# [Re] IMLE-Net: An Interpretable Multi-level Multi-channel Model for ECG Classification

Srijan,
IIIT-Hyderabad (Healthcare and AI Center)
Hyderabad, 500032, India
srijan.dokania@ihub-data.iiit.ac.in

## Abstract

*As an electrocardiogram (ECG) is necessary for detection of cardiovascular diseases, deep learning methods to classify the ECG signal is necessary to assist doctors and cardiologists. There is not much attention given towards the automatic analysis of ECG recordings from multiple channel perspective. To mitigate this issue, the authors in the paper has proposed a model which learns the patterns at the beat, rhythm and channel level of a ECG recording. PTB-XL dataset is used. This is the report of an attempt to reproduce the results of the mentioned paper [8]. The experimental results of the original and reproduced are also compared. Re-implementation code can be found at: https://github.com/Srijan221/IMLE-Net.git*

## 1. Introduction

Heart disease is becoming the main cause of mortality in the modern period. Early identification of cardiac disease is critical for improved therapy. The electrical activity of the heart is captured by an electrocardiogram (ECG), which a cardiologist can use to diagnose a variety of diseases. However, it takes some time to analyse an ECG data manually, so, it is necessary to automate the classification of ECG signals to support human diagnosis and minimize human error. While multi-channel ECG recordings provide information about the heart's electrical activity in three dimensions, the majority of the work on interpretable deep learning is on ECG models for single-channel data, which is fairly rudimentary. A 12-channel recording reliably reproduces a number of ECG characteristics, including as the QRS complex, ST-segment, and T waveforms, which are occasionally inaccurately portrayed in a single-channel recording. [5]. It is crucial to examine an ECG recording from a multi-channel viewpoint, yet analysis of ECG recordings from a multiple-channel perspective has not received adequate attention.

To address this gap, the authors propose a model that makes use of multi-channel information from a multi-channel ECG recording. This proposed model from the original paper is re-implemented in this paper, along with the 4 different existing deep learning models mentioned in the original paper to compare the accuracy differences in original and reproduced results.

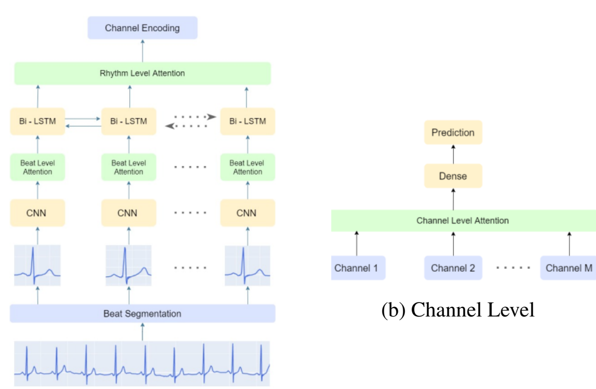
## 2. Methodology

### 2.1. Dataset

The dataset used is the PTB-XL dataset [1] which is the largest openly available dataset that provides clinical 12 channel ECG waveforms. It has total of 21837 ECG records from 18885 patients of 10 seconds length which comprises of 12 channels (I, II, III, aVL, aVR, aVF, V1–V6).

For preprocessing of the dataset, an ECG signal with several channels is divided into beat segments of length W for each channel. The authors used a sliding window approach with no overlap to segment beats of the ECG signal [4]. To get the $k$th beat segment, the beat is spanned from $(k-1)*W$ to $k*W$ over the ECG signal, here, W is the length of the window.

### 2.2. Model Architecture




(a) Individual Channel

(b) Channel Level

Figure 1. Model Architecture

The authors split their proposed model into two parts: the first part processes each channel separately to generate a channel encoding for an input ECG recording, and the second part pools all the channel encodings and generate predictions as shown in Figure 1.

After pre-processing of the signal, the beats are segmented and passed to a CNN that consists of 6 residual blocks. The CNN weights are distributed across all beat segments and channels in the model. As illustrated in Figure 1a, The beat-level attention block, a two-layer neural network, receives the output of the CNN and uses it to identify the places within a beat segment that are more crucial by assigning them greater attention scores. This helps in identifying abnormal heartbeat in a segment. This output is then generated into a ECG rhythm signal and a Bi-directional LSTM is used to learn information from the rhythms. This rhythm is transmitted to the rhythm attention block, which recognises key beat parts in the rhythm by assigning higher attention scores to them. In Equation 1, $\beta^c$ is the rhythm level attention score for the $c^{th}$ channel. For the $c^{th}$ channel, the complete ECG signal for a certain channel is encoded into a vector using the rhythm context vector, or $R^c$.

$$R_c = \sum_{i=1}^{N} \beta_i^c r_i^c \tag{1}$$

As shown in Figure 1b, these channel encodings are then transferred to the channel level block, which uses the data from several ECG channels to analyse it and provide predictions. The authors constructed channel level attention mechanism that can differentiate which channel has more characteristics for a particular heart disease. The channel level attention scores are calculated similar to beat and rhythm level attention.

## 3. Experimental Work

Multiple diagnostic statements can be annotated against an ECG recording. For the diagnostic statement labels, hierarchical groups are provided in terms of 5 superclasses and 24 subclasses. The authors in the original paper used the 5 superclass labels which are Hypertrophy (HYP), Normal ECG (NORM), Myocardial Infarction (MI), , ST/T changes (STTC), and Conduction Disturbance (CD). In Re-implementation, the proposed model is compared with reproduced models like Resnet 101 [3], ECGNet [7], Mousavi et al. [6] and Rajpurkar et al. [2]. Metrics like the mean accuracy, maximum F1 score and macro averaged Area under Receiver Operating Characteristics (ROC-AUC) are used to analyse the performance of the model. Class-wise ROC-AUC scores and Class-wise accuracy are also presented along with these measurements.

## 4. Results

The Reproduced IMLE-Net model is compared with other reproduced models on the 5 superclass-wise ROC-AUC and accuracy as depicted in Tables 1 and 2 respectively. Both the tables reveals that the reproduced IMLE-Net model outperforms the reproduced Rajpurkar et al. on most of the classes which is because of the effective usage of multi-channel information by including multi-levels in the model and channel-level attention. The overall performance of reproduced and proposed models is compared in Table 3. It further shows that the proposed model achieved the best performance across the metrics macro ROC-AUC, mean accuracy and maximum F1 score as compared to the other models. The tensorflow IMLE-Net [8] model performs better in contrast to the tensorflow model Rajpurkar et al. [2] and torch models like Resnet101 [3] and ECGNet [7] as visualised in Figure 2. The beat level attention scores, the rhythm level attention scores and the channel level attention scores are visualized in Figure 3, which shows the reproduced IMLE-Net model performance on a patient's ECG recording who has an inferior myocardial infarction (IMI) present in leads II, III and aVF. The reproduced model was able to identify the channels responsible for this particular subtype of MI in Figure 4. These findings were successfully validated by an independent clinician as well.

Table 1. ClassWise ROC-AUC comparison of Reproduced Models

| Model | CD | HYP | MI | NORM | STTC |
|---|---|---|---|---|---|
| Resnet101 [3] | 0.8516 | 0.8417 | 0.8233 | 0.8903 | 0.8886 |
| Mousavi et al. [6] | 0.8544 | 0.8507 | 0.8468 | 0.9072 | 0.8980 |
| ECGNet [7] | 0.9002 | 0.8742 | 0.8921 | 0.9136 | 0.9105 |
| Rajpurkar et al. [2] | 0.8995 | **0.8805** | 0.8893 | **0.9245** | 0.9293 |
| IMLE-Net [8] | **0.9015** | 0.8715 | **0.9050** | 0.9215 | **0.9300** |

Table 2. Class-Wise Accuracy comparison of Reproduced Models

| Model | CD | HYP | MI | NORM | STTC |
|---|---|---|---|---|---|
| Resnet101 [3] | 86.26 | 89.69 | 81.27 | 79.84 | 85.06 |
| Mousavi et al. [6] | 86.68 | 89.29 | 81.73 | 82.38 | 84.65 |
| ECGNet [7] | 87.12 | 90.78 | 84.04 | 85.86 | 86.87 |
| Rajpurkar et al. [2] | **88.62** | 90.61 | 84.51 | 85.43 | **88.44** |
| IMLE-Net [8] | 87.74 | **90.98** | **86.35** | **86.02** | 87.40 |

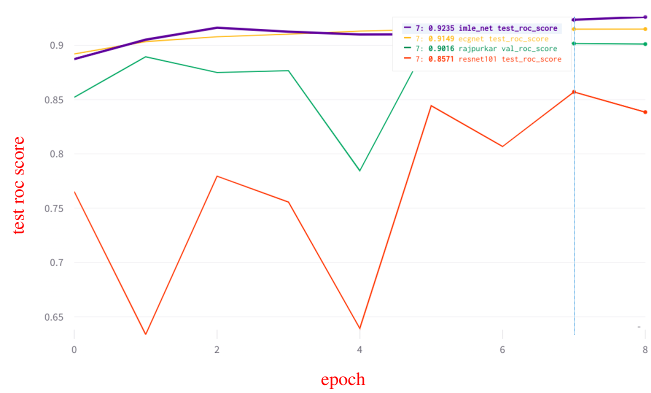



Figure 2. Comparison of Test ROC-AUC scores

Table 3. Overall Performance Comparison of Proposed and Reproduced Models

| Model | Macro ROC-AUC | | Mean Accuracy | | F1 Score (Max) | |
|---|---|---|---|---|---|---|
| | Proposed | Reproduced | Proposed | Reproduced | Proposed | Reproduced |
| Resnet101 [3] | 0.8952 | 0.8591 | 86.78 | 84.24 | 0.7558 | 0.7067 |
| Mousavi et al. [6] | 0.8654 | 0.8714 | 84.19 | 85.01 | 0.7315 | 0.7355 |
| ECGNet [7] | 0.9101 | 0.9020 | 87.35 | 87.25 | 0.7712 | 0.7704 |
| Rajpurkar et al. [2] | 0.9155 | 0.9112 | 87.91 | 87.63 | 0.7895 | 0.7884 |
| IMLE-Net [8] | **0.9216** | **0.9156** | **88.85** | **88.11** | **0.8057** | **0.7965** |


Figure 3. Visualization of normalized attention scores

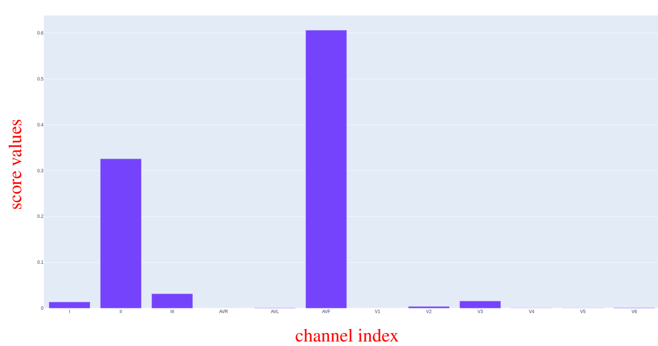



Figure 4. Channel Importance scores

## 5. Conclusion

In this paper, the model for multi-channel ECG classification is re-implemented and results of comparison between original and reproduced scores are shown. The reproduced model is trained and tested on PTB-XL dataset that outperforms the other reproduced existing models. The accuracy scores slightly deviate from the original reported values because of the limited GPU memory and overall less training of the models, to only 30 epochs in comparison to the 60 achieved in the original paper. The resulting attention scores were also visualized to represent and assess two different myocardial infarction subtypes.

In future scope, the model can be trained on not just 5 superclass but more. Also the model can be evaluated on wider range of heart diseases to get better understanding on its interpretability.

## References

[1] Barbara J Drew and Dewar D Finlay. Standardization of reduced and optimal lead sets for continuous electrocardiogram monitoring: where do we stand? *Journal of Electrocardiology*, 41(6):458–465, 2008. 1

[2] Awni Y Hannun, Pranav Rajpurkar, Masoumeh Haghpanahi, Geoffrey H Tison, Codie Bourn, Mintu P Turakhia, and Andrew Y Ng. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature medicine*, 25(1):65–69, 2019. 2, 3

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. 2, 3

[4] Shenda Hong, Cao Xiao, Tengfei Ma, Hongyan Li, and Jimeng Sun. Mina: multilevel knowledge-guided attention for modeling electrocardiography signals. *arXiv preprint arXiv:1905.11333*, 2019. 1

[5] Paul Kligfield. Value and limitations of 12-lead ecg monitoring. *Datex–Ohmeda Clin Window Web J*, 6:1–6, 2001. 1

[6] Sajad Mousavi and Fatemeh Afghah. Inter- and intra- patient ecg heartbeat classification for arrhythmia detection: A sequence to sequence deep learning approach. In *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1308–1312, 2019. 2, 3

[7] Balamurali Murugesan, Vignesh Ravichandran, Keerthi Ram, Preejith S.P., Jayaraj Joseph, Sharath M. Shankaranarayana, and Mohanasankar Sivaprakasam. Ecgnet: Deep network for arrhythmia classification. In *2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pages 1–6, 2018. 2, 3

[8] Likith Reddy, Vivek Talwar, Shanmukh Alle, Raju. S. Bapi, and U. Deva Priyakumar. Imle-net: An interpretable multi-level multi-channel model for ecg classification. In *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 1068–1074, 2021. 1, 2, 3