A Hybrid CNN-LSTM Architecture for Detection of Coronary Artery Disease from ECG

Rohan Banerjee
Tata Consultancy Services
Kolkata, India
rohan.banerjee@tcs.com

Avik Ghose
Tata Consultancy Services
Kolkata, India
avik.ghose@tcs.com

Dr. Kayapanda Muthana Mandana
Fortis Hospitals
Kolkata, India
kayapanda.mandana@fortishealthcare.com

***Abstract*—Coronary Artery Disease (CAD) causes significant global mortality. The recent development in artificial intelligence shows the feasibility of early non-invasive screening of several life-threatening cardiovascular diseases. However, such approaches have been less prolific in diagnosis of CAD due to lack of clinically known definite bio-marker. In this paper, we propose a novel neural network architecture that effectively combines two non-specific CAD markers, 1) anomalous morphology of Electrocardiogram (ECG) waveform and 2) abnormal Heart Rate Variability (HRV). A Convolutional Neural Network (CNN) structure is defined for extraction of morphological ECG features. Another composite structure is defined based on Long Short-Term Memory (LSTM) and a set of hand crafted statistical features for measuring the extent of HRV. The two independent bio-markers are subsequently combined in a hybrid CNN-LSTM architecture for classification of CAD. The proposed approach is evaluated on two datasets, a corpus, selected from the MIMIC II waveform dataset and a partially noisy in-house dataset, recorded using a low-cost ECG sensor. Results show that overall classification accuracy of 93% and 88% are achieved on the two datasets, which outperform the existing approaches.**

***Index Terms*—Coronary Artery Disease (CAD), ECG, CNN, LSTM, Hybrid Architecture**

I. Introduction

Cardiovascular diseases (CVDs) continue to be a major cause of death, representing 31% of all global deaths in the year 2016. Coronary Artery Disease (CAD) is a type of CVD that affects millions of people every year. CAD forms due to deposition of cholesterol and other fatty materials on the inner walls of coronary arteries over time. This restricts the normal blood flow in coronary arteries, which may lead to a stroke or severe cardiac arrest. Coronary angiography, the gold standard diagnosis for CAD is an invasive medical test, requiring an admission to hospital. An early detection can often prevent the fatal consequences of many life-threatening CVDs. Hence, attention for Artificial Intelligence (AI) based screening systems is rapidly growing in the developing nations where the ratio between doctors and patients is less than desired as well as the developed nations for looking after the elderly population. Such screening systems are required to be low-cost, non-invasive and should be operated with minimum involvement of a clinician. However, non-invasive diagnosis of CAD is an unsolved research problem. Although CAD is often asymptomatic at early stages, available research works pointed out several non-specific bio-markers. Heart Rate Variability

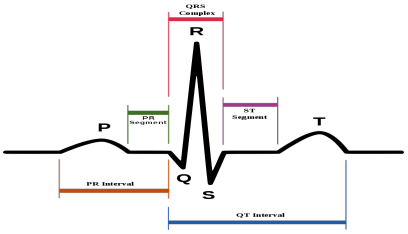
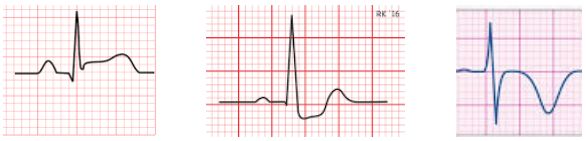


Fig. 1: Components of a normal ECG cycle. Source: Wikipedia



(a) ST elevation (b) ST depression (c) T wave inversion

Fig. 2: Morphology changes in ECG cycles due to CAD [1]

(HRV) is widely considered as an important surrogate marker. A number of articles reported that CAD patients exhibit lower and irregular HRV than non-cardiac subjects [2], [3], [4], [5]. HRV can be accurately measured from the Electrocardiogram (ECG) waveform. As shown in Fig. 1, a normal ECG cycle contains three major components. The P wave represents the depolarization of the atria, the QRS complex represents the depolarization of the ventricles and the T wave represents the repolarization of the ventricles. The PR interval measures the time, taken by an electrical impulse to travel from the sinus node through the AV node. The ST segment measures the duration when the ventricles depolarize and repolarize. The distance between the R peaks indicates the heart rate. Dua *et al.* [5] proposed novel HRV features based on recurrence plot, Poincaré plot, de-trended fluctuation analysis of the RR intervals for classification of CAD. Non-linear dynamics of ECG were analysed in [6]. The work in [2] showed significant difference in the standard deviation of the RR intervals and average heart rate between CAD and non-cardiac subjects. In another approach [7], discrete cosine transform, discrete wavelet transform and empirical mode decomposition methods were applied on ECG beats for HRV. However, abnormal HRV can be associated with a number of pathological conditions which are not cardiac in nature (e.g. mental stress).

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

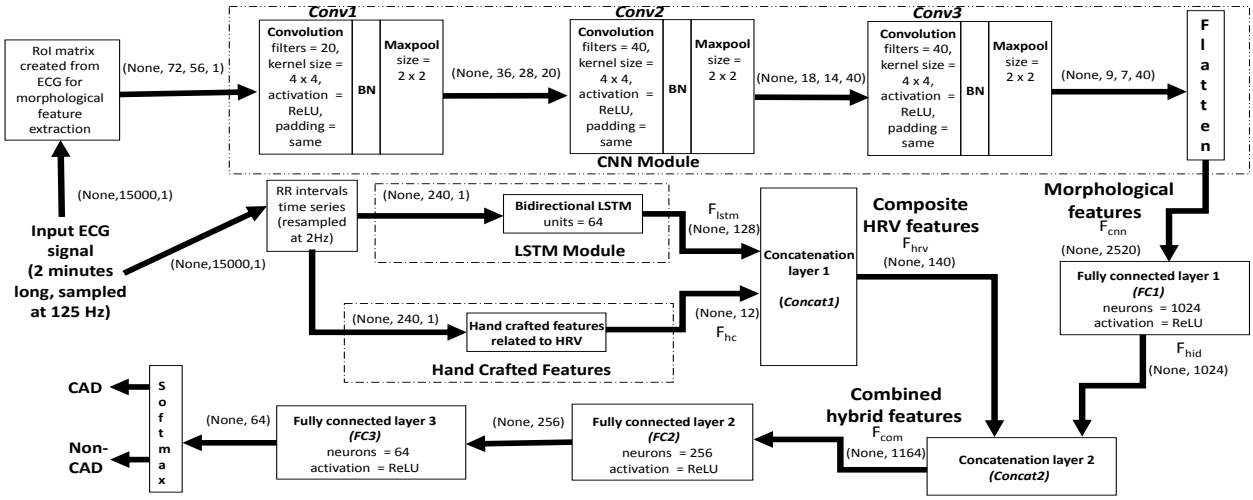


Fig. 3: Proposed hybrid CNN-LSTM neural network architecture for classification of CAD

Morphological changes in normal ECG pattern is reported as another CAD marker. As shown in Fig. 2, the changes happen in terms of ST elevation myocardial infarction (STEMI), ST depression or T wave inversion [1], [8], [9]. However, being intermittent markers, they are not guaranteed at all stages of CAD. Moreover, computing such features is challenging on noisy ECG signals. Hence, majority of the existing AI approaches analyse HRV for CAD. Recently, applications of deep learning have become popular in biomedical engineering, where the raw signal is directly fed to a deep architecture for automatic extraction of relevant features. Such approaches have been successful in detection of arrhythmias or atrial fibrillation, where the bio-markers are well-known and prominent in the input. However, there is no clinically proven definite bio-marker for CAD till date. In this paper, we propose a deep learning approach to combine two non-specific surrogate CAD markers, 1) anomalous ECG morphology and 2) irregular HRV using medical domain knowledge. The markers complement each other in terms of sensitivity and specificity of detecting CAD and their combined effect improves the overall accuracy of a screening system. Major contributions of the paper are:

- A Convolutional Neural Network (CNN) structure for efficient morphological feature extraction from ECG.
- A composite HRV vector based on Long Short-Term Memory (LSTM) network and hand crafted features.
- A hybrid CNN-LSTM structure, together with the hand crafted features to combine the two independent CAD markers in a single architecture for disease classification.
- An algorithm for the selection of relevant hand crafted features, related to HRV.
- A pre-processing algorithm for extraction of time series from clinically interpretable digital ECG images, recorded by commercially available low-cost ECG sensor.

We evaluate the proposed network on an open access hospital dataset and a second in-house dataset, recorded by a non-medical grade sensor, typically deployed in a low-cost real-world screening system. The proposed network architecture is detailed in Section II. Section III describes our experimental datasets. Section IV summarizes different network hyper-parameters and experimental results followed by a conclusion.

II. Proposed Hybrid Network Architecture

A hybrid neural network is proposed by combining two non-specific surrogate CAD markers, i.e anomalous ECG morphology and abnormal HRV in a single CNN-LSTM architecture. Block diagram of the proposed architecture is shown in Fig. 3. The output tensor dimensions are duly mentioned in brackets. The structure has three modules, 1) CNN module for extraction of morphological ECG features, 2) LSTM module for modeling of temporal dependencies among the RR intervals and 3) hand crafted statistical features, derived from the RR intervals. The output state of the LSTM is merged with the hand crafted features to create a composite HRV vector, which is further combined with the morphological feature map by the CNN. The hybrid network is trained end to end to optimize a single objective function for classification.

A. CNN Module for Morphological Feature Extraction

Acute myocardial ischemia affects the electrical activation of the heart, causing morphological changes in the normal ECG pattern. The changes mostly reflect in terms of elevation or depression of the ST segment, inversion of the T wave or fragmentation of the QRS complex [1], [8], [9]. An automatic segmentation of different ECG components is challenging. Although the existing algorithms can accurately locate the R peaks even from a noisy ECG, the performance is poor in segregating the ST region, QRS complex or the T wave. Hence, the hand crafted morphological features, computed from a segmented recording become unreliable, affecting the performance of a classical machine learning based classifier. In this paper, we define a 2D CNN structure for extraction of discriminating morphological features from an ECG dataset of labelled CAD and non-CAD subjects, without segmenting the

fundamental components. A CNN can successfully capture the spatial and temporal dependencies in an input via convolution operation through a set of filters (kernel) to create a feature map that summarizes the detected features in the input. In a multi-layer CNN, the first convolutional layer is typically responsible for capturing the low-level features. With added layers, the architecture adapts to the high-level features.

A CNN is not guaranteed to learn the desired pattern from a time series, containing multiple independent patterns. Anomalous ECG morphology is a non-specific CAD marker and its effects are often inconsistent across multiple cardiac cycles. Applying ECG data directly to 1D CNN may not result in an optimum spatio-temporal analysis and may cause overfit. We propose a novel pre-processing step to create a 2D representation of 1D ECG data by cropping a fixed region from successive cardiac cycles, where the relevant morphological markers are likely to be present and merge them vertically. An ECG time series (ecg_t) is represented as a vector of n real numbers in eqn. 1:

$$ecg_t = [ecg_1, ecg_2, ...ecg_{R_1}, ...ecg_{R_2}, ...ecg_{R_p}, ...ecg_n] \quad (1)$$

ecg_i is the i^{th} sample. In the time series, each cardiac cycle is landmarked by the corresponding R peak, whose locations are known. For example, R_p is the landmark point for p^{th} cycle and the corresponding amplitude is ecg_{R_p}. The open source R peak detection algorithm by Behar [10] is used in this paper. From every cardiac cycle, a window of length w is cropped around the R peak, which is called the RoI vector. The RoI vector (ST_p) for p^{th} cycle is denoted as:

$$ST_p = [ecg_{(R_p - w_1)}, ...ecg_{(R_p-1)}, ecg_{R_p}, ecg_{(R_p+1)}, ...ecg_{(R_p+w_2)}] \quad (2)$$

Here, $w = w_1 + w_2 + 1$. It is known from domain knowledge that the duration of the QRS complex and the T wave are around 200 ms. Duration of the RoI vector is selected 450 ms, 80 ms before and 370 ms after the reference R peak. This contains the relevant portions in an ECG cycle, where the morphology is typically affected due to CAD and removes other portions without segmenting the recording. A total of N RoI vectors from consecutive cardiac cycles are vertically arranged to construct the RoI matrix ST, $ST \in \mathbb{R}^{N \times w}$, which is applied to a 2D CNN. Fig. 4 shows an ECG signal with 3 cycles and the corresponding RoI vectors. The RoI matrix is an efficient 2D representation of an ECG recording for effective spatio-temporal feature extraction by a CNN. The input matrix is normalized between 0 and 1 using minmax normalization. The signals in our experimental datasets are sampled at 125 Hz. Every RoI vector contains 56 data points (125 Hz x 450 ms/1000 $\approx$ 56). Duration of an ECG recording is selected as 2 minutes, i.e 15,000 data points. Irrespective of the underlying cardiac condition of a person, a minimum of 72 cardiac cycles are expected in every such recording. Hence, the dimension of RoI matrix is fixed as 72 x 56. As shown in Fig. 3, the CNN contains 3 convolutional blocks (*Conv1*, *Conv2*, *Conv3*), each of them comprising a convolutional, a batch normalization (BN) and a maxpool layer. The kernel size is selected as 4x4 for all the convolutional layers. Zero padding is applied to the inputs for retaining the original dimension after convolution operation. Rectified Linear Unit (ReLU) is used for non-linear activation, $ReLU(x) = max\{0, x\}$. The number of filters is increased to 40 in the second and third convolutional layers for extraction of more detailed features, but the overall dimension is reduced by the associated maxpool layers of size 2 x 2. The feature map, created by *Conv3* is flattened and merged with the HRV parameters.

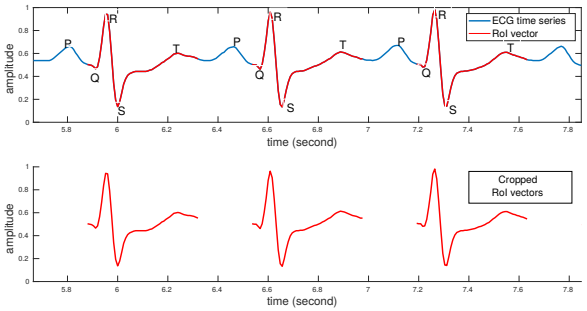


Fig. 4: A sample ECG signal, indicating the RoI vectors. 72 RoI vectors are vertically arranged to create the RoI matrix

B. LSTM Module for Analysis of RR Intervals

Lower and inherently aperiodic HRV is clinically known as another important CAD marker. HRV is measured from the RR intervals time series, extracted from ECG and is denoted by $RR_t = [RR_1, RR_2, ...RR_{n1}]$. Here, $RR_p = R_{p+1} - R_p$, R_p denotes the time location of p^{th} R peak and $n1$ is the number of RR intervals in the ECG recording. The time series does not have a uniform sampling rate owning to instantaneous variation in successive RR interval distances and hence it is fixed to 2 Hz using cubic spline interpolation technique. The resampled time series contains 240 data points corresponding to a 2 minutes long ECG. This is further scaled between 0 and 1 via minmax normalization and is applied to a bidirectional LSTM network for measuring of temporal dependencies among the RR intervals. A Recurrent Neural Network (RNN) has its internal memory for sequential modeling of a time series via extraction of temporal patterns. An LSTM is an improved class of RNN that can effectively learn a longer pattern of unknown length because of its unique cell structure (the forget gate), that enables deletion of less important information from the memory. It can also deal with the exploding and the vanishing gradient problems, typically faced by RNNs during training. For an input vector $x_t = [x_1, x_2, ...x_T]$ of length T, an LSTM cell with one forget gate computes a hidden vector sequence $h_t = [h_1, h_2, ...h_T]$ by iterating the following set of equations over time t.

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \quad (3)$$
$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \quad (4)$$
$$c_t = f_t * c_{t-1} + i_t * tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (5)$$
$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \quad (6)$$
$$h_t = o_t * tanh(c_t) \quad (7)$$

Here, W_x, W_h represent the trainable weight matrices for the input vector and the recurrent connections; b are the bias terms; σ is the logistic sigmoid function; $*$ denotes element-wise product operation. The input gate, the forget gate, the output gate and the cell activation vectors are represented by i, f, o and c. Hyperbolic tangent ($tanh$) is used for non-linear activation of the cell. A bidirectional LSTM (Bi-LSTM) is an improved version of a traditional LSTM. It trains two LSTMs on a single input sequence, one on the input sequence as the same order and the second on the reversed order, enabling to preserve the information from both past and future. In our case, 64 hidden units are used in the bidirectional LSTM structure for the analysis of RR intervals time series. Hence, the output hidden vector returns 128 parameters.

C. Hand Crafted HRV Features

Although the Bi-LSTM can model the temporal dependencies among the RR interval distances for predicting the overall aperiodicity in HRV, it is not guaranteed to quantify the complex chaotic HRV pattern manifested by many CAD patients. A number of higher order statistical parameters exist in various applications for measuring the inherent randomness of a time series. We propose a set of 12 hand crafted features, derived from the time series, RR_t and merge them with the output state of the Bi-LSTM to create a composite HRV vector. The hand crafted features are selected from a larger feature set using an empirical feature selection algorithm, detailed in Section IV-A. The optimum feature set contains two novel features, 1) self similarity and 2) average Maharaj's distance and few features adapted from literature.

*1) **Self similarity**:* In theory, self similarity of a time series measures the rate of decrease in the autocorrelation with the increase in lag between a pair of observations and is measured from the Hurst exponent (H) [11]. The parameter, H is computed using Fractional Autoregressive Integrated Moving Average (FARIMA) processes. Autoregressive Integrated Moving Average (ARIMA) models are used in statistics for analysing and forecasting of time series. In an $ARIMA(p,q,r)$ process, p denotes the order of AR, r is the order of MA and q is the degree of differencing which is measured by the number of times the time series have had past values subtracted. The parameter q is an integer for a stationary time series. If long-range dependence is suspected, q can be a non-integer, hence results in a $FARIMA$ model. The time series RR_t is fitted to a $FARIMA(0,d,0)$ by an approximation of the maximum likelihood method [12]. The Hurst parameter is calculated, as $H = d + 0.5$.

*2) **Average Maharaj's distance**:* In a time series, Maharaj's distance [13] is indicative of a moving average factor along with the number of changes in direction. Maharaj's distance can capture a desired similarity metric across spatial entities. In Autoregressive Moving Average (ARMA) model, a time series Y_t is defined in terms of autoregressive term p and moving average term r, as:

$$Y_t = \lambda + \sum_{i=1}^{p} \Psi_i Y_{t-i} + \sum_{i=1}^{r} \theta_i \epsilon_{t-i} + \epsilon_t \tag{8}$$

Here, λ is a constant, ϵ_t is white noise, Ψ_i-s are the AR parameters and θ_i-s are the MA parameters. For an $ARMA$ model, discrepancy measurement based on hypotheses testing determines whether two time series have significantly different generating processes. The output metric of this mechanism is called the Maharaj's distance and can be used to find whether two time series are similar to each other. The measured p-value, close to 1 indicates they are similar. In this paper, average Maharaj's distance (AMD_i) for the i^{th} RR interval time series is calculated with respect to the known CAD population in the training dataset as:

$$\text{AMD}_i = \sum_{j \neq i}^{nu} \text{MD}_{ij}/(nu-1) \tag{9}$$

where MD_{ij} is the Maharaj's distance of the RR intervals time series of i^{th} training or test subject from j^{th} CAD subject in the training dataset and nu is the total number of CAD subjects in the training set. This feature measures the average dissimilarity of an unknown test instance from the CAD population in the training set.

*3) **Entropy features**:* Approximate entropy ($ApEn$) and sample entropy ($SampEn$) are used in statistics to measure the irregularities in a time series. $ApEn(RR_t, q, r)$ is computed in terms of two predefined parameters, a pattern length (q) and another parameter (r), related to similarity measurement. A sequence of vectors $[x_q(1), x_q(2), ...x_q(n1-q+1)]$ in real q-dimensional space is defined from RR_t, such that $x_q(i) = [RR_i, RR_{i+1}, RR_{i+2}, ...RR_{i+q-1}]$. Two vectors $x_q(i)$ and $x_q(j)$ are similar if $|RR_{i+k} - RR_{j+k}| < r,\ for\ 0 < k < q$. A new parameter $C_{iq}(r)$ is defined as: $C_{iq}(r) = (number\ of\ x_q(j)\ similar\ to\ x_q(i))/(n1-q+1)$, $n1$ be the length of RR_t. If $C_q(r)$ indicates the mean of all $C_{iq}(r)$ for $i \in 1...n1-q+1$, $ApEn$ is defined as:

$$ApEn(RR_t, q, r) = ln[\frac{C_q(r)}{C_{q+1}(r)}] \tag{10}$$

Sample Entropy ($SampEn(RR_t, q, r)$) is defined as:

$$SampEn(RR_t, q, r) = -ln[\frac{A}{B}] \tag{11}$$

Here, A = number of vector pairs where $|x_{q+1}(i) - x_{q+1}(j)| < r$, B = number of vector pairs where $|x_q(i) - x_q(j)| < r$.
Shannon entropy (E_{sh}) of RR_t is another important parameter for measuring of unpredictability .

$$E_{sh} = -\sum_{b=1}^{N} pr_b \log pr_b \tag{12}$$

A normalized histogram of N bins is created from the distribution of RR_t. Empirical probability of b^{th} bin is denoted by pr_b. Here, $b \in 1...N$ and $\sum_{b=1}^{N} pr_b = 1$.

*4) **Other features**:* The other features include mean, variance, kurtosis, skewness and root mean square of successive difference ($RMSSD$) of RR_t. The number of successive RR interval pairs, differ by more than 20 ms ($pNN20$) and 50 ms ($pNN50$), divided by the total number of RR intervals in an ECG recording are the remaining two features. The 12 hand crafted features are normalized to zero mean and unit variance for merging with the output state of the Bi-LSTM.

D. The Hybrid CNN-LSTM Network Structure

As shown in Fig. 3, the output state vector of the Bi-LSTM (F_{lstm}) is concatenated with the hand crafted feature vector (F_{hc}) at the first concatenation layer ($concat1$) to construct the composite HRV vector (F_{hrv}). On the other hand, the morphological feature map at the output of the third convolutional block ($conv3$) is flattened as F_{cnn} and applied to a fully connected layer ($FC1$) of 1024 neurons to form a hidden vector (F_{hid}) of reduced dimension. F_{hid} is concatenated with F_{hrv} at the second concatenation layer ($concat2$) to create the hybrid feature vector (F_{com}), combining two independent CAD markers which is then applied to a pair of fully connected layers ($FC2, FC3$), having 256 and 64 neurons followed by a softmax function for binary classification. $ReLU$ activation function is used in the fully connected layers. The class labels are converted to one-hot encoding to minimize the categorical cross-entropy loss function, shown in eqn. 13:

$$J = -\frac{1}{N}\sum_{i=1}^{N}(y_i \cdot log(\hat{y_i}) + (1 - y_i) \cdot log(1 - \hat{y_i})) \quad (13)$$

N denotes the number of training instances in a batch; the true label and the predicted label of i^{th} training instance are indicated by y_i and $\hat{y_i}$. An amount of 20% drop-out is applied to the Bi-LSTM and the cell weights are constrained using L2 regularization. This protects the Bi-LSTM from over-fitting and restricts its overall output activation, ensuring the lower dimensional F_{hc} is not suppressed after merging with F_{lstm} at $concat1$. On the other hand, the hidden vector of morphological features (F_{hid}) is of much higher dimension than the composite HRV vector (F_{hrv}). Hence, the effect of F_{hrv} might get suppressed after merging with the morphological features at $concat2$. In order to avoid that, we regularize the activation of $FC1$ (where F_{cnn} is applied) to limit the number of active neurons by imposing a sparsity constraint. For i^{th} input, x_i, if $a_j(x_i)$ denotes the activation of j^{th} hidden unit at $FC1$, then $\hat{\rho_j}$ measures the average activation of j^{th} unit, as:

$$\hat{\rho_j} = \frac{1}{N}\sum_{i=1}^{N} a_j(x_i) \quad (14)$$

During training, $\hat{\rho_j}$ is forced to be similar to a sparsity parameter, ρ of small value (0.05, in our case), so that the overall activation, a_j is regularized. This is done by using the Kullback-Leibler (KL) divergence between them as a penalty term, which is added to the loss function.

$$KL(\rho||\hat{\rho_j}) = \rho \cdot log\frac{\rho}{\hat{\rho_j}} + (1 - \rho) \cdot log\frac{1 - \rho}{1 - \hat{\rho_j}} \quad (15)$$

In addition, an amount of 20% drop-out is applied to the convolutional layers. The hybrid CNN-LSTM network is trained end to end to minimize a single cumulative loss function, which is derived by combining eqn. 13 and eqn. 15, as:

$$J_t = J + \frac{\lambda}{2} \cdot \sum_{k}(W_k)^2 + \beta \cdot \sum_{j=1}^{s} KL(\rho||\hat{\rho_j}) \quad (16)$$

W_k summarizes the weights of k^{th} Bi-LSTM cell; λ is the weight decay parameter; β is the sparsity penalty term and s is the number of hidden units at $FC1$ layer.

III. Experimental Datasets

The proposed methodology is evaluated on two datasets of different patient demography, sensor device and overall signal quality. The first dataset is a corpus of CAD and non-cardiac subjects, selected from MIMIC II waveform dataset, matched subset [14]. MIMIC II is a large open access dataset, containing physiological signals (ECG, pleth, blood pressure waveform etc.) from patients during their stay at different hospitals in the USA and Europe. The recordings were made using medical-grade equipments at a sampling rate of 125 Hz. The dataset comes with the disease information for patients, which follows the International Classification of Disease, Ninth Revision, Clinical Modification (ICD-9, CM) codes. Depending upon the disease information and the availability of ECG recordings, a total of 100 CAD and an equal number of non-cardiac patients were selected. One of our co-authors, a practising cardiologist helped us in selection of the patients based on the ICD-9 codes. The data, corresponding to ECG lead II was considered for analysis. We split the selected corpus into two portions based on random selection, 80% of all subjects were kept for training and 20% for test purpose. Duration of a recording is taken as 2 minutes for analysis, which is long enough to capture the HRV. A total of 100 such non-overlapping instances were selected from every subject. This data augmentation step enhances the number of instances for training a deep learning system. The final corpus contains 16000 training and 4000 test instances from 200 subjects, with no common subject in the training and test set.

Being recorded by medical-grade equipments at hospital ICU, the signals in MIMIC II are moistly free from noise and abrupt motion artifacts. However, a real-world screening system typically deploys low-cost, non-medical grade devices. Hence, we created a second dataset in a more challenging scenario using AliveCor Kardia, a commercially available single-lead portable ECG device. The dataset was recorded at the out-patient department of a hospital in Kolkata, India under the supervision of a practising cardiologist and it was used only for testing. The data collection protocol was approved by the hospital ethics committee. A total of 80 CAD and 70 non-CAD subjects participated in the process. The disease annotation were done by the hospital clinicians based on angiography reports. The CAD population covered both borderline and severe patients, whereas the non-CAD population had no prior cardiac history at the time of data collection. The recordings in this dataset are in general noisier than MIMIC II.

A. Extraction of Time Series from an ECG Image

Similar to most of the commercially available single-lead devices, AliveCor Kardia provides the recorded ECG in a clinically interpretable digital image format, printed on a standard grid based graphical layout for easy visualization and does not provide access to the time series data. As shown in Fig. 5:(a), the layout is partitioned into small grids of 5 mm X 5 mm boxes, representing 200 ms in horizontal axis and 0.5 mV in vertical axis. The large grids represent 25 mm X 50 mm boxes. There is a reference pulse of 1 mV in amplitude,

providing the reference for zero voltage. The ECG time series is required to be extracted from the image as a pre-processing step. The following algorithm is proposed for that purpose.

*1) **Histogram analysis**:* The image is converted to gray scale. A histogram analysis reveals that the pixel values between 40 and 180 constitute the ECG signal, the large grids and the reference pulse. Zeroing the pixels, outside that range removes the small grids from the image. A sample ECG image and the quantized output are shown in Fig. 5:(a) and 5:(b).

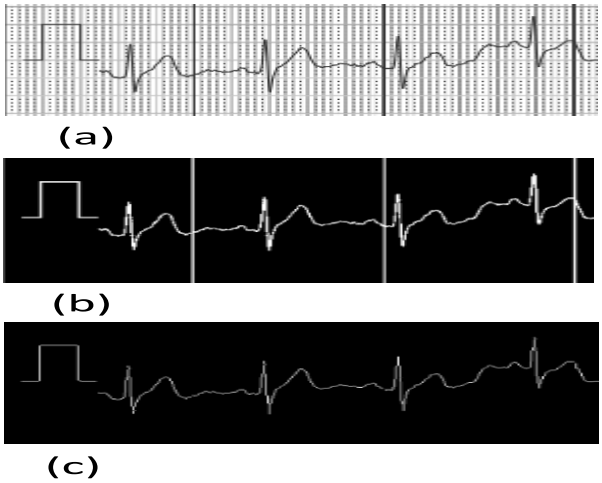


Fig. 5: (a) Sample ECG image by AliveCor Kardia, (b) Output of quantization based on histogram analysis, (c) Output of morphological operation and removal of large grid lines

*2) **Morphological operation**:* A standard thinning operation is performed on the quantized image with kernel size 5x5 till there is no update in successive thinning operations.

*3) **Extraction of large grid lines**:* The horizontal and vertical lines for the large grids are detected using Hough transform and removed. The output after the removal of large grids is shown in Fig. 5:(c).

*4) **Extraction of reference pulse**:* The reference pulse is extracted by matching with a template of the pulse and the zero reference line is identified. Template matching is done in pixel level using two-dimensional autocorrelation analysis.

*5) **Extraction of ECG signal**:* A pixel value quantization of the image, after removal of large grid lines gives the time series signal. The interpretation of the time scale and the amplitude in voltage is derived from the reference pulse.

*6) **Interpolation of missing data**:* There are some missing points in the extracted ECG time series due to various image processing operations. A cubic-spline filtering is performed locally in the regions of missing data to generate the interpolated signal. The extracted signal is resampled at 125 Hz.

IV. Data Analysis

In this section, we discuss our proposed algorithm for selection of the relevant hand crafted HRV features, different hyper-parameters used for performance optimization of the proposed network and the experimental results.

A. Selection of Hand Crafted HRV Features

Feature selection is an important part in any machine learning approach. This is a process of selecting the relevant and discriminating feature subset from a larger superset for an optimum classification performance. A heuristic approach for feature selection is proposed in this paper, which is performed on the training portion of the dataset, selected from MIMIC II, described in Section III. We start with 20 features as a superset, compute them on all the recordings in the dataset and rank the features with respect to the subject disease labels in a descending order based on relevance using Maximal Information Coefficient (MIC) scores. MIC is a statistical tool that measures the strength of association between two variables by forming grids with various sizes to find the largest normalized mutual information between them [15]. For each pair of data (x, y), if I is the mutual information for a grid G, then MIC of a set D of pairwise data with sample size n and grid size (xy), less than $B(n)$ is given by [15], as:

$$MIC(D) = max_{xy<B(n)}\{M(D)_{x,y}\} \tag{17}$$

where $B(n)$ is a function of sample size (usually $B(n) = n^{0.6}$). For different distributions of G, $M(D)$ is given by

$$M(D)_{x,y} = \frac{max\{I(D|G)\}}{\log min(x, y)} \tag{18}$$

A high MIC value indicates a relevant feature. The optimum feature set is selected from the ranked feature list in an iterative manner based on average classification performance on the same training dataset, applying subject level 5-fold cross validation. We consider the highest ranking feature as the most important feature and measure the sensitivity and specificity of detecting CAD using that. In each subsequent iteration, the next most relevant feature is added from the list and the impact in classification performance is noted. The features are applied to a neural network having a single hidden layer with 10 neurons followed by softmax function for classification. Our aim is to simultaneously achieve the optimum sensitivity and specificity. Hence, we want to maximize the arithmetic mean of the two metrics at a minimum feature dimension. Fig. 6 plots the average values of sensitivity, specificity and their numeric mean, achieved by increasing the number of relevant features. It can be observed that both sensitivity and specificity improve with the addition of more features. Their numeric mean becomes maximum at a minimum feature dimension when top 12 features are selected. Addition of more features eventually reduces the sensitivity. Hence, these 12 features are selected as the optimum hand crafted HRV parameters.

B. Selection of Network Hyper-parameters

Similar to feature selection, different hyper-parameters for the neural network are also tuned on the training dataset from MIMIC II based on trial and error. The combination of hyper-parameters producing the highest median accuracy in 5-fold cross validation is selected as optimum. The final learning model is created on the entire training set and the evaluation is done on the test data from MIMIC II and the in-house dataset,

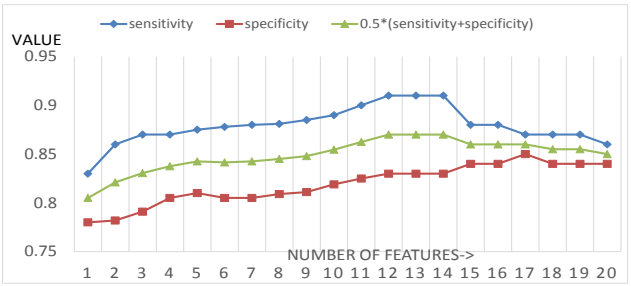


Fig. 6: Average subject level 5 fold cross validation performance on MIMIC II training data by iteratively adding features from the ranked feature list. The optimum performance is achived with 12 features. These are detailed in Section II-C

recorded using AliveCor Kardia. In order to avoid over-fit, an mount of 20% drop-out is applied to the convolutional and the LSTM layers. The weight decay parameter (λ) of the LSTM cells in eqn. 16 is selected as 0.02 and the sparsity penalty term (β) of the CNN is selected as 0.8. Initial weights of the neurons in the convolutional layers, the LSTM cells and the fully connected layers are set using Xavier initialization [16]. In this process, the weights in a layer are randomly initialized from a Gaussian distribution of zero mean and a finite variance $\frac{2}{n_{in}+n_{out}}$, where, n_{in} and n_{out} represent the number of input and output neurons. The bias terms are initially set to zero. During training, the cumulative loss function is minimized using Adam optimizer with learning rate of 0.003, mini-batch size of 128 and 200 epochs. The network is trained on a platform of Intel® Xeon(R) 16-core processor having 64 GB of RAM and a graphics processing unit. The implementation is done in python using Keras API with TensorFlow.

C. Experimental Results

1) ***Accuracy of time series extraction algorithm****:* In the in-house dataset, the ECG time series, extracted from images can be corrupted due to various image processing operations like histogram based quantization, morphological operation or interpolation of missing data. The extracted time series are compared with the actual images to evaluate the quality of reconstruction. A set of 50 recordings are randomly selected from the dataset and the cardiac cycles are manually annotated from the ECG images and the extracted time series with the help of a cardiologist for comparison between them. Table I shows that the mean squared error in the QRS intervals between them remains close to an acceptable range of 2%. The errors are even lower for the QT intervals and RR intervals.

TABLE I: Comparison between ECG images and time series

	RR intervals	QRS intervals	QT intervals
Mean squared error	0.1 %	1.95 %	0.85 %

2) ***Contribution of individual network components and the benefit of the hybrid architecture in CAD classification****:* Throughout this paper, classification performance is reported in terms of sensitivity (Se) and specificity (Sp) of detecting CAD, which are defined by true positive (TP), true negative (TN), false positive (FP) and false negative (FN), as:

$$Se = \frac{TP}{TP+FN},\ Sp = \frac{TN}{TN+FP} \tag{19}$$

Table II summarizes the contribution of different components of the proposed hybrid network by evaluating them individually for classifying CAD vis-a-vis the improvement achieved by their combined effect in the proposed hybrid CNN-LSTM architecture. For an exhaustive performance comparison, the hybrid network is split into four components based on individual output feature vectors, 1) the CNN for extraction of morphological ECG features, 2) the Bi-LSTM for analysis of RR intervals, 3) the hand crafted HRV features and 4) the composite HRV vector via merging the output of the Bi-LSTM with the hand crafted features. All of them are separately applied to fully connected layers and softmax function for creating individual neural networks and are compared with the proposed hybrid structure. Different network hyper-parameters are retuned in all cases and the best accuracy values are reported in Table II which shows the subject level 5-fold cross validation result in $mean \pm std$ format on the training data ($Training$) from MIMIC II and the accuracy on the MIMIC II test set ($D1$) and the in-house data ($D2$). In all cases, the learning model for evaluation on $D1$ and $D2$ are created on the entire $Training$ set. Anomalous ECG morphology can not be

TABLE II: Performance comparison among individual components of the proposed hybrid CNN-LSTM network

Network Structure	$Training$ set		Test set $D1$		Test set $D2$	
	Se	Sp	Se	Sp	Se	Sp
CNN (for ECG morphology)	0.83 ± 0.03	0.94 ± 0.04	0.82	0.94	0.78	0.89
Bi-LSTM (for HRV)	0.94 ± 0.04	0.80 ± 0.06	0.93	0.78	0.90	0.77
Hand crafted features (for HRV)	0.91 ± 0.05	0.83 ± 0.02	0.90	0.82	0.86	0.80
Bi-LSTM + hand crafted features (for HRV)	0.96 ± 0.05	0.85 ± 0.03	0.95	0.83	0.92	0.81
CNN + Bi-LSTM + hand crafted features (ECG morphology + HRV) (proposed hybrid CNN-LSTM)	**0.94 ± 0.04**	**0.93 ± 0.03**	**0.94**	**0.92**	**0.90**	**0.85**

considered as a definite CAD marker owing to its inconsistent manifestation. This is often not present at the onset of the disease. Thus, the morphological feature map, extracted by the convolution filters are less reliable in detecting all the CAD patients in the datasets. The borderline patients are often misclassified by this approach. However, It can accurately detect most of the non-cardiac subjects who have a consistent and normal ECG morphology. Hence, the CNN achieves a high specificity and a low sensitivity. On the other hand, irregular

HRV is common to majority of the CAD population. However, many non-CAD subjects also exhibit a similar HRV pattern due to various non-cardiac pathological conditions. Although the Bi-LSTM and the hand crafted features can individually identify majority of the CAD patients in both the datasets, they generate a large number of false positives. This results in high sensitivity and low specificity. Results show that the composite HRV vector, obtained by merging the Bi-LSTM and the hand crafted features yields an improved performance, compared to their individual efforts. Table II reveals that anomalous ECG morphology and irregular HRV are two complementary CAD markers in terms of sensitivity and specificity. An effective combination of them can significantly boost up the overall classification performance. The same is done in our proposed hybrid network structure. Hence, it yields the optimum performance for CAD classification.

3) ***Comparison with existing approaches****:* The proposed CNN-LSTM network is compared with popular prior approaches in [5] and [7]. Both of them used classical machine learning approach to quantify HRV for CAD using hand crafted statistical features. Due to unavailability of implementable existing deep learning approach, we design a baseline 1D CNN classifier. This takes the raw ECG directly as input for relevant feature extraction without a pre-processing. It has 5 convolutional layers, each of them contains 40 filters with kernel size 7 and an associated maxpool layer of size 2. The feature map is flattened and applied to fully connected layers and softmax function for classification. The comparative study is shown in Table III. The baseline CNN often fails to extract the relevant features for CAD due to their inconsistent manifestation in a prolonged recording. Hence, it produces a suboptimal sensitivity. Whereas, our proposed approach that combines two independent CAD markers in a single neural network is found to outperform the prior approaches and the baseline 1D CNN on both our test datasets. Hence, it yields the optimum accuracy for non-invasive screening of CAD.

TABLE III: Performance comparison with existing approaches

Prior art	Test set $D1$		Test set $D2$	
	Se	Sp	Se	Sp
Dua *et al.* [5] (classical machine learning)	0.80	0.75	0.77	0.71
Acharya *et al.* [7] (classical machine learning)	0.88	0.83	0.82	0.78
Baseline 1D CNN (deep learning)	0.73	0.90	0.71	0.85
Proposed approach (hybrid CNN-LSTM)	**0.94**	**0.92**	**0.90**	**0.85**

V. Conclusion

Low-cost, non-invasive screening of CAD is an important research area in medicine. Lack of definite non-invasive biomarker makes the diagnosis of CAD difficult. In this paper we propose a CNN-LSTM approach that combines two independent non-specific CAD markers in a single hybrid architecture for an improved screening system. The proposed approach has been successfully evaluated on two hospital datasets, one of them was recorded in a more practical scenario using low-cost sensor. However, both the bio-markers, considered in this paper are often not guaranteed at the onset of CAD. Hence, our approach fails to detect few of the borderline patients. The proposed architecture can be logically upgraded by incorporating other potential non-invasive CAD markers (e.g. heart sounds signals) along with categorical parameters like patient demography, life-style and family disease history which indirectly determine the cardiac risk factor of a person.

References

[1] R. Klabunde, *Cardiovascular physiology concepts*. Lippincott Williams & Wilkins, 2011.

[2] R. Krittayaphong, W. E. Cascio, K. C. Light, D. Sheffield, R. N. Golden, J. B. Finkel, G. Glekas, G. G. Koch, and D. S. Sheps, "Heart rate variability in patients with coronary artery disease: differences in patients with higher and lower depression scores," *Psychosomatic Medicine*, vol. 59, no. 3, pp. 231–235, 1997.

[3] T. Mironova, V. Mironov, V. Antufiev, E. Safronova, M. Mironov, and E. Davydova, "Heart rate variability analysis at coronary artery disease and angina pectoris," *Recent patents on cardiovascular drug discovery*, vol. 4, no. 1, pp. 45–54, 2009.

[4] T. C. Rodrigues, J. Ehrlich, C. M. Hunter, G. L. Kinney, M. Rewers, and J. K. Snell-Bergeon, "Reduced heart rate variability predicts progression of coronary artery calcification in adults with type 1 diabetes and controls without diabetes," *Diabetes technology & therapeutics*, vol. 12, no. 12, pp. 963–969, 2010.

[5] S. Dua, X. Du, S. V. SREE, and T. A. VI, "Novel classification of coronary artery disease using heart rate variability analysis," *Journal of Mechanics in Medicine and Biology*, vol. 12, no. 04, p. 1240017, 2012.

[6] K. Antanavičius, A. Bastys, J. Blužas, L. Gargasas, S. Kaminskienė, G. Urbonavičienė, and A. Vainoras, "Nonlinear dynamics analysis of electrocardiograms for detection of coronary artery disease," *Computer Methods and Programs in Biomedicine*, vol. 92, no. 2, pp. 198–204, 2008.

[7] U. R. Acharya, H. Fujita, M. Adam, O. S. Lih, V. K. Sudarshan, T. J. Hong, J. E. Koh, Y. Hagiwara, C. K. Chua, C. K. Poo, and T. R. San, "Automated characterization and classification of coronary artery disease and myocardial infarction by decomposition of ecg signals: A comparative study," *Information Sciences*, vol. 377, pp. 17 – 29, 2017.

[8] Y. Birnbaum *et al.*, "The role of the ecg in diagnosis, risk estimation, and catheterization laboratory activation in patients with acute coronary syndromes: a consensus document," *Annals of Noninvasive Electrocardiology*, vol. 19, no. 5, pp. 412–425, 2014.

[9] A. Sanaani, S. Yandrapalli, G. Jolly, R. Paudel, H. A. Cooper, and W. S. Aronow, "Correlation between electrocardiographic changes and coronary findings in patients with acute myocardial infarction and single-vessel disease," *Annals of translational medicine*, vol. 5, no. 17, 2017.

[10] A. E. Johnson, J. Behar, F. Andreotti, G. D. Clifford, and J. Oster, "R-peak estimation using multimodal lead switching," in *Computing in Cardiology 2014*, 2014, pp. 281–284.

[11] B. Qian and K. Rasheed, "Hurst exponent and financial market predictability," in *IASTED conference on Financial Engineering and Applications*, 2004, pp. 203–209.

[12] J. Haslett and A. E. Raftery, "Space-time modelling with long-memory dependence: Assessing Ireland's wind power resource," *Applied Statistics*, vol. 38, no. 1, 1989.

[13] E. Maharaj, "Clusters of time series," *Journal of Classification*, vol. 17, pp. 297–314, 2000.

[14] M. Saeed, M. Villarroel, A. Reisner, G. Clifford, L. Lehman, G. Moody, T. Heldt, T. Kyaw, B. Moody, and R. Mark, "Multiparameter intelligent monitoring in intensive care ii (mimic-ii): A public-access intensive care unit database," *Critical Care Medicine*, vol. 39, no. 5, pp. 952–960, 2011.

[15] D. N. Reshef, Y. A. Reshef, H. K. Finucane, S. R. Grossman, G. McVean, P. J. Turnbaugh, E. S. Lander, M. Mitzenmacher, and P. C. Sabeti, "Detecting novel associations in large data sets," *Science*, vol. 334, no. 6062, pp. 1518–1524, 2011.

[16] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, 2010, pp. 249–256.