

Comparison of SVM and k-NN classifiers in the estimation of the state of the arteriovenous fistula problem

Marcin Grochowina
University of Rzeszów
al. Rejtana 16, 35-310 Rzeszów, Poland
Email: gromar@ur.edu.pl

Lucyna Leniowska
University of Rzeszów
al. Rejtana 16, 35-310 Rzeszów, Poland
Email: lleniow@ur.edu.pl

Abstract—The paper presents a concise report on the comparison of the classifiers k-NN and SVM in the case of a fuzzy classification of the arterio-venous fistula based on audio recordings. What has been used in the studies are the acoustic signals taken from both healthy patients as well as those diagnosed with the narrowing of a fistula in a mild and major degree of stenosis. In the publication there have been selected two features, each presenting one- time and frequency domain, which enable a quite clear depiction of the classification result. The aim of the study is to develop a solution enabling the detection of fistula's pathologies at an early stage.

I. INTRODUCTION

THE MAINTENANCE of a properly functioning arterio-venous fistula is extremely important for those who undergo a haemodialysis. Its good condition providing an access to arterial blood with a great degree of flow enables an efficient process of extracorporeal blood filtration. The condition of a fistula is examined by a stethoscope auscultation each time before connecting the patient to a dialysis apparatus. In particular examples diagnostics may be extended to USG examination with the Doppler's facility. Unfortunately, the most abnormalities in functioning of a fistula are diagnosed when they are so advanced that they might be detected just with a use of a stethoscope. The most frequent problem is a deformation of a fistula's vessel which result in a decreasing of its diameter. The narrowing causes a decrease of an arterial blood supply and as a result it could lengthen the process of haemodialysis or even it could make it impossible. That is why it is crucial to determine the methods of an early detection of any pathology within a fistula.

According to the studies conducted, the character of a sound emitted by the blood flowing inside a fistula's vessel vary depending on the state of a fistula. The blood flowing through a fistula in a normal state is continuous and it undergoes a very slight changes in a heart rhythm period. As a result a noise of a flowing blood can be heard well all the time. In a stenosed fistula a blood flow is hindered, what effects in rhythm changes of the intensity of a sound, compatible with a heart rhythm. The amplitude of changes is then even bigger and the stenosis of a fistula is more advanced. The differences in the picture

of an acoustic signal of a properly working fistula as well as a pathologically changed fistula are presented in the fig.1.

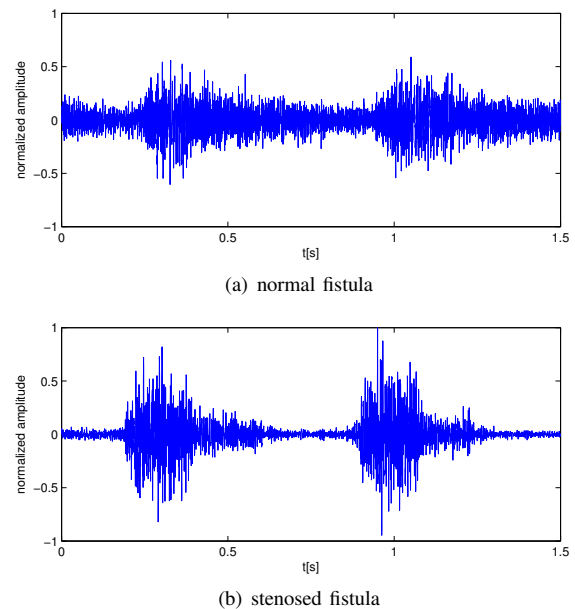


Fig. 1. An image of the time-domain audio signal emitted by the arterio-venous fistula

Not all anomalies which may occur within the fistula's area result in a decreasing of the flow. When it comes to the hyperkinetic fistula the blood flow increases excessively, what may result in cardiological complications or even strokes. However, each change of geometry of a fistula causes the change in a character of the flow of the transported blood. Usually, the flow becomes turbulent and causes changes in a frequency spectrum of a signal, what has been show in the fig.2.

Presented examples of trajectories in both the time and the frequency domain represent extreme cases, moreover, the fistula diagnosed as pathological qualifies for a reconstruction through a surgical treatment. Another problem is detecting an anomaly in an early stage, what enables a preventive

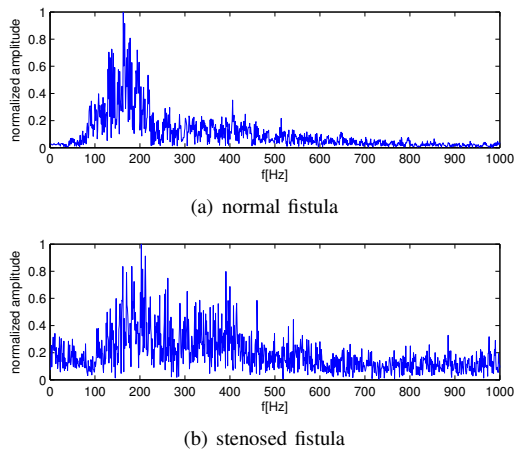


Fig. 2. The frequency spectrum of the acoustic signal emitted by the arterio-venous fistula

intervention enabling a suppression of unwanted changes. In order to do that what should be determined, apart from the collections of data from extreme cases, are concentrations of data referring to average cases with the use of which the classifier should be trained enabling their proper assessing.

II. MATERIALS AND METHODS

The research material has been collected from 9 patients, all male, age from 35 to 52. Two patients were diagnosed with an extreme malfunction (occlusion) of a fistula, another two had stenosis in various degrees of development, in the case of the last five, the fistula was normal. The examination concerned only end-to-side radio-cephalic fistulas located in a wrist. The acoustic signal was taken from the distance of 5 cm from the anastomosis point. The point of a signal intake is pictured in the fig.3

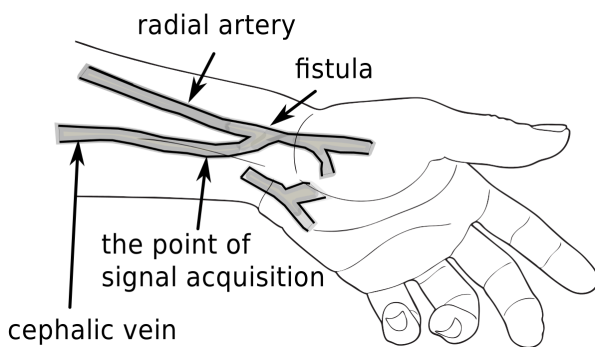


Fig. 3. The point of a signal intake

The material has been taken with the use of a specially created measuring area (fig.4) which includes:

- A header made from the ABS material which is shaped in a way that it enables the most possible anatomic position of its end during the signal intake. Inside the main body of the header electret microphone Ringford CZ034 has

been installed with responsivity -42dB ($0\text{dB}=1\text{V/Pa}$, 1KHz , 8mV/Pa) and S/N ratio better than 60dB .

- A PC computer with Xubuntu Linux 14.10 operating system, equipped with a sound card Sound Blaster PCI 64 constructed according to the Ensoniq AS1371 chip. Dedicated application serving as a registrant of a sound written in C++ with the use of the Qt library, communicating also with the equipment thanks to the ALSA sound server. The sample rate was determined on 8kHz level, with a resolution 16 bits per sample.
- Headphones used for the organoleptic monitoring of the quality of the recorded material.

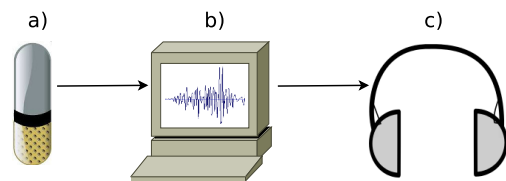


Fig. 4. data acquisition system diagram

From the taken material there have been isolated 100 samples from each group. The material has been randomly divided into a training sets including 60 samples of each kind and the testing set having 40 samples of each kind.

A common problem in classifiers building tasks is a great difference in the amount of positive and negative training patterns. In order to receive a set pattern with a similar number for each of the classes there have been registered more samples for each person with a stenosed or occluded fistula.

Registered material has been divided into parts corresponding to a single heart rhythm period, next, there has been distinguished properties being components of vector's features.

Altogether, there have been determined over 20 properties in a time domain and a frequency domain, and for the visualization needs of this elaboration, two have been chosen, each from one domain-time and frequency.

A. A property in a time domain.

When it comes to time domain, there has been chosen one property based on a difference between the maximal and the minimal value of the envelope of a signal (fig.5).

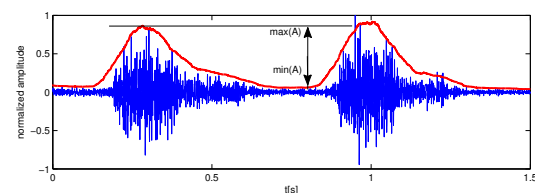


Fig. 5. A property in a time domain

In order to get an envelope signal, it has been straightened and filtrated with the use of a low-pass filter. A moving average filter with the length of 1000 elements has been used.

This property has been marked as f_1 and calculated from the equation (1).

$$f_1 = \frac{A_{min}}{A_{max}} \quad (1)$$

The initial analysis has shown that in the case of a fully efficient fistula, the flowing blood generates a sound of a slightly reducing intensity during the whole heart rhythm period. The intensity of a sound emitted by a stenosed fistula definitely characterizes with a greater fluctuation, as a result of a periodical decrease of a flowing blood speed. In connection to that, for an efficient fistula the value of its properties f_1 will be higher than in the case of a narrowed fistula.

B. A property in a frequency domain

In order to get the properties in a frequency domain, a signal had been firstly divided into fragments corresponding to the heart rhythm and then transformed to the frequency domain with using FFT transformation. In order to eliminate a spectral leakage the Hamming time window has been used. From the spectral module possessed, there have been distinguished two divisions: $FR_1 = [125Hz, 175Hz]$ and $FR_2 = [375Hz, 425Hz]$ (fig.6).

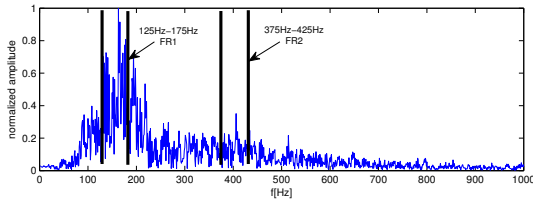


Fig. 6. A property in a frequency domain

The value of a spectral amplitude within these divisions have been summarized and calculated its property f_2 according to the equation (2).

$$f_2 = \frac{FR1}{FR2} = \frac{\sum_{f=125}^{175} fft}{\sum_{f=375}^{425} fft} \quad (2)$$

An analysis of signals coming from normal as well as stenosed fistulas, has shown that in the case of those stenosed ones, the value of a $FR2$ factor is definitely higher. Therefore, the value of a f_2 property will be smaller when the condition of a fistula will be better.

C. The use of an SVM classifier in a fuzzy classification task

The classifier used in the elaboration comes from the libSVM library in 3.2 version by Chih-Chung Chang and Chih-Jen Lin. The library is written in the C++ language owns interfaces which enable using with various programming languages, including the "m" language in Matlab and Octave environment, which were used in order to work out the results of this elaboration.

The used classifier in the classifying tasks gives back the value of -1 or 1 indicating the adjunction of an examined

feature vector in a particular category. In the regression tasks what is given back is the real value, which may go beyond the $[-1, 1]$ interval.

$$y_{SVM} = \begin{cases} -1 & \text{for } SVM_{out} < -1 \\ SVM_{out} & \text{for } SVM_{out} \in [-1; 1] \\ 1 & \text{for } SVM_{out} > 1 \end{cases} \quad (3)$$

Taking the specifics of the examined problem, it became necessary to limit the set of values given back through the interval classifier $[-1, 1]$. It has been accomplished by an application of a function described as (3) pattern, where SVM_{out} is a regression value calculated through the classifier.

D. The use of the classifier k-NN in a fuzzy classification task

Normally, the algorithm k-NN is used in classification tasks, what means it gives back a discreet value determining an adjunction to the one of the classes determined beforehand. On the account of the requirements determined by the assumptions of this elaboration, it has become necessary to create an own implementation of the k-NN algorithm in a way that it gives back a real value from the continuum between two extreme cases.

In order to do that, after having found the k-nearest neighbors and calculating their Euclidean distance from the examined point, the transformation described with a equation (4) has been applied, where $d(n)$ is a distance of the umpteenth pattern from the examined point.

$$s(n) = \frac{e^{-d(n)}}{\sum_{i=1}^k e^{-d(i)}} \quad (4)$$

In effect, the distance measurement has been changed into the proximity measure $s(n)$, simultaneously receiving

$$\sum_{n=1}^k s(n) = 1$$

to calculate definitively an expected degree of belonging of the examined point to the model classes.

$$y_{k-NN} = \sum_{n=1}^k s(n)X(n) \quad (5)$$

Assuming that all the model vectors are ascribed to the classes from the range $[-1, 1]$, the result of operation of such constructed classifier will be within the same range.

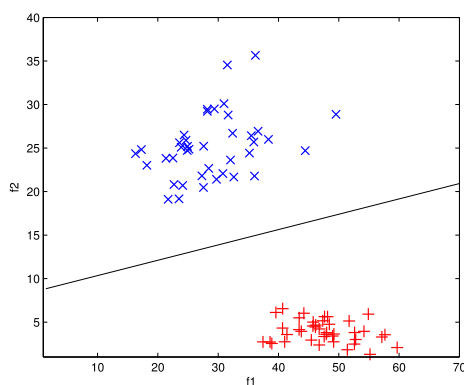
III. RESULTS AND DISCUSSION

The SVM and k-NN classifiers have been compared in three cases of using. For each case the accuracy has been determined and the analysis with the use of a confusion matrix has been conducted. In the first example the classifiers have been used in a model classification task, however in two other cases in a fuzzy classification task. For the k-NN classifier, the $k=19$ parameter has been established and used in all three cases. In the first two cases for the SVM classifier the parameter $c=1$ has been established and there hasn't been any transformations

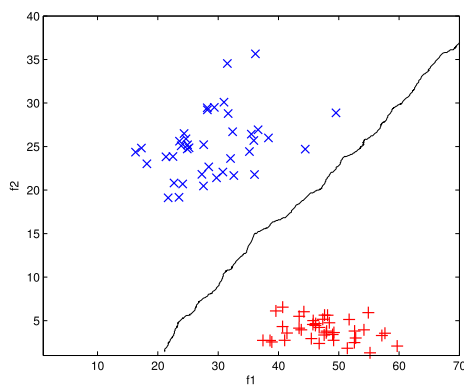
of the feature's space used. In the third case, because of the lack of a linear separation of the data, for the SVM classifier the parameter $C=100$ has been established and a kernel transformation with the use of the Gauss's function has been used.

A. Classification

In this task, the classifiers have been trained with a training set including only the extreme data, namely, ascribed to one of the two sets labeled "-1" or "1". Since the sets of data of both classes are substantially distant from each other and moreover, linearly separated, none of the classifiers had any problems with their proper division, what is presented in the fig.7.



(a) SVM



(b) k-NN

Fig. 7. Classification

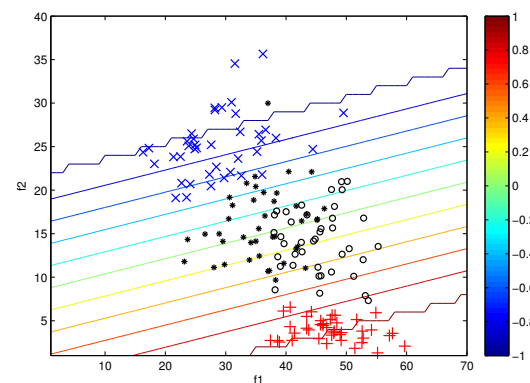
In both cases, all the examined points have been accurately recognized in both classes, what gave 100% of accuracy. The classifiers differs only in the course of hyperplane of the division of the classes, what happened to be irrelevant in this particular case.

B. The fuzzy classification - version 1

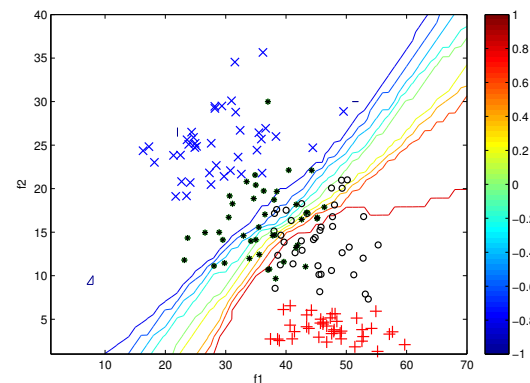
In cases of a real assessment of an arterio-venous fistula's condition classifiers cannot determine the adjunction of the

examined point to one from the two extreme classes. The task is problematic since it requires determining the position of an examined point in a space between the extreme cases. Therefore, as a result of its operation, a classifier cannot give back the value of -1 or 1. It is necessary to adjust a classifier in a way it could give back the value from the $[-1, 1]$ range. These values indicate a degree of adjunction of an examined point to the extreme classes.

In the first version of a fuzzy classification a teaching set containing only extreme cases has been applied. In effect, the functions mapping the $[f1, f2]$ features' space in the $[-1, 1]$ range has been obtained as presented graphically in the fig.8.



(a) SVM



(b) k-NN

Fig. 8. Fuzzy classification - version 1

In testing, the set containing average cases have been used, marked on a basis of an expert's assessment with the labels -0.33 and 0.33, what gave, in total, four classes of training patterns labeled with: $\{-1, -0.33, 0.33, 1\}$.

In order to make the analysis the values given back by the classifiers have been digitized to four ranges corresponding to four entering classes, and next, compared with the real adjunction of the testing patterns. The assignment of the ranges of the starting classifiers to the appropriate classes is shown in the table.I

TABLE I
ASSIGNATION OF THE RANGES OF THE STARTING CLASSIFIERS TO THE APPROPRIATE CLASSES

| range | class |
|---------------|-------|
| $[-1, -0.66)$ | -1 |
| $[-0.66, 0)$ | -0.33 |
| $[0, 0.66)$ | 0.33 |
| $[0.66, 1]$ | 1 |

The result of the operation of the classifiers has been compared in the table II and table III.

TABLE II
CONFUSION MATRIX - k-NN 0/1

| accuracy:0.59 | | Actual value | | | | class precision |
|---------------|-------|--------------|------|------|------|-----------------|
| | -1 | -0.33 | 0.33 | 1 | | |
| Predicted | -1 | 40 | 19 | 1 | 0 | 0.67 |
| | -0.33 | 0 | 7 | 3 | 0 | 0.70 |
| | 0.33 | 0 | 7 | 8 | 0 | 0.53 |
| | 1 | 0 | 9 | 28 | 40 | 0.52 |
| class recall | | 1.00 | 0.17 | 0.20 | 1.00 | |

TABLE III
CONFUSION MATRIX - SVM 0/1

| accuracy:0.78 | | Actual value | | | | class precision |
|---------------|-------|--------------|------|------|------|-----------------|
| | -1 | -0.33 | 0.33 | 1 | | |
| Predicted | -1 | 35 | 1 | 0 | 0 | 0.97 |
| | -0.33 | 5 | 24 | 11 | 0 | 0.60 |
| | 0.33 | 0 | 15 | 26 | 0 | 0.63 |
| | 1 | 0 | 9 | 3 | 40 | 0.93 |
| class recall | | 0.88 | 0.60 | 0.65 | 1.00 | |

The analysis of the confusion matrix shows definitely better parameters of the SVM classifier. Although in the case of the extreme classes labeled with -1 and 1 recall for the classifier k-NN equal 100% and it is better than for SVM, for the average classes, the values between ten and twenty percent disqualify the k-NN classifier completely. However, the level of accuracy of 80% obtained for the SVM classifier is also not very satisfying, all the more, for the classes labeled negative it has a tendency to make a better diagnosis.

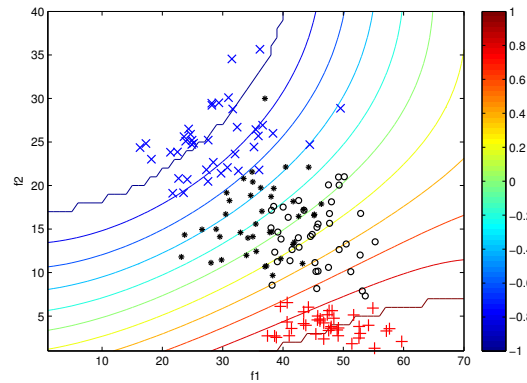
C. The fuzzy classification - version 2

In order to improve the classifiers' parameters, the next step was to insert the vectors from the sets labeled -0.33 and 0.33 into the training sets. In effect, the functions mapping the features' space $[f1, f2]$ into range $[-1, 1]$ has undergone a significant modification what has been pictured in the fig.9.

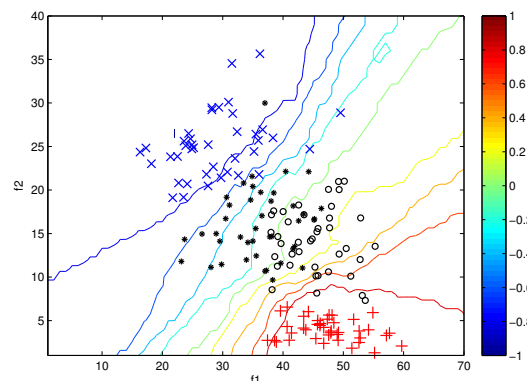
The modification improved the quality of operating of both of the classifiers. The comparison of the results of the classifiers' testing is presented in the table IV and table V.

For the SVM classifiers the improvement is slight. The accuracy increased from 78% to 81%, similarly, the class precision and the class recall parameters have undergone in some cases a slight correction down.

The results of the operation of the k-NN classifier have undergone a significant improvement. The accuracy increased from 59% to 85% and it exceeded the value obtained by the



(a) SVM1



(b) k-NN

Fig. 9. Fuzzy classification - version 2

TABLE IV
CONFUSION MATRIX - kNN FUZZY

| accuracy:0.85 | | Actual value | | | | class precision |
|---------------|-------|--------------|------|------|------|-----------------|
| | -1 | -0.33 | 0.33 | 1 | | |
| Predicted | -1 | 37 | 2 | 0 | 0 | 0.95 |
| | -0.33 | 3 | 29 | 6 | 0 | 0.76 |
| | 0.33 | 0 | 9 | 30 | 0 | 0.77 |
| | 1 | 0 | 0 | 4 | 40 | 0.91 |
| class recall | | 0.93 | 0.72 | 0.72 | 1.00 | |

SVM classifier. Only class recall for the "-1" class has slightly reduced.

In both cases what can be observed is unfortunately the problem of an accurate classification of the average classes.

IV. CONCLUSION

As a result of the conducted tests, what has been tentatively stated is that both the SVM and the k-NN classifiers can be applied in a task of an estimation of a fistula's condition and enable obtaining a relatively high co-factor of accuracy. What provides proper operation of the examined classifiers is putting into a set of pattern data not only extreme patterns determining the limits of a case's space but also, when it is possible, the whole continuum. On the basis of obtained results it cannot

TABLE V
CONFUSION MATRIX - SVM FUZZY

| accuracy:0.81 | | Actual value | | | | class precision |
|---------------|-------|--------------|------|------|------|-----------------|
| | -1 | -0.33 | 0.33 | 1 | | |
| Predicted | -1 | 37 | 1 | 0 | 0 | 0.97 |
| | -0.33 | 3 | 27 | 12 | 0 | 0.64 |
| | 0.33 | 0 | 12 | 24 | 0 | 0.67 |
| | 1 | 0 | 0 | 4 | 40 | 0.91 |
| class recall | | 0.93 | 0.68 | 0.60 | 1.00 | |

be stated which of the applied classifiers will be the best for the task. The differences in the accuracy of the classification are slight and in most cases they are within the limits of the statistical error.

In order to obtain more reliable results in continuing of the examinations, the realization of a few conclusions seems to be essential:

- it is necessary to collect a significantly greater number of an overall data since a small amount of data used in testing does not guarantee a solution of the problem,
- the signal samples have to be taken from the greater number of patients,
- it is recommended to obtain the possible biggest amount of data from the fistula's in an average state between the fully efficient to the occluded ones,
- it should be considered to widen or to change the set of the features' signal included in the analysis of the features' vector.

The application of the above-mentioned conclusions gives a great chance for the fruitful continuation of the examinations and in effect obtaining a solution enabling identification of pathological states within a fistula in the earliest state possible.

REFERENCES

- [1] Marcin Grochowina, Lucyna Leniowska, Piotr Dulkiewicz, "Application of Artificial Neural Networks for the Diagnosis of the Condition of the Arterio-venous Fistula on the Basis of Acoustic Signals," *Brain Informatics and Health, Lecture Notes in Computer Science Volume 8609*, Springer, 2014, pp 400-411.
- [2] Marcin Grochowina, Lucyna Leniowska, "Analiza parametrów akustycznych prototypu głowicy do akwizycji sygnału z przetoki tętniczko-żylniej," *Mechanika w Medycynie*, Uniwersytet Rzeszowski, 2014, pp 63-72.
- [3] Mikkel Grama, Jens Tranholm Olesen, Hans Christian Riisa, Maiuri Selvaratnama and Michalina Urbaniaka, "Stenosis detection algorithm for screening of arteriovenous fistulae," *15th Nordic-Baltic Conference on Biomedical Engineering and Medical Physics (NBC 2011)*, Springer, 2011, pp. 241-244.
- [4] Fan, Rong-En and Chen, Pai-Hsuen and Lin, Chih-Jen, "Working set selection using second order information for training support vector machines," *The Journal of Machine Learning Research vol.6*, JMLR.org, 2005, pp. 1989-1918.
- [5] Vesquez, PO and Marco, MM and Mandersson, Bengt, "Arteriovenous fistula stenosis detection using wavelets and support vector machines," *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE, IEEE, 2009*, pp. 1298-1301.
- [6] Castillo, Oscar and Melin, Patricia and Ramirez, Eduardo and Soria, José, "Hybrid intelligent system for cardiac arrhythmia classification with Fuzzy K-Nearest Neighbors and neural networks combined with a fuzzy system," *Expert Systems with Applications vol.39*, Elsevier, 2012, pp. 2947-2955.
- [7] Chih-Chung Chang and Chih-Jen Lin, "LIBSVM - A Library for Support Vector Machines," <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>
- [8] Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin "A Practical Guide to Support Vector Classification," Department of Computer Science National Taiwan University, Taipei 106, Taiwan.
- [9] Tadeusz Morzy, "Eksploracja danych - metody i algorytmy," PWN, 2013.
- [10] Pawel Delimata, Zbigniew Suraj, "Reducts Evaluation Methods Using Lazy Algorithms," *Rough Sets and Knowledge Technology, 4th International Conference, RSKT 2009, Gold Coast, Australia, July 14-16, 2009. Proceedings*, Springer, 2009, pp. 120-127.