A New Approach to Classifying Myocardial Infarction and Cardiomyopathy Using Deep Learning

Conference Paper · November 2020

DOI: 10.1109/ICISCT50599.2020.9351386

Rashid Nasimov
Tashkent State University of Economics

Sanjar Mirzahalilov
Tashkent University of Information Technology

Nigorakhon Nasimova
Tashkent University of Information Technology

A New Approach to Classifying Myocardial Infarction and Cardiomyopathy Using Deep Learning

Rashid Nasimov
Computer Systems
Tashkent University of Information Technologies named after Muhammad al-Khwarizmi
Tashkent, Uzbekistan
rashid.nasimov@tuit.uz

Bakhodir Muminov
Fundamentals of Informatics
Tashkent University of Information Technologies named after Muhammad al-Khwarizmi
Tashkent, Uzbekistan
mbbahodir@gmail.com

Sanjar Mirzahalilov
Computer Systems
Tashkent University of Information Technologies named after Muhammad al-Khwarizmi
Tashkent, Uzbekistan
mirzahalilova86@gmail.com

Nigorakhon Nasimova
Physics
Tashkent University of Information Technologies named after Muhammad al-Khwarizmi
Tashkent, Uzbekistan
nigora.amonova@gmail.com

2020 International Conference on Information Science and Communications Technologies (ICISCT) | 978-1-7281-9969-6/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICISCT50599.2020.9351386

Abstract – In this study, for the first time, an automatic diagnostic algorithm for myocardial infarction (MI) and cardiomyopathy was developed based on ECG data obtained at different periods of time. ECG data were used taken from the ECG-VIEW II database. ECG data classification was done by the developed Convolution Neural Network (CNN) model. In order to determine the effectiveness of the proposed model, diseases with similar ECG manifestations (myocardial infarction and cardiomyopathy) were selected, which are still the most misdiagnosed among physicians today. Despite the less of data and dedicated symptoms and the similarity of symptoms of the disease on the ECG, the accuracy of the network test result became high, 91,1%. The conclusion is that with a larger database and more features, or with 12 channel ECG database it is possible to further increase the accuracy of the network, as well as increase capability of the network to differentiate other diseases. Moreover, most importantly, this model has the ability to increase the diagnostic accuracy of 1-channel smartphone ECG devices.

Keywords - CNN, ECG, deep learning, classification, myocardial infarction, cardiomyopathy.

I. INTRODUCTION

Cardiomyopathy is a heart disease characterized by an enlarged, thick, or rigid heart muscle, which usually occurs even in people who do not have diseases such as high blood pressure, valves or congenital heart disease, coronary artery disease. Although the disease is similar to myocardial infarction, the difference is that the disease is not acquired, but may be hereditary. There are several types of cardiomyopathy, among which Hypertrophic cardiomyopathy (HCM) is the most dangerous, and statistics show that one in every 100 patients with this disease dies. Also, cardiomyopathy is one of the leading causes of sudden cardiac death, especially among young people at present [1]. The number of deaths due to cardiomyopathy has been increasing in recent years [2].

Echocardiography is commonly used to diagnose this type of heart disease. This is because patients with cardiomyopathy are less likely to be diagnosed with an ECG. This is due to the fact that the disease is not associated with the appearance of specific ECG signals. Analysis and statistics show that ECG signal of cardiomyopathy patients have different patterns, sometimes similar to a healthy person and sometimes mimics myocardial infarction. Therefore, the sensitivity of ECG signal features in the diagnosis of cardiomyopathy is very low [3].

Myocardial infarction is usually a condition involving damage to the cells of the heart (myocardium) due to a lack of oxygen in the heart muscle due to ruptured blood vessels. If a myocardial infarction is not diagnosed and treated in a timely manner, the disease can lead to serious complications and even death.

In this study, both ECG and ultrasound scans are used to automatically detect and diagnose MI and cardiomyopathy. Typically, ECG data provides only the information about the presence of a disease and an ultrasound scan provides detailed information about the disease, for example about the extension or location of the disease. As, sometimes cardiomyopathy shows ECG features like myocardial infarction, diagnosing based only ECG data may leads to misdiagnosis not only by computer programs and but also by qualified physicians [4]. However, the treatment methods for these two diseases are different. In particular, patients with MI can be seriously endangered if they do not receive timely medical treatment. In short, misdiagnosis can put a patient's life at risk.

For this reason, blood tests and echocardiography are usually done to accurately diagnose cardiomyopathy and myocardial infarction. However, in cases that require urgent medical attention, especially if patient location is far from hospital and first aid must be given in ambulance, the time and ability to perform these tests are limited. The results of laboratory tests of MI and cardiomyopathy are usually determined within 8-9 hours, and there are not always enough facilities and specialists to perform an ultrasound examination. In such cases, the only diagnostic tool is the ECG device and its information.

From the above argument it clearly seen that, today medicine needs an effective method which can be differentiate cardiomyopathy and myocardial infarction by ECG data. In cardiomyopathy, the symptom of myocardial infarction, such as pathologic Q waves can appear during 30 days and then disappear, while in MI they remain unchanged after appearance. In myocardial infarction, these symptoms change over time. Other symptoms like QRS change, ST elevation are noticeably different over long periods of time [5].

It can be seen that in order to make a correct diagnosis of these diseases, it is necessary to compare the patient's previous ECG data and monitor the changes over time.

Last decade, many computer applications have been developed to automatically diagnose myocardial infarction and cardiomyopathy, almost all of which are based on analyzing changes of ECG waveforms in particular time. At the time of diagnosis, computer applications distinguish the specific symptoms of the disease and then determines the presence or absence of the disease according to certain rules. However, as noted above, since the symptoms themselves sometimes are similar for two different diseases, the program cannot accurately distinguish the disease, despite it uses any advanced set of rules.

To prevent and eliminate such cases in computer based diagnosing applications, a new approach has been developed that allows diagnose the disease by comparing ECG data obtained over several periods of time. The performance of the proposed approach was also tested and the result was highly accurate. It has been observed that this approach can differentiate cardiomyopathy from myocardial infarction with high accuracy.

The paper is organized in the following order. The first section covers the structure of the database and how to prepare the data for the training process. The second chapter provides information on the structure of the CNN model, the parameters of the layers, the process of training the network and the obtained results. The third section is a discussion section, which provides a brief analysis of the research done so far on this topic, the advantages and disadvantages of the proposed approach, and information about the activities to be done in the future.

II. Database of the Network

A. *ECG-VIEW II database*

The ECG-VIEW II database was used in this research work. ECG-VIEW II is a very large ECG database collected and developed by Korean researchers between 1994 and 2013 and is openly available for use [6]. This database is provided free of charge only to all researchers conducting research in the relevant field. The database was developed based on 710,369 ECG data from a total of 371,401 patients, with ECG data collected from each patient once to 18 times, with an average of 1.6 times. Moreover, the average interval between ECG data acquisitions was 502 ± 1,008 days. ECG data is not presented in the form of a signal, but in the form of values of special features separated from the signal by Korean researchers.

The data in the dataset consists of 5 files in database, each of which is stored in ***.csv*** format in tabular form. In addition to the main files, there are 2 additional files with comments to the data. The following Tab. I shows the names of the files in the database, and a description of the data.

TABLE I. Data description and name of the files fo the dataset

File name	The data description
Person.csv	Patient's information such as ID, age, gender, ethnicity.
Electrocardiogram.csv	Extracted ECG data, person ID and time.
Drug.csv	Drug prescribed at a certain time
Diagnosis.csv	Patient diagnosis code, patient ID and time of diagnosis at the time of ECG data taken.
Laboratory.csv	Results of laboratory analysis submitted by patients at a given time (Person ID and time of laboratory analysis).
DrugCodeMaster.csv	Prescription code and name of a drug
DiagnosisCodeMaster.csv	A mapping table between local diagnostic code and ICD-10 code

The data in the Electrocardiogram.csv file is in the following order (Tab. II):

1. Patient ID number.
2. The time the ECG was recorded.
3. The code of the hospital department where the patient is registered.
4. Information about the original source of ECG data.
5. RR interval duration (ms).
6. PR interval duration (ms).
7. QRS interval duration (ms).
8. QT interval duration (ms).
9. QTc interval duration (ms).
10. P wave axis.
11. QRS wave axis.
12. T wave axis.
13. Age-adjusted Charlson comorbidity index.

The ECG-VIEW II database is presented by Korean researchers in two different forms. The database of the first form is a sample database (ECG-VIEW_II_SAMPLE) that gives a general conclusion about the database. This database contains a small amount of ECG data and can be downloaded through the website without any registration. The second form of database is the main database, which contains 710369 ECG data, which requires filling official Agreement form and mailing it to the authors to full download. If it is accepted, a link will be provided to download the full database.

B. *The study included a dataset of ECG data from patients diagnosed only with MI and cardiomyopathy.*

In the period of this research work, only ECG data related to Myocardial Infarction and Cardiomyopathy, which were included in the ECG-VIEW II database, were extracted in our database. A sample database was used because the main database did not contain data on these two diseases. Moreover, data from patients diagnosed with myocardial infarction code I21.0, I21.1, I21.3, I21.49B, I21.9 and cardiomyopathy code I42.0, I42.1, I42.2, I42.4A, I42.5A, I42.6, I42.8, I42.9 were extracted from it.

To do this, the IDs of individuals belonging to the ICD10 code of these two diseases were extracted from the diagnosis.csv file. The range of data from 5 to 12 in the corresponding row of the isolated person ID in this

Electrocardiogram.csv file was then extracted (RR, QRS, QT, QTc, P wave axis, QRS wave axis, T wave axis) (Tab. III). Because 1: 4 data are clinical data and are not used to diagnose the disease. The information in the last column is also not used in the diagnosis, but is usually calculated after the diagnosis of the disease. Therefore, these series of data were not used.

TABLE II. VIEW THE DATA IN THE SAMPLE_EXAMPLE.CSV FILE

266455,"2001-05-28 16:43:18","O","M","759","156","140","408","467","39","-3","84","10"
266455,"2004-10-21 10:33:56","O","M","833","162","146","434","475","84","4","148","11"
266455,"2006-01-23 10:35:51","O","M","984","148","140","448","450","67","0","164","11"
266455,"2007-02-26 08:54:56","O","M","882","162","146","482","512","11","3","172","11"
266455,"2008-06-16 10:51:07","O","M","822","146","150","460","506","20","4","156","11"
266455,"2009-06-01 08:31:46","O","M","952","152","152","468","478","20","6","175","11"

TABLE III. THE DATA EXTRACTED FROM THE SAMPLE_EXAMPLE.CSV FILE (INPUT DATA OF CNN)

759	156	140	408	467	39	-3	84
833	162	146	434	475	84	4	148
984	148	140	448	450	67	0	164
882	162	146	482	512	11	3	172
822	146	150	460	506	20	4	156
952	152	152	468	478	20	6	175

As mentioned above, ECG data recorded for only 6 different times were extracted for each patient. After a successful extraction process, a 6x8 matrix datasheet was generated for each patient. The data generated for each patient was saved as a *.mat* file. The total number of ECG data was 201 for myocardial infarction and 71 for cardiomyopathy. The developed database was implemented using MatLab 2018a for training on the CNN network.

III. CONVOLUTIONAL NEURAL NETWORK

A. Network architecture

The database used in this work is not just ECG data, but special features extracted from ECG data for classifying required heart disease. Typically, machine learning techniques are useful for classifying and identifying diseases using these extracted special features. Because machine learning algorithms cannot directly extract raw ECG signals, these methods require first the extracting of specific features and then the data classification process is done based on those features.

However, there are two disadvantages to using machine learning to successfully achieve targeted work. Firstly, the extracted special features in the developed database are not sufficient to diagnose MI or cardiomyopathy: data about ST elevation, J point elevation, T inversion in each of the 12 leads are not given. Secondly, the data is presented in the form of a two-dimensional (2D) matrix rather than one-dimensional and machine learning methods are not adapted to work with 2D data. This means that in order to apply the proposed algorithm in machine learning, the database needs to be redesigned. This requires additional calculations not only in the process of training the network, but also in the subsequent use of the developed method in real life.

The most optimal way to overcome the above disadvantages is to use deep learning methods, especially convolutional neural network (CNN) method. Although this deep learning approach is a specialized model for working with images [7] as well as it has been proven in many scientific studies that it can work well with 2D data [8], not just images [9, 10, 11, 12, 13]. Any data in the 2D form of matrix provided to the input layer of the CNN architecture can be considered as input data. But this requires the use of a special function that allows you to read the data when the data is in a format other than the image at the time of the formation of the Datastore. In our work, also, because the data is in the form of the *.mat* files, a special function was developed that allows the data in the *.mat* files to be read when forming the datastore.

The data in the database is divided into three parts in a ratio of 8:1:1, training data, validation data and test data respectively. In this distribution, the test data (one in 10 of the data) was randomly selected first, and then the remaining data were randomly divided in an 8: 1 ratio.

In the proposed work, a total of 16 layers of CNN networks were constructed, consisting of different layer configurations. A brief description of CNN network layers and their parameters is shown in Tab. IV.

TABLE IV. PARAMETERS OF PROPOSED CNN MODEL

Name of the layer	Layer parameter
Input	6x8
Convolution 2d	Filter size = [1 1], Number of filters = 256, Number of strides =1
Batch normalization	
Leaky RELU	
Convolution 2d	Filter size = [2 2], Number of filters = 64, Number of strides =1,
Batch normalization	
LEAKY RELU	
Convolution 2d	Filter size = [2 1], Number of filters = 2, Number of strides = 1
Batch normalization	
Leaky RELU	
Convolution 2d	Filter size = [2 2], Number of filters = 8, Number of strides = 1
Batch normalization	-
Leaky RELU	-
Fully connected layer	2
SoftMax layer	
Classification output	Crossentropyex with 'A' and 1 other class

B. *Training parametrs and results*

Adam optimizer was used for training of the CNN network. The network training parameters are given in the Tab. V below. 10 cross fold validation method was used to prevent overfitting in network training. The training and validation results are shown in Fig. 1.

TABLE V. THE PARAMETRS OF NETWORK TRAINING

Parameter name	**Parameter value**
SquaredGradientDecayFactor	0,95
MaxEpochs	30
MiniBatchSize	8
InitialLearnRate	0,001
GradientDecayFactor	0,95
L2Regularization	0,001
Epsilon	0,001

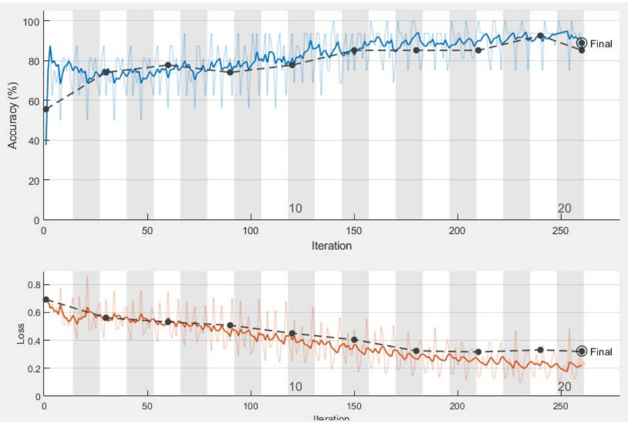

Fig. 1. The network training process on CNN

Tab. VI shows the sum of confusion matrix calculated from the test results.

TABLE VI. THE CONFUSION MATRIX OF TESTING

	Cardiomyopathy	**Myocardial infarction**
Cardiomyopathy	56	10
Myocardial infarction	14	190

Using Tab. VI, we calculated the parameters which evaluate the performance of the network using the following formulas.

*Specificity = TN/(TN+FP) *100%*

*Precision = TP/(TP+FP) *100%*

*Sensitivity = TP/(TP+FN) *100%*

F1 Score = 2(Sensitivity * Precision) / (Sensitivity + Precision) *100%*

*Accuracy = (TP+TN)/(TP+FP+FN+TN) *100%*

The obtained results are presented in Tab. VII.

TABLE VII. THE STATISTICAL PARAMETERS VALUES WICH SHOW PERFORMANCE OF THE NETWORK

Statistical Parameters Names	Cardiomyopathy (%)	Myocardial infarction (%)
Specificity	95	80
Precision	85	93
Sensitivity	80	95
F1-score	82,4	94
Accuracy	91,1	91,1

Even the 10-fold validation showed an average accuracy of 91.1% in classification two different diseases, the accuracy of the test after some training processes achieved 100%.

Discussion

So far, many algorithms have been developed for the automatic diagnosis of cardiomyopathy and myocardial infarction. Most of existing algorithms are only able to classify myocardial infarction according to location [1, 8, 14], or to distinguish only cardiomyopathy. Some studies have proposed algorithms that divide ECG signals into MI and healthy [2, 15, 16] as well as cardiomyopathy and healthy groups. Moreover, a lot of scientific work has been done in the last thirty years on the automatic detection of MI, and their advantages and disadvantages have been very well analyzed [17].

However, in a small amont of studies, we can see that an algorithm for automatic differentiation of myocardial infarction and cardiomyopathy has been proposed. An algorithm to distinguish between 2 different types of cardiomyopathy and MI, as well as ECG from a healthy person using k-NN was proposed by [18]. 5 different heart diseases were automatically determined, including MI and cardiomyopathy, with an average accuracy of 90.3% [19].

In the study we proposed, very less ECG features were used in contrast to the above studies. Also, the proposed algorithm achieved good results using fewer features and less ECG data. Moreover, the ECG features used in the proposed study are the same ECG features used in many applications used in medicine today. It means that our work can be easily implemented in current medical expert systems.

The main reason for the good results taken from this study is the development of a new approach to ECG-based diagnostics. In general, human ECGs are radically different from each other, even ECG may not be observed a standard ECG waveform during 24 hours. The ECG changes continuously as the disease progresses [20, 21, 22]. Interestingly, no computer program today is able to draw conclusions by comparing ECGs over multiple periods.

Another advantage of our work is that the data of each patient for several periods were stored in the form of a single matrix, that is, a single integrated data was formed. However, many researchers store ECG data recorded by a single patient at different periods of time in the database as different data in order to enlarge the database. In some studies, even a single complete ECG signal is broken down and stored in a database as different data. In such cases, the network becomes overfitting even the test accuracy is high. That is, the experiment results have shown that when a network is tested with similar data, its accuracy is much higher. Such cases do not occur in the proposed work.

Another advantage of the proposed approach is that proposed method can increase the efficiency of the automatic diagnostic process by using on 1 or 3 leads Smartphone ECGs. Typically, 1 or 3 leads ECG data does not contain enough features to diagnose MI. However, the use of changes in ECG data in automatic diagnosis dramatically increases the accuracy of the diagnosis.

IV. Future Work

Due to the problem of forming the required database and the lack of ECG data, the ECG-VIEWII database was used. It has been observed that this data is not sufficient to diagnose some diseases. If the database consists of 12-lead ECG data used at least 5-6 time periods, better results can be achieved and more diseases can be automatically detected. Most importantly, the reliability of the results obtained will be high.

V. Conclusion

In this paper, a new unique approach to differentiating MI and cardiomyopathy using the CNN model is proposed. The proposed method allows first draw conclusions by comparing data for a certain period of time. That's why the accuracy of the network is high. Unlike the PTB database, the ECG-View II database was used to differentiate MI and cardiomyopathy because it consisted of only extracted ECG feature, not just 12-lead ECG. In short, this approach opens a new way in automatic diagnostics of heart disease.

References

[1] Q. A. Rahman, G. T. Larisa, M. Kongkatong and other, “Utilizing ECG-based Heartbeat Classification for Hypertrophic Cardiomyopathy Identification”, IEEE Trans Nanobioscience. 2015 July; 14(5): pp. 505–512.

[2] Quazi Abidur Rahman, Larisa G. Tereshchenko, Matthew Kongkatong and others “Identifying Hypertrophic Cardiomyopathy Patients by Classifying Individual Heartbeats from 12-lead ECG Signals”, IEEE Int Conf Bioinformatics Biomed, 2014 Nov; pp. 224-229

[3] C. McGuinty, P. Dorian, K. Connelly, R. Chan, A. Adler, H. Rakowski, “How accurate is the ecg in screening for hypertrophic cardiomyopathy?” T Aves, C Landry, Canadian Journal of Cardiology Volume 34 2018, S152

[4] F Luzza, S Carerj, G Oreto “Hypertrophic cardiomyopathy with persistent ST segment elevation simulating acute myocardial infarction”, BMS journal, Heart, Volume 90, Issue 4, 90(4): 380, 2004

[5] Shamim, W., Yousufuddin, M., Cicoria, M., Gibson, D. G., Coats, A. J., & Henein, M. Y. (2002). Incremental changes in QRS duration in serial ECGs over time identify high risk elderly patients with heart failure. *Heart (British Cardiac Society)*, *88*(1), pp. 47–51.

[6] http://ecgview.org/default.asp

[7] Mohit Sewak, Md. Rezaul Karim , Pradeep Pujari “Practical Convolutional Neural Networks”, Publisher Packt, pp. 218, 2018

[8] B. Muminov, R. Nasimov, S. Mirzahalilov, N. Sayfullaeva and N. Gadoyboyeva, "Localization and Classification of Myocardial Infarction Based on Artificial Neural Network," 2020 Information Communication Technologies Conference (ICTC), Nanjing, China, 2020, pp. 245-249

[9] Muminov Bakhodir Boltaevich, Nasimov Rashid Hamid ogli, Gadoyboyeva Nigora Soibjon qizi and Mirzahalilov Sanjar Serkabay ogli. “Estimation affects of formats and resizing process to the accuracy of convolutional neural network.” 2019 International Conference on Information Science and Communications Technologies, 2019, pp. 1-5.

[10] A. Turgunov, K. Zohirov, A. Ganiyev and B. Sharopova, "Defining the Features of EMG Signals on the Forearm of the Hand Using SVM, RF, k-NN Classification Algorithms," 2020 Information Communication Technologies Conference (ICTC), Nanjing, China, 2020, pp. 260-264,

[11] R. M. Fazliddinovich and B. U. Abdumurodovich, "Parallel processing capabilities in the process of speech recognition," 2017 International Conference on Information Science and Communications Technologies (ICISCT), Tashkent, 2017, pp. 1-3.

[12] Muhammadjon Musaev, Ilyos Khujayorov, Mannon Ochilov. Image Approach to Speech Recognition on CNN. ISCSIC 2019: Proceedings of the 2019 3rd International Symposium on Computer Science and Intelligent Control. September 2019, No. 57 pp. 1–6.

[13] M. Musaev, I. Khujayorov and M. Ochilov, "The Use of Neural Networks to Improve the Recognition Accuracy of Explosive and Unvoiced Phonemes in Uzbek Language," 2020 Information Communication Technologies Conference (ICTC), Nanjing, China, 2020, pp. 231-234

[14] H. WaiLuiabKing LauChow, “Multiclass classification of myocardial infarction with convolutional and recurrent neural networks for portable ECG devices”, Informatics in Medicine Unlocked,Volume 13, 2018, pp. 26-33,

[15] H. Hamidi and A. Daraei, “Analysis and evaluation of techniques for myocardial infarction based on genetic algorithm and weight by SVM,” Information Systems & Telecommunication, p. 85, 2016.

[16] V. Seenivasagam and R. Chitra, “Myocardial infarction detection using intelligent algorithms,” Neural Network World, vol. 26, no. 1, pp. 91–110, 2016.

[17] S.Ansari, H. Andersson, “A review of automated methods for detection of myocardial ischemia and infarction using electrocardiogram and electronic health records”, IEEE reviews in biomedical engineering · October 2017

[18] A. Mohsin, O. Faust, “Automated characterization of cardiovascular diseases using wavelet transform features extracted from ecg signals”, Journal of Mechanics in Medicine and Biology, Vol. 19, No. 01, 1940009 (2019),

[19] R.K. Tripathy, L.N. Sharma, and S. Dandapat,” A new way of quantifying diagnostic information from multilead electrocardiogram for cardiac disease classification”, Healthc Technol Lett. 2014 Oct; 1(4): pp. 98–103

[20] Chen X, Hu Y, Fetics Bj, Berger Rd, Trayanova Na. 2011. unstable qt interval dynamics precedes vt onset in patients with acute myocardial infarction: a novel approach to detect instability in qt interval dynamics from clinical ecg. circ. arrhythm electrophysiol. 4, pp. 858–866.

[21] Potter SL1, Holmqvist F, Platonov PG, Steding K, Arheden H, Pahlm O, Starc V, McKenna WJ, Schlegel TT.,”Detection of hypertrophic cardiomyopathy is improved when using advanced rather than strictly conventional 12-lead electrocardiogram”, J Electrocardiol. 2010 Nov-Dec; 43(6):713-8.

[22] R. Nasimov, B. Muminov, S. Mirzahalilov, N. Nasimova “Algorithm of Automatic Differentiation of Miocardial Infarction from Cardiomyopathy based on Electrocardiogram”, The 14th IEEE International Conference on Application of Information and Communication Technologies (AICT), 07-09 October 2020y, Tashkent