Liliana BYCZKOWSKA-LIPIŃSKA[1], Agnieszka WOSIAK[2]

University of Computer Sciences and Skills (1)
Institute of Information Technology, Lodz University of Technology (2)

doi:10.15199/48.2015.12.36

Classification of Medical Data Derived from Diagnostic Devices Using Ensembles of Classifiers

Abstract. *The use of ensemble of classifiers for classification of medical data derived from diagnostic devices has been proposed in this research. The experimental studies were carried out on three datasets concerning different medical problems: arrhythmia, breast cancer and coronary artery disease using SPECT images. The comparison of single classification algorithms (kNN- IBk, C4.5 - J48, Naïve Bayes, Random Tree and SMO) with bagging, boosting and majority voting using all single classifiers was performed. Experimental studies have proved that hybrid classifiers outperformed single classification in all cases in terms of accuracy, precision, sensitivity and root squared mean error, regardless of the dataset.*

Streszczenie*. W ramach niniejszej pracy zaproponowane zostało zastosowanie komitetów klasyfikatorów w procesie klasyfikacji danych pochodzących z urządzeń medycznych. Badania eksperymentalne zostały przeprowadzone na trzech zbiorach danych dotyczących różnych problemów medycznych: arytmii, nowotworu piersi oraz choroby wieńcowej. Przeprowadzono porównanie pojedynczych technik klasyfikacji (kNN-IBk, C4.5 - J48, Naïve Bayes, Random Tree oraz SMO) z metodami hybrydowymi (bagging, boosting oraz głosowanie większościowe). Badania eksperymentalne wykazały skuteczność klasyfikacji z zastosowaniem komitetów klasyfikatorów – w wszystkich badanych przypadkach rezultaty klasyfikacji hybrydowej były lepsze od wyników najlepszego pojedynczego klasyfikatora biorąc pod uwagę dokładność, precyzję, czułość oraz błąd średniokwadratowy. (****Zastosowanie Komitetów Klasyfikatorów w Procesie Klasyfikacji Danych Pozyskanych za Pomocą Urządzeń Diagnostyki Medycznej****).*

Słowa kluczowe: eksploracyjna analiza danych, klasyfikacja, komitety klasyfikatorów, dane medyczne
Keywords: data mining analysis, classification, ensemble classifiers, medical data

Introduction

Medical devices are instruments or machines that are used to diagnosis, monitor, treat, or prevent disease or other conditions. The interpretation of data from various scanning devices is important to medical imaging, diagnosis and treatment, as well as reliability and sustainability analysis [1].

The classification process of medical data sets is an extremely important and difficult issue, due to the need to achieve the highest rates of accuracy for the results of classification.

Currently numerous scientific studies are conducted aiming to identify the most effective methods of classification, however, no method of classification, which would provide accurate results in relation to a variety of medical problems, has been developed [2]. For this reason, there is a constant need for further research.

The aim of this research was to constitute an independent contribution to the relevant literature in terms of classification process as well as a try to find a successful method of multiple classification applied to medical data derived from diagnostics devices using ensembles of classifiers.

The remainder of this paper is organized as follows. Section 2 presents literature review concerning data mining techniques applied in the analysis of the states of power transformers. Next section concerns the description of the proposed methodology. In Section 4 we describe the studies that were conducted. We introduce data collected for this application and discuss the results. Finally, in Section 5 we draw the conclusions.

Related Works

Large number of researches concerning medical data classification were discussed in the literature during the last years. This section refers an overview of studies to provide a general idea of the current state of the art.

The authors of [3] in their study, used Artificial Neural Networks (ANN) for Electrocardiographic (ECG) Arrhythmias classification. During the training process of ANN, the ECG recordings from MIT BIH Arrhythmia database were used as a reference. 24 cases out of 48 30 minutes recordings in this database were used for data extraction. The arrhythmia samples that were extracted from the database were preprocessed to create input sets to train ANNs. The Fourier Transforms of a predefined window of signals were taken as a feature extraction method. As a result, 5 types of ECG signals (Ventricular Tachycardia, Left Bundle Branch Block, Right Bundle Branch Block, Atrial Fibrillation and Normal ECG) were labeled with 82% accuracy.

In the study of [4], six types of arrhythmia beats observed in ECG signals have been analyzed by using clustering methods. The purpose of the analysis was to verify if the examined arrhythmia types form natural groups in the feature spaces. The performances of the clustering algorithms were tested using different distance metrics and algorithms. The results were examined based on the average sensitivity, specificity, selectivity and accuracy of the classifier. The results show that k-means clustering technique gave the best results. The results also showed that analyzed arrhythmia types did not form distinct clusters in examined feature spaces. On the other hand, in some cases very high specificity results were observed for some arrhythmia types. That suggested that the features could be quite useful in elimination processes in hierarchic classifiers.

In [5] random forests have been investigated for classification of SPECT images and the design of an Alzheimer's disease computer aided diagnostic system. The proposed system was based on voxel-based normalized mean square error feature extraction, the t-test with feature correlation weighting for feature selection and random forest image classification. It was shown in the paper that the generalization error for the forest converges to a limit as the number of trees in the forest becomes large. Moreover the authors stated, that the generalization error depended on the strength of the individual trees in the forest and the correlation between them. The proposed method yielded an up to 96.2% classification accuracy for Alzheimer's disease diagnosis.

The effectiveness of the combination of the classifier ensemble and the feature selection for each component classifier was described in [6]. The experiments using 1793 mammograms have been performed. Results of the classification between malignant lesions and normal tissues by the four CAD systems have been shown. Leave one-out method was applied in the experiments. Experimental results showed that the introduction of classifier ensemble is effective in improving the performance of the mammography CAD system.

Materials and Methods

The main purpose of classification is to identify to which set of categories a new observation belongs to. This is done on the basis of a training set consisting of instances that are already assigned to the known classes.

The main idea of the multiple classification methodology is to consider several individual classifiers, and combine them in order to obtain a complex classifier that outperforms every one of them [7]. In the literature, there are two terms that refer to multiple classification: "ensemble methods" and "hybrid classifiers".

Hybrid classification using committees of classifiers is now one of the most promising trends in machine learning [8, 9, 10, 11, 12]. It allows for more accurate classification analysis especially of data sets for which there are no precise classification rules. It also increases classification efficiency of diverse data, which may require testing of multiple methods before selecting one of them.

The most popular methods of hybrid classification are committees of homogeneous classifiers, which include aggregation technique models - voting and bagging, boosting [13, 14, 15], as well as random forests [16].

The idea of bagging is to create an ensemble classifiers based on bootstrap replicates of the training set. The classifier outputs are combined by the plurality vote [14].

The main idea of boosting algorithms is to combine multiple classification algorithms (weak learners) into one strong composite classifier. One of the most popular ensemble algorithm that improves the simple boosting algorithm by an iterative process is AdaBoost (Adaptive Boosting). It was first introduced in [15]. It is worth mentioning that he basic AdaBoost algorithm deals with binary classification. To handle multiple classes, AdaBoost should be extended by reducing multiclass classification to multiple two-class problems [17].

One of the most important problem to solve in hybrid classification is to develop efficient combination rules for set of classifiers. In practice majority voting schemes are implemented in different versions [8, 10]:

- unanimity, where the answer requires that all classifiers agree,
- simple majority, where the answer is given by greater than half majority of classifiers,
- plurality voting, taking the answer with the highest number of votes.

In literature the term majority voting usually refers to the last version - plurality votes.

This research concerned a comparative analysis of single and multiple techniques for classification of medical data sets derived from a variety of diagnostic tests. Experimental studies were conducted on real data.

In order to assess the performance of classification methods, following comparison criteria have been used:

- accuracy,
- sensitivity,
- specificity,
- precision, and
- root mean square error.

Experimental Analysis and Results

The main objective of the experiments was to examine the accuracy of different hybrid classifiers in comparison to the results derived from application of single classification algorithms as applied to data derived from medical diagnostic devices.

Three different datasets were used in the experimental analysis. They were obtained from the research servers providing their resources for scientific purposes. The following data sets studies were analyzed:

- cardiac arrhythmia (452 cases, 279 attributes, multi-class with 16 classes),
- cardiac SPECT images data (267 cases, 22 attributes, binary classification),
- mammography (961 cases, 6 attributes, binary classification).

The experiments conducted in this research used the WEKA (Waikato Environment for Knowledge Analysis) data mining tool [18]. The data sets were tested using a cross-validation technique by randomly choosing 10 disjoint 10% of the original data set sized samples and running the classification to each of the 10 samples using the remaining 90% of the original data set as training set and the sample itself as testing set. The results correspond to the average of each 10 classification runs.

In the first step of the experiments, single classification algorithms were applied:

- IBk (k=5),
- J48,
- NaiveBayes,
- RandomTree, and
- SMO.

The results of single classification process are summarized in Table 1.

Table 1. Single classification results.

Data set	C	IBk	J48	NB	RT	SMO
Arrhythmia	ACC	59.956	64.159	62.168	50.221	70.796
	PREC	0.539	0.608	0.625	0.495	0.673
	SENS	0.600	0.642	0.622	0.502	0.708
	RMSE	0.195	0.201	0.215	0.249	0.230
Mammography	ACC	81.998	82.726	76.899	83.350	78.564
	PREC	0.820	0.829	0.776	0.834	0.789
	SENS	0.820	0.827	0.769	0.834	0.786
	RMSE	0.370	0.359	0.419	0.375	0.463
SPECT	ACC	56.684	75.401	74.866	70.588	73.262
	PREC	0.922	0.909	0.901	0.912	0.915
	SENS	0.567	0.754	0.749	0.706	0.733
	RMSE	0.501	0.416	0.454	0.517	0.517

In the case of classification conducted by single classifiers, none of them significantly outperformed other classifiers in terms of classification accuracy for different datasets. That confirms the theorem which lies behind hybrid classification and says that there is no single pattern recognition algorithm, and can be appropriate for all the classification tasks we deal with – Wolpert's theorem [19].

The best results attained over 80% for binary mammography dataset. There is no denying that the multiclass classification of arrhythmia provided the worst results, achieving up to 70% for SMO classifier (average accuracy for arrhythmia equaled 61.46%). The average accuracy of single classification approach attained 70.78%.

Next step of the experiments concerned performing classification using hybrid classifiers.. Different combinations were applied:

- bagging with IBk, Naive Bayes, J48 and SMO as base classifiers,

– random forest as a variant of bagging with a decision tree as a base classifier,

– boosting with IBk, Naive Bayes, J48 and SMO as base classifiers, and

– voting by all single classifiers using majority as a combination rule.

The results of classifications using different methods are presented in Table 2. (bagging) and Table 3. (boosting and majority voting). The visualization of accuracies for all single classifications, bagging, boosting and majority voting is shown in Fig. 1.

Table 2. Hybrid classification results for bagging.

Data set	C	Bagging with base classifier of				RF
		IBk	J48	NB	SMO	
Arrhythmia	ACC	59.956	72.566	62.832	71.238	69.027
	PREC	0.529	0.681	0.603	0.666	0.614
	SENS	0.600	0.726	0.628	0.712	0.690
	RMSE	0.194	0.160	0.185	0.229	0.167
Mammography	ACC	82.206	83.390	79.292	79.501	82.622
	PREC	0.823	0.834	0.800	0.797	0.827
	SENS	0.822	0.834	0.793	0.795	0.826
	RMSE	0.364	0.354	0.401	0.413	0.361
SPECT	ACC	57.219	80.749	77.005	70.053	77.540
	PREC	0.912	0.915	0.903	0.912	0.918
	SENS	0.572	0.807	0.770	0.701	0.775
	RMSE	0.506	0.370	0.433	0.459	0.400

Table 3. Hybrid classification results for boosting and majority voting.

Data set	C	Boosting with base classifier of				MV
		IBk	J48	NB	SMO	
Arrhythmia	ACC	60.177	71.018	64.602	66.372	68.584
	PREC	0.547	0.679	0.663	0.644	0.171
	SENS	0.602	0.710	0.646	0.664	0.644
	RMSE	0.212	0.182	0.188	0.200	0.686
Mammography	ACC	81.478	82.414	77.003	78.564	81.270
	PREC	0.815	0.824	0.777	0.789	0.818
	SENS	0.815	0.824	0.770	0.786	0.813
	RMSE	0.376	0.357	0.405	0.399	0.370
SPECT	ACC	67.380	77.005	76.471	71.123	76.471
	PREC	0.910	0.911	0.917	0.929	0.910
	SENS	0.674	0.770	0.765	0.711	0.765
	RMSE	0.483	0.417	0.430	0.462	0.430

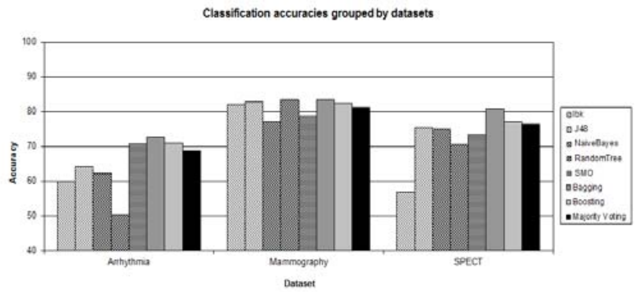

Fig. 1. The chart of classification accuracies [Source: Own work].

One can see, that in most cases hybrid classifications improved the corresponding single classifiers in terms of classification accuracy, as well as precision, sensitivity and root mean square error.

To sum up classification results, in Table 4. there are best results attained for single and hybrid classification approaches in terms of all considered criteria.

In all cases hybrid classifiers appeared to give higher accuracy than the best single classifier. However the accuracy for arrhythmia classification equaled 72.6% which is still not enough to be considered in medical diagnostics systems. The rest of results exceeded the threshold of 80% for accuracy and might be considered for further stages of computer aided diagnostic process.

Table 4. Comparison of best accuracies for single and hybrid classifiers.

Data set	Criteria	Single Classifier	Hybrid Classifier	Difference
Arrhythmia	ACC	70.796	72.566	1.770
	PREC	0.673	0.681	0.008
	SENS	0.708	0.726	0.018
	RMSE	0.230	0.160	-0.070
Mammography	ACC	83.350	83.390	0.040
	PREC	0.834	0.834	0.000
	SENS	0.834	0.834	0.000
	RMSE	0.375	0.354	-0.021
SPECT	ACC	75.401	80.749	5.348
	PREC	0.909	0.915	0.006
	SENS	0.754	0.807	0.053
	RMSE	0.416	0.370	-0.046

Conclusions

Experimental studies confirmed the efficacy of the hybrid classification for medical data sets derived from medical diagnostic equipment. It was successfully proved that the hybrid classifiers outperform single classification methods regardless of the input dataset used for the purpose of training the model.

One of the elements that is considered as the part of the whole process of data classification is appropriate feature selection method. The selection of the best discriminative features plays an important role when constructing classifiers [20]. Feature selection is usually done as a pre-processing step to improve system performance by selecting optimal features from entire datasets. However such an approach might be not suitable for hybrid classification as feature selection method should be considered in conjunction with one of the classifiers [21]. Therefore a well-suited feature selection methods will be the aim of our future works.

Moreover our further work will be also associated with the use of other algorithms and strategies of classifier selection. In addition, it is planned to implement fuzzy classification, which may be applicable to medical data sets [22, 23, 24].

REFERENCES

[1] GAVRILESCU M., VASILE M., GRÖLLER E.. Gradient-based classification and representation of features from volume data. System Theory, Control, and Computing (ICSTCC), 2011 15th International Conference on. IEEE, 2011.

[2] BYCZKOWSKA-LIPIŃSKA L., WOSIAK A.: Hybrid Classification of High-Dimensional Biomedical Tumour Datasets. Advanced and Intelligent Computations in Diagnosis and Control 386 (2015): 287.

[3] ZEYBEKOĞLU S., ÖZKAN M.: Classification of ECG Arrhythmia beats with Artificial Neural Networks. Biomedical Engineering Meeting (BIYOMUT), 2010 15th National. IEEE, 2010.

[4] GEZER B.L., KUNTALP D., KUNTALP M.: Clustering of arrhythmic ECG beats using morphological properties and windowed raw ECG data. Signal Processing and Communications Applications (SIU), 2011 IEEE 19th Conference on. IEEE, 2011.

[5] RAMÍREZ J., GÓRRIZ J. M., CHAVES R., LÓPEZ M., SALAS-GONZALEZ D., ÁLVAREZ I., SEGOVIA F.: SPECT image classification using random forests. Electronics letters 45.12 (2009): 604-605.

[6] NEMOTO M., SHIMIZU A., KOBATAKE H., TAKEO H., NAWANO, S.: Classifier ensemble for mammography CAD system combining feature selection with ensemble learning. International Congress Series. Vol. 1281. Elsevier, 2005..

[7] MICHALSKI R.S., TECUCI G.: Machine Learning, A Multistrategy Approach. J. Morgan Kaufmann, 1994.

[8] KUNCHEVA L.I.: Combining Pattern Classifiers. Methods and Algorithms. John Wiley & Sons, Inc.; Hoboken, New Jersey, 2004.

[9] ROKACH L.: Pattern Classification Using Ensemble Methods. World Scientific Publishing Co., Inc.; 2010; River Edge, NJ, USA.
[10] WOZNIAK M., GRANA M., CORCHADO E.: A survey of multiple classifier systems as hybrid systems. Information Fusion 2014; pp. 3-17; DOI: 10.1016/j.inffus.2013.04.006.
[11] GALAR M., FERNÁNDEZ A., BARRENECHEA E., BUSTINCE H., HERRERA F.: A Review on Ensembles for the Class Imbalance Problem: Bagging-, Boosting-, and Hybrid-Based Approaches. Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, 2012; 42(4):463-484. doi: 10.1109/TSMCC.2011.2161285
[12] SENI G., ELDER J.F.: Ensemble Methods in Data Mining: Improving Accuracy Through Combining Predictions. Morgan & Claypool, 2010.
[13] BREIMAN L.: Bagging predictors. Technical Report 421; Department of Statistics, University of California, Berkeley 1994.
[14] BREIMAN L.: Bagging predictors. Machine Learning; 26(2):123–140; 1996.
[15] FREUND Y., SCHAPIRE R.E.: A decision – theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences; 55(1):119–139; 1997.
[16] BREIMAN L.: Random forests. Machine Learning; 45:5–32; 2001.
[17] HAO W., JIEBO L.: Generalized multiclass AdaBoost and its applications to multimedia classification. Computer Vision and Pattern Recognition Workshop, 2006. CVPRW'06. Conference on. IEEE, 2006.
[18] WITTEN I.H., FANK E.: Data mining: practical machine learning tools and techniques, 2nd edn. Morgan Kaufmann, San Francisco (2005)
[19] WOLPERT D.H., The supervised learning no-free-lunch Theorems, Soft Computing in Industrial Applications, 6th Online World Conference on, 2001, pp. 25-42.
[20] MORITA M., OLIVEIRA L.S., SABOURIN R.: Unsupervised feature selection for ensemble of classifiers. Frontiers in Handwriting Recognition, 2004. IWFHR-9 2004. Ninth International Workshop on. IEEE, 2004.
[21] DUANGSOITHONG R., WINDEATT T.: Bootstrap feature selection for ensemble classifiers. Advances in Data Mining. Applications and Theoretical Aspects. Springer Berlin Heidelberg, 2010. 28-41.
[22] GRANA M., TORO C., eds.: Advances in knowledge-based and intelligent information and engineering systems. Vol. 243. Ios Press, 2012.
[23] GRANA M., TORO C., HOWLETT R.J., eds.: Innovation in Medicine and Healthcare 2014. Vol. 207. IOS Press, 2015.
[24] WIDJAJA M., DARMAWAN A., MULYONO S.: Fuzzy classifier of paddy growth stages based on synthetic MODIS data. In Advanced Computer Science and Information Systems (ICACSIS), IEEE International Conference on; 239-244; 2012.

***Authors**: prof. dr hab. inż. Liliana Byczkowska-Lipińska,, University of Computer Sciences and Skills, ul. Rzgowska 17 a, 93-008 Lodz, Poland e-mail: liliana.byczkowska-lipinska@p.lodz.pl, dr inż. Agnieszka Wosiak, Politechnika Łódzka, Instytut Informatyki, ul. Wólczańska 215, 90-924 Lodz, Poland, e-mail: agnieszka.wosiak@p.lodz.pl*