The Application of Deep Neural Networks for the Detection of Myocardial Infarction by Electrocardiogram

Jarred Fisher
University of Cape Town
Cape Town, South Africa
fshjar002@myuct.ac.za

ABSTRACT
A literature review was performed to establish the effectiveness of deep neural networks (DNNs) for the detection of Myocardial Infarction (MI) using electrocardiogram (ECG) data. In this review, three DNNs will be explained with respect to their architecture, performance, and applicability for MI detection. This included the simple Recurrent Neural Network (RNN) and two of its extensions, i.e. Long Short-Term Memory (LSTM), and Bidirectional LSTM (BiLSTM). Related works were analysed to establish common limitations and challenges of these DNN techniques, in addition to the datasets, data splitting and pre-processing methods utilised. This review found the simple RNN design to be an inadequate classifier for MI. However, the LSTM provided powerful detection capabilities for MI, followed closely by the BiLSTM design. The success of these architectures supports further research into their applicability for MI detection.

KEYWORDS
Deep learning, neural networks, electrocardiogram (ECG), myocardial infarction detection

1 INTRODUCTION

1.1 Background

According to the WHO's 2019 global health figures, Coronary Artery Disease (CAD) has been the primary cause of death worldwide, responsible for 16% of the world's total deaths. In the last twenty years, the increase of deaths due to CAD, has risen from 2 million to 8.9 million deaths [2]. CAD is the precursor to MI which can be described as when the blood supply to the muscular tissue of the heart is interrupted, causing irreversible damage to the cells responsible for generating contractile force in the heart [3]. MI, also called a heart attack, can frequently occur in patients with a history of CAD [23].

By detecting MI early, irreversible damage to heart tissue can be reduced [5]. First responders such as paramedics may lack the proficiency of cardiologists when it comes to interpreting ECGs or the hospital might be short-staffed of such experts [1]. This establishes precedence to develop an automated expert system (AES) that can effectively screen patients for potential MI, at a low cost and with relative ease. The strength of DNNs to learn from previous patterns without needing a medical expert to develop the decision logic, positions them as effective candidates for such an AES [11].

As a subset of machine learning, deep learning makes use of a neural network with three or more layers. The numerous layers give way to the term deep neural network. By learning from large amounts of data, these neural networks attempt to simulate the behaviour of the human brain to some extent [19]. Rather than categorizing features as two separate modules, the DNN model can integrate feature extraction and classification into one [7,8,14]. This is referred to as the "end-to-end" model (see Figure 1). Such a model can solve tasks which would otherwise require multiple components or steps [76]. The concern with numerous modules when applied to a complex task is the snowballing of errors, as any discrepancies between modules can carry over and thus affect the classification task [28].

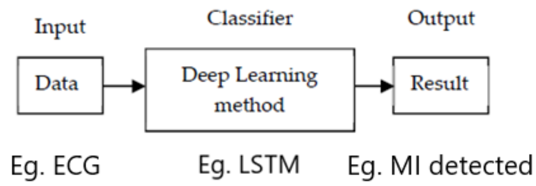

Figure 1: An example DNN end-to-end model [7]

The sequential nature of ECG signals lends itself to exploring architectures which excel at learning long term dependencies [20]. Hence, this review will focus on RNNs and two of its extensions, the LSTM and BiLSTM. The application of recurrent architectures to MI detection is a relatively new research focus, with the earliest papers cropping up only in 2019 [34,35,45,46]. This review presents an overview of these three architectures, and their comparative performances in MI detection. With the limited research that was available, LSTMs were shown to be the preferred candidate in MI detection, as will be confirmed in this review.

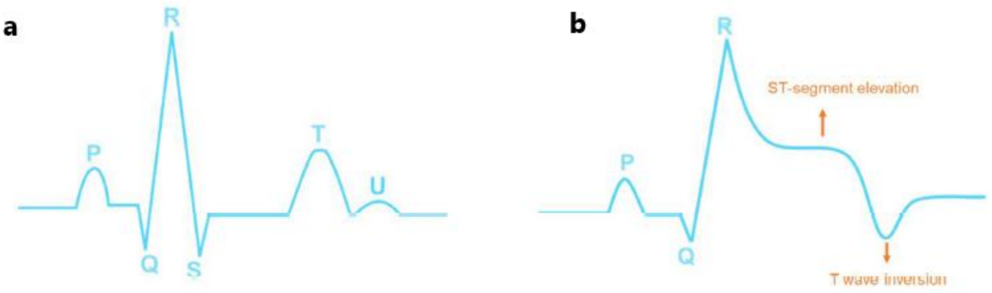

Figure 2: The ECG characteristics of normal sinus rhythm (a) and myocardial infarction (b). [4]

1.2 ECG Analysis

Several methods in detecting MI, such as coronary angiography and echocardiogram, are invasive techniques and labour intensive [10]. Furthermore, these strategies require extensive resources and may be limited to a short access window, making them impractical to screen patients for timely medical intervention [1].

ECG analysis is a cost-effective alternative. It is a non-invasive technique that records electrical signals of the heart in real time [29]. It also benefits from a simple setup process and continuous monitoring capability [18].

The rhythm of a healthy heart (referred to as normal sinus rhythm) [30] in a single cardiac cycle is a P wave followed by the QRS complex, then a T wave [17]. The sample of a normal sinus rhythm is shown in Figure 2a. Variation of the ECG signal based on its characteristic morphology and interval helps experts in disease diagnosis [8]. In the case of MI, the characteristics of the ECG signal generally include ST-segment elevation, abnormal Q wave appearance, and T-wave inversion [20]. MI classification can be seen in Figure 2b.

At first glance, ECGs present as two-dimensional image data, but they can easily be represented as one dimensional data in the form of a time series [71]. Heartbeats are inherently sequential and can be linked to one another in time [6]. This format is desirable for recurrent architectures, which expect sequential data [21].

2 SUITABLE DEEP NEURAL NETWORKS

Convolutional neural networks (CNN) are the most popular architecture for MI detection [4]. A recent study demonstrated a hybrid CNN-LSTM model to be highly effective at detecting ST-segment elevation MI and determining the occluded coronary artery from ECG data [5]. In several cases, the study's model even outperformed expert cardiologists. However, in isolation, vanilla CNNs can only deduce spatial relations and are thus restricted by their kernel size [47]. In the case of ECG analysis, this means only characteristic features of the signal's shape can be learnt. The concern is that these extracted spatial features can misdiagnose MI, as ST-segment deviation may be observed in other conditions such as acute pericarditis, left ventricular hypertrophy, left bundle-branch block, Brugada syndrome, and early repolarizations [50].

2.1 Recurrent Neural Networks

RNNs benefit from recognising input signals as time-varying and can therefore exploit the signal's temporal as well as its spatial relationships [44]. Thus, an RNN can make inferences on the beat-to-beat variations in addition to the characteristic shape of the ECG data.

Unlike traditional machine learning methods, RNN's recognise that there is often a dependent link between the input and output of a model [21]. Standard feed-forward neural networks send data through the hidden layers in only one direction, thus excluding past results from future outputs [27]. This poses an issue in problems such as natural language processing, wherein a sentence's meaning is not merely the sum of its independent words but rather the placement of the words in context [31]. RNNs are suited as they can develop a memory by forming a directed cycle wherein the previous layer's output is combined with new input to form the current layer's output [24]. The RNN design can be seen in Figure 3.

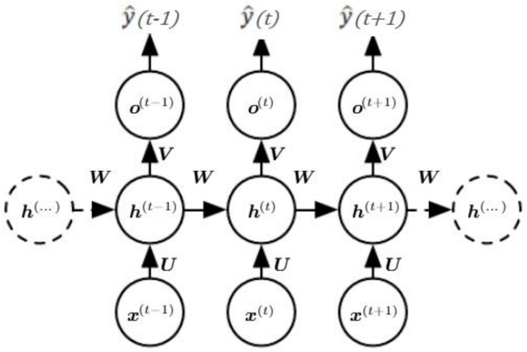

Figure 3: RNN Design [33]

The hidden state (*h*) at time step *t* for an RNN is calculated as:

$$h_t = fn(Ux_t + Wh_{t-1} + b)$$

fn is an application function (such as *tanh* or *ReLu*), x_t is the input sequence and b is the bias vector. Weight matrices U and W are applied to x_t and the previous hidden state h_{t-1}, respectively [24].

This simple RNN design as described above does suffer from a significant issue: A study demonstrated that such RNNs can manage input dependencies over long-intervals, in principle [22]. However, the study showed that training these networks with gradient descent for large datasets often leads to the vanishing or exploding gradient problem [22].

2.2 Long Short-Term memory

To address this problem the Long-Short-Term (LSTM), was introduced in 1997, as an extension of the simple RNN [27]. The LSTM model addresses the shortcomings of the simple RNN through implementing various activation function layers called gates. The gate structure determines to what extent prior information will contribute to the current output [32]. The LSTM architecture comprises of multiple memory blocks, which contain the gates as well as an internal cell memory. The memory cells represent the information that the previous memory block chose to keep [43]. The design can be seen in Figure 4.

The LSTM gate structure is composed of three gates namely the:

1) forget
2) input
3) and output gate

The output of the forget gate at time step *t* is given by the following application of the sigmoid activation function:

$$f_t = \sigma(U_f x_t + W_f h_{t-1} + b_f)$$

Note that definitions of symbols used carry over from the previous section. The forget gate's output is applied to the previous memory cell C_{t-1} by taking their element-wise product (denoted as $\odot$). The forget gate thus determines to what degree previous data must be discarded [33].

The input gate determines the amount of information that will be written to the current memory cell C_t [37] and is calculated as:

$$i_t = \sigma(U_i x_t + W_i h_{t-1} + b_i)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot tanh(U_i x_t + W_i h_{t-1} + b_i)$$

Lastly, the output gate calculates the current hidden state to be passed onto subsequent LSTM blocks [49]. This is calculated as follows:

$$o_t = \sigma(U_o x_t + W_o h_{t-1} + b_o)$$

$$h_t = o_t \odot \tanh(C_t)$$

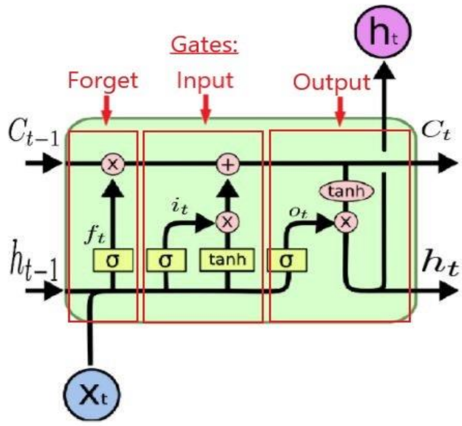

Figure 4: LSTM memory block design [43]

2.3 Bidirectional Long-Short Term Memory

The Bidirectional RNN (BRNN) was developed to incorporate not only past, but also future state information into the training algorithm [25]. BRNNs achieve this by training both a forward and a backward RNN which share a common output layer [25]. This allows for the dependencies between sequential input to be learnt in both directions [38]. This leads to improved predictions as the context of the entire input sequence is considered [13]. Notably, the neurons of the forward and backward RNNs are independent of each other [42]. Thus, the training of the BRNN follows the same paradigm as in the regular unidirectional case [25].

The Bidirectional LSTM (BiLSTM) follows the same structure as the BRNN, however the forward and backward RNNs are simply replaced with LSTM layers [12]. Two independent models are effectively trained, thus a system to combine both of their outputs is needed. The merging of their outputs can be performed via several methods such as summation, average, multiplication, or concatenation. [37] A visualisation of the BiLSTM model can be seen below in Figure 5.

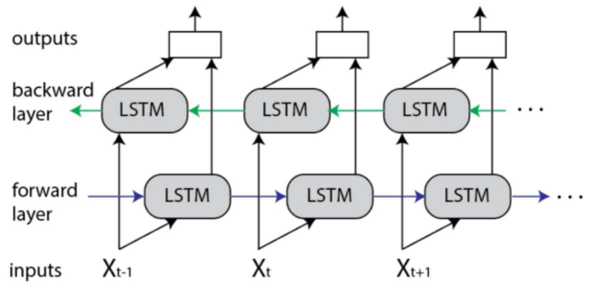

Figure 5: BiLSTM Model design [42]

BiLSTM models can outperform standard unidirectional LSTMs in time series prediction [38] and are significantly more effective in natural language processing [36]. A disadvantage of the BiLSTM is that they require longer training times, owing to the dual model design [38].

3 APPLICATIONS TO MI DETECTION

3.1 Evaluation Metrics

A study analysed data from 564 412 patients treated for MI in the UK over a nine-year period discovered that almost a third were initially misdiagnosed [39]. The misdiagnosed patients faced a seventy percent increased risk of death after thirty days, as opposed to those who were correctly diagnosed at first [39]. On the other end of the scale, overdiagnosis of MI wastes medical resources and leads to unnecessarily invasive treatments [1].

The above warrants a careful consideration of how a MI classifier should be evaluated. The most important metric for a diagnostic system tasked with detecting a dangerous condition like MI is its sensitivity [9]. This measures the system's ability to correctly identify patients who have the disease [10]. Following on closely, is the classifier's specificity as this indicates how effective the classifier is at excluding healthy patients [40]. There are many other metrics frequently used such as precision, accuracy, F1 score and area under the receiver operator curve (AUC-ROC) [6]. As is common practice, we consider positive cases to represent the disease (i.e. MI), whilst the negative is the healthy control [8].

Precision measures the ratio of samples correctly labelled as positive to all samples labelled positive [14]. Accuracy is the proportion of correctly classified samples to the entire pool of samples [8]. It is the most used measurement in MI diagnosis [4]; however, when the data exhibits significant class imbalance, accuracy may be misleading [34,35]. This is because the model can still score high by simply classifying everything as the majority class. In cases of class imbalance, the F1 score is preferred [40]. This metric gives the harmonic mean of the precision and sensitivity [11].

Four central values, True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN) are needed in calculating the metrics mentioned. The following equations summarise the various metrics:

$$\textbf{\textit{Accuracy}} = TP+TN/(TP + FP + TN + FN)$$

$$\textbf{\textit{Precision}} = TP/\ (TP+FP)$$

$$\textbf{\textit{Sensitivity}} = TP/(TP+FN)$$

$$\textbf{\textit{Specificity}} = TN/(TN+FP)$$

$$\textbf{\textit{F1 Score}} = 2*(Precision*Sensitivity)\ /\ (Precision +Sensitivity)$$

Utilizing robust evaluation strategies to make results comparable and generalizable is crucial. However, there exists a lack of consistency in the choice of such metrics used for MI detection [9].

3.2 Architectures

Recurrent architectures have achieved great success in classifying and modelling time series data [49]. The capacity of such models to process and remember information through time, would imply that heartbeat sequences can be processed with their sequential dependencies remaining intact [51]. This inherent strength of recurrent structures hypothesises its success in MI detection. *Note* that the studies explored in this section all focused on MI detection and made use of the PTB diagnostic dataset [58].

3.2.1 Recurrent Neural Networks

A study assessed the classification power of the simple RNN versus LSTMs [34]. The RNN model performed considerably worse than the LSTM, achieving a precision score of 68% compared to the LSTM which gave 91%. A limitation of the study was that minimal pre-processing on the ECG signal was performed prior to it being passed to the models.

An additional study compared the relative performances of the simple RNN to that of an LSTM and a Gated Recurrent Unit (GRU) [35]. A GRU merges several of the gates and states present in the vanilla LSTM. The result is a model that is simpler, whilst still retaining the memory capability of LSTMs [52]. The study used training-testing data splits ranging from 50:50% to 90:10% to demonstrate the effect on model performance within various splits. Optimal results came from a 90:10% split, with the vanilla LSTM model achieving the best sensitivity score of 98.49% [35]. The GRU was however a close second to the LSTM in its specificity, precision and F1-score. The RNN performance was inferior, averaging 10.2% below the LSTM's scores [35].

The results of the previous studies [34,35] support the notion that a simple RNN is inadequate at detecting

MI, as compared to its extension, the LSTM. Moreover, the RNN has sub-par performance in MI detection when compared to CNNs and more traditional ML techniques [53].

3.2.2 Long Short-Term Memory

In a study comparing CNNs, LSTMs and a hybrid CNN-LSTM model, the CNN and LSTM models demonstrated accuracy scores of 84.95% and 85.23% respectively [48]. This study however had several issues, including the absence of pre-processing and data splitting methods used, as well as only presenting accuracy scores for the models.

Near perfect classification ability was found in a study using an LSTM [46]. The classifier achieved sensitivity, specificity, precision and accuracy scores of 99.91 99.95 99.91 and 99.91%, respectively [46]. Furthermore, the study concluded that it exhibited superior performance compared to several related works which used other machine learning techniques such as CNN and support vector machines [46].

Notably, another study using an LSTM trained its model to detect MI on a single heartbeat from a single lead ECG (Lead II), in real time (< 40ms) [47]. The model achieved sensitivity, precision and F1-scores all above 90%. This study presents a novel application as detecting MI from Lead II alone is challenging but more importantly, Lead II is widely available in wearable devices and fitness machines, making the proposed model widely deployable at a low cost [47]. Other studies used eight [34,46], twelve [44,45] or fifteen leads [35,48].

The first study from Section 3.2.1 also experimented with one, two and three hidden layer LSTMs, shedding light on the impact of changing this number. The study demonstrated no notable improvement when the LSTM's hidden layer count was increased [34].

3.2.3 Bidirectional Long Short-Term Memory

This review found limited research to have been done in the application of BiLSTMs to MI detection. In the one study found, the BiLSTM achieved sensitivity, specificity and accuracy values of 93.86, 85 and 92.54% respectively [45]. The study experimented with the use of a heartbeat attention mechanism, comparing the BiLSTM model with and without the mechanism. Improvements were seen by including the mechanism, raising the metrics on average by 3.2% [45]. The attention mechanism allows the model to weight ECG segments which provide valuable diagnostic markers in MI detection [44].

As prior research would indicate [15,36,38], the ability of BiLSTMs to train on both the forward and backward sequence of the data suggests it would outperform simpler models like the LSTM and RNN. The limited research provides an opportunity for future studies to ascertain its comparative performance versus other deep learning models in MI detection.

4 ECG DATASETS

Without high quality datasets, deep learning techniques cannot be trained optimally, and their classifying ability cannot be generalised for use in clinical settings [7]. In the case of supervised learning, samples need to be expertly annotated to distinguish healthy control samples from those who are not [67]. This, in addition to the dataset needing to be publicly available, restricts the inclusion of viable datasets for MI detection to the following:

1. **Physikalisch-Technische Bundesanstalt (PTB) Diagnostic ECG Database**
2. **PTB-XL**
3. **Long Term ST Database (LTST)**
4. **ECG-VIEW II**

Researchers frequently analyse ECGs from the PTB diagnostic database since it contains cardiologist annotations of ECGs for various heart conditions, including MI [58]. The smallest dataset is the LTST [60], consisting of only 86 recordings from 80 patients whilst the largest database is the ECG-ViEW II database comprising 979 273 recordings from 461 178 patients obtained over a 19-year research period [61]. The PTB-XL database was made publicly available in 2020 and is a large dataset (21 837 recordings from 18 885 patients) and is extensively annotated, boasting the highest number of records explicitly labelled MI [59]. Further details of the four datasets can be found in Table 1.

Other public ECG datasets used in MI detection include the ESCDB [62] and STAFFIII [63]. They are however excluded due to a combination of small sample sizes (< 100 patients), minimal ECG leads (two or three), containing non-MI ST-segment deviations or no healthy control identified in the data. There are also no prior studies that used these datasets for MI detection with an end-to-end recurrent architecture.

4.1 Pre-processing

A strength of DNNs is that they can learn from raw data which has seen minimal pre-processing. This explains the relative success in two studies this review identified which did not employ any noise removal [34,35]. Notwithstanding, ECGs are weak electric signals generated from the body's surface, which are relatively unstable and subject to distortion [46]. To name a few sources, ECGs can be distorted by high-frequency noise [54], electromagnetic interference from power lines [55] as well as baseline wander attributed to patient breathing and movement [16]. Studies commonly employ waveform transform to remove these artifacts [12, 14, 23, 45, 56], however other filters were found to be used such as moving average [47], median plus notch [46] and bandpass [44].

Regardless of denoising steps taken, the ECG signals must then be resampled at a consistent rate [23]. This involves transforming the ECG beats into uniform segments, which can then be passed to the classifier [14].

Heartbeats are generally segmented based on the R-wave, owing to its high amplitude [56]. The Pan Tompkins algorithm is commonly employed to achieve this [57].

Sliding windows with the same number of heartbeats can then be generated from the relative positions of the R-peaks [45,46]. A more popular technique is to determine a sliding window based on a particular length of time [34,35,44,47]. After the data is processed, it must then be separated into groups for training the model and evaluating its performance [47].

4.2 Data Splitting

To determine the performance of DL methods, data splitting plays a crucial role [68]. Over fitting is a prime concern if all the data is used to train the model. This would result in poor performance when the model is presented with new data [67]. Depending on the task at hand and the structure of the data, various partitioning methods are employed to remedy this [69].

A traditional method such as standard cross validation assumes an absence of a dependent relationship between observations [64]. At first glance, this is of great concern to ECG analysis, as the beat-to-beat variation would be lost if the signals are sliced up arbitrarily. This would result in the model's weights and hyperparameters being tuned incorrectly, leading to poor classification power [72]. However, the segmenting and windowing as described in the previous section takes care of this, as the ECG samples are then independently and identically distributed. The use of a sliding window closely aligns with that of the walk-forward validation technique which is more effective at evaluating models operating on time series data [65].

Application of a simple training-validation-testing data splitting is popular [12,34,35,45,46], however some studies choose to exclude the use of a validation set [34,35]. Furthermore, the proportion of these splits vary amongst studies [4]. Each split has a specific function: Training data is the initial input and is the how the model develops its classification or predictive ability [26]. The validation split is for preliminary evaluation and is thus separated from the training data. The validation set determines the model's performance on unseen data, whilst still optimising the model's hyperparameters [66]. The test set emulates a real world setting and provides an unbiased evaluation of the final model's generalisability [70].

Another common option is to use, K-fold cross-validation [44,47]. In this technique, the samples are randomly divided into k equally sized subsets [41]. Of the k groups, one is held back to be used as the validation set, whilst the rest are used for training the model [30]. This process is then repeated k times with the mean of the k results used as a measure of the final model's skill [72]. Importantly, each unique group is used exactly once to validate the model [75]. K-fold cross-validation is a popular technique as it generally results in a more realistic and unbiased estimate of the model's performance as opposed to the simple training-testing split method [74].

5 DISCUSSION

Results from notable studies which incorporated an end-to-end recurrent architecture for MI detection have been summarised in Table 2. The simple RNN design performed the worst in both inter and intra study comparisons, achieving the lowest sensitivity, precision and F1-scores of 83%, 68%, 75% respectively [34]. LSTM based models were the top performer across the studies, with one study achieving sensitivity, specificity, precision and accuracy values of 99.91% 99.95% 99.91% 99.91% respectively [46]. This review found the BiLSTM model for MI detection to only have been studied in one paper [45] and it performed considerably better than the simple RNN. However, the BiLSTM only outperformed three of the six studies which incorporated an LSTM. Lastly, the average ratio of MI to HC samples was 4.81:1. This represents a moderate class imbalance across the studies.

6 CONCLUSIONS

This review established several conclusions relating to the literature on MI detection. All the studies developed models which achieved high performance for MI classification. A primary concern however is that every study trained their models on the public PTB Diagnostic ECG database. The models are thus over-fitted to this dataset, bringing into question their robustness and generalisability to real world settings. This leaves room for further research to be done employing ECG datasets such as the LTST, PTB-XL or ECG_VIEW II.

The AUC-ROC is a valuable metric in the medical field and is robust to class imbalance [77]. Moderate class imbalance was identified, with samples skewed towards those with MI. However, of the studies identified, none made use of the AUC-ROC. Relying on single threshold metrics such as accuracy do not shed light on the classifier's usefulness in a clinical setting [78]. This, in addition to a more consistent choice of evaluation metrics is encouraged for future research.

Lastly, this review established that the simple RNN design had the worst performance compared to the LSTM and BiLSTM. The LSTM was the top contender achieving near-perfect classification results in several studies. Notably, limited research exists for BiLSTMs in MI detection and no single study has compared its relative performance against other architectures, warranting further research. The success of LSTMs over BiLSTM identified in this review may be due to the more widespread use of LSTMs in the literature.

REFERENCES

[1] Xiao, Xu, Y., Pelter, M. M., Mortara, D. W., & Hu, X. (2018). A Deep Learning Approach to Examine Ischemic ST Changes in Ambulatory ECG Recordings. AMIA Summits on Translational Science Proceedings, 2017, 256–262.

[2] Global Health Estimates: Life expectancy and leading causes of death and disability: 2019. https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates. Accessed: 2022- 04- 10.

[3] Manfroi, W., Peukert, C., Berti, C., Noer, C., Gutierres, D. and Silva, F. 2002. Acute Myocardial Infarction: The First Manifestation of Ischemic Heart Disease and Relation to Risk Factors. Arquivos Brasileiros de Cardiologia. 78, 4 (2002), 392-395.
[4] Xiong, P., Lee, S., & Chan, G. (2022). Deep Learning for Detecting and Locating Myocardial Infarction by Electrocardiogram: A Literature Review. Frontiers In Cardiovascular Medicine, 9. doi: 10.3389/fcvm.2022.860032
[5] Wu, L. et al. 2022. Deep Learning Networks Accurately Detect ST-Segment Elevation Myocardial Infarction and Culprit Vessel. Frontiers in Cardiovascular Medicine. 9, (2022).
[6] Somani, S., Russak, A., Richter, F., Zhao, S., Vaid, A., & Chaudhry, F. et al. (2021). Deep learning and the electrocardiogram: review of the current state-of-the-art. EP Europace, 23(8), 1179-1191. doi: 10.1093/europace/euaa377
[7] Rim, B., Sung, N., Min, S. and Hong, M. 2020. Deep Learning in Physiological Signal Data: A Survey. Sensors. 20, 4 (2020), 969.
[8] K, Sanjana., V, Sowmya., E.A., Gopalakrishnan. and K.P., Soman. 2020. Explainable artificial intelligence for heart rate variability in ECG signal. Healthcare Technology Letters. 7, 6 (2020), 146-154.
[9] Bond, R., Finlay, D., Al-Zaiti, S. and Macfarlane, P. 2021. Machine learning with electrocardiograms: A call for guidelines and best practices for ‘stress testing’ algorithms. Journal of Electrocardiology. 69, (2021), 1-6.
[10] Huang, P., Tseng, Y., Tsai, C., Chen, J., Yang, S., Chiu, F., Chen, Z., Hwang, J., Chuang, E., Wang, Y. and Tsai, C. 2022. An Artificial Intelligence-Enabled ECG Algorithm for the Prediction and Localization of Angiography-Proven Coronary Artery Disease. Biomedicines. 10, 2 (2022), 394.
[11] Hannun, A., Rajpurkar, P., Haghpanahi, M., Tison, G., Bourn, C., Turakhia, M. and Ng, A. 2019. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nature Medicine. 25, 1 (2019), 65-69.
[12] Yildirim, Ö. 2018. A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification. Computers in Biology and Medicine. 96, (2018), 189-202.
[13] Faust, O., Shenfield, A., Kareem, M., San, T., Fujita, H. and Acharya, U. 2018. Automated detection of atrial fibrillation using long short-term memory network with RR interval signals. Computers in Biology and Medicine. 102, (2018), 327-335.
[14] Baloglu, U., Talo, M., Yildirim, O., Tan, R. and Acharya, U. 2019. Classification of myocardial infarction with multi-lead ECG signals and deep CNN. Pattern Recognition Letters. 122, (2019), 23-30.
[15] Baldi, P., Brunak, S., Frasconi, P., Soda, G. and Pollastri, G. 1999. Exploiting the past and the future in protein secondary structure prediction. Bioinformatics. 15, 11 (1999), 937-946.
[16] Blanco-Velasco, M., Weng, B. and Barner, K. 2008. ECG signal denoising and baseline wander correction based on the empirical mode decomposition. Computers in Biology and Medicine. 38, 1 (2008), 1-13.
[17]]Kutlu, Y., Allahverdi, N. and Atlan, G. 2016. Arrhythmia Classification using Waveform ECG Signals. International Conference on Advanced Technology & Sciences (Konya, 2016), 3-8.
[18] Siontis, K., Noseworthy, P., Attia, Z. and Friedman, P. 2021. Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. Nature Reviews Cardiology. 18, 7 (2021), 465-478.
[19] IBM Cloud Education 2020. What is Deep Learning? https://www.ibm.com/cloud/learn/deep-learning. Accessed: 2022- 04- 15.
[20] Jayachandran, E., Joseph K., P. and Acharya U., R. 2009. Analysis of Myocardial Infarction Using Discrete Wavelet Transform. Journal of Medical Systems. 34, 6 (2009), 985-992.
[21] Schmidhuber, J. 2015. Deep learning in neural networks: An overview. Neural Networks. 61, (2015), 85-117.
[22] Bengio, Y., Simard, P. and Frasconi, P. 1994. Learning long-term dependencies with gradient descent is difficult. IEEE Transactions on Neural Networks. 5, 2 (1994), 157-166.
[23] Tan, J., Hagiwara, Y., Pang, W., Lim, I., Oh, S., Adam, M., Tan, R., Chen, M. and Acharya, U. 2018. Application of stacked convolutional and long short-term memory network for accurate identification of CAD ECG signals. Computers in Biology and Medicine. 94, (2018), 19-26.
[24] Pascanu R., Gulcehre C., Cho K., Bengio Y. 2013. How to Construct Deep Recurrent Neural Networks. arXiv.1312.6026. Retrieved from https://arxiv.org/abs/1312.6026
[25] Schuster, M. and Paliwal, K. 1997. Bidirectional recurrent neural networks. IEEE Transactions on Signal Processing. 45, 11 (1997), 2673-2681.
[26] Joseph, V. and Vakayil, A. 2021. SPlit: An Optimal Method for Data Splitting. Technometrics. 64, 2 (2021), 166-176.
[27] Hochreiter, S. and Schmidhuber, J. 1997. Long Short-Term Memory. Neural Computation. 9, 8 (1997), 1735-1780.
[28] Oh, S., Ng, E., Tan, R. and Acharya, U. 2018. Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats. Computers in Biology and Medicine. 102, (2018), 278-287.
[29] Electrocardiogram (ECG or EKG) - Mayo Clinic: 2022. https://www.mayoclinic.org/tests-procedures/ekg/about/pac-20384983. Accessed: 2022- 04- 23.
[30] Jung, Y. 2018. Multiple predicting K-fold cross-validation for model selection. Journal of Nonparametric Statistics. 30, 1 (2018), 197-215.
[31] IBM Cloud Education. 2020. What are Recurrent Neural Networks? https://www.ibm.com/cloud/learn/recurrent-neural-networks. Accessed: 2022- 04-19.
[32] Vu, K. 2020. A Friendly Introduction to Graph Neural Networks - KDnuggets: https://www.kdnuggets.com/2020/11/friendly-introduction-graph-neural-networks.html. Accessed: 2022- 04- 23.
[33] Goodfellow, I., Bengio, Y. and Courville, A. 2016. Deep learning. MIT Press.
[34] Darmawahyuni, A., Nurmaini, S. and Sukemi 2019. Deep Learning with Long Short-Term Memory for Enhancement Myocardial Infarction Classification. 2019 6th International Conference on Instrumentation, Control, and Automation (ICA). (2019), 19-23.
[35] Darmawahyuni, A., Nurmaini, S., Sukemi, Caesarendra, W., Bhayyu, V., Rachmatullah, M. and Firdaus 2019. Deep Learning with a Recurrent Network Structure in the Sequence Modeling of Imbalanced Data for ECG-Rhythm Classifier. Algorithms. 12, 6 (2019), 118.
[36] Graves, A. and Schmidhuber, J. 2005. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. Neural Networks. 18, 5-6 (2005), 602-610.
[37] Alhamid, M. 2021. LSTM and Bidirectional LSTM for Regression: https://towardsdatascience.com/lstm-and-bidirectional-lstm-for-regression-4fddf910c655. Accessed: 2022- 04- 26.
[38] Siami-Namini, S., Tavakoli, N. and Siami Namin, A. 2019. The Performance of LSTM and BiLSTM in Forecasting Time Series. 2019 IEEE International Conference on Big Data (Texas, 2019), 3285-3292.
[39] Wu, J., Gale, C., Hall, M., Dondo, T., Metcalfe, E., Oliver, G., Batin, P., Hemingway, H., Timmis, A. and West, R. 2016. Editor’s Choice - Impact of initial hospital diagnosis on mortality for acute myocardial infarction: A national cohort study. European Heart Journal: Acute Cardiovascular Care. 7, 2 (2016), 139-148.
[40] Accuracy, precision, specificity & sensitivity: 2018. https://labtestsonline.org.uk/articles/accuracy-precision-specificity-sensitivity. Accessed: 2022- 04- 27.
[41] Soper, D. 2021. Greed Is Good: Rapid Hyperparameter Optimization and Model Selection Using Greedy k-Fold Cross Validation. Electronics. 10, 16 (2021), 1-5.
[42] Basaldella, M., Antolli, E., Serra, G. and Tasso, C. 2017. Bidirectional LSTM Recurrent Neural Network for Keyphrase Extraction. Communications in Computer and Information Science. (2017), 180-187.
[43] Olah, C. 2015. Understanding LSTM Networks -- colah's blog: https://colah.github.io/posts/2015-08-Understanding-LSTMs/. Accessed: 2022- 04-22.
[44] Jyotishi, D. and Dandapat, S. 2020. An Attention Based Hierarchical LSTM Model for Detection of Myocardial Infarction. 2020 IEEE 17th India Council International Conference (INDICON). (2020), 1-5.
[45] Zhang, Y. and Li, J. 2019. Application of Heartbeat-Attention Mechanism for Detection of Myocardial Infarction Using 12-Lead ECG Records. Applied Sciences. 9, 16 (2019), 3328.
[46] Zhang, X., Li, R., Hu, Q., Zhou, B. and Wang, Z. 2019. A New Automatic Approach to Distinguish Myocardial Infarction Based on LSTM. 2019 8th International Symposium on Next Generation Electronics (ISNE). (2019), 1-3.
[47] Martin, H., Izquierdo, W., Cabrerizo, M., Cabrera, A. and Adjouadi, M. 2021. Near real-time single-beat myocardial infarction detection from single-lead electrocardiogram using Long Short-Term Memory Neural Network. Biomedical Signal Processing and Control. 68, (2021), 102683.
[48] Manimekalai, K. 2020. Deep Learning Methods in Classification of Myocardial Infarction by employing ECG Signals. Indian Journal of Science and Technology. 13, 28 (2020), 2823-2832.
[49] Greff, K., Srivastava, R., Koutnik, J., Steunebrink, B. and Schmidhuber, J. 2017. LSTM: A Search Space Odyssey. IEEE Transactions on Neural Networks and Learning Systems. 28, 10 (2017), 2222-2232.
[50] Wang, K., Asinger, R. and Marriott, H. 2003. ST-Segment Elevation in Conditions Other Than Acute Myocardial Infarction. New England Journal of Medicine. 349, 22 (2003), 2128-2135.
[51] Lui, H. and Chow, K. 2018. Multiclass classification of myocardial infarction with convolutional and recurrent neural networks for portable ECG devices. Informatics in Medicine Unlocked. 13, (2018), 26-33.
[52] Cho, K., Merrienboer, B.V, Bahdanau, D., Bengio, Y. 2014. On the Properties of Neural Machine Translation: Encoder–Decoder Approaches arXiv:1409.1259. Retrieved from https://arxiv.org/abs/1409.1259
[53] Joloudari, J.H et al. 2022 Application of Artificial Intelligence Techniques for Automated Detection of Myocardial Infarction: a Review.
[54] Vidhya, R. and Jerritta, S. 2022. Pre-processing ECG signals for smart home material application. Materials Today: Proceedings. 49, (2022), 2955-2961.
[55] Hoglinger, M. 2016. ECG Preprocessing. PhD dissertation, Johannes Kepler University Linz. Linz, Austria
[56] Li, R., Shen, S., Zhang, X., Chen, G., Li, R., Zhou, B. and Wang, Z. 2019. ECG Beat Classification Based on Deep Bidirectional Long Short-Term Memory Recurrent Neural Network. Proceedings of the 2nd International Conference on Healthcare Science and Engineering. 536, (2019), 45-56.

[57] Pan, J. and Tompkins, W. 1985. A Real-Time QRS Detection Algorithm. IEEE Transactions on Biomedical Engineering. BME-32, 3 (1985), 230-236.
[58] Bousseljot, R., Kreiseler, D. and Schnabel, A. 2009. Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das Internet. Biomedizinische Technik/Biomedical Engineering. (2009), 317-318.
[59] Wagner, P., Strodthoff, N., Bousseljot, R., Kreiseler, D., Lunze, F., Samek, W. and Schaeffter, T. 2020. PTB-XL, a large publicly available electrocardiography dataset. Scientific Data. 7, 1 (2020).
[60] Jager, F., Taddei, A., Moody, G., Emdin, M., Antolič, G., Dorn, R., Smrdel, A., Marchesi, C. and Mark, R. 2003. Long-term ST database: A reference for the development and evaluation of automated ischaemia detectors and for the study of the dynamics of myocardial ischaemia. Medical and Biological Engineering and Computing. 41, 2 (2003), 172-182.
[61] Kim, Y., Shin, D., Park, M., Lee, S., Jeon, M., Yoon, D. and Park, R. 2017. ECG-ViEW II, a freely accessible electrocardiogram database. PLOS ONE. 12, 4 (2017), e0176222.
[62] Taddei, A., Distante, G., Emdin, M., Pisandi, P., Moody, G., Zeelenberg, C. and Marchesi, C. 1992. The European ST-T database: standard for evaluating systems for the analysis of ST-T changes in ambulatory electrocardiography. European Heart Journal. 13, 9 (1992), 1164-1172.
[63] Martinez, J., Pahlm, O., Ringborn, M., Warren, S., Laguna, P. and Sornmo, L. 2017. The STAFF III Database: ECGs Recorded During Acutely Induced Myocardial Ischemia. Computing in Cardiology Conference (CinC). 44, (2017), 1-4.
[64] Brownlee, J. 2016. How To Backtest Machine Learning Models for Time Series Forecasting: https://machinelearningmastery.com/backtest-machine-learning-models-time-series-forecasting. Accessed: 2022- 04- 31.
[65] Kouassi, K.H, Moodley, D. 2020. An analysis of deep neural networks for predicting trends in time series data. arXiv:2009.07943. Retrieved from https://arxiv.org/abs/2009.07943
[66] Brownlee, J. 2017. What is the Difference Between Test and Validation Datasets? https://machinelearningmastery.com/difference-test-validation-datasets. Accessed: 2022- 04- 30.
[67] Xu, Y. and Goodacre, R. 2018. On Splitting Training and Validation Set: A Comparative Study of Cross-Validation, Bootstrap and Systematic Sampling for Estimating the Generalization Performance of Supervised Learning. Journal of Analysis and Testing. 2, 3 (2018), 249-262.
[68] Nguyen, Q., Ly, H., Ho, L., Al-Ansari, N., Le, H., Tran, V., Prakash, I. and Pham, B. 2021. Influence of Data Splitting on Performance of Machine Learning Models in Prediction of Shear Strength of Soil. Mathematical Problems in Engineering. 2021, (2021), 1-15.
[69] Warnes, Z. 2021. How to Select a Data Splitting Method: https://towardsdatascience.com/how-to-select-a-data-splitting-method-4cf6bc6991da. Accessed: 2022- 04- 28.
[70] Birba, D. 2020. A Comparative study of data splitting algorithms for machine learning model selection. KTH Royal Institute of Technology.
[71] Tseng, K., Li, J., Tang, Y., Yang, C. and Lin, F. 2020. Healthcare knowledge of relationship between time series electrocardiogram and cigarette smoking using clinical records. BMC Medical Informatics and Decision Making. 20, S3 (2020), 2-5.
[72] Passos, D. and Mishra, P. 2022. A tutorial on automatic hyperparameter tuning of deep spectral modelling for regression and classification tasks. Chemometrics and Intelligent Laboratory Systems. 223, 104520 (2022), 1-13.
[73] Rodriguez, J., Perez, A. and Lozano, J. 2010. Sensitivity Analysis of k-Fold Cross Validation in Prediction Error Estimation. IEEE Transactions on Pattern Analysis and Machine Intelligence. 32, 3 (2010), 569-575.
[74] Brownlee, J. 2018. A Gentle Introduction to k-fold Cross-Validation: https://machinelearningmastery.com/k-fold-cross-validation/. Accessed: 2022- 05-01
[75] Wong, T. and Yeh, P. 2020. Reliable Accuracy Estimates from K-Fold Cross Validation. IEEE Transactions on Knowledge and Data Engineering. 32, 8 (2020), 1586-1594.
[76] Donahue, J., Hendricks, L., Rohrbach, M., Venugopalan, S., Guadarrama, S., Saenko, K. and Darrell, T. 2017. Long-Term Recurrent Convolutional Networks for Visual Recognition and Description. IEEE Transactions on Pattern Analysis and Machine Intelligence. 39, 4 (2017), 677-691.
[77] Zou, K., O'Malley, A. and Mauri, L. 2007. Receiver-Operating Characteristic Analysis for Evaluating Diagnostic Tests and Predictive Models. Circulation. 115, 5 (2007), 654-657.
[78] ROC curves – what are they and how are they used?: 2011. https://acutecaretesting.org/en/articles/roc-curves-what-are-they-and-how-are-theyused#:~:text=As%20the%20area%20under%20an,stands%20for%20Receiver%20Operating%20Characteristic. Accessed: 2022- 05- 01.

Table 1: Selected open-access ECG datasets used in MI detection and their properties

Dataset	Study period	Country	ECG Information	Patient Count	Record Count	Strengths	Limitations
LTST [60]	2003-2007	Slovenia	**Leads**: 2 or 3 **Length**:21-24h **Frequency**: 250Hz **Resolution**: 12 bits with +-10milivolts **Annotations**: locations of the PQ junction (the isoelectric level) and the J point, ST level time series or the ST deviation time series	80	86	- Detailed clinical notes and ST deviation trend plots are provided for all 86 records.	- Only 2 or 3 leads used -Data sample size is small. - MI and healthy control not explicitly labelled.
ECG-VIEWII [61]	1994-2013	South Korea	**Leads**: 12 **Patient demographics**: Aged 0 – 100, mean 42,6, interquartile range 19,2 Male: Female ratio = 50:50% **Annotations**: Test date, clinical department, RR interval, PR interval, QRS duration, QT interval, QTc interval, P axis, QRS axis, and T axis.	461 178	979 273	- Comprises long-term follow-up data - Real-world ECG data of patients who took medication to treat a variety of diseases. - Large sample size -Healthy control patients labelled (94 326)	-MI not explicitly labelled - Contains data from only one large hospital.
PTB [58]	2000	Germany	**Leads**: 12 leads + 3 Frank leads **Patient demographics**: Aged 17–87, mean 57.2; 209 men, mean age 55.5, and 81 women, mean age 61.6 **Length**: 2 min **Frequency**: 1 kHz-10 kHz **Resolution**: 16 bits with 0,5 V/LSB (2,000 A/D units per mV)	290	549	- 15 leads ECG used. - MI (148) and healthy control patients (52) explicitly labelled. - Higher resolution than LTST	-Small sample size. - Data is only recorded from a single site.
PTB-XL [59]	1989-1996	Switzerland	**Leads**: 12 **Patient demographics**: (Male: Female ratio = 52:48% (Ages: from 0 to 95 years Median 62 and interquartile range of 22) **Length**: 10 s	18 885	21 837	- Largest open access 12-lead ECG-waveform dataset. - Recorded at multiple sites. - MI (5486) and healthy control records (9528) explicitly labelled.	

			Frequency: 500Hz **Metadata**: demographics, infarction features, likelihoods for diagnostic ECG statements and annotated signal properties			- Suggests folds for splitting training and testing data. This incentivises standardization for evaluating model performance	

Table 2: Results and Properties of Notable Recurrent Architecture-based MI detection on ECGs

Study	Year	Data Properties	Models	Data Split	Evaluation Metrics				
					Sensitivity (%)	**Specificity (%)**	**Precision (%)**	**Accuracy (%)**	**F1-Score (%)**
[34]	2019	**Dataset**: PTB **Leads: 8** **Data extracted:** MI: 10144, HC: 2215 (ECG sequences)	RNN 1,2,3 hidden LSTM layers	Train:Test = **80:20%**	RNN: **83** 1 Hidden LSTM layer: **91** 2 Hidden LSTM layers: **91** 3 Hidden LSTM layers: **91**		RNN: **68** 1 Hidden LSTM Layer: **91** 2 Hidden LSTM Layer: **90** 3 Hidden LSTM Layer: **91**		RNN: **75** 1 Hidden LSTM Layer: **90** 2 Hidden LSTM Layers: **90** 3 Hidden LSTM Layers: **90**
[35]	2019	**Dataset:** PTB **Leads:** 15 **Data Extracted:** MI: 10144, HC: 2215 (ECG sequences)	RNN, LSTM and GRU	Best Train:Test = **90:10%**	RNN: **85.81** LSTM: **98.49** GRU: **87.07**	RNN: **87.92** LSTM: **97.97** GRU: **98.1**	RNN: **89.56** LSTM: **95.67** GRU: **94.89**		RNN: **84.97** LSTM: **96.32** GRU: **94.08**
[44]	2020	**Dataset**: PTB **Leads:** 12 **Data extracted:** MI: 20337, HC: 4652 (ECG sequences)	Attention Based Hierarchical LSTM	**5-Fold** cross validation	**99.92**	**99.73**		**99.88**	
[45]	2019	**Dataset**: PTB **Leads:** 12 **Data extracted:** MI: 369, HC: 79 (records)	BiLSTM with and without heartbeat-attention mechanism (HAM)	Train:Test = **70:30%**	BiLSTM w/o HAM: **93.86** BiLSTM with HAM: **95.58**	BiLSTM w/o HAM: **85.00** BiLSTM with HAM: **90.48**		BiLSTM w/o HAM: **92.54** BiLSTM with HAM: **94.77**	
[46]	2019	**Dataset**: PTB **Leads: 8** **Data extracted:** 54753 heartbeats	LSTM	Train:Test = **90:10%**	**99.91**	**99.95**	**99.91**	**99.91**	
[47]	2021	**Dataset:** PTB **Leads:** Only Lead II **Data Extracted:** MI: 50732, HC: 10123 (ECG sequences)	3 layers LSTM	**10-fold** Cross Validation	**91.88**	**80.81**	**95.30**	**89.56**	**93.45**
[48]	2020	**Dataset:** PTB **Leads:** 15 **Data extracted:** MI: 3897, HC: 690 (ECG sequences)	CNN, LSTM, EDN	Not given				CNN: **84.95** LSTM: **85.23** EDN: **88.89**	

Table 2 Key:

GRU = Gated Recurrent Unit

EDN = Enhanced Deep Neural Network. This was the study's proposed model which is based on a hybrid CNN – LSTM model.

MI = Myocardial Infarction**, HC =** Healthy Control

HAM = Heartbeat Attention Mechanism

RNN: Refers to the simple RNN without any memory/ gate modifications

Note: Studies which focused on other ML techniques or solely used a hybrid technique (e.g. CNN-LSTM) were excluded from this comparison for consistency. Additionally, only reported metrics were included.