

# Binary Classification of Heart Failures Using $k$ -NN with Various Distance Metrics

Yevhenii Udovychenko, Anton Popov, Illya Chaikovsky

**Abstract**—Magnetocardiography is a sensitive technique of measuring low magnetic fields generated by heart functioning, which is used for diagnostics of large number of cardiovascular diseases. In this paper,  $k$ -nearest neighbor ( $k$ -NN) technique is used for binary classification of myocardium current density distribution maps (CDDM) from patients with negative T-peak, male and female patients with microvessels (diffuse) abnormalities and sportsmen, which are compared with normal control subjects. Number of neighbors for  $k$ -NN classifier was selected to obtain highest classification characteristics. Specificity, accuracy, precision and sensitivity of classification as functions of number of neighbors in  $k$ -NN are obtained for classification with several distance measures: Mahalanobis, Cityblock, Euclidean and Chebyshev. Increase of the accuracy of classification for all groups up to 10% was obtained using Cityblock distance metric in binary  $k$ -NN classifier with 19 - 27 neighbors, comparing to other metrics. Obtained results are acceptable for further patient's state evaluation.

**Keywords**—magnetocardiography; current density imaging; current density distribution map;  $k$ -NN classification, negative T-peak, heart failure diagnostics, Mahalanobis distance, Cityblock distance, Euclidean distance, Chebyshev distance

## I. INTRODUCTION

MAGNETOCARDIOGRAPHY (MCG) is a modality of heart activity study using the measurement of magnetic field emerged during heart functioning. It can be performed using a superconducting quantum interference device sensor which converts magnetic flux to voltage, and is the most sensitive sensor to detect magnetism [1]. Magnetic measurements are not only non-invasive, but also contactless. Thereby the occurrence of artifacts due to insufficient reliability of contacts between electrodes and skin is prevented. MCG is more sensitive to intra- and extracellular activation currents than electrocardiography (ECG) [2-3]. The noninvasive magnetocardiographic mapping is used for diagnostics and analysis of ischemic heart disease [2], Wolf-Parkinson-White syndrome [4] and other diseases accompanied by lesions of current flow in the heart muscles.

One of the approaches of studying magnetic field in human

heart is analysis of current density distribution map (CDDM) over the myocardial surface. This is the approach to noninvasive measurement of the electrical current density inside a conductive sample. Current density imaging measures the local magnetic field vector generated by the current flowing inside the tissue and uses a vector curl operation to calculate the current density. Since current density imaging can be performed contactlessly, it is inherently noninvasive (when compared with multiple-electrode measurements) and does not suffer from the limitations of ECG technique.

Therefore, it is possible to accurately measure the detailed current density information inside the body using current density imaging followed by solving of inverse problem [5]. Furthermore, current flow information can be calculated from the current density data using streamline analysis [6].

Heretofore, great efforts were made mostly for getting informative representation of current density distribution in myocardium [7-8]. Attempts of 3-D modeling and three-dimensional imaging of cardiac electrophysiological activity based on current density images were made [9]. Fairly good results have been achieved in classification of current density vector maps, including classification, based on clusters analysis for coronary artery disease and ischemic heart disease detection [10], [11]. Nevertheless, classification of MCG data, especially CDDMs is on its early stage.

Aim of the present study is the development of the classifier between normal subjects and patients having heart failures using  $k$ -nearest neighbor algorithm, and analysis of its characteristics, such as accuracy, precision, sensitivity and specificity.

## II. $k$ -NN CLASSIFICATION OF CDDMS

### A. Forming CDDMs using MCG Data

MCG is measured spatially over the patient's chest resulting in data readings from the square grid on the thorax. For spatial data capture during MCG registration, the observation points are selected as the intersection nodes of a square grid, which has binding to anatomical landmarks on the thorax selected by the medical doctor. Spatial resolution is the same for each subject. Model displaying the orientation of CDDM with respect to patient's body is shown in Fig. 1.

Instant distribution maps of magnetic field are constructed using two-dimensional interpolation and synchronous averaging of MCG measurements in time and over the surface. Instant CDDM can be obtained from MCG maps after converting the map of magnetic field distribution by solving inverse problem [3].

This paper is an extended version of paper "k-NN Binary Classification of Heart Failures Using Myocardial Current Density Distribution Maps" [14], presented in Signal Processing Symposium, Debe, Poland, June 10 - 12, 2015.

Y. Udovychenko, A. Popov are with the Physical and Biomedical Electronics Department of National Technical University of Ukraine "Kyiv Polytechnic Institute", Kyiv, Ukraine,  
(email: yevhenii.udovychenko@yahoo.gr, popov.kpi@gmail.com).

I. Chaikovsky is with the Glushkov Institute of Cybernetics of NAS of Ukraine, Kyiv, Ukraine, (email: illya.chaikovsky@gmail.com).

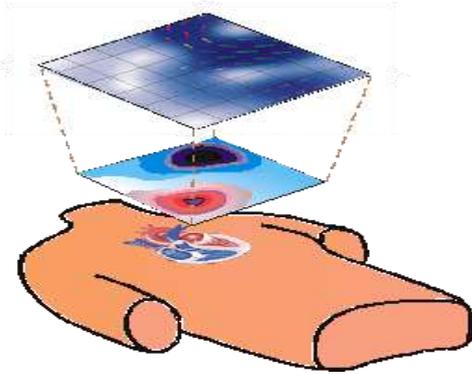


Fig. 1. Model describing MCG imaging process

Each patient belongs to a particular group by heart state. Each of these patients groups is characterized by CDDMs of specific structure.

The structure of CDDM in our understanding is the mutual location of zones with high and low density of currents.

Data obtained with MCG is an array of pairs of numbers. Each pair of numbers represents the coordinates of current density radius vector in each point. For each point, magnitude and angle to horizontal axis of current density vector is found. Thus each CDDM can be represented as normalized grayscale image with dimensions of  $M \times N$  pixels, where white color corresponds the highest current density, and black responds the lowest current density over the patient chest. From the calculated data, vectors representing the direction of current can be represented as well (Fig. 2). In this study, each CDDM has resolution of  $10 \times 10$  pixels; two-dimensional interpolation can be applied for better visualization.

CDDMs usually are calculated for several time instants with 10 ms period during  $T$  wave of electrocardiogram  $QT$  interval. This enables to correlate ECG data with equivalent CDDM characteristics [12]

### B. Feature Extraction

In this study each CDDM is divided into 4 equal parts (quarters), as depicted in Fig. 3.

Each element of CDDM map can be described by two parameters: brightness, which corresponds to the current density in particular point, and angle of magnetic field vector in each point. Given these two data sets, the following characteristics were calculated for each quarter: mean value, variance, kurtosis and skewness of the current density and magnetic field vector separately. In the result, 32 features were calculated for each map.

### C. $k$ -Nearest Neighbor Algorithm

One of the methods for pattern classification is the  $k$ -nearest neighbors ( $k$ -NN) technique. It classifies each unlabeled object according to the majority class label of its  $k$ -nearest neighbors in the training set.

In [13] it was shown that using different distances for  $k$ -NN classification decreased the error rates for different classification problems such as face recognition, spoken letter recognition, text categorization etc. It was also shown in the same study that  $k$ -NN classifier with correctly chosen distance metric shows better results even comparing to Support Vector Machines used for the same classification tasks. The result of

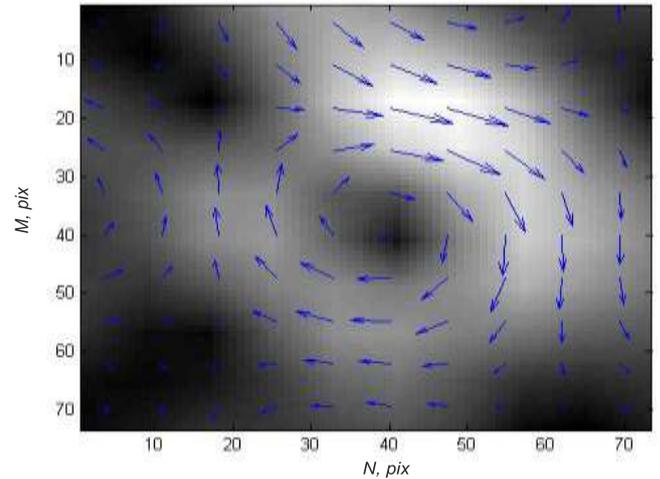


Fig. 2. Example of current density distribution map

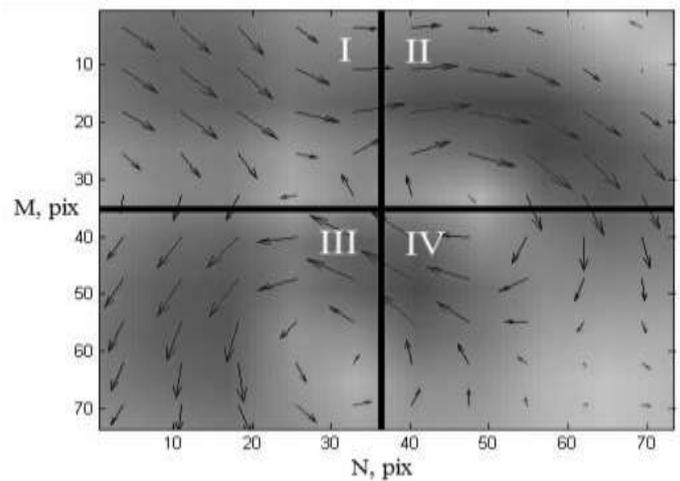


Fig. 3. Current density distribution map divided into 4 parts

$k$ -NN classification depends significantly on the metric used to compute distances between different feature vectors.

In our previous study [14] we used  $k$ -NN classifier with Euclidian metric. In this study we tried to improve classification results by employing other available metrics: Mahalanobis, Cityblock and Chebychev to the task of classification between CDDMs containing heart failures and without them.

Let's consider  $X$  as 1 by 32 feature vector of classified CDDM, and  $Y$  as feature vector of each CDDM in the training set. In our study binary classifiers with 4 different distance metrics were constructed. For classifier with Euclidian metric distance between two points  $X_s$  and  $Y_t$ , which coordinates are values of  $X$  and  $Y$ , respectively, is defined as follows:

$$d_{st}^2 = (x_s - y_t)(x_s - y_t)' \quad (1)$$

For Mahalanobis metric this value is defined as follows:

$$d_{st}^2 = (x_s - y_t)C^{-1}(x_s - y_t)', \quad (2)$$

where  $C$  is the covariance matrix.

For Cityblock metric:

$$d_{st} = \sum_{j=1}^n |x_{sj} - y_{tj}|, \quad (3)$$

where  $n$  is the length of vectors  $X$  and  $Y$ , In our case  $n = 32$  - number of features.

For Chebychev metric:

$$d_{st} = \max_j \{|x_{sj} - y_{tj}|\} \tag{4}$$

### III. EXPERIMENTAL RESULTS

#### A. Experiment Setup and Data Collection

For the present study 800 CDDMs from 5 groups of patients with different heart states were used: healthy volunteers, patients with negative  $T$ -peak, male and female patients with microvessels (diffuse) abnormalities and sportsmen. Sportsmen have been defined as separate group, since in healthy condition they have differences in heart anatomy and activity in rest, because of prolonged extensive training resulted in changes in cardiovascular activity.

Four different binary classifiers were developed. Each classifier had 2 possible categorical variables, which meant the group of patient's state: normal or disease. For each classifier half of CDDMs from two groups were used for training set, and all the rest CDDMs were used for testing set. The same number of CDDMs were used to examine the quality of classification.

To evaluate classification quality, the following characteristics were used: sensitivity ( $TPR$ ), specificity ( $SPC$ ), precision ( $PPV$ ) and accuracy ( $ACC$ ). These values are defined as follows:

$$TPR = \frac{TP}{P} \tag{5}$$

$$SPC = \frac{TN}{N} \tag{6}$$

$$PPV = \frac{TP}{TP + FP} \tag{7}$$

$$ACC = \frac{TP + TN}{P + N} \tag{8}$$

where  $TP$  – number of CDDMs of persons with heart failure, identified correctly;  $TN$  – number of CDDMs of persons without heart failure identified correctly;  $FP$  – number of CDDMs of persons with heart failure, identified incorrectly;  $P$  – total amount of CDDMs of persons with heart failure,  $N$  – total amount of CDDMs of persons without heart failure.

#### B. Dependence of classification results on different metrics

After  $k$ -NN classifier has been constructed, its quality was examined by classifying of CDDMs from test sample. This operation was made for different numbers of neighbors in a range from 1 to 35 with the purpose of finding the optimal number of neighbors to obtain best performance characteristics. For each number of neighbours, parameters mentioned above were calculated for evaluating the quality of classification.

Figure 4 shows the dependence of accuracy on number of neighbors for different metrics for group “Negative\_T”. As we can see, using a Cityblock metric gives the best result for almost all numbers of neighbors, showing the value of accuracy higher up to 4% comparing to Euclidian metric, used in our previous study [14]. It gives an opportunity to obtain accuracy of classification up to 90%. Chebychev metric shows almost the same result as the Euclidian, but when the number of neighbors is larger than 20, the accuracy shown by  $k$ -NN classifier with

Chebychev metric is up to 3% worse than using Euclidian metric, and, respectively, up to 6% worse result than using Cityblock metrics. The worst result was obtained using Mahalanobis metric. Accuracy of classification using  $k$ -NN classifier with Mahalanobis metric lies in a range of 40-60% and cannot compete with classifiers employing other metrics.

Figures 5 and 6 shows the dependence of sensitivity and specificity on number of neighbors, respectively. From obtained results we can see that  $k$ -NN classifiers with Euclidian and Chebychev metrics shows almost equal sensitivity and specificity for all number of neighbors. Cityblock metrics gives better values than both Euclidian and Cityblock metrics.  $k$ -NN classifier with Mahalanobis metric shows much worse result, than using other metrics, and does not depend on number of neighbors strongly. It should be mentioned that for all metrics sensitivity is much higher than specificity. In case of  $k$ -NN classifier with Cityblock metric sensitivity even lies in a range of 95-100% when number of neighbors is larger than 5. It means that all metrics gives better result in classifying CDDMs of persons with failure, but sometimes misclassifies CDDMs of healthy persons.

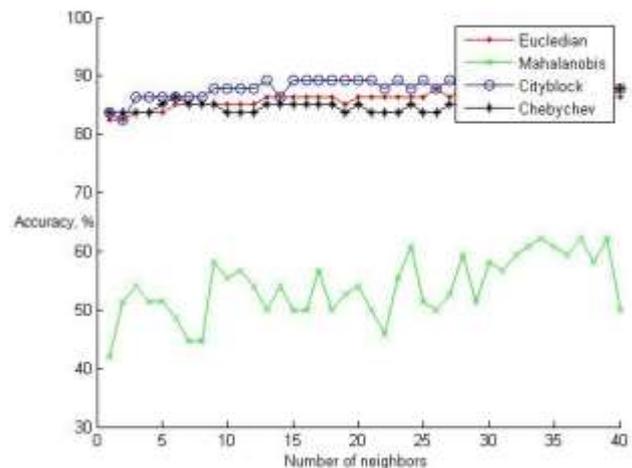


Fig. 4. Dependence of accuracy on number of neighbors for different distance metrics for group “Negative\_T”

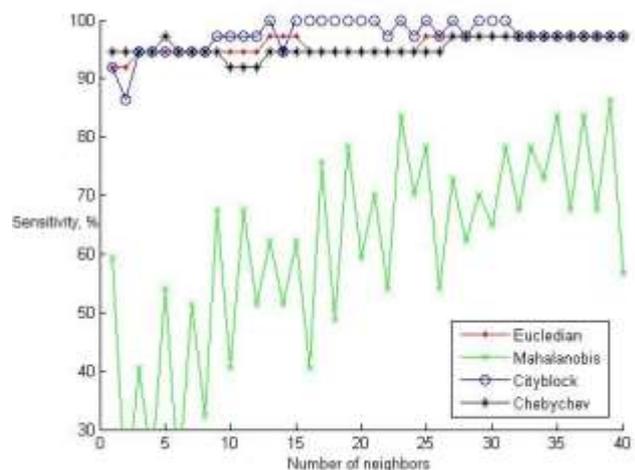


Fig. 5. Dependence of sensitivity on number of neighbors for different distance metrics for group “Negative\_T”

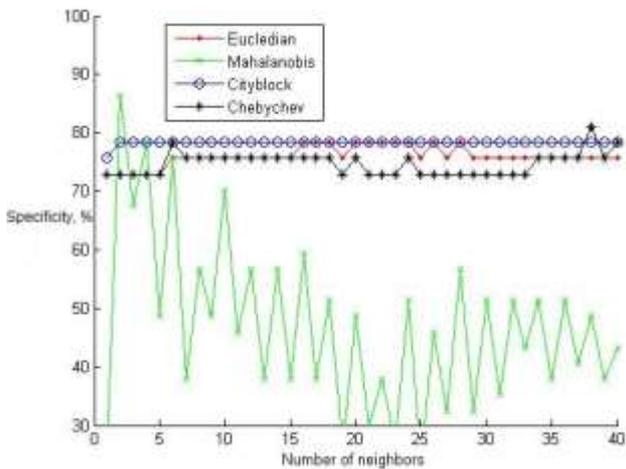


Fig. 6. Dependence of specificity on number of neighbors for different distance metrics for group "Negative\_T"

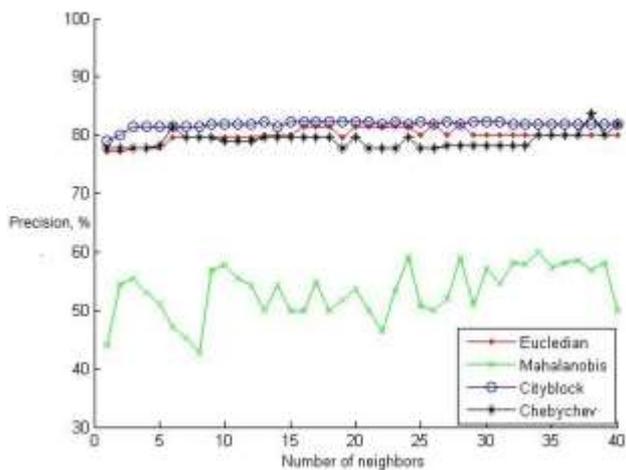


Fig. 7. Dependence of precision on number of neighbors for different distance metrics for group "Negative\_T"

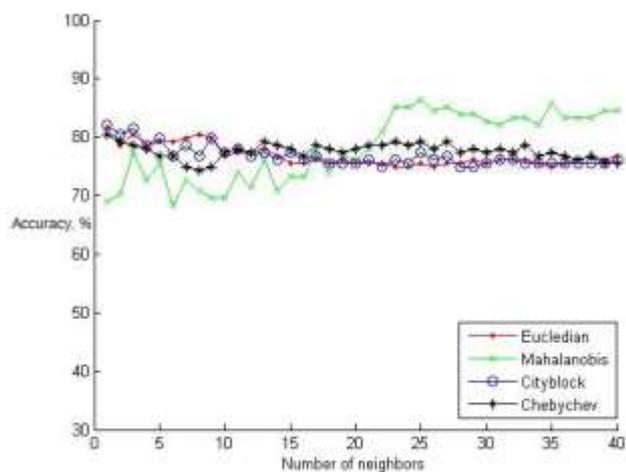


Fig. 8. Dependence of accuracy on number of neighbors for different distance metrics for group "Sport"

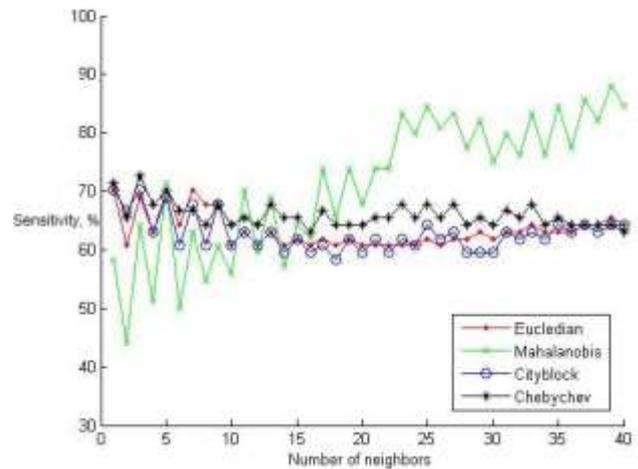


Fig. 9. Dependence of sensitivity on number of neighbors for different distance metrics for group "Sport"

Figure 7 shows the dependence of precision on number of neighbors. The obtained result confirms that  $k$ -NN classifier with Cityblock metric shows the best result for almost all number of neighbors and Mahalanobis metric shows much worse result than the other metrics.

Figure 8 shows the dependence of accuracy on number of neighbors for different metrics for group "Sport". As it can be seen from the figure,  $k$ -NN classifiers with Euclidean, Cityblock and Chebychev metrics shows almost the same result for all numbers of neighbors. Chebychev metric is a slightly better when the number of neighbors is larger than 10

- it shows accuracy larger up to 3% than Euclidean and Cityblock metrics, but shows worse result when number of neighbors is less than 10. For Euclidean, Cityblock and Chebychev metrics we can conclude that the value of accuracy does not depend strongly on number of neighbors and lies in a range of 75 - 80%.  $k$ -NN classifier with Mahalanobis metric shows more interesting result: accuracy of classification is much less than for other metrics mentioned above when number of neighbors is less than 17, but it shows much better result when number of neighbors is larger than 22. In this case the value of accuracy lies in the range of 81 - 86%, which means increasing of classification accuracy up to 8% comparing to other metrics.

Figures 9 and 10 shows the dependence of sensitivity and specificity on number of neighbors, respectively. As we can see, classifiers with Euclidean, Cityblock and Chebychev metrics shows almost the same specificity, only when the number of neighbors is less than 10, specificity of classification using Chebychev metric is up to 10% less than for Euclidean and Cityblock metrics. In the same time,

classifier with Chebychev metric shows better sensitivity than the one with Euclidean or Cityblock metrics, for almost all numbers of neighbors. It should be mentioned that classifiers with each of these 3 metrics have much larger value of specificity than sensitivity, which means that they are better in classifying CDDMs of persons without failure, but could make mistakes classifying CDDMs of patients with failure. As we can see from Fig. 8, using Mahalanobis metrics in  $k$ -NN classifier

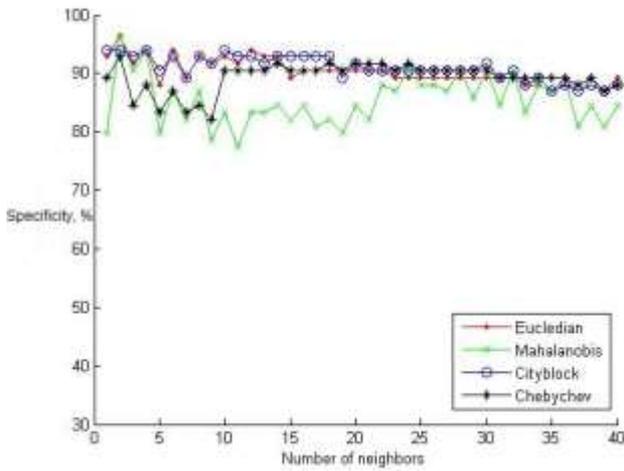


Fig. 10. Dependence of specificity on number of neighbors for different distance metrics for group "Sport"

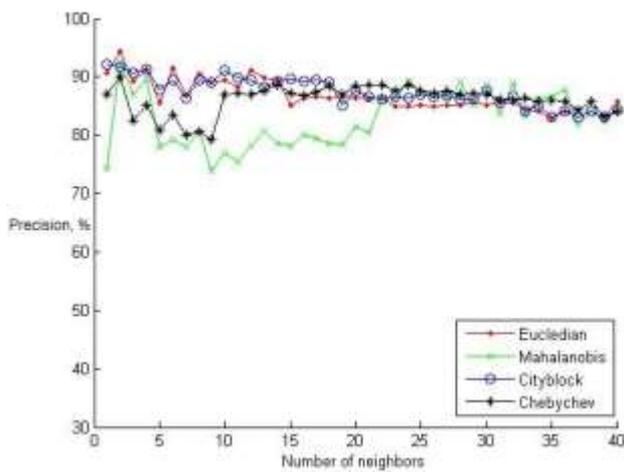


Fig. 11. Dependence of precision on number of neighbors for different distance metrics for group "Sport"

can solve this problem by increasing sensitivity up to 20% comparing to classifiers with other metrics, when the number of neighbors is larger than 15. Thus Mahalanobis metric gives better result in classifying CDDMs of patients with failure and this is the reason of such a good accuracy of classification using classifier with this metric.

Figure 11 shows the dependence of precision on number of neighbors. As we can see, all metrics give almost the same result when number of neighbors is larger than 22, but value of precision is much smaller when number of neighbors is less than 22 for Mahalanobis metric and less than 10 for Chebychev metric.

Figure 12 shows the dependence of accuracy on number of neighbors for different metrics. As it can be seen, all the metrics result in almost same value of accuracy for all numbers of neighbors, increasing with number of neighbors increase. Slightly better than other metrics, result shows Cityblock metric, as in case of classifying CDDMs for group "Negative\_T". Classifier with Cityblock metric increases accuracy of classification up to 4% giving 85-90% accuracy when number of neighbors is larger than 10.

Figure 13 shows the dependence of accuracy on number of neighbors for different metrics. As it can be seen, all metrics

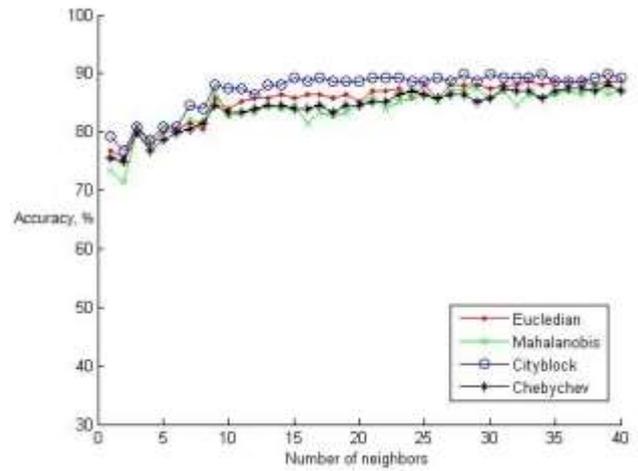


Fig. 12. Dependence of accuracy on number of neighbors for different distance metrics for group "Diff\_F"

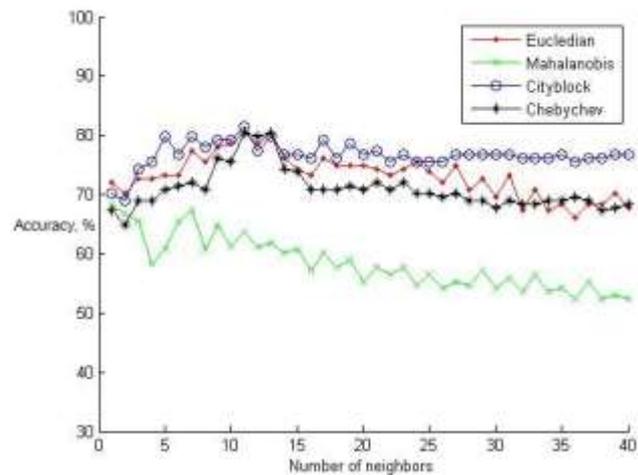


Fig. 13. Dependence of accuracy on number of neighbors for different distance metrics for group "Diff\_M"

give almost same value of accuracy, except classifier with Mahalanobis metric, which shows accuracy up to 20% less than other classifiers. For all numbers of neighbors classifier with Cityblock metric gives the best result. But the highest values of accuracy, shown by Cityblock, Chebychev and Euclidean metrics are almost equal and are observed for the same numbers of neighbours: 10-15.

#### IV. CONCLUSIONS AND DISCUSSION

Summing up the obtained results, it could be concluded that using Cityblock distance metric in binary  $k$ -NN classifier increases the accuracy of classification for all groups of heart failures up to 10% comparing to Mahalanobis, Chebychev and Euclidean metrics. As far as for Cityblock metric the form of the curve that characterizes dependence of accuracy on number of neighbors is almost the same as for Euclidean metric, conclusions about optimal number of neighbors are the same for this classifiers. This number lies in a range of 19–27 neighbors.  $k$ -NN classifier with Chebychev metric shows almost the same result, as with Euclidean one and is worse for classification than classifier with Cityblock metric. Classifier with Mahalanobis metric has much lower accuracy of classification than classifiers with other metrics for all groups

of patients, except “Sport”. For this group accuracy of classification, shown by classifier with this metric, is significantly (up to 10%) higher than accuracy of classification of others classifiers, for number of neighbors larger than 22. This makes Mahalanobis metric suitable only for some specific classification purposes.

Obtained results may be improved by CDDMs pre-processing and modifying of other parameters of  $k$ -NN classifier. The proposed classification may improve the accuracy of patient’s state evaluation and may be used in conjunction with other techniques, for example, correlation analysis.

#### REFERENCES

- [1] Hyun Kyoon Lim *et al.*, “Magnetocardiogram difference between healthy subjects and ischemic heart disease patients” in *IEEE Transactions on Magnetics*, Vol. 45, June 2009, pp. 2890-2893.
- [2] Tsukada K. *et al.*, “Magnetocardiographic mapping characteristic for diagnosis of ischemic heart disease” in *Computers in Cardiology 2000*, Cambridge, MA, Sept. 2000, pp. 505-508.
- [3] Chaikovsky I. *et al.*, “Magnetocardiography in clinical practice: algorithms and technologies for data analysis” in *Medical Science 3-4*, June 2011, pp. 21-38.
- [4] Jazbinsek V. *et al.*, “Magnetocardiographic localization of accessory conduction pathway in patients suffering from WPW syndrome”, *Computers in Cardiology 1995*, Vienna, Austria, Sept. 1995, pp. 417-420.
- [5] Romanovych S. *et al.*, “Imaging of heart biomagnetic sources by current lines in a plane using the magnetic moments method” in *Proceedings of the EMBEC99*, Part 1, 1991, pp. 410-411.
- [6] Ogata K. *et al.*, “Projecting cardiac-current images onto a 3-D standard heart model” in *Engineering in Medicine and Biology Society*, 2003, Vol. 1, Sept. 2003, pp. 517-520.
- [7] Kobayashi K. *et al.*, “Visualization of the Current-Density Distribution for MCG With WPW Syndrome Patients Using Independent Component Analysis in *IEEE Transactions on Magnetics*, Vol. 40, July 2004, pp. 2970-2972.
- [8] Brockmeier K. *et al.*, “Magnetocardiography and 32-Lead Potential Mapping: Repolarization in Normal Subject During Pharmacologically Induced Stress”, *Journal of Cardiovascular Electrophysiology*, Vol. 8 No. 6, June. 1999, pp. 615-626.
- [9] Leder U. *et al.*, “Noninvasive Biomagnetic Imaging in Coronary Artery Disease Based on Individual Current Density Maps of the Heart”, *International Journal of Cardiology* 64 (1998), pp. 417-420.
- [10] Fainzilberg L. *et al.*, “Sensitivity and specificity of magnetocardiography, using computerized classification of current density vectors maps, in ischemic patients with normal ECG and echocardiogram”, *International Congress Series 1300* (2007), pp. 468-471.
- [11] Chaikovsky I. *et al.*, “Detection of coronary artery disease in patients with normal or unspecifically changed ECG on the basis of magnetocardiography”, *Proceedings of the 12-th International Conference on Biomagnetism*(2000), pp. 565-568.
- [12] Udovychenko Y., *et al.*, “Current Density Distribution Maps Threshold Processing” in *2014 IEEE 34th International Scientific Conference on Electronics and Nanotechnology (ELNANO)*, Apr. 2014, pp. 313 – 315.
- [13] Kilian Q. *et al.*, “Distance Metric Learning for Large Margin Nearest Neighbor Classification”, *Journal of Machine Learning Research* 10 (2009), Sept. 2009, pp. 207-244.
- [14] Udovychenko Y., *et al.*, “k-NN Binary Classification of Heart Failures Using Myocardial Current Density Distribution Maps”, *Signal Processing Symposium (SPSymo)*, June 2015, pp. 98 – 102.