Available online at www.elixirpublishers.com (Elixir International Journal)

Digital Processing

Elixir Digital Processing 80 (2015) 30954-30959

ISSN: 2229-712X

Comparing the Accuracy of Classification Algorithms for Automatic Medical Image Annotation by Using an Improved Scale Invariant Feature Transform

Elahe Dorri Dolat Abadi and Ramin Nasiri

Department of Computer Engineering, Central Tehran Branch, Islamic Azad University, Tehran, Iran.

ARTICLE INFO

Article history:

Received: 11 January 2015;

Received in revised form:

25 February 2015;

Accepted: 5 March 2015;

Keywords

Automatic image annotation,
Scale invariant feature transform,
SMO, Nave Bayes,
J48, Decision Tree,
Random Forest, Bagging,
Ada Boost Classifiers.

ABSTRACT

Automatic annotation is in fact the process of classifying medical using global and local features of standard image codes (IRMA) while being extracted. This includes four technical data axes of providing image (modality), direction, anatomy, and biological system. A number of recent researches have been conducted on the extraction of the scale invariant feature transform for automatic annotation, but until now no complete comparison has been conducted on the accuracy of the different classifications in resolution and annotation of the images based on the scale invariant feature transform. The results from the known and famous classifiers used on the four characteristics of anatomy, direction, biological system and modality are presented in this paper, which show that Sequential Minimal Optimization is the most efficient classification group as far as accuracy is concerned.

Introduction

Automatic annotation of medical image is an emerging technology and now still is considered as an important tool for the physicians in their daily activities, which is the subject of discussion for many of the radiology physicians. And reasonable studies have been conducted in this context. Nowadays hospitals create a huge amount of the data, which is the calculated by the size of the radiological groups producing multiple tera-bytes of data year.

Apart from this, the manual annotation causes errors in the labels, such that part of the available knowledge is no more accessible to physicians, Guleld et al in 2002. [1].This calls for individuals requirements to be able to develop proper algorithms to automatically annotate the medical images. [2]. image annotation could be done in a manual, semi-automatic and automatic method. In the manual method of annotation, human power is used. The accuracy of this kind of annotation is high, but it is a boring process in the long run, and the users mostly ask to use other alternative methods. [3]. Content based image retrieval (CBIR) systems distinguish images based on their visual feature such as color, texture, and shape.[4] but mostly among the content features of lower level (color- texture-shape) and high level meaningful features used by humans, for description Of the image, there is a semantic gap. [5]. To map bridge the semantic gap, the automatic annotation is used. In the automatic annotation method, we try to produce annotation through machines; the accuracy of this method is lower than the other methods. In this method, the classification of images is totally based on features extracted by using image processing techniques and machine learning algorithms and the use of training data is done. [5].In the semi-automated method, the user's participation is required for the image annotation process. With regard to the quality of human modification, compared with manually annotated, this process has improved. [3].

Tian et al in [6], proposed combining extracted features, local binary pattern methods and MPEG-7, and SVM Classifier for automatic annotation of medical images. The best method results presented in this paper was obtained by using the 330 elements extracted feature vector. Thomas et al [7], in another system, proposed a combination of local binary pattern features [8] and produced SIFT and SVM Classifier. The method presented in this paper had the best results in the ImageCLEF2008. Dmitrovski et al [9], developed a hierarchical system for multi-label annotation of medical images. They have used different methods of feature extraction & combination and combinational classifiers of bagging and random forests.

Vailaya et al [10], used non-parametric Bayesian approach for classifying the images. Instead of segmenting, they directly cluster and calculate the image features for conditional probabilities. Dzeroski et al [11], used a combination of machine learning algorithms (Boosting) to annotate medical images. This method is based on the combined results of the weak classifiers to generate strong classifiers with very high accuracy. The purpose of this method is to combine several classifiers with less precision in order to make a highly accurate classifier.

The proposed method of this paper is based on two steps of annotation, i.e., feature extraction and classification of images. For the feature extraction part, improved scale invariant feature transform, which is one of the most powerful and most used local image features, is used; and for the classified images part, the images are classified using various machine learning algorithms. In this paper conclusions employing scale invariant feature transform of various classifiers on the four classes (modality- direction- anatomy- biological system) are investigated.

Section 2 describes the methodology used to test the data set that is an introduction of used data set, feature extraction techniques, and evaluation of the various classifiers such as support vector machine - Decision trees - Bayesian methods - Bagging and Boosting combination all methods, etc. The accuracy of the experimental results obtained from the application of these well-known provisions on the four characteristics (modality-direction- anatomy- biological system)

Tele:
E-mail addresses: elahe.dorri@yahoo.com

Testing method

This section explains the implied data set and introduces the features method extracted from the image and the reviews the different classifiers for each system data axis of annotation. The purpose of this paper is to obtain the highest accuracy in the area of classifier algorithms for automatic medical image annotation, using scale invariant feature transform.

Data set

The Image Retrieval on Medical Application (IRMA) data set is a database to create an automatic medical image annotation system which is presented by the IRMA group from the Anchen University Hospital, Germany [12]. This set is used for calling ImageCLEF medical image annotation, and has dealt with comparing the performed work from 2005 to 2009. [13]. Image classification work was started in 2005 with 57 classes; this figure raised to 116 by 2006, but since 2007 image annotation includes a 13 character IRMA code. An ImageCLEF2007 dataset which includes a 10000 training image and 10000 test images has been used for this paper.

IRMA is a 13 character code and is used to describe a class or annotation of a medical image. The Schema of IRMA code has four axes, each of which has a three to four positions. Each position has a value of a 0 to 9, a to z set given to it, in which the amount of "0" denotes "unspecified" and determines the end of a Path along the axis. These four axes are:

1- Technical axis (T, image modality) explains the method used to obtain the image.

2- Directional axis (D, body orientation) describes the direction of photographing the body organ.

3- Anatomical axis (A, body region) indicates the body organ presented in the image.

4- Biological axis (B, biological system) describes the biological system of the organ presented in the image including cardiovascular-spinal and muscular. IRMA code can be shown as follows:

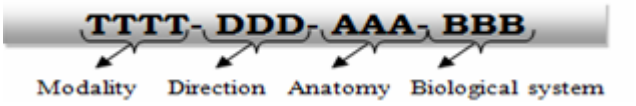


Figure 1. Schema of IRMA code [9, 2].

Examples of the annotation medical images along with the 13 character IRMA code has been shown in Figure 2.

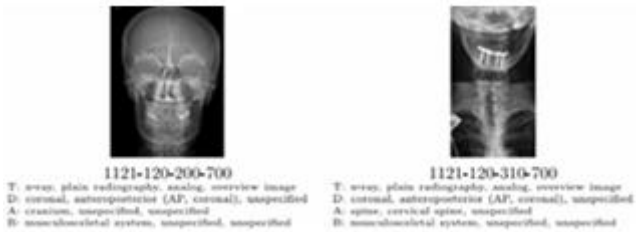


Figure 2. Examples of annotated images of ImageCLEF2007 [14].

For our project, Matlab 2013 software is used to implement the extraction of the scale invariant feature transform, and RapidMiner and Weka software has been used to measure the accuracy of the classified algorithms for automatic medical image annotation in the various information axes of the image (modality, direction, anatomy, and biological system) on the extracted data**.**

Feature extraction

In the area of image classification and retrieval, Those are shown with low level features. Since the image is formed as a set of pixels, therefore the first step to understand the image would be to extract the visual features of the pixels. Providing the appropriate features helps to improve the process of the learning techniques immensely. In each image, by using various methods, the extractable features are classified into two groups, as names Global and Local. The global features are usually not sensitive to change in spatial or locality of different images. Local features are more appropriate for explaining the details of the image. For example, the histogram color chart can be used to show or explain the global color contents of the image. As a sample, an image can be interpreted in which 40% is blue, 37% yellow, etc. Therefore, the image can be shown globally based on parts; but the choice is usually for regional Segmentation. [5]. the method of feature extraction used in this research is explained here with.

Scale invariant feature transform

One tools for image local feature description is Scale Invariant Feature Transform (SIFT). It is insensitive to the rotation conversions and image stretch and has good accuracy in object recognition, face recognition, etc. [15].

This descriptor works on the basis of extracted feature points on the images. Extraction of key points of image is a good representative for describing that object. But the number of extracted key points in images are high that requires more calculations. This problem in images with higher complexity is more apparent. The purpose of this paper is to decrease the number of feature points, using K-means algorithms clustering technique that improves the accuracy of classification and the efficiency of computational time. [9].

In this paper, a modified version of descriptor SIFT (ModSIFT) is used. In this version, SIFT rotation-invariance, not related to medical image classification, is removed and key points extraction are considered in one octave.

Local feature extraction method (SIFT) of image, is considered in this case: [9,2].

1) Extracting of 30 key points randomly from each medical image of the standard dataset training group using SIFT improved algorithm.

2) Clustering of extracted key points in the previous step using the K-Means algorithm to 500 clusters.

3) Producing a representative for each cluster in step 2 called Visual-Words.

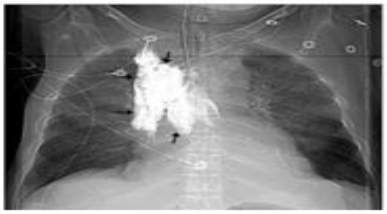
4) Determining the state of belonging of extracted feature points of test image set generated in the previous step.

(In the phase of testing new image, at first it is divided into 2 × 2 spaces and, for each blocks, 1500 key points are extracted. Then it becomes clear that each feature point belongs to which cluster. And instead of its feature vector, its (visual-word) is taken into consideration).

5) 500 bins histogram from feature extraction of improved SIFT of each block of image.

6) Creating the final 2000bins histogram from the combination of improved SIFT feature extraction of each block of the image.

Figure 11. One sample image along with the implementation of SIFT local feature extraction histogram



A). Sample image

Table 1. Result of the recognition of the modality axis using local ModSIFT feature.

feature	%Accuracy modality	Classifiers
ModSIFT	77/40	Bayes>Navebayes
ModSIFT	58/30	Trees>DecisionStump
ModSIFT	86/80	Trees>Decision Tree
ModSIFT	86/4	Trees>j48
ModSIFT	68/50	Trees>Random Tree
ModSIFT	80/50	Trees>Random Forest
ModSIFT	92/7	Bagging/WEKA
ModSIFT	86/80	AdaBoostM2
ModSIFT	94/40	SMO

Table 2. Result of the recognition of the direction axis using local ModSIFT feature.

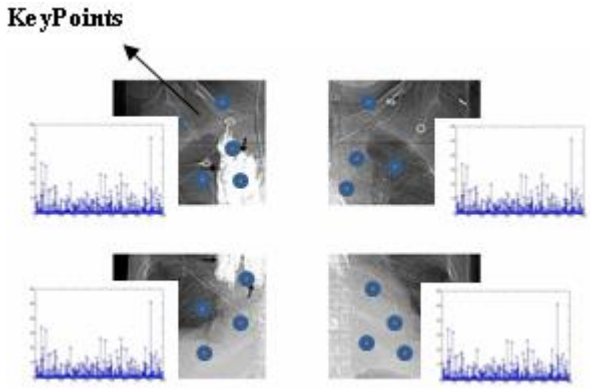
feature	%Accuracy direction	Classifiers
ModSIFT	43/20	Bayes>Navebayes
ModSIFT	25/20	Trees>DecisionStump
ModSIFT	25/20	Trees>Decision Tree
ModSIFT	48/7	Trees>j48
ModSIFT	25/20	Trees>Random Tree
ModSIFT	25/20	Trees>Random Forest
ModSIFT	58/7	Bagging/WEKA
ModSIFT	20	AdaBoostM2
ModSIFT	70	SMO

Table 3. Result of the recognition of the anatomy axis using local ModSIFT feature.

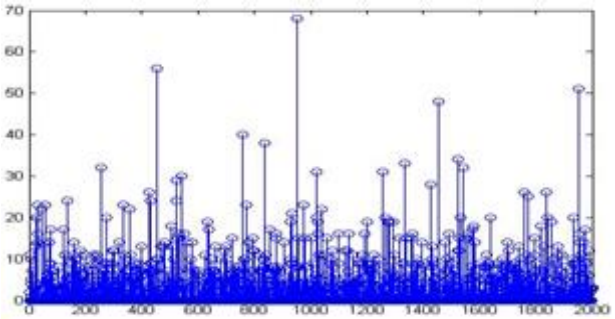
feature	%Accuracy anatomy	Classifiers
ModSIFT	62	Bayes>Navebayes
ModSIFT	38/90	Trees>DecisionStump
ModSIFT	38/90	Trees>Decision Tree
ModSIFT	44/18	Trees>j48
ModSIFT	38/90	Trees>Random Tree
ModSIFT	38/90	Trees>Random Forest
ModSIFT	57/41	Bagging/WEKA
ModSIFT	38/90	AdaBoostM2
ModSIFT	64/90	SMO

Table 4. Result of the recognition of the biological system axis using local ModSIFT feature.

feature	%Accuracy Biological system	Classifiers
ModSIFT	70/80	Bayes>Navebayes
ModSIFT	56/20	Trees>DecisionStump
ModSIFT	85/40	Trees>Decision Tree
ModSIFT	81/5	Trees>j48
ModSIFT	67/50	Trees>Random Tree
ModSIFT	79/80	Trees>Random Forest
ModSIFT	90/8	Bagging/WEKA
ModSIFT	85/40	AdaBoostM2
ModSIFT	91/07	SMO



B). Creating 500 bins histogram for each block test image.



C). Creating 2000 bins histogram from ModSIFT local feature extraction combination for each test image bock.

Figure 3. Displaying final histogram from ModSIFT local feature extraction

Algorithms of medical image classification

In this section we introduce different algorithms in order to compare the accuracy of the classifiers for automatic medical annotations. They are briefly explained as follows:

Bayesian group

In the multi label annotation, an image is labeled by multi content or multi class. The meaning of multi-label method is connected with the learning discussion of multi-instance multi-label (MIML). In MIML an image is shown with a bag of features or a bag of regions. A typical MIML, using probabilistic tools, such as the Bayesian method, is achieved. The Bayesian methods work by finding the posterior probability that an image belongs to any particular concept. The Bayesian method explicitly needs all the hypotheses in the model, and it is then used optimally to extract classifier rules. The Bayesian method, with the help of training data for calculating the probability from each class with regards to special vector is made of a new sample. [5].

If a set of images {I1,I2, . . . , In**}** is taken into consideration from a set of {Cn,….,C2,C1}, the Bayesian models are trying to procure posterior probability from conditional and priors probabilities. We assume that an image I, with feature vector **x** is shown. We also have the prior probability P(Ci) and density of conditional probability P(X/Ci).

The probability that an unknown image I belongs to a Ci class is obtained from the following relation. [5].

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)} \quad (1)$$

Nave Bayes

This is one of the oldest official classification algorithms from the Bayesian group. This classification is a family from the probabilistic simple classifications based on using the Bayes' theorem with independent and strong hypotheses (Nave) between the features. [16].

The Nave Bayes model is known under different names such as simple Bayes and independent Bayes. All these names used from the Bayes' theorem in the decision making rules in classification. Nave Bayes has been studied widely since 1950 and was introduced with popular names in the text retrieval society from early 1960. This is a popular method (basic) for classification of texts. Nave Bayes is scalable and needs only a few linear parameters for a few variables (predications/features) in a Learning problem.

The advantage of Nave Bayes is that it needs only a few of the training data to estimate the average parameters and variances for classification.

Trees group

The decision tree is a tool of classification or a multi-level decision making. Based on some of the decisions made in each of the inner nodes, a decision tree (DT) can be a binary tree or n-ary. A decision tree classifies a sample by arranging them through a tree to an appropriate leaf node. Each node shows a number of features from the sample, and each branch is connected to some probability for this features. A number of DT algorithms are used in this literature which include j48, Decision Stump, C4.5, etc. these algorithms of DTs are different as far as the kind of feature, selection criteria, result, etc., are concerned.[5].

Decision stump

The decision stump classifier is a learning machine model comprising a one level decision tree. This means that it is a decision tree with an inner node (root) which immediately connects with the terminal nodes (leaf). The decision stump predicts with one of the input features. Sometimes it's called rules.

The decision stump classifier is often used like a basic learner or a weak learner in the complex learning machine technique such as Boosting and Bagging.

J48

J48 is another algorithm from the decision tree set, and makes a pruned or non-pruned tree from the C4.5.C4.5 tree starts from the root which shows all the data and in return divides them into smaller sets. This is done by testing each of the two groups. This process continues until all the samples are grouped together, at this time the tree stops growing.

Group support vector machine

The group support vector machine (SVM) is a supervised classifier which can learn from the samples and make decisions for the new samples. SVM can classify the data in a linear and non-linear method. In producing the SVM model in a linear manner, by finding an optimal hyperplane, the data is separated without error and maximum distance between the planes and the nearest training points (supporting vectors). Figure 4 shows this process. All the samples have label -1 on one side of the hyperplane and all the samples have +1 label on the other side. The training samples close to the hyperplane, is named supporting vector. The number of these supporting vectors are in comparison with the size of the training set is small and decide on the border, hyperplane and as a result the decision surface.

In a case where the data cannot be separated in a linear manner, they are mapped in a more dimension space so that they can be separated in a linear manner in a new space.

For the implementation of SVM, with the purpose of maximizing the distance between the classes, the (SMO) version is used. The sequential minimum optimization (SMO) algorithm is a simple algorithm which can solve the Quadratic

programming (QP) problems, which arise during the SVM training, quickly.

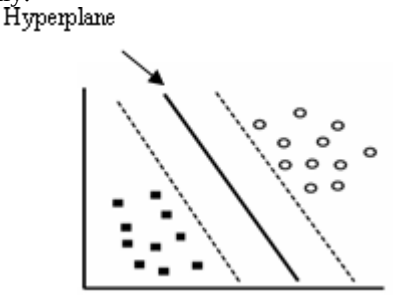


Figure 4. production of linear SVM model [17].

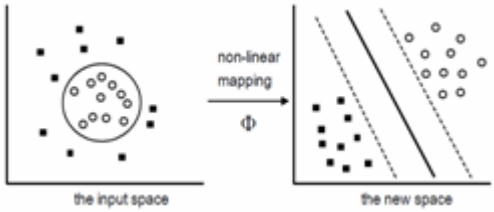


Figure 5. production of no- linear SVM model [17].

The SMO algorithm was introduced by John- Platt in 1998 in Microsoft's researches. [18]. The SMO, larger QP problems are broken into smaller possible QP problems. These small QP problems are solved analytically. The time for calculating the SMO is determined through SVM evaluation. Therefore, SMO is very fast for the linear SVM and the dispersed data set. The main idea of SMO is that after normalizing the data, the classes are mutually separated by the same method SVM.

Combined Group Classifiers

Recently, in the pattern recognition and machine learning domains, a combination of a number of classifiers has been known as an active research area. They can be called group or modular classifiers. The purpose of these modular classification is to obtain highly accurate results by combining weak classifier results.[19]. Voting, Bagging and Boosting methods are a combination of techniques that we introduce here:

Majority voting method

Majority voting is the simplest method to combine the classifiers. In this method, the binary output of K classifier is separately combined together. Then the class with the highest number of votes will be chosen as the final classification decision. In general, the final classification decision of the majority vote is taken from K + 1/2. Figure 6 shows the overall architecture of the modular classifier.

Figure 6. The overall architecture of the modular classifier. [17].

In figure 6, a number of different nervous networks (for example EXPERTS) is trained and the inputs are shared; the output of each of these is combined to create the final output.

Bagging method

To improve the results of single classifier, researchers often focus on combination approaches such as Bagging. In the Bagging method, several networks are independently trained through the method of Bootstrap. [20]. Figure 7 shows the Bagging method of classification.these is combined to create the final output.

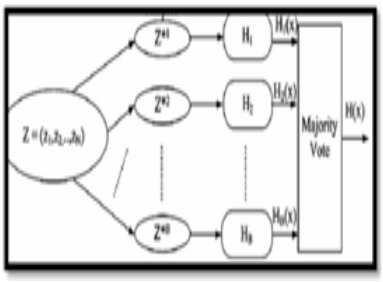


Figure 7: Classification of Bagging. [21].

In Figure 7, at first a number of training samples are generated. Each sample is given to a weaker classifier. The final classifier decision is made by the majority vote on a combination of weak classifier result output.

Boosting method

Boosting is a general method for improving the performance of each learning algorithm. This approach can be used to significantly reduce errors in any "weak" learning algorithm, which constantly produces a strong classifier. Various versions of learning algorithms based on Boosting is available; Adaptive Boosting is the most popular among them (AdaBoost). For classification with two classes or two sets, 'AdaBoostM1' , 'LogitBoost' , 'GentelBoost' 'RobustBoost', 'LPBoost', 'TotalBoost', and 'Bag' can be used. For classifiers with three or more classes only 'AdaBoostM2', 'LPBoost', 'TotalBoost', 'Bag' can be used. In this paper, due to the classification of more than three classes, a copy of 'AdaBoostM2' is used. AdaBoost is a combination of Bagging and Boosting.

AdaBoost, performs the learning classifier on the weighted versions of the training set. Samples which are wrongly classified at each step will be assigned more weight in later stages. The final classifier is the linear combination of the basic classifiers. [22].

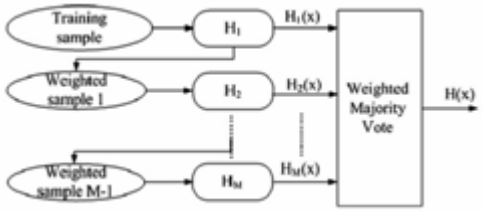


Figure 8: The overall architecture of the Adaboost classifier. [21].

Random Forest

Random Forest is another combination classifier method for classification and regression which is constructed by several decisions trees during training and its output classes is obtained by individual tree outputs. This method combines Bagging with random selection of features to provide the most diverse set of decision trees. The advantage of this classifier is its very high accuracy among current algorithm classifiers and it runs effectively on large databases. This classifier handles thousands of input variables and is an efficient method to estimate data and accuracy when a large proportion of the data is missing. [9, 23].

Results and Evaluation

The overall recognition rate is a common method and is widely used for evaluation measures. It is a fraction of the test images which is predicted correctly by the 13 character IRMA code [9].

$$accuracy = \frac{correctly\ annoteted\ images}{total\ tested\ images} \quad (2)$$

In this method of evaluation, as a first step, one train the classifiers with training images, and after using the test images, the accuracy of the classification algorithm based on the extraction criterion is tested.

Each of the four axes of the 13 character IRMA code is reviewed, to define that for how many images of total tested images, different parts of the code have been correctly identified. For example, from among the 150 tested images in 95% of the images the first code was correctly identified, 92% of the images the second code was correctly identified; the correct identification result was 96% and 90% for the third and fourth codes respectively.

Tables1, 2, 3 and 4 show the results of the classification of (modality-direction-anatomy-biological system) using scale invariant feature transform and distinct classifiers.

Conclusions

To annotate medical images, the aim is to produce four data axes including modality, direction, anatomy, and biological systems.

In this paper, the method of image feature extraction (SIFT) is introduced, and distinct classifiers for each data axes of annotation are reviewed in order to compare the accuracy of the classifier algorithms for automatic medical image annotation. The results indicate that the SMO classifier group provides the best results in classification (modality, direction, anatomy, and biological systems).

References

[1] M. O. Gueld, M. Kohnen, D. Keysers, H. Schubert, B. B. Wein, J. Bredno, and T. M. Lehmann, "Quality of DICOM header information for image categorization," in Medical Imaging 2002, pp. 280-287, 2002.

[2] T. Tommasi, F. Orabona, and B. Caputo, "Discriminative cue integration for medical image annotation," Pattern Recognition Letters, vol. 29, pp. 1996-2002, 2008.

[3] S. K. Kharkate and N. J. Janwe, "Automatic Image Annotation: A Review," International Journal of Computer Science & Applications (TIJCSA **,(**vol. 1, 2013.

[4] T. Pavlidis, "Limitations of content-based image retrieval," in Invited Plenary Talk at the 19th Internat. Conf. on Pattern Recognition, Tampa, Florida, December, pp. 8-11, 2008.

[5] D. Zhang, M. M. Islam, and G. Lu, "A review on automatic image annotation techniques," Pattern Recognition, vol. 45, pp. 346-362, 2012.

[6] G. Tian, H. Fu, and D. D. Feng, "Automatic medical image categorization and annotation using LBP and MPEG-7 edge histograms," in Information Technology and Applications in Biomedicine, 2008. ITAB 2008. International Conference on, pp. 51-53, 2008.

[7] T. Tommasi, F. Orabona, and B. Caputo, "An SVM confidence-based approach to medical image annotation," in Evaluating Systems for Multilingual and Multimodal Information Access, ed: Springer, pp. 696-703, 2009.

[8] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 24, pp. 971-987, 2002.

[9] I. Dimitrovski, D. Kocev, S. Loskovska, and S. Džeroski, "Hierarchical annotation of medical images," Pattern Recognition, vol. 44, pp. 2436-2449, 2011**.**

[10] A. Vailaya, M. A. Figueiredo, A. K. Jain, and H.-J. Zhang, "Image classification for content-based indexing," Image Processing, IEEE Transactions on, vol. 10, pp. 117-130, 2001**.**

[11] T. G. Dietterich, "Ensemble methods in machine learning," in Multiple classifier systems, ed: Springer, pp. 1-15, 2000.

[12] T. M. Lehmann **,**H. Schubert, D. Keysers, M. Kohnen, and B. B. Wein, "The IRMA code for unique classification of medical images," in Medical Imaging 2003, pp. 440-451, 2003.

[13] ImageCLEF – The CLEF Cross Language Image Retrieval Track, . Available: wwww:imageclef.org **,**2009.

[14] T. Tommasi, B. Caputo, P. Welter, M. O. Güld, and T. M. Deserno, "Overview of the CLEF 2009 medical image annotation track," in Multilingual Information Access Evaluation II. Multimedia Experiments, ed: Springer, pp. 85-93, 2010.

[15] D. G. Lowe, "Distinctive image features from scale-invariant keypoints **",**International journal of computer vision, vol. 60, pp. 91-110, 2004**.**

[16] D. J. Hand and K. Yu, "Idiot's Bayes—not so stupid after all?," International statistical review, vol. 69, pp. 385-398, 2001**.**

[17] C.-F. Tsai and C. Hung, "Automatically annotating images with keywords: A review of image annotation systems," Recent Patents on Computer Science, vol. 1, pp. 55-68, 2008**.**

[18] J. Platt, "Sequential minimal optimization: A fast algorithm for training support vector machines," 1998**.**

[19] S. Džeroski, P. Panov, and B. Ženko, Machine Learning, Ensemble Methods in: Springer, 2009**.**

[20] L. Breiman, "Bagging predictors," Machine learning, vol. 24, pp. 123-140, 1996.

[21] A. J. Ferreira and M. A. Figueiredo, "Boosting algorithms: A review of methods, theory, and applications," in Ensemble Machine Learning, ed: Springer, pp. 35-85, 2012.

[22] S **.**Tulyakov, S. Jaeger, V. Govindaraju, and D. Doermann, "Review of classifier combination methods," in Machine Learning in Document Analysis and Recognition, ed: Springer, pp. 361-386, 2008.

[23] T. M. Khoshgoftaar, M. Golawala, and J. Van Hulse, "An empirical study of learning from imbalanced data using random forest," in Tools with Artificial Intelligence, 2007. ICTAI 2007. 19th IEEE International Conference on, pp. 310-317, 2007.