Autonomous Myocardial Infarction Detection from Electrocardiogram with a Multi Label Classification Approach

Vishwa Mohan Singh VISHWA@DOZEE.IO
Vibhor Saran VIBHOR@DOZEE.IO
Pooja Kadambi POOJA@DOZEE.IO
Turtle Shell Technologies Pvt. Ltd., India

Editors: Emtiyaz Khan and Mehmet Gönen

Abstract

Myocardial Infarctions (MI) or heart attacks are among the most common medical emergencies globally. Such an episode often has mild or varied symptoms, making it hard to diagnose and respond in a timely manner. An electrocardiogram (ECG) is used to analyze the heart's electrical activity and, through this help, clinicians detect and localize a heart attack. However, interpretation of the ECG is made manually by trained professionals. In order to make this diagnosis more efficient, multiple methods have tried to automate the MI detection and localization process. In this work, we aim to create a more effective method of MI detection by restructuring the localization as a multi-label classification (MLC) problem, in which one set of attributes can belong to one or more classes. For this classification, features like the ST-deviation, T wave amplitude, and R-S ratios have been extracted and fed into the MLC model, which in our case, is a chain classifier of random forest. This proposed model will have five classes as the target, which represent the locations where an MI can occur. Our method achieves the best overall hamming accuracy of 81.49% in a k-fold cross validation test, with the highest accuracy for an individual class being 97.72% for anterior.

Keywords: Electrocardiography; Myocardial Infarction; Multi-Label Classification; Machine Learning

1. Introduction

Myocardial Infarction (MI), commonly known as heart attack, is one of the leading causes of death across the world. According to statistics given by Thomas (2020), in 2018 alone, over 30.3 million people in the USA were diagnosed with myocardial infarction, and every year, around 647,000 people die from the same. As stated in the article by Beckerman (2020), an MI occurs when there is a halt in the blood flow to some crucial areas in the heart. People who are suffering from high cholesterol, high blood pressure or are overweight have a very high chance of having a heart attack. In patients suspected of MI, understanding the severity and location of the infarction is important. Electrocardiography (ECG) is one of the most commonly used methods to screen for the presence of an MI, the methodology and guidelines for which are detailed in the work by Crawford et al. (1999). The trace obtained from the ECG is analyzed by experts to confirm a diagnosis. This slows down the decision-making and healthcare response, which is crucial in such a situation. There is a

need for systems that can aid the path to diagnosis so that the treatment can be initiated sooner. One potential approach is including automation in the ECG evaluation process.

In recent years, machine learning-based methods have been really useful in analyzing and adding interpretation to ECG. However, there has not been sufficient research on techniques that gives an effective method for localization of the infarction. Referring to the article by Lome (2020), this information can be derived by looking at signal artifacts in an ECG reading like abnormal ST-deviation and T-wave inversion in the leads corresponding to the wall. A 12-lead ECG reading can be used to get the infarction status from the lateral, anterior, inferior, septal, and to some extent, the posterior wall. It is also possible for multiple infarctions to happen in multiple walls at once. This makes the signal analysis problem a multi-label classification task. This paper proposes a method that can analyze a 12-lead ECG data to detect if there is an infarction present and, if present, to determine its location(s) using a multi-label classifier model.

2. Literature Survey

2.1. Infarction Detection

An ECG is standard practice to check the heart's rhythm using electrodes to capture the electric signals produced by the heart during particular events. As stated in the publication by Gacek and Pedrycz (2011), in a standard ECG pattern, the P wave represents the atrial contraction, the QRS complex represents the ventricular contraction, and the T wave shows the ventricular relaxation phases. All the crucial segments of an ECG reading can be seen below in Fig. 1 (a). The two parts of an ECG waveform that are typically considered to

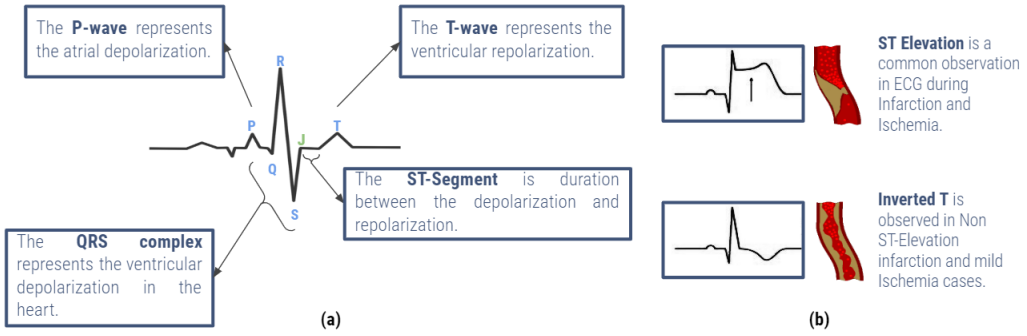


Figure 1: (a) Points and segments in an ECG readings. (b) ST-elevation and T-wave inversion with their effects of the blood flow.

diagnose an infarction episode are the ST segment and the T wave. If there is an elevation in the ST segment, i.e., if the J-point is significantly above the iso-electric reference, it implies that there is an occlusion to the blood flow to the heart. These types of infarction are classified as ST-Elevation Myocardial Infarction (STEMI). Another indicator of infarction is an inverted T-wave, which is an indicator of ischemia in the vessels. These anomalies are labeled as Non-ST-Elevation Myocardial Infarction (NSTEMI), which is discussed in detail in the work by Daga et al. (2011). Both the observations are shown in Fig. 1 (b).

These two markers may be used for a first-level screening of the presence or absence of the infarction.

There are three major vessels where the occlusion can occur, namely the Right Coronary Artery, Left Anterior Descending, and the Left Circumflex. The blood flow can stop in one or more than one vessels at the same time. Since they are located on different walls of the heart, a multi-lead reading is taken in order to get the condition of hearts from multiple perspectives. Work by Macfarlane et al. (1990) describes the interpretation method of a 12-Lead system. Every lead essentially looks at a particular wall of the heart. These directions are shown bellows in Fig. 2 (b). During an infarction episode occurring in a particular wall, the corresponding ECG leads will show an ST elevation, a T-wave inversion, and in some cases of STEMI, the opposite lead will also show an ST-Depression. This is shown in Fig. 2 (a).

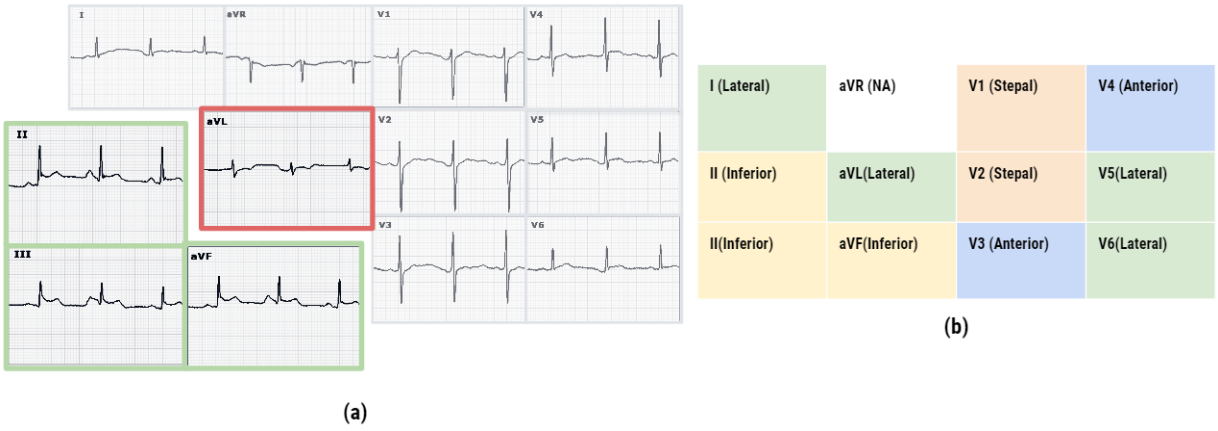


Figure 2: (a) A case of Inferior wall STEMI, where the Inferior lead (green) shows ST elevation and the opposite lead aVL (red) shows ST depression. (b) Leads and the walls they are overlooking.

2.2. Current Methodology

To automate the diagnosis process using ECG, there has been previously described methods like machine learning and deep learning. The work by Tsien et al. (1998) is one of the foundational steps in this field that uses a classification tree and regression model in order to differentiate between MI waveforms and Normal waveforms based on the mentioned markers. Since then, the methods have improved and have become capable of extracting these features directly from the ECG readings. Most of these works are being carried out on the open-source Physikalisch Technische Bundesanstalt dataset published by Bousseljot et al. (1995), available on PhysioNet from Goldberger et al. (2000). This contains the recordings of people with infarction episodes along with the label for the location of the wall/s. In research by Feng et al. (2019), and Haddadi et al. (2019), the authors have used methods like Recurrent Neural Network, Convolution Neural Network, and manual feature

extraction for a statistical model to do a binary classification of infarction vs. normal. The model proposed by Padhy and Dandapat (2017) has demonstrated an accuracy of 95.3% on this binary classification task. However, some works try to go beyond detection and try to find the location of the infarction. The model proposed by Arif et al. (2010) groups infarction classes together to reduce the number of targets of the model and achieves an accuracy of 70.87%. Work done by Baloglu et al. (2019) and Arif et al. (2012) achieves an accuracy of 99.68% and 98.3% respectively on an 11 class problem. Work done by Jahmunah et al. (2022) achieves a similarly high accuracy with a CNN base architecture and uses GradCAM to make the results explainable.

2.3. Research Gap and Motivation

In the aforementioned literature, it is evident that a binary classification of MI is possible. Accurate localization, however, still remains a challenge. All the work done so far treat this problem as a multi-class classification problem which, in the case of the PTB dataset, gives a total of 11 target classes. This makes creating a model complex and computationally heavy. Although the models by Arif et al. (2012), Baloglu et al. (2019) and Jahmunah et al. (2022) achieve high accuracies, it is a result of a preprocessing method that is susceptible to overfitting. This is because rather than doing the train-test split on the subjects, it has been done on individual beats extracted from the ECG recordings. This results in the train and test set having segments from the same subjects, and the model starts remembering the pattern from specific patients rather than learning to identify MI. An accuracy jumping to around 99% is highly unlikely on obscure classes like posterior, lateral, and their variants without overfitting, especially with only a single lead signal as the input, as demonstrated in work by Baloglu et al. (2019). Furthermore, the GradCAM results shown in the work by Jahmunah et al. (2022) has a lot of inconsistencies with Figure 2 (b) in terms of which lead is in the focus for which kind of infarction. For example, to identify a lateral infraction, the model focuses more on the inferior-facing lead II compared to the lateral-facing I, aVL, V5 or V6. This further reinforces the notion that the patterns that the model is learning might be incorrect and a result of overfitting. Our experiments have found that models trained on similar splits tend to collapse on other datasets or unseen subjects, giving an accuracy of 99% on the unseen beats of the seen subjects and 35% on the unseen subjects. Along with this, no work addresses the class imbalance that is seen in MI datasets that hinders the training process drastically.

Many of the 11 classes are a combination of one or more core classes. This creates another problem in the training phase impacting common class prediction. If the model classifies a combination class as a single class within the combination, we can say that the model has partially predicted the condition. However, the loss function will give the model a score of 0, and it will not be incentivized for getting only a single label right. This reduces the models capable of predicting the most common classes effectively. However, the hybrid nature of some classes allows us to convert the problem from an 11-class multi-class classification (MCC) to a 5-class multi-label classification (MLC), an approach that has been described in the work by Tsoumakas and Katakis (2007). This restructuring addresses the class imbalance problem to some extent and prevents an obscure and hard-to-predict class from lowering the model's performance in predicting other classes.

The multi-label classification problem is most effectively managed by the chain-classification model proposed by Read et al. (2009) and its variants. The aim of the presented research is to create and test a model that is capable of predicting the infarction classes as a multi-label vector from the features extracted from an ECG segment. This will lead to a more reliable prediction of the individual classes without being affected by the obscure ones.

Thus fundamentally there lacks a comprehensive model that maintains accuracy across the different presentations of MIs whether common or obscure. This is clinically relevant as it will impact the sensitivity and specificity of the model as an automated evaluation tool. Since the use of such models is in critical patients there is low tolerance to false negatives and false positives as well as delays in analysis.

3. Methodology

3.1. Dataset and Feature Extraction

The dataset used in our study is the aforementioned PTB-DB dataset, which provides ten classes of arrhythmia along with control subjects. To convert this 11-class MCC problem to a 5-class MLC problem, we need to convert classes into a vector of length five with the fundamental classes set as 1. The class distribution before and after conversion to the MLC problem is shown below in Fig. 3. Note that the class "Control" has been used as a separate model for the basic detection of infarction presence or absence.

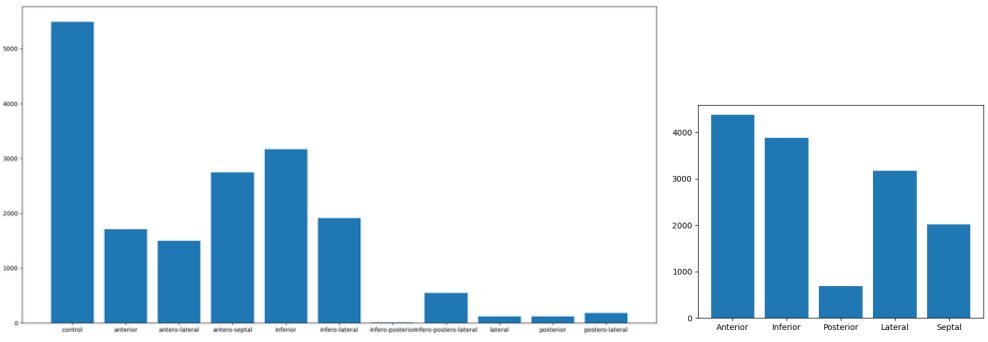


Figure 3: Class Distribution of the original PTB dataset (left) and the class distribution of PTB dataset when converted to a multi-label classification problem (right)

In order to extract the ST segment and T wave features, an algorithm is needed to extract the core points (P, Q, R, S points, and J inflection) shown in Fig. 1 from the ECG reading. The method to find these have been adapted from a method presented by Kumar and Singh (2018). The steps for the feature extraction process from a 10-second ECG signal were as follows:

Preprocessing: The first step is to remove artifacts like extraneous noise and baseline created by breathing. To remove the disturbance, we have used the Savitzky–Golay filter, which, as described in the work by Schafer (2011), smoothens the signal by finding a polynomial fit within a window without hindering crucial aspects like waveform

amplitude and frequency. The filter is represented by the following equation:

$$(y_k)_s = \frac{\sum_{i=-n}^{i=n} A_i . y_{k+i}}{\sum_{i=-n}^{i=n} A_i} \tag{1}$$

where n is the number of points in the raw signal and A is the set of derived integers used by the filter. To remove the baseline, a Butterworth band-pass filter was used to extract the signal in the 0-1 Hz range. This extracted signal was removed from the raw signal. The following Fig. 4 shows the difference between the input raw and output filtered signal.

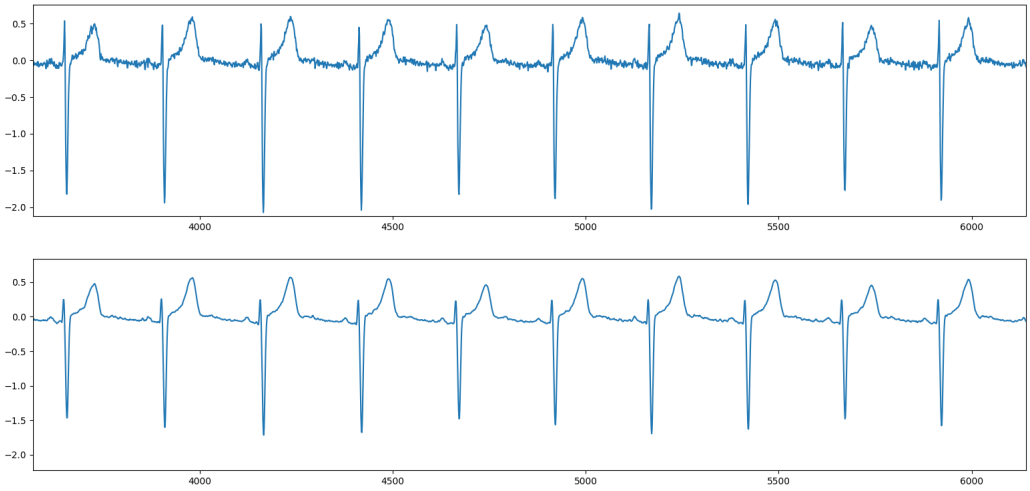


Figure 4: The raw signal (top) and the filtered signal after applying Savitzky–Golay filter and baseline removal (bottom)

QRS complex detection: To detect the R-peak, the algorithm proposed by Saran et al. (2018) has been used. Here, the points in the cluster with the highest average peak amplitude have been selected as the R-peaks. The Q and S points are the minimas located in the range [(R - 200ms) to R] and [R to R + 200ms], respectively.

P-wave and T-wave detection: The P-wave is the maxima located to the left of the Q point in the range [Q-200ms to Q]. The onset and offset of P-wave are within 80ms on the left and right of the P-point, respectively. The T-wave is calculated using the P-wave offset as the baseline as the wave can be inverted in some classes. To execute this T-wave detection, both the minima and the maxima are located in the range [S to S + 200ms], and the point which was the farthest away in terms of amplitude from the baseline was selected as the T peak. The onset and offset points of the T-peak were selected in a manner similar to the S-wave in step 2.

J point and ST-segment detection : The J-point is essentially the first inflection point after the S-point. To get the inflection points, the double differential of the wave was calculated. All the points with the double differential value of 0 were selected. The ST-Segment is the region from the J-point to the onset point of the next T-wave. The following Fig. 5 shows the point detection on the lead II of a record taken from the PTB dataset. The Blue Point represents the QRS complex along with the P-peak and T-peak. The orange point is the J-point, and the green and red ones are the onset and offset points, respectively.

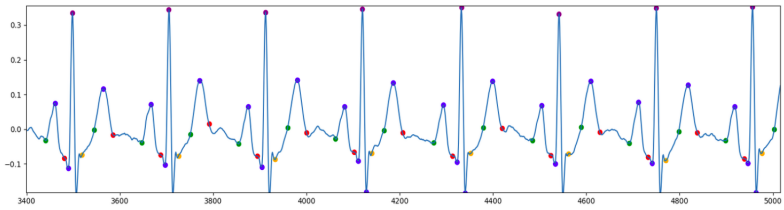


Figure 5: Point detection on a segment of Lead-II signal taken from a record of PTB dataset

The information given in the publications by Katsaris et al. (1993) and Aldrich et al. (1987) gives us a basic guideline to localize the infarction based on the 12-lead reading. For the 10-classes of infarction present in the dataset, the following Fig. 6 summarises these guidelines in the form of a decision tree that can be used to identify the type of infarction. From this, we can determine that there are three important features that are needed for this diagnosis: ST Deviation, R-S ratio (for high R), and T wave inversion. To create the input vector, these three features were extracted from every lead of a record. For ST-deviation, we took the mean of the potential difference between all the ST-segment and the iso-electric reference, which in this case was taken as the T-P segment as recommended by Maeba (2012). For T-wave inversion, we take the mean amplitude of T waves relative to the respective P-wave offset. The last feature, the R-S ratio, helps with the identification of posterior wall ischemia and infarction, according to the article by Healio (2020). For this, we take the mean of the ratio of the potential difference between the R point and the iso-electric reference and the potential difference between the S point and the iso-electric reference. This gives us an input feature vector size of $N \times 36$ where N is the number of ECG segments.

3.2. Classification flow and Model

In order to keep the control patients out of the localization detection part, we have divided the flow into two parts. We first focus on performing a binary classification to find whether an infarction is present in the given feature set. For these, various models were tested based on accuracy, precision and recall, and F1 score. The performance of all will be discussed later in the results section.
If infarction is detected, the feature vector is then passed to a multi-label classification model. For this, the bi-directional chain classifier proposed by Singh and Kulkarni (2020)

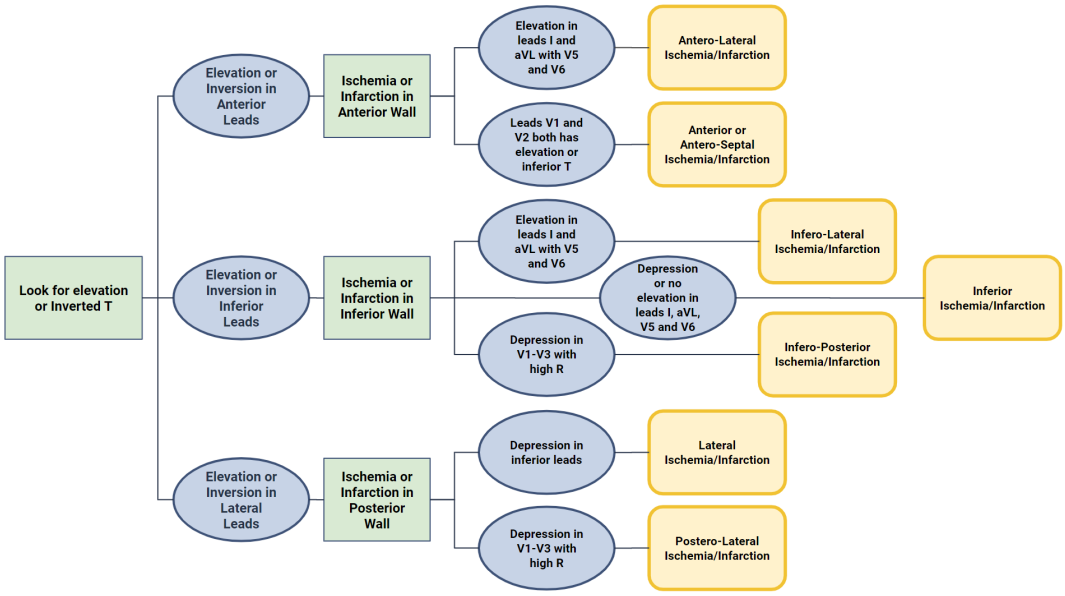


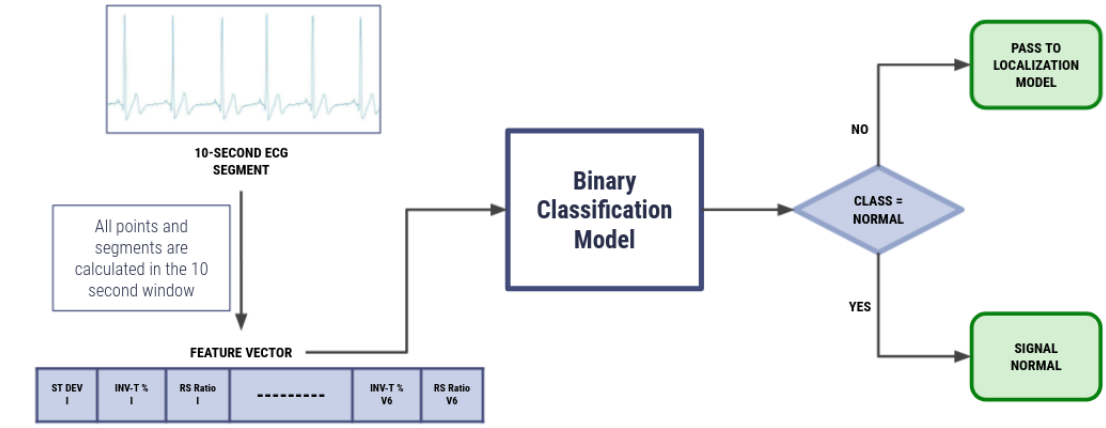
Figure 6: Decision Tree for localizing the infarction in an ECG reading based on the ST-deviation, T-wave inversion and presence of high R

was used, which is a variant of the chain classifier proposed by Read et al. (2009). A chain classifier helps in capturing the inter-dependency amongst the labels present in a multi-label vector. This helps with boosting the performance of the model. In our case, the base classifier used in the chain architecture is a random forest binary classifier with a 10 decision trees each. Fig. 7 shows the flow of the full model.

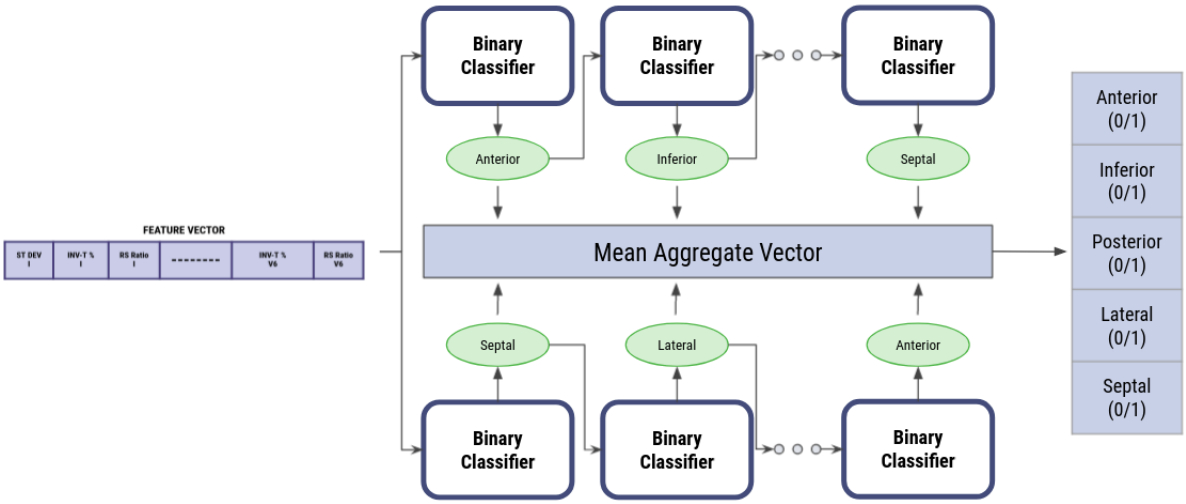
3.3. Evaluation Methodology

For the binary classification part, multiple models, including random forest, K-Nearest Neighbours by Peterson (2009), Support Vector Machine by Noble (2006) and X-Gradient Boosting by Friedman (2002) were tested on metrics like accuracy, precision, and recall. For the MLC models, we have tested a classic (Chain) and bi-direction chain model (BD Chain). The chains have been created using Random Forest Classifiers as they were found to be the best performing model. A simple Artificial Neural Network (ANN) model with batch normalization was also tested as a control method. The MLC model has been tested in two parts. The first test is to measure the overall performance of the model, and the second test is to get the performance of the model on individual classes. For the overall performance, the following metrics have been used:

Hamming Accuracy: Described in the work by Wu and Zhu (2020), this metric gives a score on the basis of how many labels the model got right. This is, therefore, a more appropriate measure compared to accuracy for an MLC problem. If Y_i is the predicted label vector, S_i is the actual label vector, and $|N|$ is the size of the dataset,



(a)



(b)

Figure 7: (a) Binary Classification model to detect the presence of infarction and (b) the bi-directional chain model to localize the infarction location

the hamming accuracy is defined as :

$$Accuracy = \frac{1}{|N|}\sum_{i=1}^{|N|}\frac{|S_i \cap Y_i|}{|S_i \cup Y_i|} \tag{2}$$

Log Loss: Proposed in the publication by Vovk (2015) and also known as cross-entropy, log loss's primary goal is to penalize the model more where the error is more severe, i.e., the prediction probability is far from the true label. For $l_j \in S_i$, $\lambda \in W_i^{norm}$ and where $|L|$ is the number of labels, the log loss is defined as:

$$Loss = \frac{1}{|N|}\sum_{i=1}^{|N|}\sum_{j=1}^{|L|} -max(A, B) \tag{3}$$

Where $A = log(\lambda_j)l_j + log(1-\lambda_j)(1-l_j)$ and $B = log(\frac{1}{|N|})$.

F1$_{macro}$: Described in the publication by Opitz and Burst (2019), $F1_{macro}$ score, which is the harmonic mean between the precision and recall. If p_j and r_j are the precision and recall, the $F1_{macro}$ score is defined as:

$$F1_{macro} = \frac{1}{|L|}\sum_{j=1}^{|L|}\frac{2 * p_j * r_j}{p_j + r_j} \tag{4}$$

A K-Fold cross-validation, presented in the work by Moreno-Torres et al. (2012), was also performed on the 3 MLC models to check the consistency of performance.

4. Results

The following Table 1 shows the performance of various ML model on the binary classification problem.

Table 1: Performance comparison of models on the binary classification

Model	Accuracy	Precision	Recall	F1
SVM	73.62	73.73	99.35	84.64
KNN	69.14	82.77	73.06	77.61
RF	93.84	92.24	**100.0**	95.96
XGB	**94.43**	**94.64**	97.95	**96.26**

Next, we compare the MLC model for the localization of MI. To ensure the consistency in performance, 5-fold cross-validation was performed on the three selected models. The following Table 2 shows the acquired results.

Table 2: Performance Comparison of MLC models with 5-fold cross validation

	Chain-RF			BD Chain-RF			ANN		
Fold	Ham.	Log	F1	Ham.	Log	F1	Ham.	Log	F1
Fold 1	69.11	2.139	0.6443	**69.84**	**1.842**	**0.6586**	31.55	3.286	0.1563
Fold 2	72.54	1.612	0.5651	**73.74**	**1.542**	**0.5874**	29.36	3.202	0.2128
Fold 3	81.39	1.421	0.7269	**81.49**	**1.382**	**0.7503**	58.83	2.269	0.3815
Fold 4	**75.36**	2.357	**0.5975**	72.96	**2.136**	0.5833	58.70	2.401	0.3622
Fold 5	**71.85**	1.981	0.6027	70.80	**1.723**	**0.6070**	43.21	2.329	0.2989

Another important aspect that is needed to be analyzed is the class-wise performance of these models since there is an imbalance in their distribution. The following Table 3 shows the class-wise results from the Fold 3 models.

Table 3: Performance Comparison of MLC models on individual classes present in the multi-label representation

	Chain-RF			BD Chain-RF		
Class	Accuracy	Precision	Recall	Accuracy	Precision	Recall
Anterior	96.26	**96.50**	91.93	**97.72**	96.35	**96.35**
Inferior	92.57	**95.94**	92.88	**92.61**	**95.94**	**92.94**
Posterior	91.29	17.67	**100.0**	**91.47**	**19.40**	**100.0**
Lateral	83.58	53.84	53.84	**84.90**	**65.22**	**78.12**
Septal	92.52	**78.07**	81.61	**93.66**	70.60	**94.52**

Since our method is an alternative to a multi-class approach, we need to compare the performance of the MLC model to the MCC model (here, a basic random forest classifier) to check the prediction distribution. The following Fig. 8 shows the confusion matrix of both the models' performance on data of 80 subjects that were not included in the training set.

To round-off the results, we compared the performance of our method with the existing literature. The following Table 4 shows the comparison of our Bi-Directional Chain method with the existing multi-class classification approaches using our subject based train-test split.

Figure 8: The confusion matrix of both the models trying to predict from the 9 classes available in segments extracted from the selected 80 subjects. Evidently, the MCC model tends to overfit and predict most classes as anterior which affects the accuracy of all the other classes. The MCC model has far less errors relatively, and even in those cases, the errors is not far off the true label eg. inferior has mostly been miss-classified as infero-lateral

Table 4: Comparison of our method and the existing literature tesed on the PTB-DB dataset

Method	Accuracy	Precision	Recall
Bi-Directional Chain [Our method]	**62.52**	**62.52**	**62.52**
10 Layer CNN [Baloglu et al. (2019)]	33.89	34.14	33.89
Multi-Headed ConvLSTM1D [similar to Feng et al. (2019)]	37.56	35.72	37.56
KNN Classifier [Arif et al. (2012)]	18.04	27.61	20.52

5. Discussion and Conclusion

5.1. Result Discussion

The following section discusses the takeaways from our obtained results. Dividing the task into detection and localization handles a major part of the class imbalance during training and reduces the size of the target vector. From our results presented in Table 1, we find that the X-Gradient Boosting model works the best with an accuracy of 94.43 %. This is beneficial for both the MLC and the MCC model. The confusion matrix shown in Fig. 8 along with the comparison shown in Table 4 has shown that the MLC works better at recognizing the infarction case compared to the existing MCC methods. This can be

attributed to the fact that an MCC, while training, is not incentivized to predict at least some of the multiple walls correctly. The MLC approach dodges this, and it allows the model to give an accurate or at least a near-accurate prediction.

The k-fold analysis in Table 2 shows the performance of the selected MLC models. From the results, we find that in terms of hamming accuracy, the performance of both models is tantamount. However, the bi-directional chain makes less severe errors in all cases, which can be seen by the log loss values. This can be attributed to the bagging ensemble structure of the bi-directional chain. A disadvantage that the bi-directional possesses is its higher computational cost. The basic ANN model does not perform adequately in any case and therefore is unsuitable for dealing with the localization problem.

Furthermore, Table 3 shows the performance of both the chain models on the individual classes. Both models perform well with the majority of the classes, giving the best accuracy in cases of Anterior MI. However, they struggle in obscure classes like posterior and its variants with a low best precision score of 19.40, which hinders their overall performance. This problem can be solved with the inclusion of leads V7-V9 in the input data since these are the leads that can look at the posterior wall more effectively.

From our results, we can conclude that predicting these 5 classes independently of each other, using different model weights for each, lead to a better accuracy and a more generalized model. Therefore, an MLC approach serves the problem better than an MCC approach. And the presence of high inter-label dependency makes a chain architecture more suitable than an alternative like ANN. From our experiments, we determine that the with our feature extraction method, the combination of X-Gradient Boosting for binary classification and bi-directional chains for multi-label classification gives the best overall results.

5.2. Limitations

Although the MLC approach delivers good results on our dataset, there are some drawbacks to the current state of the model. The main issue is the inconsistency in the accuracy of individual classes, which stems from the class imbalance in the dataset. This is especially apparent in the case of Posterior MI, which has lower occurrences in the dataset because of its rarity and the inability of a 12-lead ECG to analyze the posterior wall effectively. There are also factors that are not taken into consideration, including gender, medical history, and comorbidity of the subject, that can assist the model in making more accurate classifications. A more focused data collection protocol that takes these factors into account can help train a better model.

Another issue that can be caused by insufficient samples belonging to a particular class in potential overfitting. Since ECG carries good biometric information, it is possible for the models under training to start memorizing the subjects' characteristics instead of the MI patterns if the number of available samples is insufficient or if the base model itself is very susceptible to overfitting. A part of this problem can be resolved by focusing on the architecture of the base model that deals with high variance more effectively than the ensemble used in our work.

5.3. Conclusion

In this work, we have proposed an MLC-based method to detect and localize myocardial infarction. The work has established that structuring localization as an MLC problem makes it easier for the model to recognize the type of infarction based on features like the average ST Deviation, T- wave amplitude, and the R-S ratio. Reducing the 10-second ECG signal to these three statistical features makes our model computationally lighter compared to many deep learning approaches. This makes the proposed model more practical for scenarios where time is critical, and the data burden for the model's predictions is of high volume. Identifying the location of the MI in a multi-label format improves the model's performance for individual classes. Further studies are needed to validate that this method can be utilized by healthcare providers in cases of suspected MI. This method does not aim to propose a diagnosis but mainly to help providers in their analysis and decision-making.

References

Harry R Aldrich, Nancy B Hindman, Tomoaki Hinohara, Michael G Jones, Jane Boswick, Kerry L Lee, Wanda Bride, Robert M Califf, and Galen S Wagner. Identification of the optimal electrocardiographic leads for detecting acute epicardial injury in acute myocardial infarction. *The American journal of cardiology*, 59(1):20–23, 1987.

Muhammad Arif, Ijaz A Malagore, and Fayyaz A Afsar. Automatic detection and localization of myocardial infarction using back propagation neural networks. pages 1–4, 2010.

Muhammad Arif, Ijaz A Malagore, and Fayyaz A Afsar. Detection and localization of myocardial infarction using k-nearest neighbor classifier. *Journal of medical systems*, 36 (1):279–289, 2012.

Ulas Baran Baloglu, Muhammed Talo, Ozal Yildirim, Ru San Tan, and U Rajendra Acharya. Classification of myocardial infarction with multi-lead ecg signals and deep cnn. *Pattern Recognition Letters*, 122:23–30, 2019.

James Beckerman. Heart attack: Symptoms, causes, treatment, and prevention. *WebMD*, 2020.

R Bousseljot, D Kreiseler, and A Schnabel. Nutzung der ekg-signaldatenbank cardiodat der ptb über das internet. 1995.

Michael H Crawford, Steven J Bernstein, Prakash C Deedwania, John P DiMarco, Kevin J Ferrick, Arthur Garson Jr, Lee A Green, H Leon Greene, Michael J Silka, Peter H Stone, et al. Acc/aha guidelines for ambulatory electrocardiography: executive summary and recommendations: a report of the american college of cardiology/american heart association task force on practice guidelines (committee to revise the guidelines for ambulatory electrocardiography) developed in collaboration with the north american society for pacing and electrophysiology. *Circulation*, 100(8):886–893, 1999.

Lal C Daga, Upendra Kaul, and Aijaz Mansoor. Approach to stemi and nstemi. *J Assoc Physicians India*, 59(Suppl 12):19–25, 2011.

Kai Feng, Xitian Pi, Hongying Liu, and Kai Sun. Myocardial infarction classification based on convolutional neural network and recurrent neural network. *Applied Sciences*, 9(9): 1879, 2019.

Jerome H Friedman. Stochastic gradient boosting. *Computational statistics & data analysis*, 38(4):367–378, 2002.

Adam Gacek and Witold Pedrycz. Ecg signal processing, classification and interpretation: a comprehensive framework of computational intelligence. 2011.

Ary L Goldberger, Luis AN Amaral, Leon Glass, Jeffrey M Hausdorff, Plamen Ch Ivanov, Roger G Mark, Joseph E Mietus, George B Moody, Chung-Kang Peng, and H Eugene Stanley. Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *circulation*, 101(23):e215–e220, 2000.

Rachid Haddadi, Elhassane Abdelmounim, Mustapha El Hanine, and Abdelaziz Belaguid. A wavelet-based ecg delineation and automated diagnosis of myocardial infarction in ptb database. 2019.

Healio. Posterior wall myocardial infarction (mi) ecg review. *healio*, 2020.

V Jahmunah, EYK Ng, Ru-San Tan, Shu Lih Oh, and U Rajendra Acharya. Explainable detection of myocardial infarction using deep learning models with grad-cam technique on ecg signals. *Computers in Biology and Medicine*, 146:105550, 2022.

Georgios A Katsaris, Evagelos I Tsaritsaniotis, Ioannis P Tsounos, Kostas D Panisois, Ioannis A Katsaris, Ioannis K Kaprinis, and Stefanos X Roussis. Surface electrocardiogram in the detection of myocardial ischemia during percutaneous coronary angioplasty. *Angiology*, 44(10):797–802, 1993.

Amit Kumar and Mandeep Singh. Statistical analysis of st segments in ecg signals for detection of ischaemic episodes. *Transactions of the Institute of Measurement and Control*, 40(3):819–830, 2018.

Steven Lome. Top 5 mi ecg patterns you must know. *healio*, 2020.

PW Macfarlane, B Devine, S Latif, S McLaughlin, DB Shoat, and MP Watts. Methodology of ecg interpretation in the glasgow program. *Methods of information in medicine*, 29 (04):354–361, 1990.

Hirofumi Maeba. Isoelectric reference for pericarditis: Tp may be better than pr. *Cardiology*, 123(1):39, 2012.

Jose García Moreno-Torres, José A Sáez, and Francisco Herrera. Study on the impact of partition-induced dataset shift on k-fold cross-validation. *IEEE Transactions on Neural Networks and Learning Systems*, 23(8):1304–1312, 2012.

William S Noble. What is a support vector machine? *Nature biotechnology*, 24(12):1565–1567, 2006.

Juri Opitz and Sebastian Burst. Macro f1 and macro f1. *arXiv preprint arXiv:1911.03347*, 2019.

Sibasankar Padhy and Samarendra Dandapat. Third-order tensor based analysis of multilead ecg for classification of myocardial infarction. *Biomedical Signal Processing and Control*, 31:71–78, 2017.

Leif E Peterson. K-nearest neighbor. *Scholarpedia*, 4(2):1883, 2009.

Jesse Read, Bernhard Pfahringer, Geoff Holmes, and Eibe Frank. Classifier chains for multi-label classification. pages 254–269, 2009.

Vibhor Saran, Gulshan Kumar, Udit Dhawan, and Gaurav Parchani. Unsupervised identification of cardiac contraction through ballistocardiography. pages 42–47, 2018.

Ronald W Schafer. What is a savitzky-golay filter?[lecture notes]. *IEEE Signal processing magazine*, 28(4):111–117, 2011.

Vishwa Mohan Singh and Omkaresh Kulkarni. Multi-label classification with classifier chains of ann models. pages 1–6, 2020.

Jen Thomas. Heart disease: Facts, statistics, and you. *Healthline*, 2020.

Christine L Tsien, Hamish SF Fraser, William J Long, and R Lee Kennedy. Using classification tree and logistic regression methods to diagnose myocardial infarction. *MEDINFO'98*, pages 493–497, 1998.

Grigorios Tsoumakas and Ioannis Katakis. Multi-label classification: An overview. *International Journal of Data Warehousing and Mining (IJDWM)*, 3(3):1–13, 2007.

Vladimir Vovk. The fundamental nature of the log loss function. pages 307–318, 2015.

Guoqiang Wu and Jun Zhu. Multi-label classification: do hamming loss and subset accuracy really conflict with each other? *arXiv preprint arXiv:2011.07805*, 2020.