	–	250 × 48	–	–	–	0
13	1D Same convolution	ReLU	250 × 24	2 × 1	24	1	2328
14	Concatenate Layer 13 & Layer 8 outputs	–	250 × 48	–	–	–	0
15	1D Same convolution	ReLU	250 × 24	3 × 1	24	1	3480
16	1D Same convolution	ReLU	250 × 24	3 × 1	24	1	1752
17	Upsampling	–	500 × 24	–	–	–	0
18	1D Same convolution	ReLU	500 × 12	2 × 1	12	1	588
19	Concatenate Layer 18 & Layer 5 outputs	–	500 × 24	–	–	–	0
20	1D Same convolution	ReLU	500 × 12	3 × 1	12	1	876
21	1D Same convolution	ReLU	500 × 12	3 × 1	12	1	444
22	Upsampling	–	1000 × 12	–	–	–	0
23	1D Same convolution	ReLU	1000 × 6	2 × 1	6	1	150
24	Concatenate Layer 22 & Layer 2 outputs	–	1000 × 12	–	–	–	0
25	1D Same convolution	ReLU	1000 × 6	3 × 1	6	1	222
	1D Same convolution	ReLU	1000 × 6	5 × 1	6	1	366
26(i)	1D Same convolution (Confidence map)	Softmax	1000 × 5	1 × 1	5	1	35
	1D Same convolution (Peak prediction)	Sigmoid	1000 × 1	1 × 1	1	1	7
26(ii)	Global Average Pooling (Conditions within segment)	–	1 × 5	–	–	–	0
						Total	24162

achievable by plotting the feature maps from the final convolution layer.

Since a global average pooling operation does not utilize any learnable parameters, the model is forced to optimize through the learnable parameters from other layers. Consequently, such a layer is less likely to overfit as compared with a fully connected structure. Additionally, by averaging the spatial information, it renders the model more invariant to spatial translations.

7. Training and evaluation

The network was evaluated using ten-fold a cross-validation strategy. A stratified sampling method was used to subdivide the ECG dataset into 10 equal portions in accordance with the conditions present within the segments. This is done so as to ensure that each fold will have approximately the same number of segments and conditions combination. Training of the network model used 9 portions of the ECG

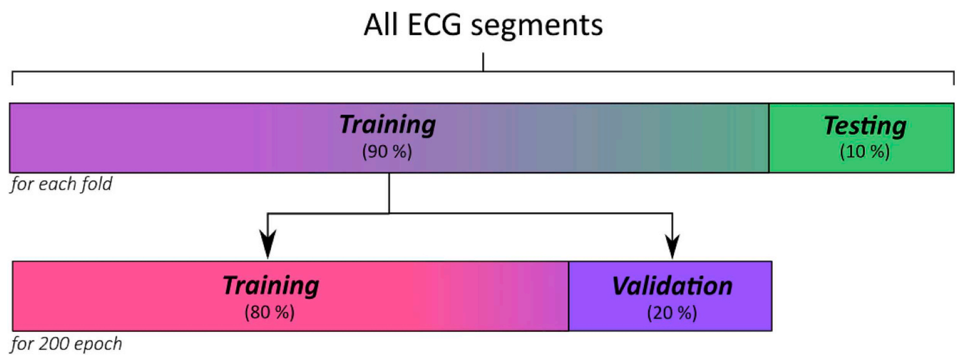

Fig. 3. Data distribution of the ECG segments used for training and testing the proposed network.

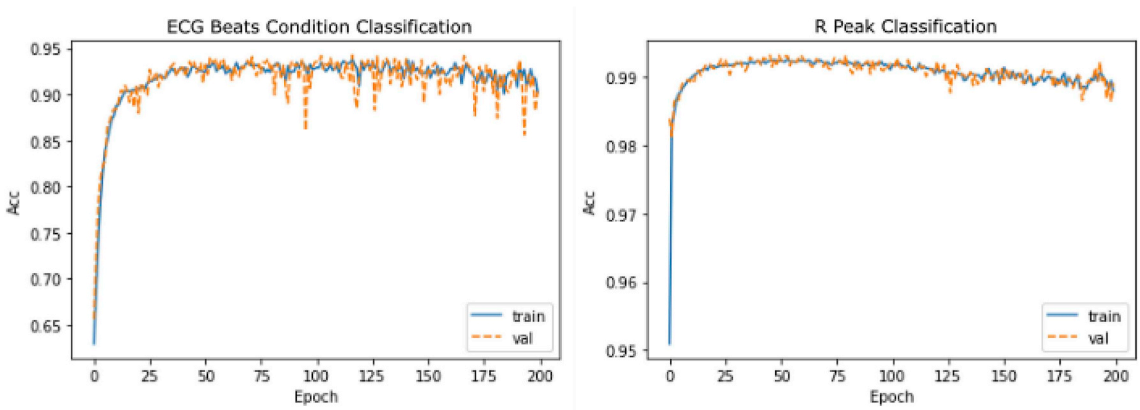

Fig. 4. Accuracy plots from the multiple classification heads of the U-net model.

Table 4
Average performance of the best U-net model across all 10 folds.

Classification Heads	Acc % ± SD		Sens % ± SD		Spec % ± SD		PPV % ± SD	
Conditions for ECG Beats	*97.32*	*± 0.66*	*94.44*	*± 2.74*	*98.26*	*± 0.24*	*94.7*	*± 0.76*
R peak	*99.3*	*± 0.21*	*29.55*	*± 5.31*	*100*	*± 0*	*98.76*	*± 0.89*

segments; testing, the remaining portion. The procedure was repeated 10 times. Each time the model was reinitialized and tested with a different data subset. In order to examine the progress of training the models, 20% of the training set was isolated for validation. Details for the ten-fold cross-validation are presented in Fig. 3.

For each training fold, weights of the networks were initialized with the Xavier algorithm [55]. The model was trained and the back-propagation process was accelerated using the algorithm with the Adam optimizer [56]. The modified U-net model was trained with a batch size of 20, and the learning the rate (η) was 0.0005. To mitigate class imbalance, a weighted class variable was introduced to the losses calculation. The performance for each fold was evaluated using accuracy (ACC), sensitivity (SEN), specificity (SPEC) and positive predictive value (PPV).

8. Results

The network was developed in Python using Keras [57] for easy prototyping and TensorFlow as the backend deep learning library [58]. The specifications of the workstation used for training the models consisted of two Intel Xeon 2.40 GHz (E5620) processors and a 24 GB RAM. The U-net model was trained for 200 epochs; each training epoch took approximately 120.13 s to run. The accuracy curves for beat classification and R peak detection averaged across all 10 folds are depicted in Fig. 4. Accuracy for classifying the beat-wise condition was evaluated based on the annotations provided at the R peaks.

The proposed U-net model was able to generalize well from the training data without any additional network regularization, and without signs of overfitting, as evidenced by learning curves that overlapped closely (Fig. 4). A few factors may have contributed to the observed good generalization ability of the modified U-net. First, the data used for training and testing the U-net were diversified and complex owing to the multitude of conditions each ECG segment could contain. Second, reduction in the number of learnable parameters mitigated the risks of the model overfitting. This was achieved through (1) use of kernels in the U-net model that were smaller in size relative to CNN-LSTM models; and (2) inclusion of global average pooling, itself a structural regularizer that does not utilize any learnable parameter, which forced the model to learn from averaged information instead.

Training and validation accuracies obtained from the two classification heads were fairly stable. Accuracy curves for ECG beats classification plateaued after 50 epochs while those for R peak identification declined slightly after training for 75 epochs. The overall cross validation performances for the modified U-net are summarized in Table 4.

The proposed U-net was able to identify the beats condition and predict the R peaks with accuracies of 97.32% and 99.3%, respectively. The confusion matrices for beats classification and R peak prediction are presented in Fig. 5 and Fig. 6, respectively. The proposed model is able to correctly identify most of the ECG beats conditions (normal, LBBB, RBBB and PVC) except for APB beats, where almost half were

Table 5
Automated detection of the arrhythmias using conventional techniques.

Author	Method	Database	Classifier	Performance(%)
[12]	ECG morphology and heart rate features	MITDB	SVM	ACC: 98.39, SEN: 99.87, PPV: 99.69
[13]	beat to beat interval and entropies extracted from Wavelet Packet Decomposition (WPD)		Random forest	ACC: 94.61
[14]	DWT + PCA HOS + ICA	MITDB	SVM	ACC: 98.91, SEN: 98.91, SPEC: 97.85
[15]	DWT + PCA DWT + ICA	MITDB	SVM	ACC: 98.8, SEN: 98.50, SPEC: 99.69, PPV: 98.91
[16]	DWT sub bands + ICA	MITDB	Probabilistic Neural Network (PNN)	ACC: 99.28, SEN: 97.97, SPEC: 99.83, PPV: 99.21
[17]	HOS + PCA	MITDB	Least Square-Support Vector Machine (LS-SVM)	ACC: 93.48, SEN: 99.27, SPEC: 98.31
[18]	PCA	MITDB	LS-SVM	ACC: 98.11, SEN: 99.90, SPEC: 99.10, PPV: 99.61
[19]	Hermite model of HOS	MITDB	1-Nearest Neighborhood	SPEC: 99.67, SEN: 98.66
[20]	WPD + HOS	MITDB	SVM	ACC: 98.48, SEN: 98.90, SPEC: 98.04, PPV: 98.13
[21]	Morphological and heart rate based features	MITDB	Linear Discriminant Analysis	Accuracy: 96.23 SENS: Normal 98.97, LBBB 91.07, RBBB 95.09, PVC 92.63, APB 84.68
[22]	time elapse between R peaks and ICA		PNN	ACC: 98.71
[23]	HOS and Hermite coefficients extracted from the QRS wave	MITDB	SVM	ACC: 98.71

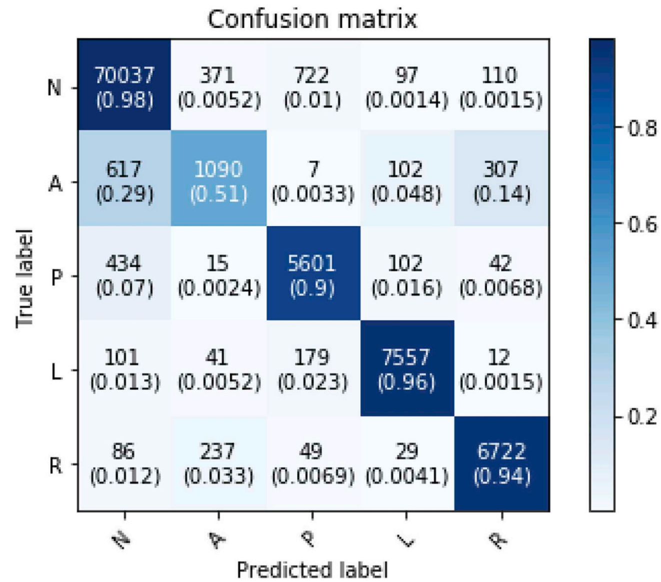

Fig. 5. Confusion matrix of the classified ECG beats (N = Normal, L = LBBB, R = RBBB, A = APB, P = PVC).

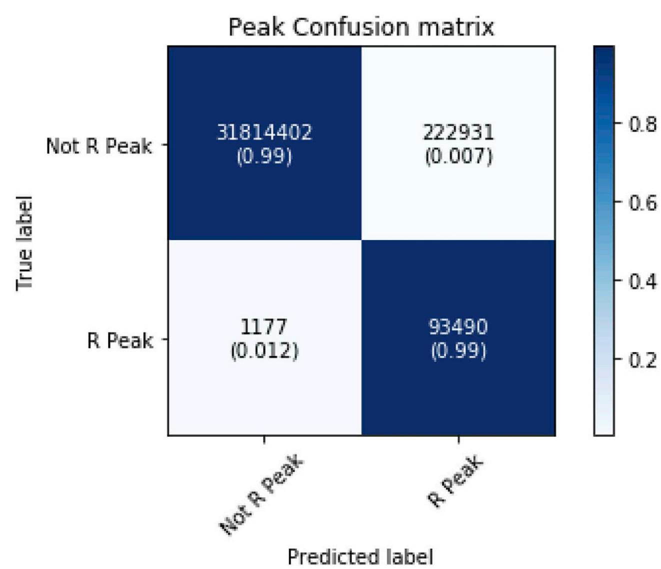

Fig. 6. Confusion matrix for R peak prediction.

misclassified as either normal or RBBB. The former may be due to (1) true absence of a preceding P wave, where APB originates from the atrioventricular junction rather than the atrium; (2) failure to detect low amplitude ectopic P waves; and (3) superimposition of ectopic P waves on the preceding T waves, which the network failed to differentiate. The latter may be contributed by the fact that many APB beats might have been aberrantly conducted as well, and therefore exhibit predominant RBBB morphology. This cannot reconcile with the expert cardiologist annotation and consequent U-net model, which were predicated on distinct mutually exclusive ECG diagnostic classification. Finally, APBs beats were proportionately least represented in the dataset (Table 2), which could arguably have affected recognition training of the underlying morphological features compared to the other conditions.

The algorithm was able to identify both the non-R peak and R peak samples at an accuracy of 99%. The low sensitivity of 29.55% for the R peak prediction was caused by the misclassification of surrounding

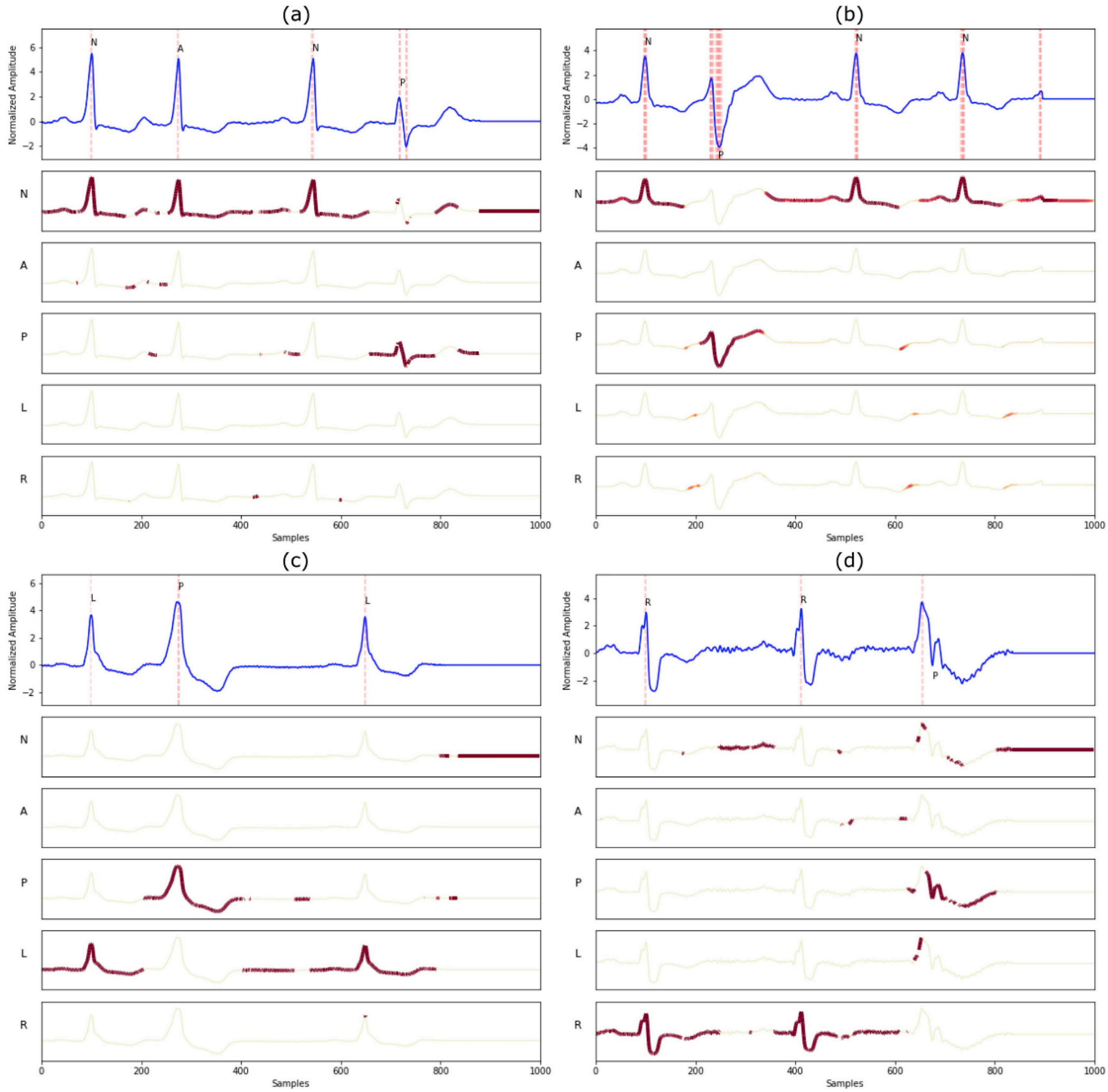

Fig. 7. Annotated ECG segments (blue) along with the predicted R peaks (red vertical dotted lines). Below each ECG segment is the corresponding class activation maps produced by the modified U-net model. The highly activated areas are the ones depicted in red.

samples at the R peak. This is more likely to occur when the beat has a wide QRS complex with a small positive R wave followed by a large negative S wave deflection, where the algorithm would classify one or both extrema, and sometimes the samples in between, as R peaks (see Fig. 7a and b).

Fig. 7 depicts four test ECG segments with the corresponding activation map below each. Each of these activation maps corresponds to a condition in which the most discriminative regions within the segments are highlighted in red. Through visualization of the activation maps, it is clear that the U-net has the ability to identify most classes except for the APB class, with several subsequences of the Normal, PVC, LBBB and RBBB beats highlighted in the appropriate regions. The model has demonstrated good localization capability, underscoring the correct attentions taken by the network during classification.

9. Discussion

Over the years many automated diagnostic systems have been developed on ECG arrhythmia classification [51]. Table 5 summarizes published studies on automated detection of arrhythmia using conventional machine learning techniques; and Table 6, deep learning techniques. It should be noted that not all arrhythmias studied were identical.

Conventional machine algorithms often require complex feature engineering and extensive domain knowledge. Features need to be carefully selected before they can be the used for prediction. Dimensionality reduction techniques are frequently applied to the features for easier processing. A few of these machine learning studies used linear features [12,13,21], nonlinear features [14–17,19,20,23], and wavelet transformed coefficients [16] for classification.

Table 6
Automated detection of arrhythmias using deep learning approach.

Author	Method	Database	Analyzed data	Deep learning structure	Performance(%)
[42]	–	MITDB	Normal (8245) APB (1004) PVC (6246) LBBB (344) RBBB (660) **Total 1000 samples sequence (16499)**	CNN-LSTM	ACC: 98.10 SENS: 97.50, SPEC: 98.70
[31]	Discrete wavelet transform (DWT)	MITDB	Normal (2190) LBBB (1870) RBBB (1356) PVC (510) Paced (1450) **Total (7376)**	Bidirectional long short-term memory (Bi-LSTM) networks	ACC: 99.39
[27]	–	MITDB AFDB CUDB	Normal Atrial fibrillation Atrial flutter Ventricular Fibrillation **Two seconds (21709)** **Five seconds (8683)**	CNN	Two seconds ACC: 92.50, SENS: 98.09, SPEC: 93.13 Five seconds ACC: 94.90, SENS: 99.13, SPEC: 81.44
[28]	–		Normal (90592) Supra-ventricular ectopic (2781) Ventricular ectopic (7235) Fusion (802) Unknown (8039) **Total (109449)**	CNN	ACC: 94.03, SENS: 96.71, SPEC: 91.54
[29]	–	MITDB	Normal Supra-ventricular ectopic Ventricular ectopic Fusion Unknown **Total (100389)**	CNN	ACC: 92.70
[30]	–	MITDB	Normal Supra-ventricular ectopic Ventricular ectopic Fusion Unknown **Total (83648)**	CNN	ACC: 99.00, SENS: 93.90, SPEC: 98.90
This work	–	MITDB	Normal 71337 APB 2123 PVC 6194 LBBB 7890 RBBB 7123 **Total 94667**	modified U-net	ACC: 97.32% (arrhythmia)and 99.3%(R-peaks)

Yeh et al. [21] applied linear discriminant analysis for classification using only morphological features extracted from the ECG component. The difference operation method (DOM), comprising a few threshold filters, was used in identifying the PQRST components on the ECG. Upon identifying the PQRST waves, the morphological features were extracted. The study obtained an accuracy of 96.23% using just four features. In another morphology based study, Karimifard et al. [19] extracted cumulants from the ECG beats using the Hermite model. The method was effective in suppressing morphological differences within classes, thereby attenuating the effects of time, amplitude shift, and noise.

Martis et al. [18] implemented principal component analysis (PCA) on ECG beats and achieved a 98% classification accuracy. Only a dozen PCA components were selected and used for training the least squares-support vector machine classifier. The following year, Martis et al. [17] tested the effect of coupling PCA with Higher Order Statistics (HOS) for features extraction, to identify the morphological differences in arrhythmic ECG beats. The same reduction was used by Li et al. [15] to extract the components from discrete wavelet transformed (DWT) signals. The study was able to attain a 97.3% classification accuracy through implementation of various reduction methods on the wavelet transformed ECG. All of these conventional machine studies confirm that arrhythmias underscore the utility of the CAD system to identify conduction abnormalities in the ECG and improve the efficacy of arrhythmia detection.

LSTM network was employed in a recent study by Tan et al. [41] to diagnose coronary artery disease using ECG segments. The process involves splitting the 5-s ECG signals into shorter segments and then performing convolution operations on the short segments, after which the LSTM was used to map the convolved segments into temporal features for classification. The model achieved a diagnostic accuracy of 99.85%.

Yildirim et al. [31] explored the use of LSTM on decomposed ECG beats for arrhythmia diagnosis. The investigators applied discrete wavelet transformation on the ECG cycle, and used features mapped by the bidirectional LSTM for arrhythmia classification. The study attained a 99.39% accuracy in classification.

Many researchers either assumed that each ECG segment consists of only one arrhythmia type [27,33,41,42] or made the prediction based only on a single beat segment that had to be identified first [29–31]. To obviate these artificial restrictions, we modified the U-net model to handle ECG segments with mixed arrhythmia types, by having multiple classification heads for (1) detecting the R peaks; and (2) simultaneously identifying the conditions in the time series. A global average pooling layer was added to the final layer of the U-net to obtain the class activation maps for each condition. The proposed U-net model demonstrated good generalization ability without overfitting during training. The performance of the modified U-net is promising: the accuracy for classifying conditions for individual beats compared to the annotations provided at the R peaks was 97.3%; and accuracy for detecting the R peak, 99.3%. The class activations maps in Fig. 7 suggest that the model is capable of differentiating the segments into subsequences and associating them with the correct conditions.

The advantages of the newly proposed U-net model are:

1. The proposed system is fully automated.
2. Observer bias is eliminated.
3. End-to-end solution, requires minimal processing.
4. Standalone classification heads for R peak detection and classification of ECG conditions.
5. Localized predictions.
6. Robustness of the system is assessed by cross validation testing.
7. Generated maps faithfully reflect the cardiac conditions in each ECG cardiac cycle.

The drawbacks of the newly proposed algorithm are:

1. Subtle changes and overlapping of waves lead to misclassification of the APB class.
2. The training phase is computationally intensive and slow.
3. Limited availability of APB class data compared to other conditions.
4. The model is trained and tested using an imbalanced dataset.
5. Predictions for R peak are not precise.

The ultimate benefits of implementing a deep learning network are to minimize the number of preprocessing techniques required, and to allow the system to be trained end-to-end. The newly developed U-net model is unique compared to other deep learning techniques, as it does not make any assumption regarding the input segments. Theoretically, all of the operations used in the U-net have the ability to handle variable size data unlike the fully connected layers that are used in conventional CNN models, which can only address input of a fixed size.

10. Conclusion

Detection of heart malfunction is critical, as prolonged arrhythmia can lead to deterioration of heart function and other cardiovascular complications. An effective screening system can aid physicians in diagnosing the conditions early, thereby providing the patients with the proper care and timely intervention. The current standard for arrhythmia screening involves visual examination and manual interpretation of ECG records by clinicians. The process is labor intensive, mundane, and vulnerable to inter-observer variability. Moreover, changes within ECG signals can be subtle, and can escape detection by the average person. Hence, computer automated systems may assist in the early screening of arrhythmic ECGs. To the best of our knowledge, no one has hitherto explored the application of using an autoencoder on the ECG for beat-wise analysis. This is the first paper to use a U-net autoencoder for beat-wise arrhythmia detection. The model is able to classify the arrhythmia conditions with 97.32% accuracy and identify the R peaks with 99.3% accuracy without noise elimination. Class activations maps obtained from the model have shown promising results for differentiating the segments into subsequences and associating them with the correct conditions. In the future, the model can be tested on variable length signals for analysis.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2018.12.012.

References

[1] U. Nations, World Population Ageing 2017 Highlights, Department of Economic and Social Affairs, New York, 2017.
[2] G.V. Chow, J.E. Marine, J.L. Fleg, Epidemiology of arrhythmias and conduction disorders in older adults, Clin. Geriatr. Med. 28 (4) (2012) 539–553.
[3] K. Mak, The normal physiology of aging, in: K.-Y. Tan (Ed.), Colorectal Cancer in the Elderly, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013, pp. 1–8.
[4] P. Anversa, et al., Myocyte cell loss and myocyte cellular hyperplasia in the hypertrophied aging rat heart, Circ. Res. 67 (4) (1990) 871–885.
[5] G. Engström, et al., Cardiac arrhythmias and stroke, Stroke 31 (12) (2000) 2925.
[6] P. Zimetbaum, A. Goldman, Ambulatory arrhythmia monitoring, Choosing the Right Device 122 (16) (2010) 1629–1636.
[7] J.F. Schneider, et al., Newly acquired left bundle-branch block: the framingham study, Ann. Intern. Med. 90 (3) (1979) 303–310.
[8] G.J. Fahy, et al., Natural history of isolated bundle branch block, Am. J. Cardiol. 77 (14) (1996) 1185–1190.
[9] I.S. Thrainsdottir, et al., The epidemiology of right bundle branch block and its association with cardiovascular morbidity — the Reykjavik Study, Eur. Heart J. 14 (12) (1993) 1590–1596.
[10] Z. Binici, et al., Excessive supraventricular ectopic activity and increased risk of atrial fibrillation and stroke, Circulation 121 (17) (2010) 1904.
[11] Fleg, J.L. and H.L. Kennedy, Cardiac arrhythmias in a healthy elderly population. Chest. 81(3): p. 302-307.
[12] S. Sahoo, et al., Multiresolution wavelet transform based feature extraction and ECG classification to detect cardiac abnormalities, Measurement 108 (2017) 55–66.
[13] T. Li, M. Zhou, ECG classification using wavelet packet entropy and random forests, Entropy 18 (8) (2016).
[14] F.A. Elhaj, et al., Arrhythmia recognition and classification using combined linear and nonlinear features of ECG signals, Comput. Methods Progr. Biomed. 127 (2016) 52–63.
[15] H. Li, et al., Arrhythmia classification based on multi-domain feature extraction for an ECG recognition system, Sensors 16 (10) (2016) 1744.
[16] R.J. Martis, U.R. Acharya, L.C. Min, ECG beat classification using PCA, LDA, ICA and Discrete Wavelet Transform, Biomed. Signal Process. Control 8 (5) (2013) 437–448.
[17] R.J. Martis, et al., Cardiac decision making using higher order spectra, Biomed. Signal Process. Control 8 (2) (2013) 193–203.
[18] R.J. Martis, et al., Application of principal component analysis to ECG signals for automated diagnosis of cardiac health, Expert Syst. Appl. 39 (14) (2012) 11792–11800.
[19] S. Karimifard, A. Ahmadian, A robust method for diagnosis of morphological arrhythmias based on Hermitian model of higher-order statistics, Biomed. Eng. Online 10 (2011) 22-22.
[20] R.J. Martis, et al., Application of higher order cumulants to ECG signals for the cardiac health diagnosis, Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2011, 2011.
[21] Y.-C. Yeh, W.-J. Wang, C.W. Chiou, Cardiac arrhythmia diagnosis method using linear discriminant analysis on ECG signals, Measurement 42 (5) (2009) 778–789.
[22] S.-N. Yu, K.-T. Chou, Integration of independent component analysis and neural networks for ECG beat classification, Expert Syst. Appl. 34 (4) (2008) 2841–2846.
[23] S. Osowski, L.T. Hoai, T. Markiewicz, Support vector machine-based expert system for reliable heartbeat recognition, IEEE (Inst. Electr. Electron. Eng.) Trans. Biomed. Eng. 51 (4) (2004) 582–589.
[24] J.-G. Lee, et al., Deep learning in medical imaging: general Overview, Korean J. Radiol. 18 (4) (2017) 570–584.
[25] Z. Akkus, et al., Deep learning for brain MRI segmentation: state of the art and future directions, J. Digit. Imag. 30 (4) (2017) 449–459.
[26] T. Pham, et al., Predicting healthcare trajectories from medical records: a deep learning approach, J. Biomed. Inf. 69 (2017) 218–229.
[27] U.R. Acharya, et al., Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network, Inf. Sci. 405 (2017) 81–90.
[28] U.R. Acharya, et al., A deep convolutional neural network model to classify heartbeats, Comput. Biol. Med. 89 (2017) 389–396.
[29] M. Zubair, J. Kim, C. Yoon, An automated ECG beat classification system using convolutional neural networks, 2016 6th International Conference on IT Convergence and Security (ICITCS), 2016.
[30] S. Kiranyaz, T. Ince, M. Gabbouj, Real-time patient-specific ECG classification by 1-D convolutional neural networks, IEEE (Inst. Electr. Electron. Eng.) Trans. Biomed. Eng. 63 (3) (2016) 664–675.
[31] Ö. Yildirim, A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification, Comput. Biol. Med. 96 (2018) 189–202.
[32] Y. Qian, et al., Very deep convolutional neural networks for noise robust speech recognition, IEEE/ACM Transactions on Audio, Speech, and Language Processing 24 (12) (2016) 2263–2276.

[33] U.R. Acharya, et al., Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals, Inf. Sci. 415–416 (2017) 190–198.
[34] Y. Wu, et al., Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation, (2016).
[35] M. Kim, et al., Speaker-Independent silent speech recognition from flesh-point Articulatory movements using an LSTM neural network, IEEE/ACM Transactions on Audio, Speech, and Language Processing 25 (12) (2017) 2323–2336.
[36] E. Song, F.K. Soong, H.G. Kang, Effective spectral and excitation modeling techniques for LSTM-RNN-based speech synthesis systems, IEEE/ACM Transactions on Audio, Speech, and Language Processing 25 (11) (2017) 2152–2161.
[37] M. Sundermeyer, H. Ney, R. Schlüter, From feedforward to recurrent LSTM neural networks for language modeling, IEEE/ACM Transactions on Audio, Speech, and Language Processing 23 (3) (2015) 517–529.
[38] K. Greff, et al., LSTM: a search space odyssey, IEEE Transactions on Neural Networks and Learning Systems 28 (10) (2017) 2222–2232.
[39] W. Yang, et al., Improved deep convolutional neural network for online handwritten Chinese character recognition using domain-specific knowledge, 2015 13th International Conference on Document Analysis and Recognition (ICDAR), 2015.
[40] X.Y. Zhang, et al., End-to-End online writer identification with recurrent neural network, IEEE Transactions on Human-Machine Systems 47 (2) (2017) 285–292.
[41] J.H. Tan, et al., Application of stacked convolutional and long short-term memory network for accurate identification of CAD ECG signals, Comput. Biol. Med. 94 (2018) 19–26.
[42] S.L. Oh, et al., Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats, Comput. Biol. Med. 102 (2018) 278–287.
[43] G.E. Hinton, R.R. Salakhutdinov, Reducing the dimensionality of data with neural networks, Science 313 (5786) (2006) 504–507.
[44] O. Yildirim, R.S. Tan, U.R. Acharya, An efficient compression of ECG signals using deep convolutional autoencoders, Cognit. Syst. Res. 52 (2018) 198–211.
[45] D.D. Testa, M. Rossi, Lightweight lossy compression of biometric patterns via denoising autoencoders, IEEE Signal Process. Lett. 22 (12) (2015) 2304–2308.
[46] P. Xiong, et al., A stacked contractive denoising auto-encoder for ECG signal denoising, Physiol. Meas. 37 (12) (2016) 2214.
[47] Y. Guo, et al., A review of semantic segmentation using deep neural networks, International Journal of Multimedia Information Retrievall, 7 (2) (2018) 87–93.
[48] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, International Conference on Medical Image Computing and Computer-assisted Intervention, Springer, 2015.
[49] Y. Lei, et al., A skin segmentation algorithm based on stacked autoencoders, IEEE Trans. Multimed. 19 (4) (2017) 740–749.
[50] F. Xing, et al., Deep learning in microscopy image analysis: a survey, IEEE Transactions on Neural Networks and Learning Systems 29 (10) (2017) 4550–4568.
[51] A.L. Goldberger, et al., PhysioBank, PhysioToolkit, and PhysioNet, Components of a New Research Resource for Complex Physiologic Signals 101 (23) (2000) e215–e220.
[52] M.B. Conover, Understanding Electrocardiography, Mosby, 2002.
[53] Y. LeCun, et al., Efficient BackProp, Neural Networks: Tricks of the Trade, This Book Is an Outgrowth of a 1996 NIPS Workshop, Springer-Verlag, 1998, pp. 9–50.
[54] M. Lin, Q. Chen, S. Yan, Network In Network, (2013) arXiv preprint arXiv:1312.4400.
[55] Y. Bengio, X. Glorot, Understanding the Difficulty of Training Deep Feed Forward Neural Networks, (2010), pp. 249–256.
[56] D.P. Kingma, J. Ba, Adam: a Method for Stochastic Optimization, (2014) arXiv preprint arXiv:1412.6980.
[57] F. Chollet, Keras: the python Deep Learning Library, Astrophysics Source Code Library, 2018.
[58] M. Abadi, et al., TensorFlow: a system for large-scale machine learning, OSDI (2016).
[59] O. Yildirim, P. Plawiak, R.S. Tan, U.R. Acharya, Arrhythmia detection using deep convolutional neural network with long duration ECG signals, Comput. Biol. Med. 102 (2018) 411–420.
[60] P. Pławiak, Novel methodology of cardiac health recognition based on ECG signals and evolutionary-neural system, Expert Syst. Appl. 92 (2018) 334–349.