MED-NET: A Novel Approach to ECG Anomaly Detection Using LSTM Auto Encoders

Koustav Dutta[1], Rasmita Lenka[1], Soumya Ranjan Nayak[2], Asimananda Khandual[3], Akash Kumar Bhoi[4]

[1]Kalinga Institute of Industrial Technology, Deemed to be University, Bhubaneswar, Odisha, India.

[2]Amity School of Engineering and Technology, Amity University Uttar Pradesh, Noida, India.

[3]Department of Textile, College of Engineering and Technology, Bhubaneswar, Odisha, India

[4]Department of Electrical & Electronics Engineering,Sikkim Manipal Institute of Technology (SMIT), Sikkim Manipal University (SMU), 737136 Majitar, Sikkim, India

[1]koustavdutta.dgp@gmail.com, [1]rasmitafet@kiit.ac.in, [2]nayak.soumya17@gmail.com, [3]asimte@cet.edu.in, [4]akash730@gmail.com

Biography of Authors

Koustav Dutta received the B.Tech degree in Electronics &Electrical Engineering from KIIT Deemed to be University, in 2020. He had been working as an Undergraduate Researcher and Product Developer in KIIT Design and Thinking Lab for the past 4 years and as the Head Student Researcher and Coordinator in IEEE Robotics & Automation Society. He has a teaching experience of 3 years as a Technical Trainer of IoT and Machine Learning at E-Labs in SOEE at KIIT University, Bhubaneswar and Microsoft Developer Community, Bhubaneswar. His research area broadly includes Machine Learning, Deep Learning, and Robotics, Internet of Things and Image Processing and its applications.

Rasmita lenka received the B.TECH degree in Electronics & Comm. Engineering from Biju Pattnaik University of Technology, in 2004, M.TECH degree in Communication System from Veer Surendra Sai University of Technology (VSSUT) ODISHA in 2010. She is continuing her Ph.D in KIIT Deemed to Be University. She has 14 years of teaching experience and currently working as an Assistant Professor in SOEE at KIIT University, Bhubaneswar. Her research area broadly includes signal processing and Image processing and its application

Soumya Ranjan Nayak was born in Odisha in 1984. He received the B.Tech, M.Tech and Ph.D degree in computer science and engineering from Biju Patnaik University of Technology, Odisha, India. He served as full time research scholar in college of engineering and technology, State govt. Autonomous Engineering College, Odisha, India, from 2013-2017 academic years under MHRD Govt. of India (TEQIP-II) fellowship. He currently is working as an Assistant Professor at Amity University, India. He has published more than 50 articles in prestigious international peer review journals and conferences in the area of Fractal Graphics, Pattern Recognition and Color and Texture Analysis. He has authored more than two research edited book with international repute publisher. His currently focus on fundamental color image processing. His research interest includes: Wireless sensor network using AI, image analysis on fractal geometry, color and texture analysis jointly and separately.

Asimananda Khandual

Textile Engineering from ITT in 1999 and the M.Sc.(Tech.) and Ph.D.(Tech.) degree in Textile Chemistry from Institute of Chemical Technology,ICT, Mumbai, India (2002-2006). He received the Best Master student award at ICT, Mumbai in 2002-03. He is currently working as an Assistant Professor in Textile Engineering, College of Engineering and Technology, Bhubaneshwar. He did postdoctoral research in the department of Computing and the Institute of Textiles and Clothing at Hong Kong Polytechnic University (2007-12). His research interest includes computational colour science and measurement, colour and illumination modeling, Device spectral characterization, Colour Image processing, colour profiling of devices and optics, nano finishing, textile chemical processing and e-textiles. He has two U.S. patents and more than 50 national international publications and four book chapters. He has the experience as the principal investigator for sponsored projects from Defence Research and Development Organisation (DRDO), AICTE, and Indian Railways.

Akash Kumar Bhoi has completed his B.Tech. (Biomedical Engineering) from Trident Academy of Technology, BPUT University, Odisha and M.Tech. (Biomedical Instrumentation) from Karunya University, Coimbatore in the year2009 and 2011 respectively. In 2019 he is awarded Ph.D. in the area of Biomedical Signal Processing from Dept. of Electrical and Electronics Engineering, Sikkim Manipal Institute of Technology, Sikkim Manipal University, India. He is working as Assistant Professor-I in the Department of Electrical and Electronics Engineering at Sikkim Manipal Institute of Technology (SMIT), India since 2012. He is the University Ph.D. Course Coordinator for “Research & Publication Ethics (RPE)”. He is member of IEEE, ISEIS and IAENG, associate member of IEI, UACEE and editorial board member reviewer of Indian and international journals. He is also a regular reviewer of journals of repute namely IEEE, Springer, Elsevier, Taylor and Francis, Inderscience etc. His areas of research are biomedical signal processing, medical image processing, sensors and transducers, and medical instrumentation. He has published several papers in national and international journals and conferences. He has 80+ documents registered in Scopus database. He has also served on numerous organizing panels for the international conferences and workshops. He is currently editing several books with Springer Nature and serving as Guest editor for special issues of the journal like Springer Nature and Inderscience.

Abstract: Time Series data is generated in various sectors of day to day life. Among all, one of the most important areas of generation and processing of time series data plays vital role in medical domain analysis. In this specific context, various continuous time series dependent EEG and ECG (Electrocardiogram) signals are the most important types of medical signals produced and monitored by doctors. Till date, heavy reliance is on the doctors regarding the detailed analysis of these signals for understanding, monitoring and detecting the anomaly is cumbersome. Thus, this paper proposes a highly novel and robust approach to analyse and detect ECG signals for tracking of anomalies in the signals using Hybrid Deep Learning Architectures (HDLA). The proposed scheme achieves by implementing self-supervised pattern recognition according to the mechanism of Long Short-Term memory networks (LSTM) in terms of auto Encoder and Decoder. Finally, the proposed scheme is tested on Physio-net dataset. The outcome of model can also handle noise associated with ECG-based time series signal, its achieved accuracy is extremely high and over fitting problems is solved in a robust and efficient manner.

.

Keywords: Bio-signals, Encoder, Decoder, LSTM, Auto encoder, ECG, Anomaly, Time Series, Hybrid Model, Reconstruction Error.

1 Introduction

The intense upsurge in death rates due to several heart diseases has spurred an interest in evolving mobile App based real time ECG monitoring systems. Heart (cardiovascular) diseases can be classified into different categories such as coronary heart diseases, cardiac arrest (heart stroke), anomaly detection, peripheral artery diseases and congenital diseases (abnormality condition before birth) [1]. Many of these illnesses and syndromes are progressive, and their prevalence and dominance increase with age. In order to analyze those symptoms medically, ECG signals are the most important parameter to be considered as Non-invasive tools to characterize the statistical characteristics such as, heart rate variability, entropy samples, and the Histogram density and variable coefficients etc. The previous work has been conducted based upon both frequency and time domain as witnessed in literature kind review [2]. However, conventional methods follow frequency domain analysis which mainly lack false alarm results, miss diagnosis, and prone to noise [3]. Hence, there exists a huge scope to analysis of ECG signals in the time domain by implanting hybrid deep learning models.

In this paper, Time Series data Reconstruction for the ECG Signals using Self-Supervised Pattern Recognition Algorithm

[4] is proposed in terms of Encoder and Decoder Mechanism with Long Short-Term Memory Networks (LSTM) [5] is designed for the development of the Pattern Recognition System. Self-Supervised Learning is an elegant subset of unsupervised learning where you can generate output labels 'intrinsically' from data objects by exposing a relation between parts of the object, or different views of the object. In this type of algorithmic implementation, the patterns in the data are learnt by the Neural Network based on similarity of various parameters, unlike Supervised Learning techniques, where the data is labelled by the output variables and the algorithm learns the parameters and features from those labelled data, but, in real-world scenarios in the industrial or healthcare domain , where extracting labeled data is almost impossible or painful task, Self-Supervised Learning Algorithms play an important role and help in extracting meaningful insights and thus, perform the work of pattern recognition of the data. The LSTM network can be structured by combining Encoder-Decoder LSTM to allow the model to accept both variable-length input sequences and to predict or output variable-length output sequences. An LSTM model encoder reads the time series-based input sequence step-by-step in this architecture. The secret state or result of this model reflects an internal awareness of the whole input sequence as a fixed-length vector following the acceptance of the whole input sequence. This vector is then given as input to the decoder model that interprets it as it generates every phase of the output series. In this way, the time series-based ECG Signal data is accepted, processed and thus, a pattern is learnt from the sequential data in time series form and thus, anomaly detection of the ECG Signal is done. A time series is a sequence of numerical data points in successive order. Time series analysis comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data. Time series forecasting is the use of a model to predict future values based on previously observed values. The Hybrid model used in the process is Long Short-Term Memory Network combine with Auto encoders for the reconstruction of the data and thus, anomaly detection is done with effective and most accuracy value.

The remaining part of this paper is represented in a continuous manner as follows. In section 2, overviews of previous work are elaborated. Materials and Method is described in section 3. Section 4 explains the test analyzes for both the graphical and the calculation table. Finally, it provides the concluding remarks in section 5.

2. Existing process: Overview of previous work

The electric, mechanical, thermal and other signals measured in human body or other organic tissues over time are considered as bio signals [6]. In 1895, Willem Einthoven identified electrocardiography as a clinically practical, non-invasive method. These signals were significant in medical research and solely were carried out initially by Bio signals analysis [7]. Low-level signal processing was used to reduce noise and filter in the early 1980s. Timeline models and supervised expert systems for the extraction of attributes were used later in the 90s and diagnosis-based predictive classifiers were used. Automated bio signals treatment has been a key component of computer-aided diagnostics (CAD) and clinical decision-making in recent decades. Existing approaches, however, are not efficient for the use of portable devices for high-dimensional, complicated and real-world data on noise. As a result, the main goal of today's research is for the prediction of real-time signal analysis to increase the accuracy, robustness and speed of the diagnostic systems. Algorithms support the automated and effective analysis of medical data by artificial intelligence and by machine perception algorithm. A deep, multilayered network architecture-based machine learning approach was also applied later on in wide form in the biomedical field. Deep learning [19] in health imaging thus replaces functional extraction produced by learning from the raw input data, feeding into several hidden layers and eventually resulting in various end-to-end learning parameters. [8]. These signals change constantly and indicate the human body's health. Many researches in the area of ECG anomaly analysis have been performed in the past Zhinang [9]. He presented a new, multi-faceted approach to the ECG signal visualization and analysis. The algorithms will consider the intricate structure and show data from the 1 D Manifest in a 2D framework in their paper, for example "Dimensionality Reduction for electrocardiogram anomaly detection: A Manifold Approach". The key issue with this method, however, was the data leakage mechanism, which in effect affects the intensity and consistency of the solution. Again, in her paper "Long-Term Memory Network for Time Series Anomaly Detection". Malholatra [10] presented an approach that was based heavily on long-term reliability on previous ECG signals. However, again, due the Vanishing Gradient issue that affected the network, the data could not be learned robustly and therefore adequate results could not be reached. Various works of study have used the ECG signal anomaly detection mining technique in time series. Anomaly detection algorithms typically identify the most unusual subsequences in long series as an anomaly or discord. Most programs are using discrete approaches to eliminate unwanted inferences and to calculate the separation between subsequences through pruning techniques using distance measurements. Sanchez, Bustos [11] recently proposed an algorithm with concept of HOT SAX algorithm, which would allow effective discovery of discord in time series data. The algorithm aimed at reducing the algorithm's time complexity. However, an ECG dataset was used without any result of an anomaly or parameter configurations in the experiment. The questionnaires are provided to doctors who work closely with ECG machines to study ECG problems of artifacts and false alarm results. In ECG the errors are also consider in terms of measurement parameters. In the article [24] idea of Short-term Fourier Transforms (STFT) and Continuous Wavelet Transformation (CWT) are implemented as graphical Method for calculating noise in filtered ECG signals and validating filtering proposed strategies results. Calculation of mathematical assessment parameters here uses SNR, Mean Square Error (MSE), Root Mean Square Error (RMSE), Peak Noise Ratio (PSNR) and Peak to Peak Amplitude (P2P) upon filtering ECG signals by Graphical findings (Frequency Domain Analysis STFT and CWT). It was found that ECG machines frequently misinterpret ECG problems as an artifact and cause unnecessary alarms on bedside monitors for patients and physicians equally. Non-experienced physicians need to check with cardiologists to reassess the findings manually, taking the ECG indications of 12 leads into account. In the article [16] a new method is discussed combined in terms of analytical model for electrocardiogram QRS complex, ST section, and T wave (i.e., QT complex) to determine the initiation or frequency of cardiovascular disorders. The methodologies suggested also distinguish healthy, arrhythmic and ischemic patients Furthermore, clinical analysis is required and the patient ECG may be re-registered. Therefore, false alarm results not only waste cardiologists' precious time, but also cause inexperienced physicians to get misdiagnosed. As Figure (1) below shows, existing anomaly detection algorithms still have false alarm findings [25].

Figure 1.An ECG sample with an anomaly beat

3. Materials and Proposed Methodology

In this section, the complete dataset and the proposed methodology adopted in this paper are presented carefully. This section consists mostly of two main sections, where first part relates to the dataset used in the proposed scheme and second part to the core idea of the proposed model. Original data from the BIDMC Congestive Heart Failure Database (CHFDB) physio-net [12] is originally named. It is registered as "chf07". After data collection the data has been pre-treated in two steps: (1) each heartbeat is extracted, (2) every heartbeat is interpolated to equal length and, lastly, 5,000 heartbeats are chosen at random. The patient has severe congestive cardiac failure and automatic annotation obtained the class values. The dataset includes 5,000 examples from the time series (received with ECG) and 140-time steps. Each sequence corresponds to a single heartbeat from a single patient with congestive heart failure. In general, there are 5 types of heartbeats (classes) such as: (i) Normal (N), (ii) -on-T Premature Ventricular Contraction (R-on-T PVC), (iii) Premature Ventricular Contraction (PVC), (iv) Supra-ventricular Premature or Ectopic Beat (SP or EB), (v) Unclassified Beat (UB).1. R-on-T Premature Ventricular Contraction (R-on-T PVC): The "R-on-T phenomenon" is the superimposition of an ectopic beat on the **T** wave of a preceding beat. Early observations suggested that **R-on-T** was likely to initiate sustained ventricular tachyarrhythmia that can result in ventricular arrhythmias leading to cardiac arrest.

2. Premature Ventricular Contraction: PVCs may be *unifocal* (see above), *multifocal* or *multiformed*. Multifocal PVCs have different sites of origin, which means their coupling intervals (measured from the previous QRS complexes) are usually different. Multiformed PVCs usually have the same coupling intervals (because they originate in the same ectopic site but their conduction through the ventricles differs). Multiformed PVCs are common in digitalis intoxication, Low blood oxygen, which could happen if someone have chronic obstructive pulmonary disease (COPD) or pneumonia.

3. Supra-ventricular Premature or Ectopic Beat (SP or EB): Supraventricular premature beats are atrial contractions triggered by ectopic foci rather than the senatorial node. They arise within the atria (atrial premature beats) or, through retrograde conduction, in the atrioventricular node (junctional premature beats). Premature beats may be found in healthy individuals as well as patients with underlying heart disease. Unless patients exhibit severe symptoms (e.g., tachycardia), those experiencing premature beats do not require treatment.

4. Unclassified Beat (UB): An 'Unclassified' finding may be caused by other arrhythmias, unusually fast or slow heart rates, poor quality recording, or for just falling outside the boundaries of the algorithmic requirements.

We considered here as class value with 140 features extracted from 5000 original samples. Sampling and preparation of dataset could have been done on per 15 seconds basis and the effect on both accuracy and response time could have been recorded and thus, analyzed. Previously, the response time of the algorithm was 259 ms and the accuracy achieved using Adam Optimizer was 97.93 % , however, by using a sampling method on 15 seconds basis, the response time of the algorithm would have surely increased but, on the other hand, a compromise would have to be made for accuracy which would be somewhat decreased. So, a trade-off between accuracy and response time is present in the implementation process of the algorithm. Every cardiac rhythm or heartbeat takes about 0,8 seconds for each rhythm is assumed completely a healthy heart and a normal rate of 70 to 75 beats per minute. Frequency range is from: 60–100 (Humans) Duration: 0.6–1 second (Humans).Once data set is generated it is used as input to the hybrid mode. The implementation and architecture development of the model is discussed in below paragraph. Secondly, we focus on core portion of proposed hybrid model for an efficient way for Anomaly Detection in an ECG Signals. A robust Self-Supervised Pattern Recognition Algorithm consisting of Stacked Long Short-Term Memory Network Layers along with Auto-encoder Networks has been the pillar of this Hybrid Model. The Hybrid Architecture is efficient in dealing with any type of noisy time series-based signals. The architecture consists of Sequential Layers, LSTM Layers, Dense Layers, Repeat Vector Layers, Time Distributed Layers and Auto-encoder Network (Encoder & Decoder Networks). The complete block diagram of the proposed scheme is depicted in Figure. 2.

Figure 2. Proposed Hybrid Model Architecture

The proposed algorithm consists of three major parts such as (i) Stacked Long Short-Term Memory Networks (LSTM), (ii) Auto encoder Network, and (iii) Hybrid Model (LSTM-AE): which combines and produces a hybrid model. The individual design and working of blocks are described in individual sub sections.

3.1. Long Short-Term Memory Networks (LSTM)

Each section displays the workflow of the LSTM. This is a special RNN developed by Hochreiter [13] to learn long-term dependencies. LSTMs were precisely designed to prevent long-term dependence. Knowledge is also a classic technique over long stretches of time, not something they seek to learn. Both replicated neural networks are chain-repeating neural networks [14]. This recurring module is very basic in regular RNNs, such as a single tanh layer. LSTMs do have this chain-like structure, but the repeated module has another structure. There are four communicating in a very unique manner, instead of just one neural network layer. The internal structure of LSTM is presented in Figure.3 considered [26].

Figure 3. Architecture of Long Short-Term Memory Network (LSTM)

In Indeed, LSTM networks have memory block connections to layers seen in Figures 4 and 5 instead of neurons. The block features that make it smarter than a neuron and a new sequence memory. A system has gates for node status and output track. Each machine operates on an input sequence, and each gate within a machine uses the sigmoid signal system (A sigmoid function is a mathematical function having a characteristic "S"-shaped curve or sigmoid curve. The main reason why we use sigmoid function is because it exists between 0 to 1. Therefore, it is especially used for models where we have to predict the probability as an output.) to control the malfunction, which allows for state adjustments and conditional integration of information into the unit. There are three types of gates [27] in a memory network [14] such as (i) Forget Gate: conditionally determines the information to remove from the network; (ii) Input Gate: conditionally determines the values to change the memory state from the data; (iii) Output Gate: conditionally decides whether to correct from the unit's data and memory. Every unit is like a mini-state machine where unit gates gain weights as seen in Figure (4), (5), and (6) respectively. A LSTM autoencoder [15] is a sequence data autoencoder with an Encoder-decoder LSTM architecture. For a given sequence data set, an encoder-decoder LSTM is set to read, encrypt, decode and reconstruct the input series. The basic idea of using auto-encoders for time series feature extraction based on two parts such as (i) Encoder: it will learn to extract the fine and minute temporal features from the time series based ECG Signals, where X is the original ECG Signal fed as an input, and (ii) Decoder: it will try to recreate and reconstruct the same ECG Signals as equivalent to input. At this point, we have Y in F(X) = Y and try to generate the input X for which we will get the output. There may be multiple time series signals for which we may get the reconstructed signal. Hence its gives rise to some loss in the process which we target to minimize and generate the perfect input time series signal's features. The different layers of the architecture [28] are presented in Figure.5.

Figure 4. Stacked LSTM- Autoencoder Architecture

Figure 5. Details step by step Stacked LSTM Autoencoder Network

At each time step, Normal Time Series data of ECG Signals which is a 2D Tensor (columns x rows) in the shape sequence length x number of features (140x1 in our case) is fed into the 1st LSTM Layer. The first layer of Encoder Part is an LSTM Layer consisting of 128 hidden neurons to capture the temporal dependencies of the data. The various non-linearity present in the signal are being captured keeping the sequential dependencies intact. The Return Sequences system is enabled in the layer which makes each cell per time step to emit a signal.Return Sequences return the hidden state output for each input time step. This helps the LSTM Neural network to extract each and every temporal information from the sequential time series data and thus, learn the parameters effectively in order to identify the patterns in the time-series data. Using Return Sequences set to True, the information at each and every step of the signal can be extracted and learnt. The activation function used after this layer is ReLU (Rectified Linear Unit) [ReLU: max (0, x)], which helps in extracting the non-linear temporal features from the signal. Next, the sequential data is further encoded with the help of another LSTM Layer consisting of 64 hidden neurons in order to extract all the features of the time-sequential ECG data, thus, the Encoder Network is prepared and the Return Sequences system is disabled in the layer and thus, the output is passed on from the last cell of the layer. ReLU (Rectified Linear Unit) Activation function, which helps in extracting the further non-linear features from the ECG Signal's time series-based data is again used after this layer. Further, the network uses a repeat-vector method to generate a 2D grid for the next row. The Repeat Vector layer serves as an encoder-decoder link. The 2D array input is prepared for the first LSTM decoder row, thereby planning a Latent Vector Representation.It is called a Latent variable because it cannot be access during train time (which means manipulate it).In a normal Feed Forward NN it cannot manipulate the values output by hidden layers. So, the term latent basically can be attributed, and finally we map higher dimensional data to a lower dimension data with no prior convictions of how the mapping will be done. The NN trains itself for the best configuration. We cannot manipulate this lower dimensional data as it was like hidden from user. Thus, the decoder layer is designed for encoding to unfold the data. The decoder layers are then stacked in the reverse direction of the decoder. Now, the operation of Decoding is done in order to perform the Signal Reconstruction process. The Encoded temporal feature map is accepted as input in this first layer of the Decoder part. In the decoder part of the network, the data is fed forwarded into another LSTM layer consisting of 64 hidden neurons in order to extract back the original temporal features. It is the mirror image of the layer 2 of the -network. Next, another LSTM Layer is included in the network which again, is actually a mirror image of the layer 1 of the network, i.e., it consists of 128 hidden neurons which helps in the re-creation of the fine details of the time series-based ECG Signals. The Activation Function used after this layer is ReLU (Rectified Linear Unit) which helps in extracting the non-linear encoded units from the signal).ReLU Activation Function is used in the process of getting detailed and complex features from the image.

ReLU stands for rectified linear unit, and is a type of activation function. Mathematically, it is defined as $y = max(0, x)$. ReLU is the most commonly used activation function in neural networks. ReLU is linear (identity) for all positive values, and zero for all negative values. ReLU is used repeatedly in the process as it doesn't suffer from Vanishing Gradient problem.The Vanishing Gradient Problem is encountered when training artificial neural networks with gradient-based learning methods and back propagation [21]. In such methods, each of the neural network's weights receive an update proportional to the partial derivative of the error function with respect to the current weight in each iteration of training. The problem is that in some cases, the gradient will be vanishingly small, effectively preventing the change in weight from its value. In the worst case, this may completely stop the neural network from further training. As one example of the problem cause, traditional activation functions such as the hyperbolic tangent function have gradients in the range (0, 1), and backpropagation computes gradients by the chain rule. This has the effect of multiplying *n* of these small numbers to compute gradients of the "front" layers in an *n*-layer network, meaning that the gradient (error signal) decreases exponentially with *n* while the front layers train very slowly. Also, the neurons are selectively activated and deactivated in order to get detailed information from the temporal feature map.Specifically, the learning rate is a configurable hyper parameter used in the training of neural networks that has a small positive value, often in the range between 0.0 and 1.0. The learning rate controls how quickly the model is adapted to the problem. Its value used in this network is 0.001. Thus, basically, the Learning Rate parameter helps to decide the rate at which the derivative of the Loss Function of the Neural Network will reach zero and attain the Global Minima Position thus, optimizing the entire network

In the last layer of the Hybrid Architecture, Time Distributed Layer is present. The distributed time layer generates a length vector equal to the number of features that have been generated in the previous layer. Layer 5 outputs 128 features in this network. The time distributed layer then generates a 128 long vector and duplicates it 140 times (= n features) and thus recreates the re-created ECG Signal time series (input) data.

3.1.1. Forget Gate: The In our LSTM, the first step is to determine what details we'll throw away from the cell state. This is handled by a sigmoid layer called the "forget gate layer," which looks at $h_{t-1}$and x_t, and outputs a number between 0 and 1 for each number in C_{t-1}. ‘1’ represents "absolutely keep this" while ‘0’ represents "fully get rid of it," as seen in equation (1).

Figure 6. Representation of Forget Gate

3.1.2. Input Gate: The next step is to determine which new data should be processed in the cell state. That's two bits. Next, a sigmoid layer called ‘input gate layer’ defines the values are modified by equation (2). Next, a ***tanh*** layer produces a vector to add new candidate values, C_t, to the entity. In the next step, we must combine these two to construct an update as given in equation (3) which is clearly depicted on Figure 7.

$$f_t = \sigma\left(W_f .\left[h_t - 1, x_t\right] + b_f\right) \qquad (1)$$

$$i_t = \sigma\left(W_i .\left[h_t - 1, x_t\right] + b_i\right) \qquad (2)$$

$$c_t = \tanh\left(Wc.\left[h_t - 1, x_t\right] + b_c\right)(3)$$

Figure 7. Representation of Input Gate

3.1.3. Intermediate State: This is the time to upgrade the old cell status, C_{t-1}, into to the current cell state C_t. The previous steps have already determined what have to implement in next step. We multiply the old state f_t, forgetting the things that we had already decided to forget. Then we add the new candidate $i_t * C_t$ which is how much we have chosen to update single state rating, that presented in Figure 8.

$$C_t \;\; = f_t * c_{t-1} \; + \; i_t * c_t (4)$$

Figure 8. Representation of Intermediate State

3.1.4. Output Gate: Finally, we will decide what the outcome is going to be obtained. It focuses on the filtered cell state. Therefore, we run a sigmoid layer that defines which part of the cell condition we’re going to consider. We then set the cell state to ***tanh*** *(*to push the values between −1 and 1) so that the output of the sigmoid gate multiplies and the desired output are decided. The details representation as shown in Figure 9.

Figure 9. Representation of Output Gate

$$O_t \;\; = \sigma(W_o \bullet [h_{t-1}] + b_o)(5)$$

$$h_t \;\; = O_t * tanh(C_t)(6)$$

But LSTM faces the problem regarding computational efficiency and response time. In order to solve this issue, we implement GRU (Gated Recurrent Units) are gating mechanism in recurrent neural networks, introduced in 2014 by Cho et al. The GRU is like a long short-term memory (LSTM) with a forget gate but has fewer parameters than LSTM, as it lacks an output gate. GRU's performance on certain tasks of polyphonic music modelling, speech signal modelling and natural language processing was found to be similar to that of LSTM. GRUs have been shown to exhibit even better performance on certain smaller and less frequent datasets and the computational efficiency and response time of the GRU Network is also better than LSTM. However, for very long-term time-series signals, temporal information insights are not so well learnt and integrated into the network by GRUs.

3.2. Auto encoders: Along with the Long Short-Term Speicher Network, auto encoders are used to create a hybrid deep learning architecture. An auto encoder [17] is a type of artificial neural network [10] for efficient, unattended data coding. The aim of an auto encoder is to acquire a representation (encoding) of a collection data by teaching the network to disregard signal noise, usually to minimize dimensionality also called as dimensionality reduction. Auto encoders are a branch of neural network which attempt to compress the information of the input variables into a reduced dimensional space 'z' and then recreate the input data set. Basically auto-encoder consist of two parts an encoder and a decoder. The encoder takes the raw input and encode it and form a intermediate vector which parsed into the target data. The basic model [18] has many versions in order to oblige the learned representation of the data to assume useful properties. A feed forward, non-recurring neural network is the best path for an auto encoder, and is similar to a single-layer representation that has multilayer perceptron (MLP) with an input layer, output layer and one or more hidden layer connecting it. The output layer has the same number of nodes (neurons) as the input layer and to retrieve its inputs (minimizing). Auto encoders are not unattended learning models (do not need defined inputs for learning)[20].The encoder 'encodes' the data which is of higher dimensional into a latent (hidden) representation space 'z', which is much less than original higher dimension. This is typically referred to as a 'bottleneck' because the encoder must learn an efficient compression of the data into this lower-dimensional space. We can sampled from this distribution to get noisy values of the representations z. The decoder gets as input to the latent representation of the space 'z' and outputs are Bernoulli parameters of original size. A variation auto encoder assumes that the source data has some sort of underlying probability distribution (such as Gaussian) and then attempts to find the parameters of the distribution. Implementing a vibrational auto encoder(probabilistic generalization approach) is much more challenging than implementing an auto encoder but, the feature extraction process is more detailed and better than stacked auto encoders and in some cases, the accuracy achieved is also high. This can be described in the Equation as transitions ± and ± bias. (7), (8) and (9) respectively. In the simplest case, given a hidden layer, the auto encoder stage takes the input as seen in Eq. (10) and (11), respectively

$$\phi : \chi \rightarrow F \quad (7)$$

$$\psi : F \rightarrow \chi \quad (8)$$

$$\phi, \psi = \arg\min \left\| x - (\psi \circ \phi)\, X \right\|^2 \quad (9)$$

$$x \epsilon R^p \ = \chi and\ maps\ it\ into\ h \in R^4 = F \quad (10)$$

$$h = \sigma(Wx + b) (11)$$

The knowledge *'h'* is usually called encoded data, latent variables, or latent representation for this time sequence. Here *'σ'*is an activation function of the entity, like a sigmoid component or a rectified linear form. '*W*' is a weight matrix and '*b'* is a bias vector. Weights and biases are initialized automatically and modified iteratively during back propagation testing. Instead the auto encoder stage of the '*h'* maps to reconstruct *x′* in the same form as '*x'* in Eq. (12).

$$x' = \sigma' \left(W'h + b' \right) \quad (12)$$

Where σ′, W′, and b′ are the corresponding activation function, weight, and bias factor of the encoder. Auto encoders are equipped to eliminate reconstruction errors (like squared errors), also called "losses" in Eq. (13). Where x' is typically averaged over certain input training sets. The preparation of an auto encoder is performed by the propagation of the error backwards as is the case for a normal feedback neural network. Figure.10[29]

$$L(x, x') = \left\| x - x' \right\| = \left\| x - \sigma'\left(W'\sigma\left(Wx + b\right)\right) + b' \right\|^2 \quad (13)$$

4. Result Analysis and Discussion

Detailed result analysis of the proposed scheme are presented in this section. The architecture consists of a long-term, auto encoder architecture-based memory network (LSTM) [22] to achieve the proposed hybrid deep learning and obtain the desired performance. Because LSTM is powerful to work with sequential time series data, derive detailed information and temporal features from the sequential ECG Signals data. LSTM is then used to obtain the initial ECG signal data as input layer. Along with Stacked LSTM Architecture, Auto encoders are used by binary means to represent temporal attributes in a latent matrix. In this paper around 140 characteristics are processed and extracted from present dataset of 5000 samples. Exploratory Data Analysis is being carried out on the dataset and thus, the following classes of heartbeats based on ECG Signals are produced and labeled below in Figure. 11. The different nature of the time series signals of the various types of Heart Beats is given below in Figure. 12.

Figure 11. Histogram of various heartbeats

Figure 12. Time Series nature of the signals of different types of Heart Beat

The goal of the fitting network variation is to connect this output to the data. The network itself guarantees that the measurements of input and output are matched. After getting the recreated time series-based ECG data as output in the final layer, this output is compared with the original ECG time series-based input data in order to calculate the Reconstruction Error or Loss (L1 Loss Function). L1 is the Loss function which reduces the error representing the total of all the absolute differences between the actual value and the forecast value. Back-propagation is performed in order to reduce the loss with the help of Adaptive Moment Optimizer and several other optimization functions like RMSProp and Stochastic Gradient.

SGD- Stochastic gradient descent is an iterative method for optimizing an objective function with suitable smoothness properties (e.g. differentiable or sub-differentiable). It can be regarded as a stochastic approximation of gradient descent optimization, since it replaces the actual gradient from the entire data set by an estimation (calculated from a randomly selected subset of the data). Especially in high-dimensional optimization problems this reduces the computational burden, achieving faster iterations in trade for a lower convergence rate. The accuracy achieved using this Optimization Algorithm is less than Adam Optimizer and it is also one of the most traditional and oldest optimization algorithms.

RMSProp is an optimizer that utilizes the magnitude of recent gradients to normalize the gradients. We always keep a moving average over the root mean squared (hence Rms) gradients, by which we divide the current gradient. This Optimization Algorithm helps to achieve better accuracy than SGD Optimizer but it is lesser than the accuracy achieved using Adam Optimizer. In order to update each and every element in the weights and parameters used in the network. Thus, in the process, the loss goes on decreasing and finally a best reconstructed time series ECG data is obtained. The number of Epochs used during the training is 150.An epoch is a measure of number training vectors that are used once for updating the weights. For batch training all of the training samples passes through the learning algorithm simultaneously in one **epoch** before weights are updated. The Reconstruction Error (Loss) as shown in given Eq. (14), and its corresponding graphical notation presented in Figure. 13.

L1 Loss Function =

$$L = \sum_{i=1}^{n} |y_{true} - y_{predicted}| \quad (14)$$

Figure 13. Training Loss over the time steps

The Hybrid Model was then used to verify system output through the reconstruction of ECG time series-based training results. The reconstruction failure threshold is calculated in conjunction with the figure indicated below, which is used to evaluate the output and precision of the model, to explain the trend in a supervised way and hence attempt to recreate a time series-based anomaly that includes ECG results.: (i) R-on-T Premature Ventricular Contraction (R-on-T PVC), (ii) Premature Ventricular Contraction (PVC), (iii) Supra-ventricular Premature or Ectopic Beat (SP or EB), (iv) Unclassified Beat (UB)). In the process, it is analyzed from the graph, the data whose reconstruction error loss is above the threshold are regarded as Anomaly containing ECG data. The figures of the Reconstruction Error (loss) got by application of the model to reconstruct the ECG training data and the anomaly containing ECG data are given below in Figure. 14, and Figure. 15 respectively.

Figure 14. Reconstruction Error (loss) for normal training data

Figure15.Reconstruction Error (loss) for Anomaly (test) data

Using the threshold loss, the anomaly data in the ECG signal can be detected robustly. If the reconstruction loss for an example is below the threshold, we'll classify it as a normal heartbeat. Alternatively, if the loss is higher than the threshold, we'll classify it as an anomaly. The accuracy is used as a metric in order to analyses the performance of the Self-supervised Hybrid Model in the process of Anomaly Detection of the ECG Signals. A comparative analysis report of accuracies achieved with the help of various network optimization techniques is given below in Table 1

Table. 1. Performance Evaluation on various optimization techniques

Therefore, from the above table, it is found that the Adaptive Momentum [23] (Adam) Optimization Algorithm is actually a combination of Stochastic Gradient Descent with Momentum (SGDM) and RMSProp Algorithms.The Adam optimization algorithm is an extension to stochastic gradient descent that has recently seen broader adoption for deep learning applications. Adam was presented by Diederik Kingma from OpenAI and Jimmy Ba from the University of Toronto in their 2015 ICLR paper (poster) titled “Adam: A Method for Stochastic Optimization as one of the best optimizers chosen by the research community for non-convex optimization problems for better computationally efficiency. Hyper-parameters have intuitive interpretation and typically require little tuning. It helps in finding the best optimization of weights and parameters in order to achieve an accuracy of 97.93 %, which is by far the most robust and efficient algorithm to obtain such a “State-of-Art” accuracy in the process of Anomaly Detection in ECG Signals. But further we can use other metrics for evaluation. F1 Score - In statistical analysis of classification, the F_1 score (also F-score or F-measure) is a measure of a test's accuracy. It considers both the precision *p* and the recall *r* of the test to compute the score: *p* is the number of correct positive results divided by the number of all positive results returned by the classifier, and *r* is the number of correct positive results divided by the number of all relevant samples (all samples that should have been identified as positive). The F_1 score is the harmonic mean of the precision and recall, where an F_1 score reaches its best value at 1 (perfect precision and recall). The F_1 score is also known as the Sorensen–Dice coefficient or Dice similarity coefficient (DSC). The comparative analysis of the Reconstruction Error (loss) between Normal Heart Beat (ECG Signal) and Anomaly containing Heart Beat (ECG signal) is given below in Figure. 16.

Figure 16. Comparative analysis of Reconstruction Loss between Normal & Anomalies Heart Beat

5. Conclusion

This paper opens the doors to a new dimension for application of Self-Supervised Learning techniques in detection of heart Anomaly by implementing hybrid deep learning models using long short-term memory networks (LSTM) with Auto encoders. The outcome of this proposed model also able to handle the persistent problem of False Anomaly Detection of ECG signals. Besides, the model can handle noise associated with the time series-based ECG signals and represent the accuracy which is achieved mostly 97.93 %. Hence, proposed model is more prominent and robust in recent medical ECG analysis. The application of LSTM Auto-encoders has not limited to the only recreation of time series based signals, as per the ongoing research conducted, therefore, work is being done for application of these Self-Supervised Pattern Recognition systems in case of analysis and detection of abnormalities and anomalies in ECG Signal which will bring out a great change in the study of Neurological data. The development of Recurrent Neural Networks and variation of Auto-encoders (RNN-VAE) as the hybrid model could robustly handle the sequential data with the help of the RNN. Moreover, the work is also being done in the field of neural machine translation systems as it could also make use of these Self-Supervised pattern recognition algorithms to understand the patterns.

REFERENCES

[1] S Shibolet, MC Lancaster, Y Danon. Heat stroke: a review. *Aviation Space, Environmental Medicine*. 1976 Mar; 47(3):280-301.

[2] L. Cohen . (1995) Time-frequency analysis.L. Cohen, *Time-Frequency Analysis, Prentice Hall, Upper Saddle River, NJ, USA*, 1995.

[3] JörgHillenbrand,Timo A. Weiss,Friedrich K. Jondral,"Calculation of detection and false alarm probabilities in spectrum pooling systems", *IEEE Communications Letters,* 2005; 9(4): 349-351.

[4] L. Jin and J. Dong, "Classification of normal and abnormal ECG records using lead convolutional neural network and rule inference," *Science China Information Sciences*, 2017; 60(7), 1-3.

[5] Lishuang Li, Liuke Jin, Yuxin Jiang ,DegenHuang,"Recognizing biomedical named entities based on the sentence vector/twin word embeddings conditioned bidirectional LSTM*",Chinese computational linguistics and natural language processing based on naturally annotated big data. Springer, Cham*,2000 pp 165-176.

[6] P. A. Karthick, S. Ramakrishnan, "Surface electromyography based muscle fatigue progression analysis using modified B distribution time frequency features". *Biomed Signal Process Control,* 2016; 26:42-51.

[7] H Sivaraks ,C. A Ratanamahatana," Robust and Accurate Anomaly Detection in ECG Artifacts Using Time Series Motif Discovery". *Computational and Mathematical Methods in Medicine*, 2015; 1-20.

[8] V. Bajaj, R. B. Pachori, "Automatic classification of sleep stages based on the time-frequency image of EEG signals". *Computer Methods and Programs in Biomedicine*. 2013; 112(03):320-328.

[9] T. M. Deserno, N. Marx, *"Computational Electrocardiography: revisiting holter ECG monitoring". Methods Inf Med.* 2016; 55(04):305-11.

[10] P. Malhotra, L. Vig, G. Shroff, P. Agarwal, "Long Short Term Memory Networks for Anomaly Detection in Time Series". *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium),* 2015; 22-24.

[11] H. Sanchez, B. Bustos, "Anomaly detection in streaming time series based on bounding boxes," *Similarity Search and Applications,* 2017; 8821:201-213.

[12] http://timeseriesclassification.com/description.php?Dataset=ECG5000, Access on 12th June 2020.

[13] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink and J. Schmidhuber, "LSTM: A Search Space Odyssey," *IEEE Transactions on Neural Networks and Learning Systems*, 2017; 28(10):2222-2232.

[14] S. Hochreiter, Schmidhuber, J Bridging, long time lags by weight guessing and "long short-term memory." *In F. L. Silva, J. C. Principe, & L. B. Almeida (Eds.), Spatiotemporal models in biological and artificial systems*, 1996; 65-72.

[15] J. Masci, U. Meier, D. Cireşan, J. Schmidhuber, "Stacked Convolutional Auto-Encoders for Hierarchical Feature Extraction. *Lecture Notes in Computer Science, vol 6791.*

[16] A. K. Bhoi, K. S. Sherpa, B. Khandelwal, "Arrhythmia and ischemia classification and clustering using QRS-ST-T (QT) analysis of electrocardiogram*". Cluster Computing*, 2018; 21(1), 1033-1044.

[17] P. Vincent, H. Larochelle, Y. Bengio, P. A. Manzagol. 2008. Extracting and composing robust features with denoisingautoencoders. *International conference on Machine learning (ICML '08). Association for Computing Machinery, New York, NY, USA,* 2008; 1096-1103.

[18] H.Wang; T.Pan; M. K Ahsan,"Hand-drawn electronic component recognition using deep learning algorithm" *,International Journal of Computer Applications in Technology (IJCAT),* 62(1), 2020.

[19] A.Q.Yang;,X.H Yu,T.Li Su, X.Bo Jin, J.L. Kong." Broad learning system for human activity recognition using sensor data*",International Journal of Computer Applications in Technology (IJCAT),* 61 (4), 2019.

[20] C. Darken, J. Chang, and J. Moody. "Learning rate schedules for faster stochastic gradient search", *IEEE Neural Networks for Signal Processing,*1992;1-11.

[21] R Rojas,"TheBackpropagation Algorithm", *Neural Networks Springer Berlin", (Heidelberg),* 1996, 149-182.

[22]D. Park, Y. Hoshi and C. C. Kemp, "A Multimodal Anomaly Detector for Robot-Assisted Feeding Using an LSTM-Based VariationalAutoencoder,*" IEEE Robotics and Automation Letters,* 2018; 3(3):1544-1551.

[23] K. Hara, D. Saito and H. Shouno, "Analysis of function of rectified linear unit used in deep learning," *International Joint Conference on Neural Networks (IJCNN), Killarney*, 2015; 1-8.

[24] A. K. Bhoi, K. S. Sherpa, B. Khandelwal, "Comparative Analysis of Filters for Cancellation of Power-line-interference of ECG Signal*". International Journal Bioautomation,* 2019; 23(3): 259-282.

[25] https://images.app.goo.gl/LvQZCgHwLiaBuZnX9. Access on 16th July 2020.

[26] https://images.app.goo.gl/PF9mE7vcaFHjihad7.Access on 16th July 2020.

[27] https://colah.github.io/posts/2015-08-Understanding-LSTMs/ Access on 16th July 2020

[28]https://images.app.goo.gl/5EUQurSeuC5keJuF6.Access on 16th July 2020.

[29]https://images.app.goo.gl/tYpXsWb2pc2dJsRw5 Access on 16th July 2020.

Figures in Menu Script

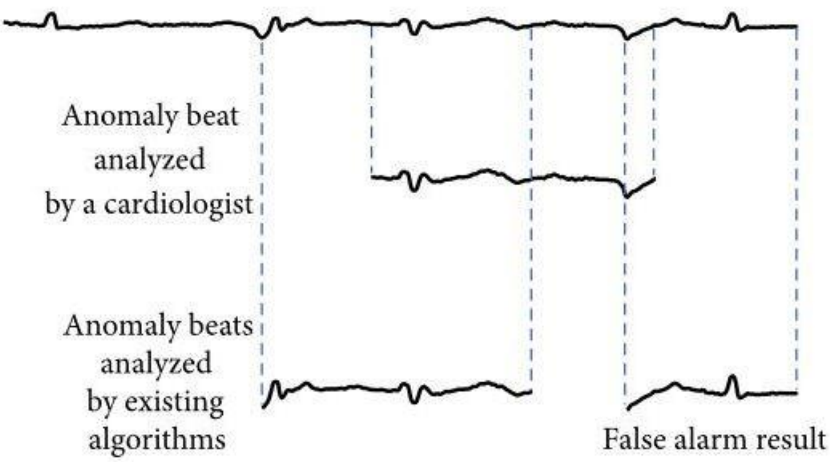


Figure 1. An ECG sample with an anomaly beat

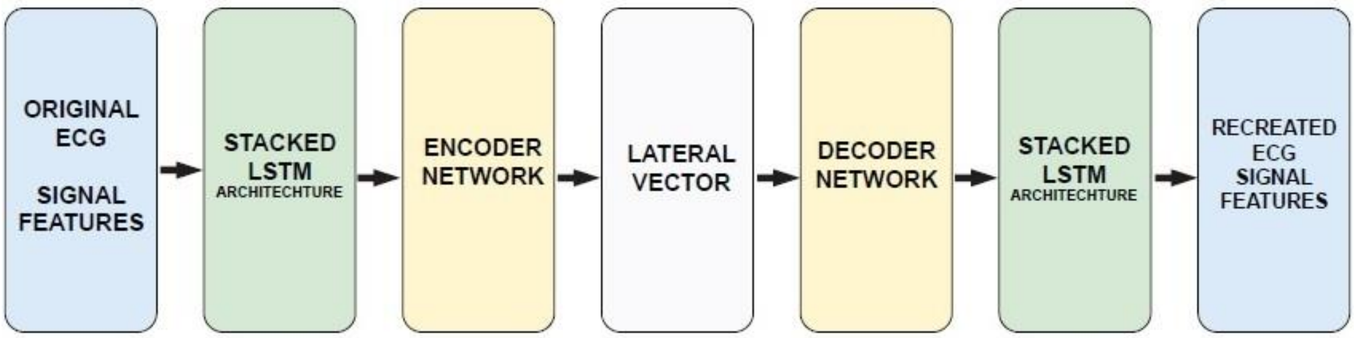


Figure 2. Proposed Hybrid Model Architecture

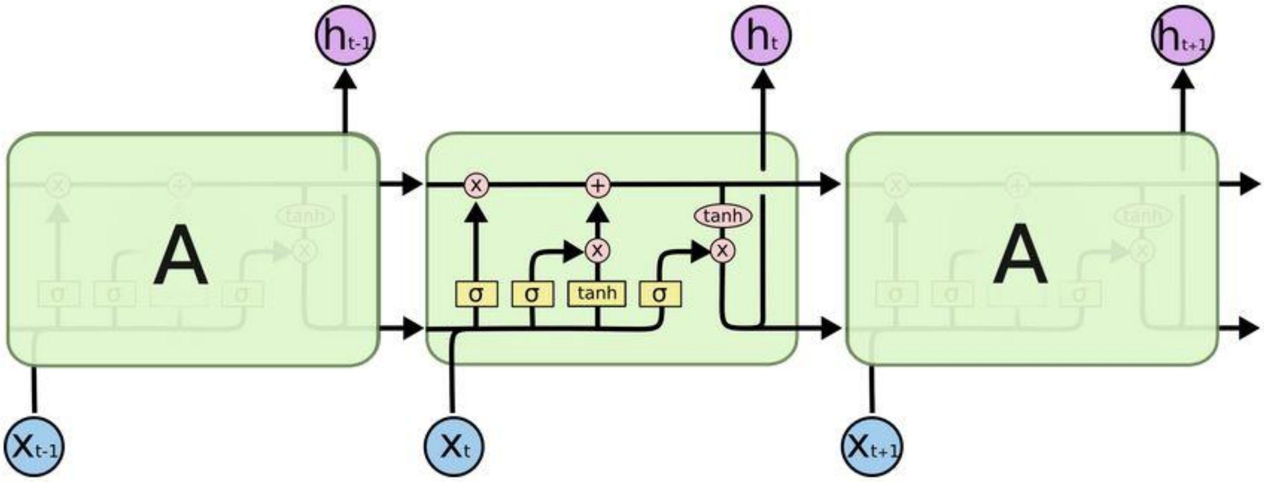
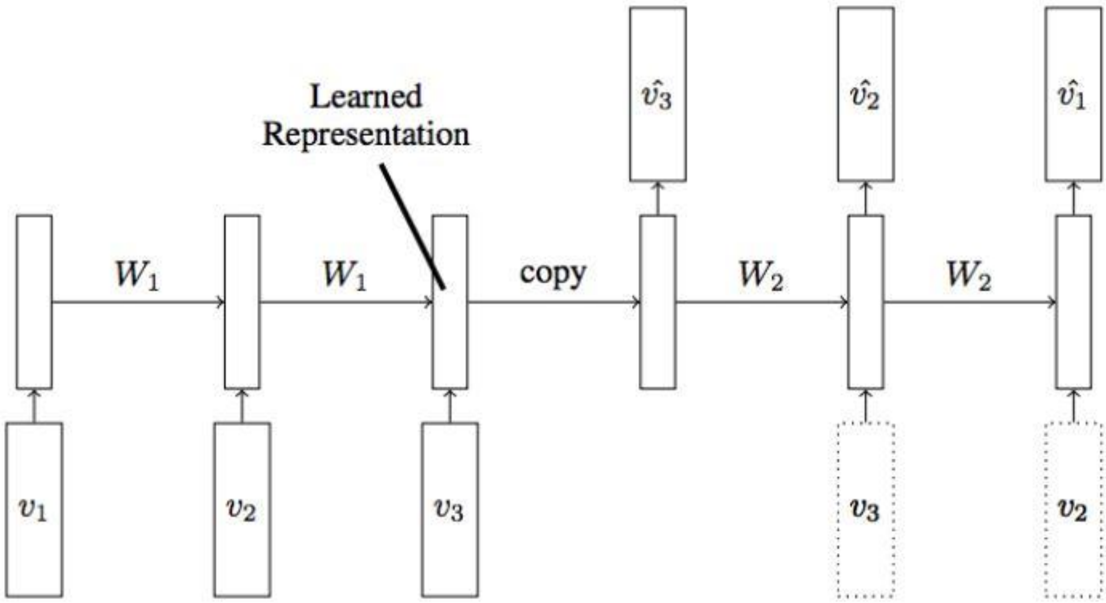


Figure 3. Architecture of Long Short-Term Memory Network (LSTM)



. Figure 4. Stacked LSTM - Autoencoder Architecture

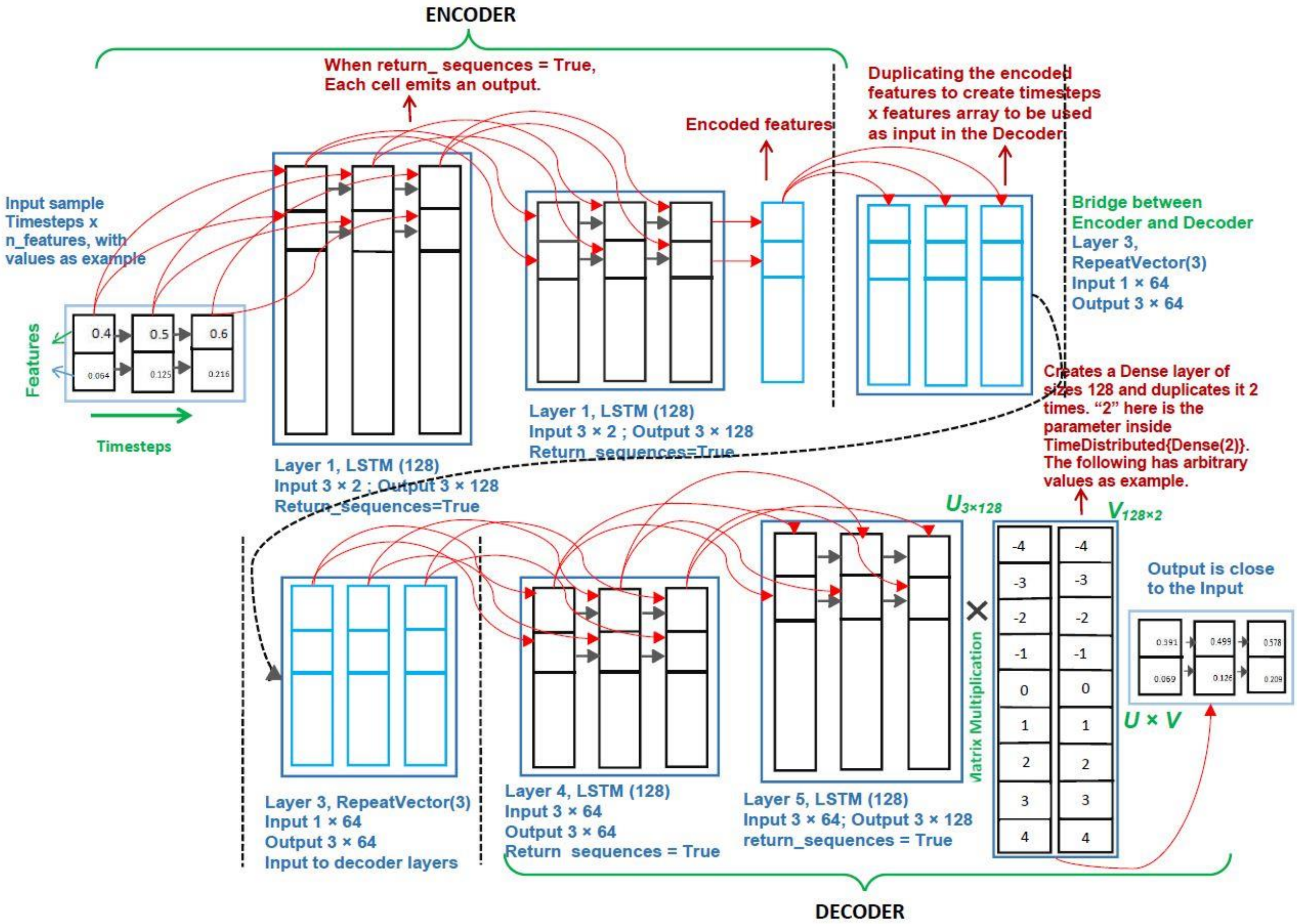


Figure 5. Detailed step-by-step Stacked LSTM – Auto encoder Network

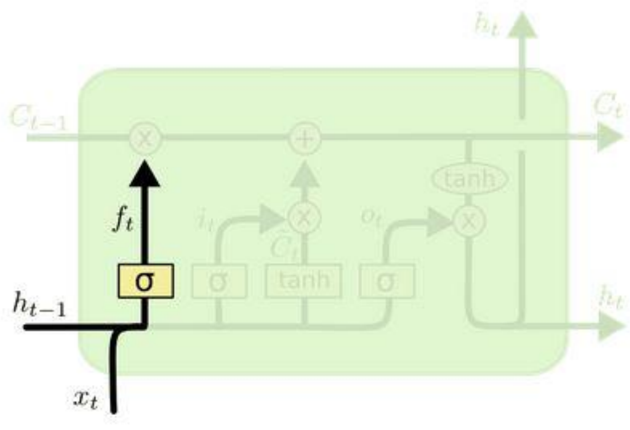


Figure 6. Representation of Forget Gate

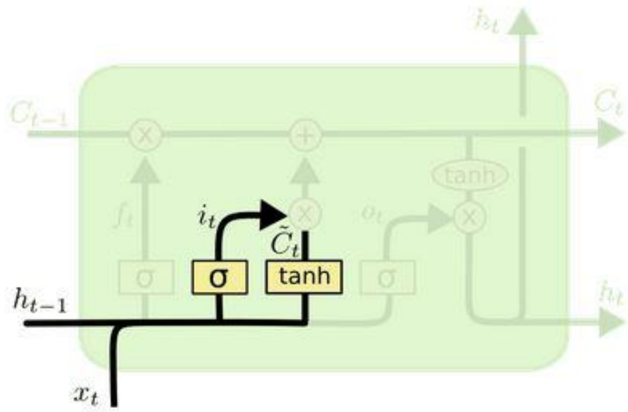


Figure 7. Representation of Input Gate

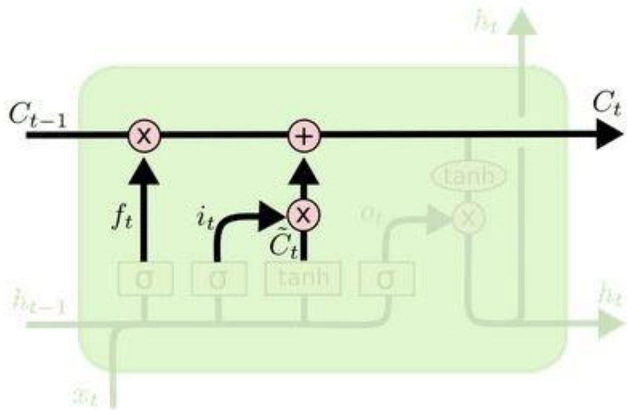


Figure 8. Representation of Intermediate State

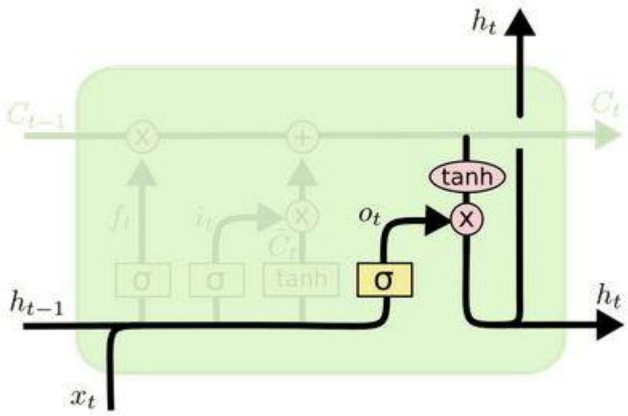


Figure 9. Representation of Output Gate

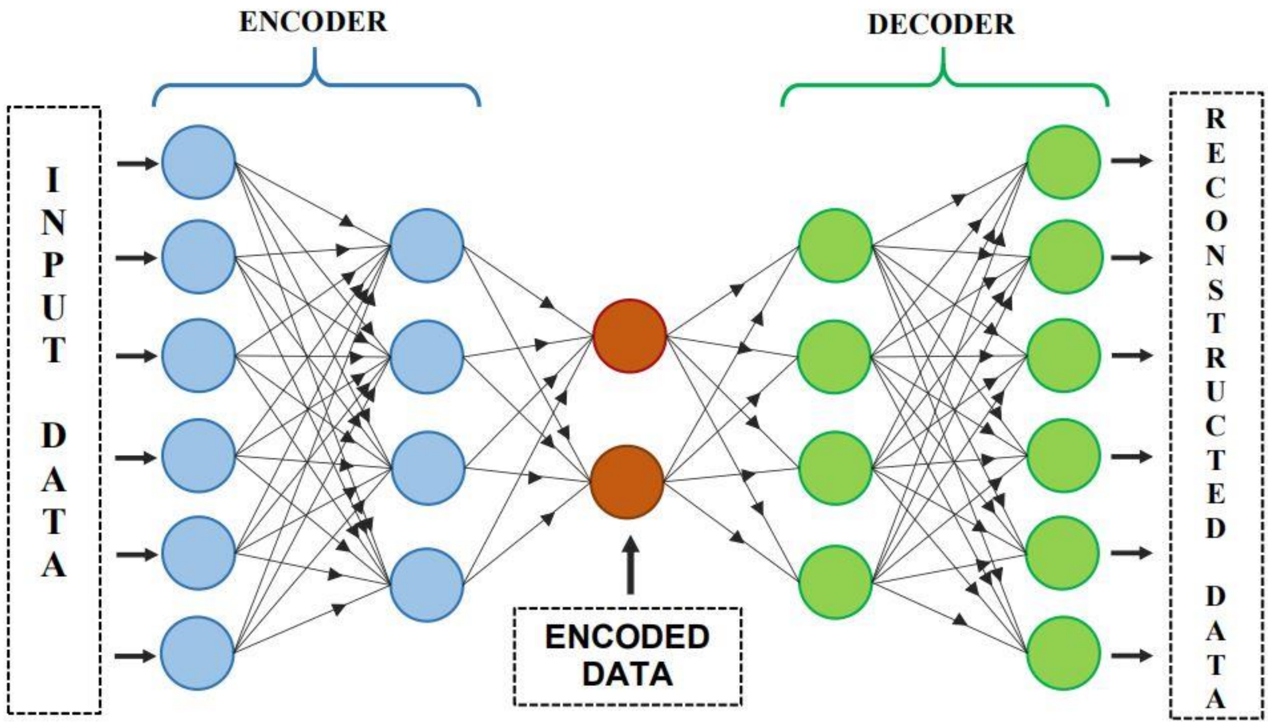


Figure 10. Proposed Autoencoder Architecture

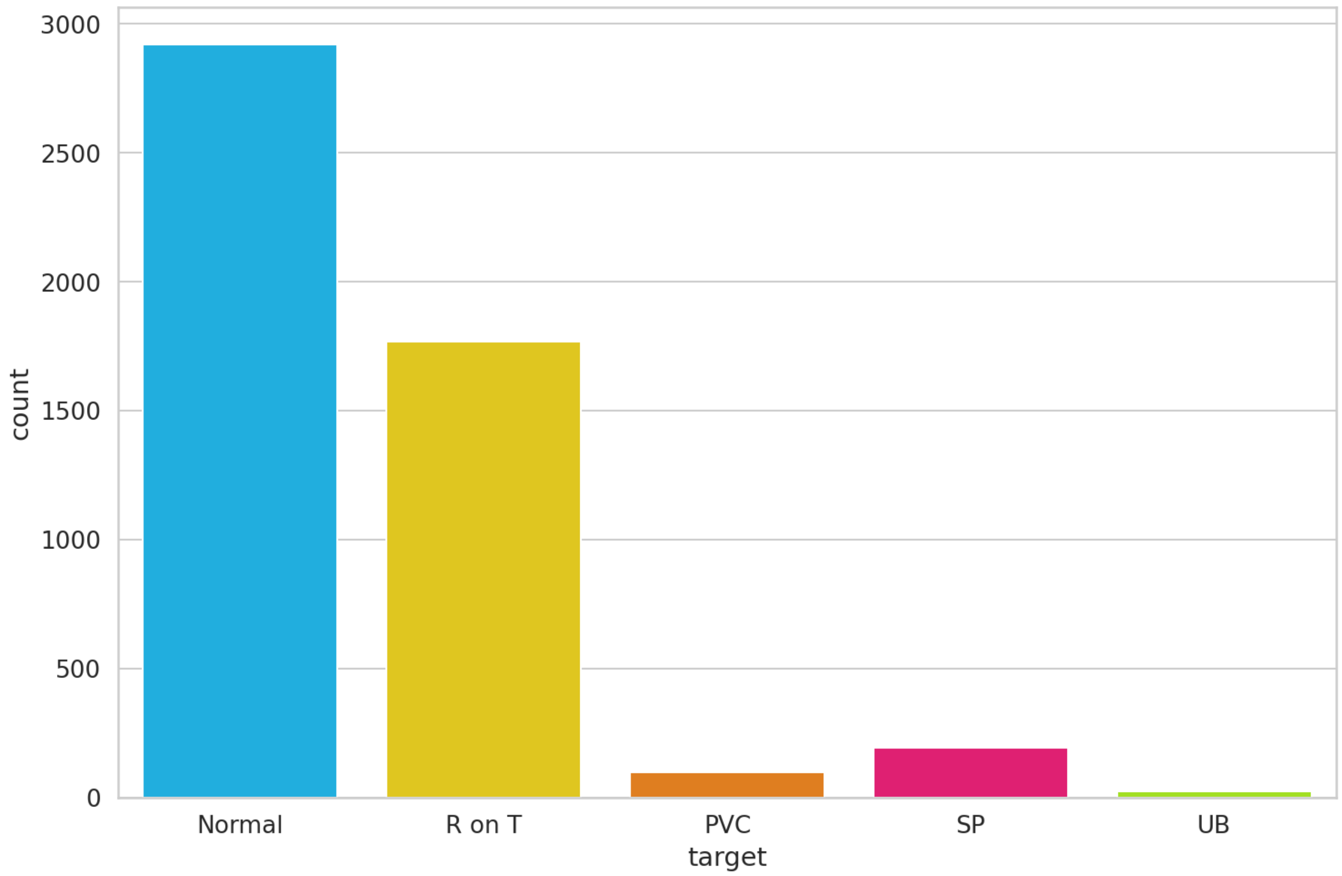


Figure 11. Histogram of various heartbeats

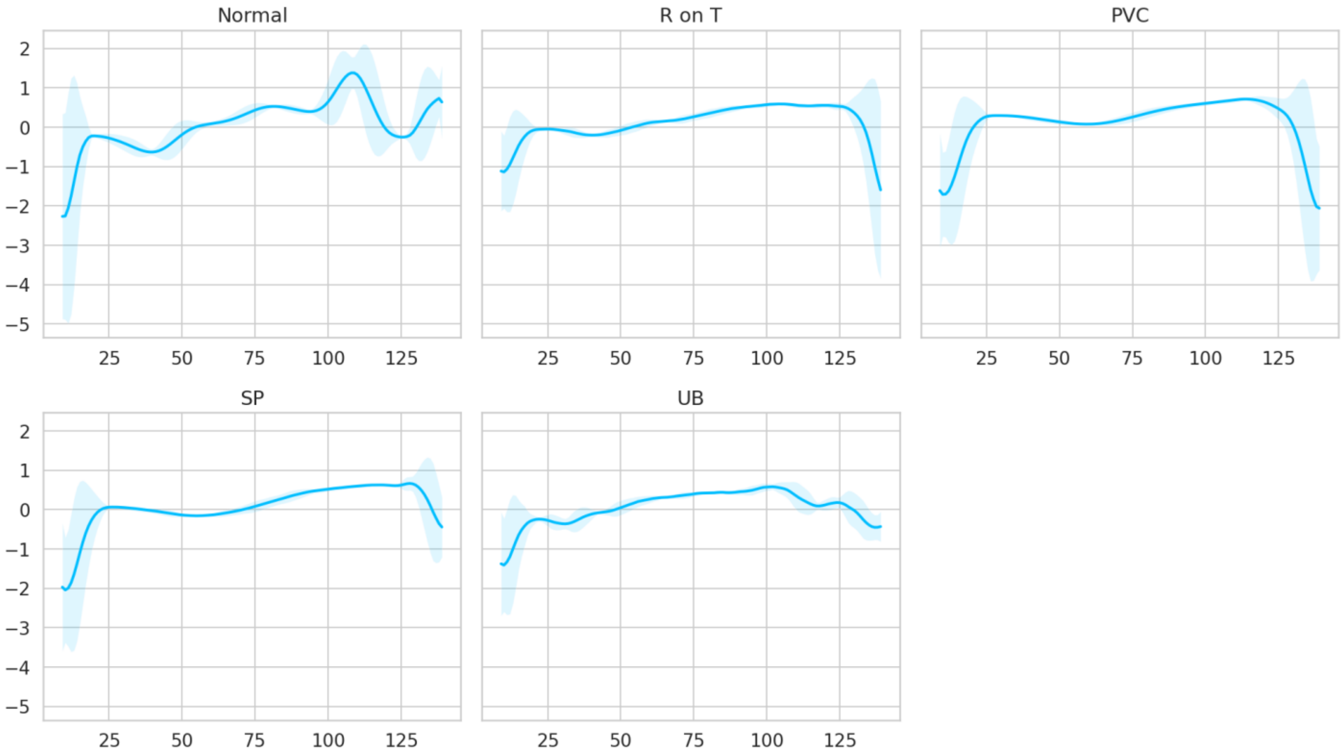


Figure 12. Time Series nature of the signals of different types of Heart Beat

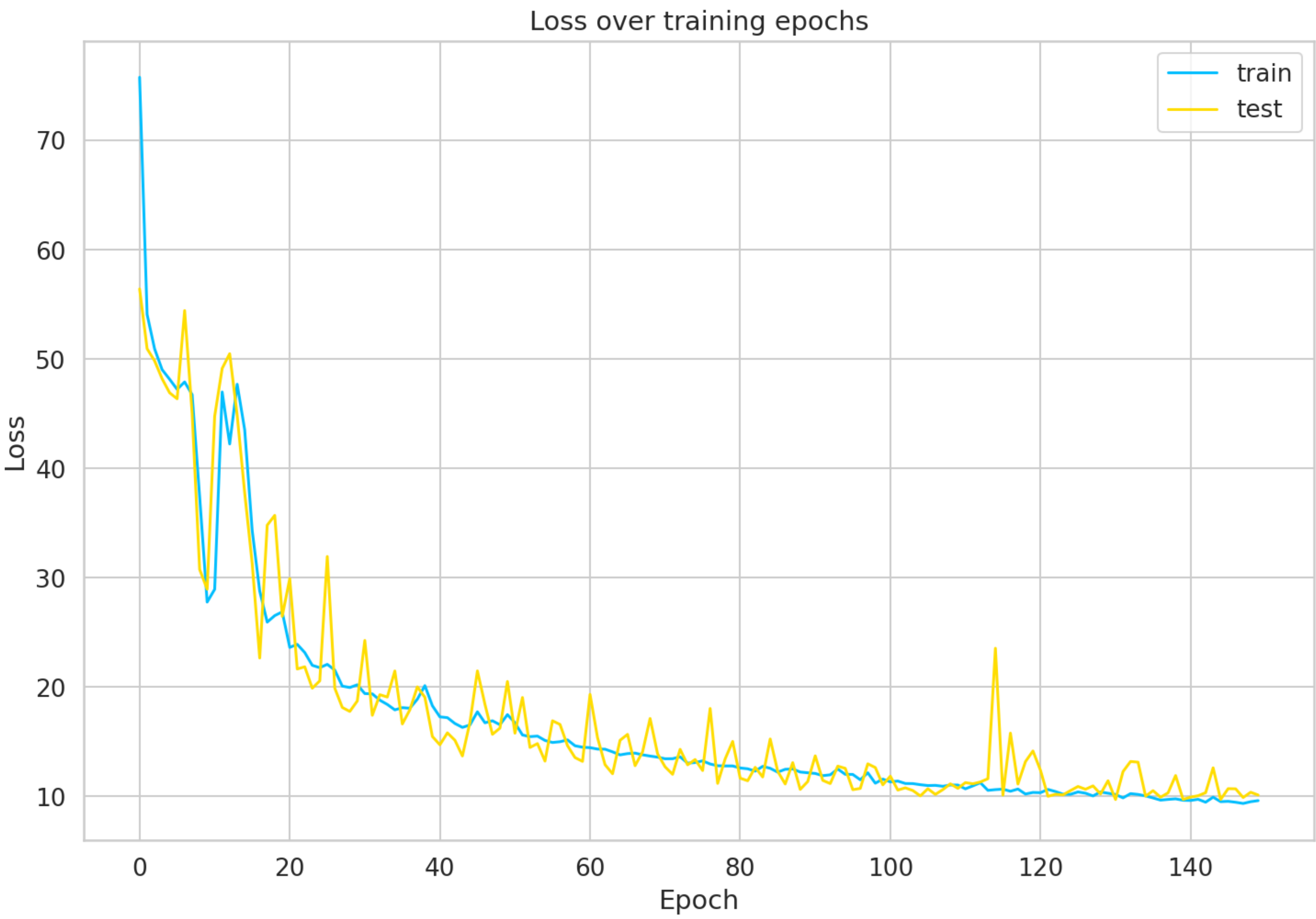


Figure 13. Training Loss over the time steps

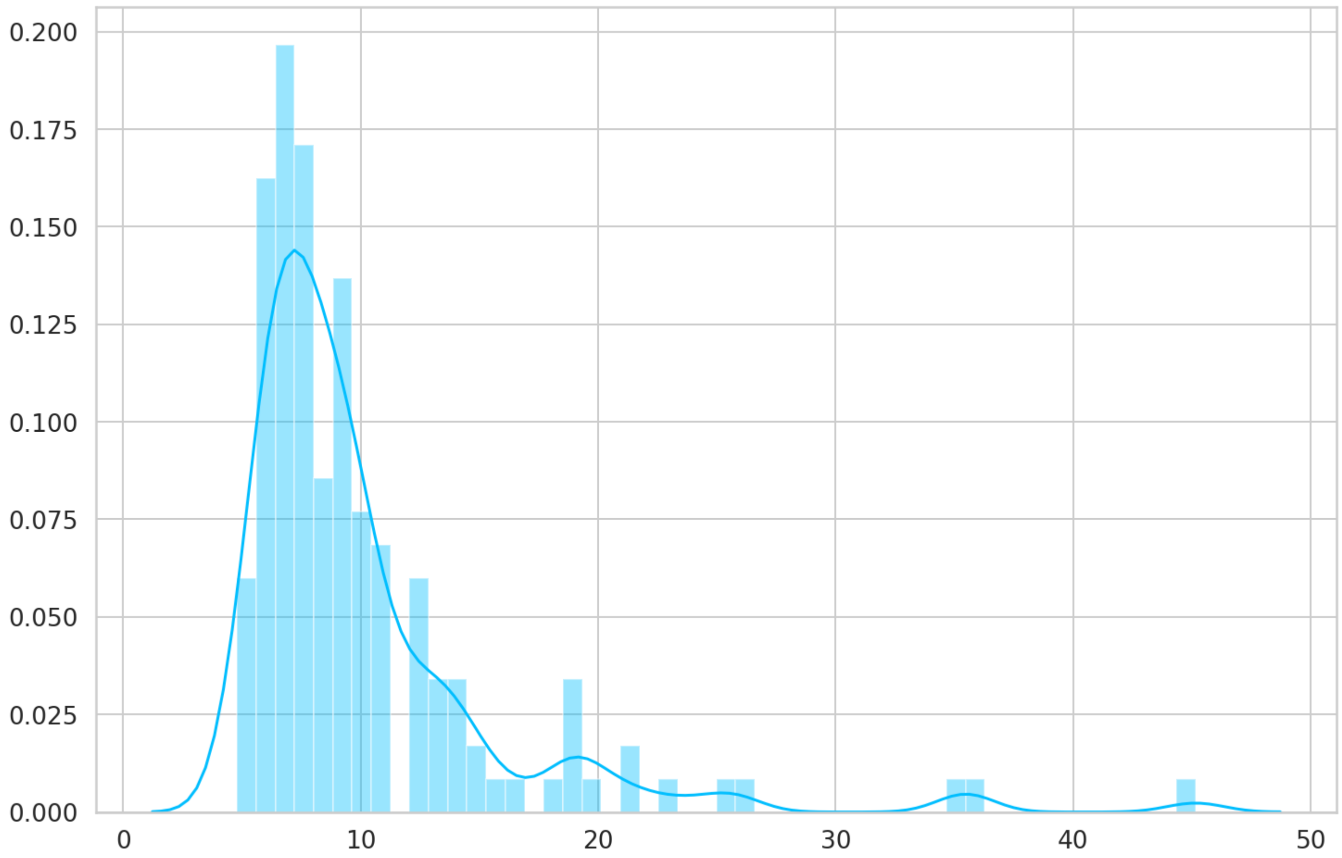


Figure 14. Reconstruction Error (loss) for normal training data

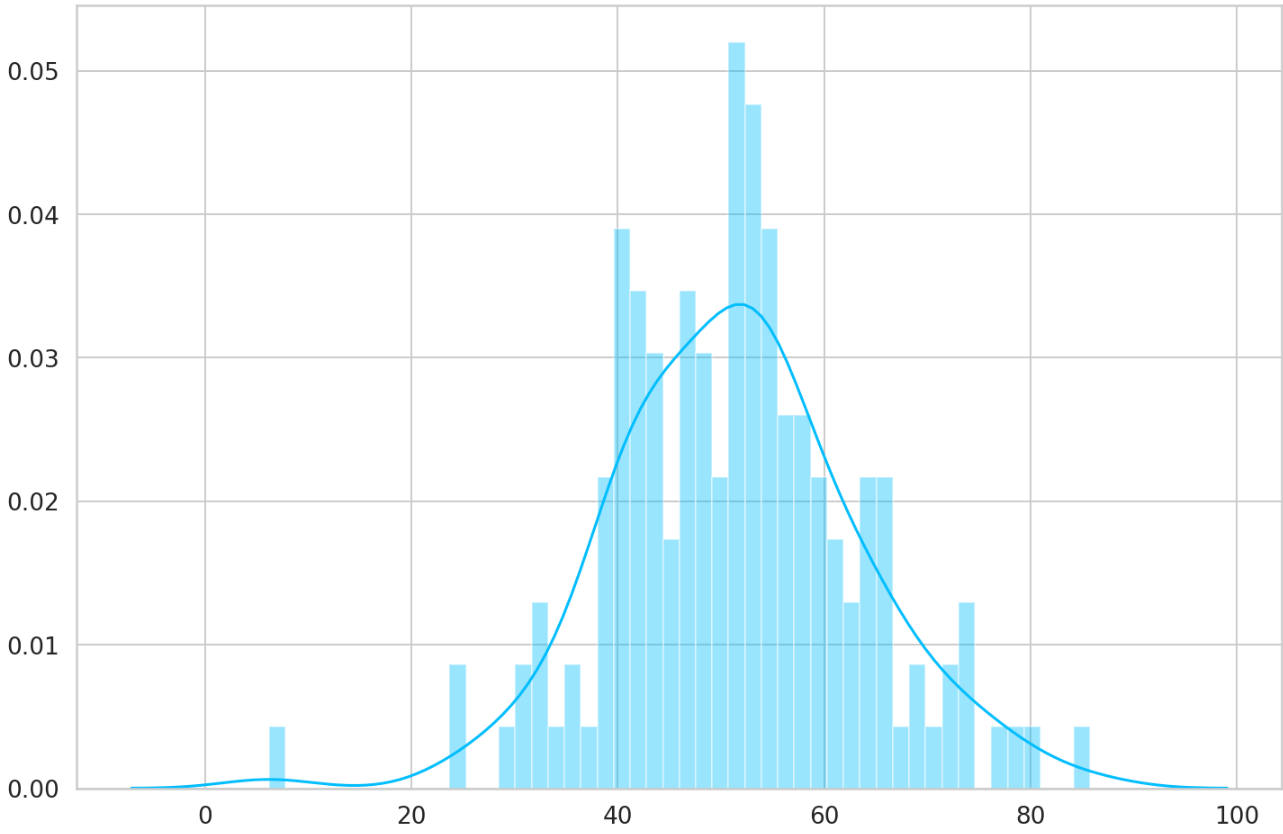
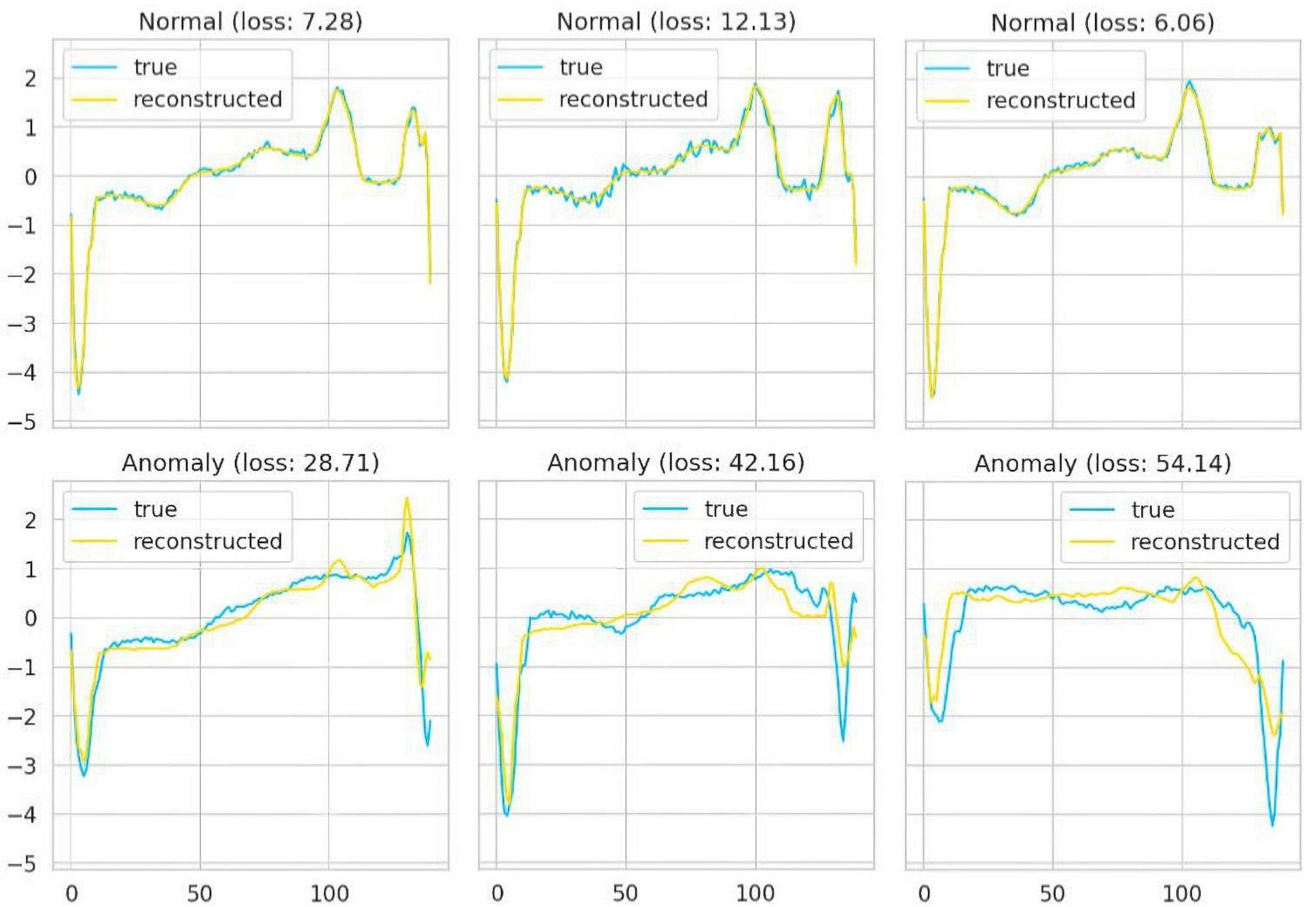


Figure 15. Reconstruction Error (loss) for Anomaly (test) data



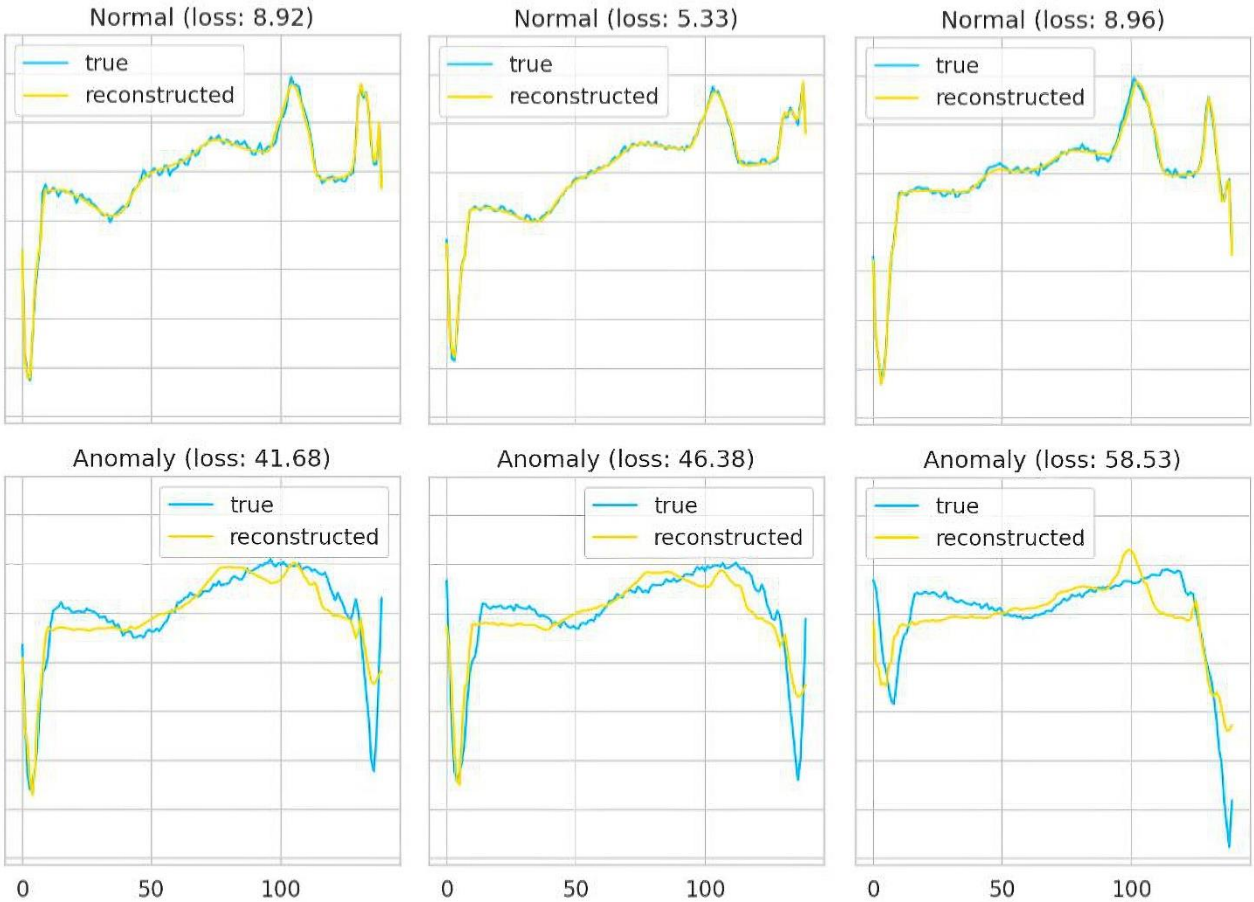


Figure 16. Comparative analysis of Reconstruction Loss between Normal & Anomalies Heart Beat

1. Tables in Manuscript

TABLE. 1: PERFORMANCE EVALUATION ON VARIOUS OPTIMIZATION TECHNIQUES

Optimization Technique or Algorithm	Accuracy Score
Adaptive Momentum	97.93 %
Adaptive Gradient	95.23 %
Adaptive Delta	91.45 %
RMSProp	91.34 %
Stochastic Gradient Descent with Momentum	85.67 %