Algorithms and methods of image classification for automated medical systems

I V Kovalev1,4,5, I G Shelomentseva1,3, N V Yakasova1,2 and A A Voroshilova4

1 Siberian Federal University, 79, Svobodny ave., Krasnoyarsk, 660041, Russia
2 Khakassia state University named after N. F. Katanov, 90, Lenin Ave., Abakan, Republic of Khakassia, 655017, Russia
3 Krasnoyarsk State Medical University named after Professor V.F. Voino-Yasenetsky, 1, Partizan Zheleznyak ave., Krasnoyarsk, 660022, Russia
4 Krasnoyarsk Regional Science and Technology City Hall, 61, Uritskogo Str., Krasnoyarsk, 660049, Russia
5 Krasnoyarsk State Agrarian University, 90, Mira ave., Krasnoyarsk, 660049, Russia

E-mail: inga.shell@yandex.ru

Abstract Modern medicine in the diagnosis of diseases increasingly relies on the use of automated image recognition systems. Automation is of particular importance in the diagnosis of socially significant diseases, when the quality and speed of diagnosis is important. The pathology processes imposes certain restrictions on the methods used, which are associated both with image quality and with the speed of analysis. The article discusses the algorithms and methods for classifying medical images as the basis for automated diagnostic systems; Criteria for choosing among them those methods that satisfy the requirements of the task are determined. Based on a comparison of compliance with the criteria of the methods, neural network methods were selected: a neuro-fuzzy network, a convolutional neural network, and self-organizing maps.

1. Introduction

Digitalization of medicine is closely related to research in the field of digital image processing and analysis. The need to attract computer technology directly follows from the needs of automation and optimization of the medical diagnostic process, the speed of diagnosis, the prevention of complications. The basis of automated medical image analysis systems is the classification subsystem.

The processing and analysis of medical data has been studied for a long time, for which a large number of classification algorithms and methods have been proposed. In [1], a group of researchers led by Professor A.V. Lapko conducts the following systematization of methods and algorithms for image classification: learning methods with a teacher, without a teacher, with an non-ideal teacher, algorithms using the principles of optimal systems, nonparametric algorithms and hard-to-formalize algorithms, taking as a criterion a priori information about classes and the decisive boundary between them.

Learning methods without a teacher (automatic classification algorithms) are characterized by the absence of a priori information about classes and the equation of the dividing surface between classes. Methods of learning with a teacher contain not only knowledge about classes and the dividing surface,
but also are characterized by the presence of a sample with precise instructions from the teacher and a sign of the effectiveness of the equation parameters. Methods with a non-ideal teacher contain fuzzy instructions in their learning set about points belonging to classes in the presence of other parameters, or are determined by the presence of only a priori information about classes and the measure of proximity between signals in the feature space.

Collective algorithms contain information about the parameters of the dividing surface equation, but do not contain a priori information about classes. The statistical classification algorithms include probabilistic knowledge of the class and the parameters of the equation of the dividing surface with the criteria of efficiency and optimality.

The authors separately consider the adaptive Bayesian approach, which is applicable in three cases: when there are probabilistic characteristics of classes and an optimality criterion, but the criterion of compactness of points and the number of classes are unknown, when there is a form of the equation of a separating surface, but the criterion of compactness of points and the number of classes are unknown, or when there are equation parameters and common performance criteria.

Another classification of the methods of intellectual processing of medical data distinguishes only 2 groups - learning methods with a teacher, which include decision trees, neural networks, discriminant analysis methods, boosting methods, a naive Bayesian classifier and methods of classification without a teachers, where the authors included clustering methods and self-organizing maps of Kohonen [2].

Pathological processes for the analysis of which medical diagnostic systems are used can have blurred borders, different localization, are characterized by discreteness of tissue damage, and may overlap with affected or healthy tissues. Pathological processes for the analysis of which medical diagnostic systems are used can have blurred borders, different localization, are characterized by discreteness of tissue damage, and may overlap with affected or healthy tissues. The choice of algorithms or methods is determined by the specific task.

In a review article [3], devoted to the use of various models of neural networks for image-based diagnostics, the problems of using other methods of intelligent analysis — neuro-fuzzy networks, genetic algorithms and simulation tools as «promising modern technological tools» are discussed.

Based on the task of creating an automated medical diagnostic system for analyzing microscopy according to the Ziehl-Neelsen method, it is necessary to consider well-known classification methods and algorithms as applied to the task, and determine the basic criteria for their comparison.

2. Materials

The study materials are a set of microscopic images of pathology, colored by the Ziehl-Neelsen method and obtained using a trinocular microscope Micromed 1 var. 3-20 at 10x60 magnification with a digital camera Toupcam UCMOS01300KPA with a resolution of 0.3 MP. Treatment of sputum by the Ziehl-Neelsen method involves treatment with carbolic fuchsin with further discoloration with a 5% solution of sulfuric acid or 3% hydrochloric alcohol and finishing with a 0.25% solution of methylene blue. Thus, the red dye is removed from the present microflora, except for acid-resistant mycobacteria. Microscopy provides additional restrictions on the execution time, which should not exceed 5 minutes.

3. Methods

In the problem we are considering, there is information about the number of classes, there is a training sample, teacher instructions may be fuzzy. Also, in the application of statistical methods, it is possible to determine the probabilistic characteristics of classes and the optimality criterion. Based on the classifications considered, we distinguish the following main groups of methods - automatic classification, learning with a teacher (ideal and non-ideal), and statistical classification.

The learning methods with the teacher classically include artificial neural networks - a multilayer perceptron, a probabilistic neural network (PNN), a radial basis neural network (RBF), and a fuzzy neural network (FNN). Among the methods of learning without a teacher, they most often consider the
Kohonen self-organizing map (SOM), the k-means clustering method, and the support vector method. Statistical classification methods include the naive Bayesian classifier, the EM algorithm, nonparametric classification methods, and discriminant analysis (table 1).

| Group                              | Methods                                                                 |
|------------------------------------|-------------------------------------------------------------------------|
| Unsupervised learning              | self-organizing Kohonen map (SOM), K-means clustering method, support vector method, FOREL, DBSCAN, ISODATA fuzzy neural network (FNN) |
| Not-ideal supervised learning      | multilayer perceptron, probabilistic neural network (PNN), radial basis neural network (RBF) |
| Supervised learning                | naive Bayesian classifier, EM algorithm, nonparametric classification methods, discriminant analysis |
| Statistical Methods                |                                                                         |

### 3.1. Supervised learning
One of the most famous methods used in the theory of pattern recognition is the multilayer perceptron (MLP). A multilayer perceptron is an artificial neural network (ANN) consisting of input layer, one or more hidden layers, and an output layer of neurons. A multi-layer perceptron is characterized by the use of nonlinear activation functions and an error back propagation algorithm. The advantage of using the perceptron is its algorithmic versatility and compact structure, which allows identify patterns in any data set. The disadvantages of using a perceptron are technical difficulties related to memory capacity and training speed.

The main idea of RBF neural networks is that nonlinear transformation of network parameters into a higher-dimensional space will increase linear separability. The architecture of hidden elements of the RBF network is built on the basis of radial basis functions, which are based on the multivariate Gaussian distribution [4]. The advantage of using RBF neural networks is the presence of only one hidden layer, the simplicity of the algorithm for optimizing weight coefficients, and high training speed. The limitations of RBF neural networks include sensitivity to noise and the complexity of configuring parameters of the functions.

Probabilistic neural network (PNN) is an implementation of kernel approximation techniques designed as a two-layer neural network. For example, the Gaussian function is often chosen as the kernel function. Advantages of the PNN network are probabilistic interpretation of the result and high training speed. A significant disadvantage of PNN is their size, since this network contains all the training data. PNN networks are useful in the selection of informative features and in test experiments, since since a large number of tests can be performed quickly due to the short training time [5].

Convolutional neural network (CNN) consists of many layers, whose task is to perform convolution with the selected kernel, get a feature map, reduce the dimension of layers, or perform classification [6]. The advantages of convolutional neural networks over conventional neural networks are smaller count of parameters and a higher training speed, as well as the possibility of parallel computing. The disadvantages of using a convolutional network include the probable loss of information during the downsampling, the fluctuation of the solution when the object is rotated and the lighting conditions changes [7].

Extreme Learning Machine (EML) is a method based on direct neural networks with one hidden layer [8]. The elements of the hidden layer can be artificial neurons and other classical algorithms of the theory of pattern recognition - Fourier and Laplace transforms, sigmoidal, wavelet and threshold networks, RBF networks, neuro-fuzzy networks. The advantage of the EML algorithm is a faster training speed and better generalization performance compared to the basic hidden layer algorithms. The disadvantages of the algorithm are the initialization by random or pseudo-random numbers, which gives the dependence of the algorithm on the initial data.
3.2. Not-ideal supervised learning

Fuzzy classification and clustering are methods of not-ideal supervised learning. Fuzzy neural networks combine the advantages of fuzzy inference systems and neural networks. The structure of neuro-fuzzy networks consists of several layers: the outputs implementation fuzzy logic based on the concepts of fuzzy inference, t-norm and t-conorms, the other layers implement the mechanisms of neural networks [9]. The advantages of a fuzzy neural network are structure transparency, the use of neural mechanisms to determinate fuzzy parameters, and a pre-defined network size. The disadvantage of fuzzy neural networks is the necessity to determine the input data [10].

3.3. Statistical methods

The basis of statistical classification methods is the Bayesian approach, which is based on the statement that if the distribution densities of the classes are known, then the dividing surface can be written in an explicit analytical form, which is easily implemented algorithmically [11]. Also, the construction of the Bayesian classifier is based on the error probability minimization, which determines its optimality for using as a standard of comparison for other classification methods. However, in practice, the distribution densities of the classes are unknown and it have to be determined by parametric, nonparametric and hybrid methods on the test data.

Naive Bayesian classifier is based on the assumption that the used features of objects are independent random variables. This assumption simplifies the construction of the classifier, since in this case it is required to estimate not one large n-dimensional probability density, but n one-dimensional densities, which entails efficiency in terms of processing speed and ease of implementation. The disadvantages of the naive Bayesian classifier will be a large classification error, because in practice there are few classification problems for which objects really contain independent features.

EM algorithm consists of two steps: the E step and the M step. At step E, the expected value of the likelihood function is calculated provided that the set of hidden variables is considered to be the set of observed variables. At the M step, the maximum likelihood score is calculated, which increases the expected likelihood of step E. In the future, the E step uses this value for its own calculations. The K-average method can be used for initial values. EM algorithm is performed until convergence [12].

The advantages of using the EM algorithm in its rigorous mathematical basis, noise resistance, fast convergence at suitable initial values. Among the disadvantages of this method are the dependence on initialization, work with data of only a normal distribution, and instability in the vicinity of a local minimum [13].

The task of discriminant analysis (DA) in determining the belonging of an object to a certain group based on the analysis of the average value of a variable or group of independent variables. In this case, the groups are set in advance, and the main task is to classify a new object on the basis of an a posteriori classification of primary information [14]. The disadvantages of discriminant analysis are the fact that the results of a priori classification will always be worse than a posteriori, as well as the restriction of the primary sample to a normal distribution. Among the advantages of discriminant analysis, we note the possibility of assessing the stability of the classification, restoring missing values, and a deeper understanding of the differences between classes of objects.

3.4. Unsupervised learning

DBSCAN is a density clustering algorithm that allows finding clusters of arbitrary shape, each of which is a collection of vertices of a connected graph [15]. Two points will be the vertices of the graph if the metric distance between them does not exceed a given value. If an object does not contain neighbors, then it is recognized as an outlier. The advantages of this algorithm will be noise resistance, search for previously undefined data clusters, a balanced computing process, and the main disadvantages are the dependence of the quality of the algorithm on a priori data and initial values, sensitivity to dimensional curse.
Linear (threshold) classifier method constructs a linear dividing surface, presented either as a hyperplane or as a piecewise linear surface. The advantages of a linear classifier can be called the ease of its implementation using programming languages, a wide range of tasks. Disadvantages - with incorrectly selected initial data, slow convergence is possible, the problem of a local minimum, in addition, this algorithm is difficult to retrain. Special cases of the linear classifier are support vector methods and a single-layer perceptron.

The support vector machine (SVM) is designed to separate image objects into classes based on the distance mark of the dividing surface from all points under consideration [16]. In contrast to neural network methods, the support vector method finds several classification options and selects the most optimal one, I achieve zero error on the training set. The main disadvantage of this method is noise instability.

Decision trees are a binary structure consisting of branches, leaves and nodes, where the leaves contain the target classes, and the nodes are the conditions for branch transitions [17]. When building a decision tree, it is important to define criteria for selecting attributes, criteria for stopping tree construction, and criteria for clipping branches. When building a decision tree, it is important to define criteria for selecting attributes, criteria for stopping tree construction, and criteria for clipping branches. Decision trees are characterized by visibility, learning speed, ease of constructing disjunctive normal form, and the ability to build nonparametric models. However, there is a problem of building a tree on a large data set to ensure the necessary classification accuracy.

Kohonen self-organizing map contains only two layers — the input and output layers, organized in the form of a topological map, most often two-dimensional [18]. Learning is an iterative procedure for grouping N-dimensional input values into two-dimensional output clusters. With each learning epoch, the neighborhoods and neurons that are closest to the centers of the sought clusters are determined, and they are corrected using the example of the training sample. Then a new epoch is started, with a reduced value for the learning speed and neighborhood dimension for more accurate clustering. The advantages of the Kohonen network are its features such as the breadth of the tasks to be solved, the visualization of multidimensional data, noise resistance, and fast learning speed.

ISODATA clustering algorithm is based on dividing the set of images into classes, based on the proximity of points by spectral distance [19]. Unlike the K–means and C-means clustering algorithms, the ISODATA algorithm does not require initial setting of the algorithm parameters. The k – means algorithm constantly redefines the centers of clusters and the affiliation of vectors in accordance with the principle of minimum of their square deviation. The c-means algorithm searches for fuzzy clusters and their centers with the necessary accuracy using the same algorithm.

FOREL is another clustering algorithm that is based on the minimality criterion for the sum of distances from the object to the centers of clusters found at the moment. The advantage of this algorithm will be convergence, visibility, the ability to check and calculate intermediate values [20].

K nearest neighbors algorithm (k-NN) is another simple clustering algorithm based on Euclidean distance and estimation of similarity of objects [21]. Simplicity of implementation of this algorithm is its advantage, and low efficiency and accuracy are disadvantages. The disadvantage of the considered clustering algorithms is their instability, that is, dependence on a given initial approximation and low performance.

To solve the problems of artificial intelligence, genetic algorithms have recently been widely used, which are built on the cyclic repetition of the main evolutionary steps - reproduction, mutation, crossbreeding, selection. A feature of the genetic algorithm is the use of the crossover operation, which is understood as the fixing in further populations of those genes that most closely correspond to the optimization problem [22]. In comparison with the known methods for solving a specific problem, the genetic algorithm has a greater time complexity and a sufficiently high computational resource consumption. Another disadvantage of genetic algorithms is the fact that it does not guarantee that the solution to the problem is a global, not a local optimum. The advantages of using the genetic algorithm are its universality and ability to find a solution, albeit an approximate one, under conditions of uncertainty.
There is a practice of using genetic algorithms to configure or search for optimal parameters of neural and neuro-fuzzy networks. The combination of genetic algorithms and neural networks is known in the literature under the abbreviation COGANN (Combinations of Genetic Algorithms and Neural Networks). For example, in [23], for the diagnosis of tuberculosis, the authors used a combined approach that includes fuzzy logic, neural networks, and genetic algorithms. The task of the genetic algorithm was to optimize the vector of signs to N signs sufficient for the correct diagnosis of the disease.

4. Conclusion

The considered methods have different sets of qualities, advantages and disadvantages, in conjunction with the statement of the problem, this requires formulating a number of criteria for choosing a classification method for an automated diagnostic medical system (table 2).

- (1) ability to work with images – ability to get an image input, without preprocessing, highlighting features and areas;
- (2) high speed of work on fuzzy data - in accordance with the statement of the problem, the image should be analyzed for a limited time;
- (3) noise resistance - the ability of the method to give a reliable result with significant image noise;
- (4) independence of the operating time from the quality (fuzziness, noise, etc.) of a particular image - the exclusion of the impact of image quality on the time of its processing;
- (5) universality in the construction of the dividing surface of any complexity - reflects the ability of the system to work with complexly organized classes, which in real-life tasks, including in the analysis of medical images, is quite common;
- (6) ease of implementation - a criterion that can be described as the ability to quickly implement a method using ready-made tools.

Table 2. Comparison of classifiers for the selected properties.

| Algorithm or method | (1) | (2) | (3) | (4) | (5) | (6) |
|---------------------|-----|-----|-----|-----|-----|-----|
| MPL                 | +   | +   | +   | +   |     |     |
| RBF                 | -   | +   |     |     |     |     |
| SOM                 | +   | +   | +   | +   |     |     |
| CNN                 | +   | +   | +   | +   | -   | -   |
| PNN                 |     | +   |     |     |     |     |
| FNN                 |     |     |     | +   | +   | +   |
| DBSCAN              |     |     |     |     | +   | +   |
| Linear classifier   | -   | -   |     |     |     |     |
| Decision trees      | -   | -   |     |     | +   |     |
| ISODATA             | -   | -   |     |     | +   |     |
| FOREL               |     |     |     |     |     | +   |
| Naive Bayes         | -   | -   |     |     |     |     |
| k-NN                | +   |     |     |     |     |     |
| SVM                 |     |     |     |     | +   |     |
| EM algorithm        | +   |     |     |     |     |     |
| DA                  |     |     |     |     |     |     |
| EML                 | +   |     |     |     |     |     |

Table 2 clearly shows that neural network methods have advantages according to the selected criteria. Analysis for the maximum number of matches to the criteria allows you to identify the most appropriate classification methods to solve the problem - neuro-fuzzy network, self-organizing maps and convolutional neural network.

Indeed, convolutional networks were originally created for working with images, they do a good job with fuzzy images, demonstrate a high classification speed, which does not depend on noise and
blurry images. Neuro-fuzzy networks, although they are not able to work directly with the image, have advantages in working with noisy and fuzzy images.

References
[1] Lapko A V, Polikarpov L S, Manchuk V T 1996 *Automation of scientific research in medicine* (Novosibirsk: Nauka) p 170
[2] Baranov A A, Namazova-Baranova L S, Smirnov I V, Devyatkin D A, Shelmanov A O, Vishneva E A, 2015 *Trudy ISA RAN* **65(2)** 81–93
[3] Payal D, Purva S 2018 *Tuberculosis* **108** 1-9
[4] Kruglov V V, Borisov V V 2002 *Artificial Neural Networks. Theory and Practice* (Moscow: Hotline-Telecom) 39-41
[5] Islam M, Chen G, Jin S 2019 *American Journal of Neural Networks and Applications* **5(1)** 7-11
[6] LeCun Y, Bottou L, Bengio Y, and Haffner P 1998 *Proceedings of the IEEE* **86(11)** 2278–324
[7] Zhang A, Lipton Z, Li M and Smola A 2020 *Dive into Deep Learning Preprint* d2l.ai
[8] Huang G B 2012 *IEEE Transactions on Systems, Man and Cybernetics* **B(42-2)** 513–29
[9] Terano T, Asai K, Sugeno M 1989 *Applied Fuzzy Systems* (Academic Press) p 314
[10] Huang M L, Chen H Y, Huang J J 2007 *Expert Systems with Applications* **32** 458–68
[11] Dubois D, Prade H 2003 *Artificial Intelligence* **148** 1-9
[12] Achcar J A, Ruffino Netto 2003 *A Rev. bras. epidemiol* **6(4)** 380-7
[13] Gupta M R, Chen, Y 2010 *Foundations and Trends in Signal Processing* **4(3)** 223–96
[14] Choi S C, Rodin E Y 1986 *Statistical Methods of Discrimination and Classification* (Pergamon) p 143
[15] Ester M. et al. 1996 *Kdd* **96(34)** 226-31
[16] Fradkin D, Muchnik I 2006 *Discrete Methods in Epidemiology* **70** 13–20
[17] Gerald LB, Tang S, Bruce F, Redden D, Kimerling ME, Brook N, Dunlap N, Bailey WC. Am J 2002 *Respir Crit Care Med* **166(8)** 1122-7
[18] Malone J, McGarry K, Wermter S, Bowerman C 2006 *Neural Computing and Applications* **15** 9-17
[19] Tou JT, Gonzalez RC 1981 *Pattern Recognition Principles* (Addison Wesley, Reading, MA) p 378
[20] Shapiro L G, Stockman G C 2001 *Computer Vision* (Pearson Education) p 608
[21] Gonzales RC, Woods R E 2002 *Digital Image Processing* (Pearson Education) p 965
[22] Holland J H 2000 *Evolutionary computation* **8** 373-91
[23] Omisore MO 2017 *Applied Computing and Informatics* **13** 27–37