Classification of Magnetic Resonance Images

Katarina Trojacanec[1], Gjorgji Madzarov[1], Dejan Gjorgjevikj[1], and Suzana Loskovska[1]
[1] *Faculty of Electrical Engineering and Information Technologies, ,,Ruger Boskovik" bb., Skopje, Republic of Macedonia*
katarina.trojacanec@feit.ukim.edu.mk, madzarovg@feit.ukim.edu.mk, dejan@feit.ukim.edu.mk, suze@feit.ukim.edu.mk

Abstract. *The aim of the paper is to compare classification error of the classifiers applied to magnetic resonance images for each descriptor used for feature extraction. We compared several Support Vector Machine (SVM) techniques, neural networks and k nearest neighbor classifier for classification of Magnetic Resonance Images (MRIs). Different descriptors are applied to provide feature extraction from the images. The dataset used for classification contains magnetic resonance images classified in 9 classes.*

Keywords. Classification, Support Vector Machines (SVMs), Magnetic Resonance Images (MRIs), neural networks.

1. Introduction

Magnetic resonance is widely used in recent years as a valuable technique in surgical and clinical environment. MR Imaging has become a useful modality since it provides plentiful image information and high sensitivity. MRI characteristics play crucial role in medical clinical diagnosis, providing abundant information of the tissues.

The necessity of efficient and automated way for magnetic resonance images analysis is continuously increasing. Manual or even semi-automated image classification is impractical for large amount of images. Moreover, it is highly subjective and non-reproducible. To avoid the human error in manual interpretation of medical image content, when large numbers of images are analyzed, a fully automated approach of image classification is required.

Magnetic Resonance Images classification is of great importance and widely used in research and clinical studies [1][2][3][4][5]. Many researchers have used different classification techniques for MRI data, such as Bayes classifier [1][6], k-Nearest Neighbors (kNN) classifier [7], Artificial Neural Networks (ANN) [5], Support Vector Machines (SVMs) [8] and Expectation Maximization (EM) as a statistical classification scheme. For instance, SVM based method for automated segmentation and classification of brain MRI is proposed in [8]. In [9] support vector machines classifier is applied on breast multi-spectral magnetic resonance images.

It is clear that MRI classification is very sensitive problem, primarily because of the overlapping tissue intensity distributions. This problem arises from the inherent limitations of the image acquisition process, such as intensity inhomogenity (also known as bias field), noise, and partial volume effect. These specific characteristics of MRIs induce to be used as much as possible precisely classifiers.

The paper is organized as follows. Section 2 provides a brief introduction to Support Vector Machines, Artificial Neural Networks and k nearest neighbor classifier used for pattern recognition, while section 3 describes the MRI dataset used for this research. The experimental results from MR images classification using SVMs, a multilayer neural network, and k-nn are presented in section 4. Finally, concluding remarks are given in section 5.

2. Classification Techniques

Algorithms for pattern recognition are in continuous development and improvement. The aim of pattern recognition is efficient data classification on the basis of a priori knowledge or statistical information extracted from the data. Hence, the capability of generalization is one of the most important characteristics of classifiers [10].

Three types of classifiers are used for MRI classification in this paper: support vector machines, artificial neural networks and k-nearest neighbour. Next subsections briefly describe their basic characteristics.

2.1. Artificial Neural Networks

Artificial neural networks (ANN) [11] have been widely used in many different pattern recognition areas for many years. Their functionality is based on the incredible functionality of biological neural system. In general, ANN are composed of many nonlinear computational elements similar to the biological neurons. The neural networks processing capability depends basically on the intensity of the connections between neurons within the neural network. In fact, during the adaptation process or the training process from the training samples, the weights of these connections between neurons are computed. In this paper, a multilayer perceptron with one hidden layer and 25 units within it is used as a technique for MRI classification.

2.2. Support Vector Machines

Support Vector Machines are based on the idea to look for the hyper-plane that maximizes the margin between two classes. In fact, SVM classifier in its basis is a binary classifier [10][12][13]. One of the limitations of SVM classifiers is exactly the nature of their basic concept – the ability for binary classification only. However, SVM classifiers could be extended to be able to solve multiclass problems as well. Next subsections briefly describe the approaches for extending SVM classifier, used in this paper as MRI classifiers.

2.2.1. One-against-all

One of the strategies for adapting binary SVM classifiers for solving multiclass problems is one-against-all (OvA) scheme. It includes decomposition of the M-class problem (M>2) into series of two-class problems. The basic concept is to construct M SVMs where the i-th classifier separates the class i from all other (M-1) classes. All M classifiers are then trained to make difference between the examples that belong to the class and those that belong to all other classes [14].

This strategy has a few advantages such as its precision, the possibility for easy implementation and the speed in the training phase and the recognition process. That is the reason for its wide use.

2.2.2. One-against-one

The other strategy for extending binary SVM classifiers for multiclass problem consists of combining multiple SVMs in one-against-one (OvO) scheme. The basic idea in this strategy is building particular SVM classifier for each pair of classes.

For the M-class problem, the number of SVMs required in this case is M(M-1)/2. During the testing phase, each of the trained SVMs votes for the winning class. At the end, each example gets the label with the bigger number of votes.

The OvO strategy has big computational cost because the decisions of bigger number of classifiers should be calculated [15], especially when solving multiclass problem with big number of classes.

2.2.3. Support Vector Machines in binary tree architecture

The first step in the algorithm that represents SVMs in binary tree architecture (SVM-BTA) is creating distance matrix that represents the similarity between each pair of classes. The arbitrary class is then chosen and kept into the so called selected list. After that, the closest class to the selected one is chosen and added to the list. The algorithm is repeated continuously looking for the closest class until the last class is added to the selected list.

The classes included in the selected list are then separated into two groups of positive and negative examples. The number of classes in both groups should be approximately the same to keep the tree balanced and provide the optimal results.

The recognition of each pattern starts at the root of the tree. At each node of the binary tree a decision is being made about the assignment of the input pattern into one of the two possible groups represented by transferring the pattern to the left or right sub-tree. This is repeated downward the tree until the sample reaches a terminal node that represents the class it has been assigned to [16].

2.2.4. SVM utilizing Binary Decision Tree

SVMs utilizing Binary Decision Tree (SVM-BDT) involves hierarchical clustering with the aim to provide grouping of the classes on the basis of the similarity of the examples [17].

In the classifier training phase, the classes are grouped into two different groups of positive and negative examples. The grouping starts at the root of the tree. After calculating the distance matrix, as shown in the previews subsection, the two classes with the biggest distance value are selected. Each class is joined to only one particular group of classes. After that, the class with the smallest distance to one of the previously selected groups is joined to the nearest group. The distance matrix is then recalculated. Thus, the distance matrix is always representing the distance between the groups and the classes that are not already joined nor to the first neither to the second group. This procedure repeats until all groups are part of the appropriate group.

Once SVM classifier in the root of the tree is trained, the classes belong to one of the groups are joined to the root of the left sub-tree and the classes contained to the other group are joined to the root of the right sub-tree. The procedure repeats through all nodes of the tree until the terminal nodes are visited.

2.2.5. SVM utilizing Balanced Binary Decision Tree

This algorithm beside SVM utilizing Balanced Binary Decision Tree (SVM-BBDA) [17] is very similar with the previous one with one difference. It always tries to keep the tree balanced. In other word, the distribution of the classes through the groups is uniform. Rather than choosing only one class in each iteration of the previously described algorithm and then recalculating the distance matrix, this algorithm tends to choose two classes on the bases of their distances to both of the groups (the closest one to one of the groups and the closest one to the other group) in each iteration and then to recalculate the distances in the distance matrix.

2.3. K Nearest Neighbors Classifier

The k nearest neighbors algorithm (k-nn) algorithm [17] has been widely used method for classifying objects based on closest training examples. K-nn is one of the most effective and simplest machine learning algorithms. In the process of classification the voting process is included. The given sample is classified by voting of its k-nearest neighbors. K is a parameter which can be adjusted. When k is 1 the object is assigned to the class of its nearest neighbor. The neighbors are taken from a set of objects for which the class label is previously known.

The same method can be used for regression, by simply assigning the property value for the object to be the average of the values of its k nearest neighbors. There are cases in which is useful to assign weights to the votes, i.e. the closer the neighbor is the more valuable his vote is. But there are many variations of this technique.

The main drawback of this technique is that classes which have a number of examples, far greater than other classes, tend to dominate the prediction process, i.e. objects which we want to classify, have a greater probability to be labeled as members of the dominant classes.

3. Description of the dataset

In the paper we make analysis when applying the described algorithms for multiclass classification of Magnetic Resonance Images. The dataset used for analysis consists of magnetic resonance images provided by [18] and [19]. The dataset includes brain and abdomen MRIs and MRIs from gynecology domain. A brief textual description is available for each image from the dataset. The provided magnetic resonance images did not have any organization. We applied text based retrieval to organize the images, firstly, according to the part of the body they represent, i.e. brain, abdomen, gynecology. Then, we divided each of these classes on the bases of pathology present in the image characteristic for the specified class. The hierarchy that represents this classification is depicted on Fig. 1.

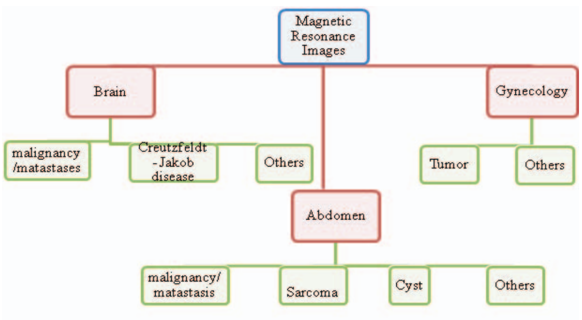

Figure 1. Hierarchy organization of the dataset

As we can see from the Fig. 1, the first level of the hierarchy contains three categories: Brain, Abdomen and Gynecology. There are three subclasses contained in the Brain class. The first one contains images taken from patients in whom malignancy, metastases or tumor has been diagnosed in the part of their brain. The second subclass represents MRIs where Creutzfeldt-Jakob disease is present. The last subclass, Others, includes images with none of the mentioned pathologies and/or images where no pathological region has been detected. The Abdomen class was divided into four subclasses. The first class contains images with presence of malignancy, metastases or tumor in the abdominal part of the human body, while the second class represents the images with presence of sarcoma. The third subclass includes MRIs that denote presence of cysts in the abdominal part of the examined patients. All other abdominal MRIs are classified in the fourth subclass of the Abdomen class. In the third, Gynecology, class two separated subclasses could be obtained, according to the presence or absence of tumor, respectively. Therefore, the examined magnetic resonance images could be classified into nine classes, presented by the leaf nodes in the hierarchy from Fig. 1.

There are 1870 magnetic resonance images in the dataset in total. The training set consists of 1247 MRIs, while the test set consists of 623 MRIs. Table 1 depicts the distribution of the number of images through the classes.

Table 1. Distribution of the number of images through the classes

Level 1	Level 2	Class No.	Training set	Test set	Total
Abdomen	malignancy /matastases	0	67	34	101
	Sarcoma	1	28	14	42
	Cyst	2	36	18	54
	Others	3	455	228	683
Brain	malignancy /matastases	4	53	27	80
	Creutzfeldt - Jakobdisease	5	13	7	20
	Others	6	343	171	514
Gynec ology	Tumor	7	56	27	83
	Others	8	196	97	293
Total			**1247**	**623**	**1870**

4. Experimental Results

In this paper, we make a comparison between SVM classifiers based on different strategies for multiclass classification, neural networks and k-nearest neighbours to determine and signify the best classifier for our dataset of magnetic resonance images. The main reason for using SVM classifiers is because of their good generalization and high precision capabilities. The reason why we use multilayer ANN is because it is capable of tolerating the noise, distortion and incompleteness of the data. We use k-nn method because of its simplicity and efficiency.

Two main processes are characteristic for our examination, the feature extraction process and the classification process. The feature extraction from the visual MR image content was performed using seven descriptors: Edge Histogram Descriptor (EHD) [22], Homogeneous Texture Descriptor (HTD) [22], Region-based Shape Descriptor (RSD) [22], Wavelet transformations [23], Moment Invariants Descriptor (MID) [24], Directional Edge Histogram Descriptor (DEHD) [24], and Directional Edge Histogram Moments Descriptor (DEHMD) [24].

The result of the feature extraction process is separate feature vector obtained for each of the images belongs to both, the train and the test set, for each descriptor. The feature vectors are then normalized using min-max normalization technique.

During the second process, the classification process, we examined five SVM-based algorithms and for MRI classification, that we implemented using the Torch library [20]:

- One-against-all (OvA) scheme
- One-against-one (OvO) scheme
- SVMs in binary tree architecture (SVM-BTA)
- SVM utilizing Binary Decision Tree (SVM-BDT)
- SVM utilizing Balanced Binary Decision Tree (SVM-BBDT)

Additionally, to classify magnetic resonance images we used a multilayer perceptron with one hidden layer and 25 units within it that we implemented using the Torch library [20]. For the k-nn classifier we used its Weka implementation [21].

In fact, the minimal classification error, obtained when each of the classifiers was used for MRI classification, is depicted in Table 2 and Table 3. The feature vectors that describe the images from the dataset, calculated by using different kind of descriptors, were separately passed through the classifiers. Thus, the classification error calculated for each classifier in the case of Edge Histogram Descriptor, Homogeneous Texture Descriptor and Region Based Descriptor are depicted in Table 2. Similarly, the classification error provided by each classifier in the case of Wavelet transformations, Moment invariant descriptor, Directional edge histogram descriptor, as well as Directional edge histogram moments descriptor are presented in Table 3.

Table 2. Classification error

Classification Error (%)	EHD	HTD	RSD
SVM OvA	17,66	47,51	41,73
SVM OvO	18,14	47,51	44,63
SVM-BTA	18,62	44,94	43,02
SVM-BDT	18,78	46,71	42,54
SVM-BBDT	18,46	45,26	43,98
ANN	25,2	45,9	47,6
k-nn	18,29	50,56	43,82

Table 3. Classification error

Classification Error (%)	Wavelets	MID	DEH	DEHM
SVM OvA	44,12	56,02	46,22	60,19
SVM OvO	42,35	55,86	48,8	68,22
SVM-BTA	40,58	55,7	45,43	58,91
SVM-BDT	43	56,34	55,06	59,87
SVM-BBDT	42	55,54	46,73	58,59
ANN	44,12	84,2	56,5	61,31
k-nn	44,28	51,36	49,44	61

According to the results shown on the Teble 2 and Table 3 we can notice that in the case when EHD and RSD are used for feature extraction separately, SVM classifier based on one-against-all strategy has minimal classification error of 17,66% and 41,73%, respectively. When one of HTD, Wavelets or DEH is used to describe MRIs, the best classification results were obtained from SVM-BTA classifier. The obtained classification error for these cases are 44,94%, 40,58% and 45,43% classification error. In the case of MID – the best classification was performed by k-nn classifier (51,36% classification error), while in the case of DEHMD for feature extraction, the minimal classification error was computed when the classification was performed by SVM – BBDT classifier (classification error of 58,59%).

According to the obtained results, we can conclude that for our MRI dataset used in this examination, the best results based on the classification error as an evaluation technique were provided by SVM classifier with one-against-all scheme, and Edge histogram descriptor used for feature extraction from the images. The best classification error provided from our examination is 17,66%.

Additionally, we made a comparison between the two of the classifiers (SVM classifier based on one-against-all strategy and k-nn classifier) based on the root mean squared error. The results based on this evaluation technique are depicted in Table 4 and Table 5. According to these results the k-nn classifier shows the best results in the cases of all descriptors. Because of the specificity of SVM binary tree algorithms (only the probability of predicted class is known), ANN algorithm, and SVM one-against-one method with majority voting, we were not able to obtain their root mean squared error. However our future work is in a direction of providing the implementation that enables calculating the probability distribution of all classes for the purposes of SVM binary tree classifiers.

Table 4. Root mean squared error

Root mean squared error	EHD	HTD	RBD
SVM OvA	0,223299	0,328547	0,320703
k - nn	0,1822	0,2724	0,2603

Table 5. Root mean squared error

Root mean squared error	Wavelets	MID	DEH	DEHM
SVM OvA	0,332026	0,352719	0,33733	0,350024
k-nn	0,2622	0,2798	0,2752	0,2913

5. Conclusion

Support vector machines, neural networks and k-nn method have crucial place in pattern recognition because of their accuracy, precision and effectiveness even in the cases where small train test is used. In this paper, we compared five SVM algorithms based on different classification techniques for multiclass classification, the multilayer perceptron and k nearest neighbor classifier were for Magnetic Resonance Images classification. According to the provided examination, we can conclude that the best classification error was achieved using the one-against-all strategy in the case of Edge histogram descriptor used for feature extraction. The classification error in this case was 17.66%. Comparing k-nn classifier and SVM classifier based on root mean squared error, k-nn classifier provides better results in the cases of all descriptors.

MRI classification is very specific and sensitive process because of the specific nature of the images. Thus, every improvement is very significant in this domain and a great challenge.

9. References

[1] Collins, D., Montagnat, J., Zijdenbos, A., Evans, A., Arnold, D., Automated

estimation of brain volume in multiple sclerosis with BICCR, In: Insana, M. F., Leahy, R. M. (Eds.), Proc. of IPMI 2001. Vol. 2082 of LNCS. Springer-Verlag, pp. 141-147, Jun. 2001.
[2] MacDonald, D., Kabani, N., Avis, D., Evans, A. C., Automated 3-D extraction of inner and outer surfaces of cerebral cortex from MRI. Neuroimage 12 (3),340-56, 2000.
[3] Paus, T., Zijdenbos, A., Worsley, K., Collins, D. L., Blumenthal, J., Giedd, J. N., Rapoport, J. L., Evans, A. C., Structural maturation of neural pathways in children and adolescents: in vivo study. Science 283 (5409), 1908-11, Mar. 1999.
[4] Rapoport, J. L., Giedd, J. N., Blumenthal, J., Hamburger, S., Jeries, N., Fernandez, T., Nicolson, R., Bedwell, J., Lenane, M., Zijdenbos, A., Paus, T., Evans, A., Progressive cortical change during adolescence in childhood-onset schizophrenia. A longitudinal magnetic resonance imaging study. Arch Gen Psychiatry 56 (7), 649-54, Jul. 1999.
[5] Zijdenbos, A. P., Forghani, R., Evans, A. C., Automatic 'pipeline' analysis of 3D MRI data for clinical trials: Application to multiple sclerosis. IEEE Trans Med Imaging 21 (10), 1280-91, Oct. 2002.
[6] Kamber, M., Shinghal, R., Collins, D. L., Francis, G. S., Evans, A. C., Model-based 3-D segmentation of multiple sclerosis lesions in magnetic resonance brain images. IEEE Trans Med Imaging 14 (3), 442-53, 1995.
[7] Wareld, S. K., Kaus, M., Jolesz, F. A., Kikinis, R., Adaptive, template moderated, spatially varying statistical classication. Med Image Anal 4 (1), 43-55, Mar 2000.
[8] H. Selvaraj, S. Thamarai Selvi, D. Selvathi, L. Gewali, Brain MRI Slices Classification Using Least Squares Support Vector Machine, International Journal of Intelligent Computing in Medical Sciences and Image Processing, Vol. 1, No. 1, Issue 1, 2007.
[9] Chuin-Mu Wang, Xiao-Xing Mai, Geng-Cheng Lin, Chio-Tan Kuo, Classification for Breast MRI Using Support Vector Machine, Proceedings of IEEE 8th International Conference on Computer and Information Technology Workshops, 2008.
[10] C. J. C. Burges. A tutorial on support vector machines for pattern recognition, data mining and knowledge discovery 2. 1998.
[11] Bishop C.M., Neural Networks for Pattern Recognition, Clarendon Press, Oxford, 1995, pp. 364-369.
[12] V. Vapnik. The Nature of Statistical Learning Theory, 2nd Edition Springer, New York, 1999.
[13] T. Joachims. Making large-scale SVM learning practical. In B. Sch¨olkopf, C. J. C. Burges, and A. J. Smola, editors, Advances in Kernel Methods—Support Vector Learning, pages 169–184, Cambridge, MA, MIT Press, (1999).
[14] Yi Liu Zheng, Y.F., One-against-all multi-class SVM classification using reliability measures, Neural Networks, 2005. IJCNN '05. Proceedings. 2005 IEEE International Joint Conference, Volume 2, 849- 854, 31 July-4 Aug. 2005.
[15] Gidudu Anthony, Hulley Gregg, Marwala Tshilidzi, Image Classification Using SVMs: One-against-One Vs One-against-All, Proccedings of the 28th Asian Conference on Remote Sensing, 2007.
[16] Gjorgji Madzarov, Dejan Gjorgjevikj, Ivan Chorbev, Multi-class Classification using Support Vector Machines in Binary Tree Architecture, International Scientific Conference, Gabrovo, 2008.
[17] Gjorgji Madzarov, Dejan Gjorgjevikj and Ivan Chorbev, A Multi-class SVM Classifier Utilizing Binary Decision Tree, Informatica 33, pages 233-241, (2009).
[18] http://www.imageclef.org/ImageCLEF2008 [01.19.2010].
[19] http://www.info-radiologie.ch/index-english.php [01.19.2010].
[20] R. Collobert, S. Bengio, J. Mariethoz, Torch: a modular machine learning software library, Technical Report IDIAP-RR 02-46, IDIAP, 2002.
[21] http://www.cs.waikato.ac.nz/ml/weka/ [01/15/2010].
[22] Manjunath, B.S., Salembier, P., Sikora, T.: Introduction to MPEG-7: Multimedia Content Description Interface. John Wiley and Sons, pp. 214-229,2002.
[23] Y. M. Latha, B.C.Jinaga, V.S.K.Reddy, Content Based Color Image Retrieval via Wavelet Transforms, IJCSNS, International Journal of Computer Science and Network Security, VOL.7 No.12, December 2007.
[24] M. Hu. Visual pattern recognition by moment invariants. IEEE Trans. Information Theory, 8(2):179–187, February 1962.