

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Hybrid Intelligent System for Diagnosing Breast Pre-Cancerous and Cancerous Conditions Based on Image Analysis

Oleh M. Berezsky

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/intechopen.72576>

Abstract

Modern diagnostic technologies are automated microscopy systems (AMSs). In this research study, the authors analyzed the modern AMS methods and algorithms. Criteria-based comparative analysis of AMS has been made, and their advantages and disadvantages have been identified at the three main levels of image processing. This allowed determining the main direction of such systems development, that is, designing the hybrid intelligent AMS. The work of an expert physician implies visual image interpretation, selection of qualitative features of micro-objects, the formation of diagnostic rules based on expert knowledge, and making diagnoses. Knowledge introduction model contains a productive model, in which knowledge is presented in the form of rules expressed in productive pseudo code if-then. Logic inference machine is a module designed to logically derive the facts and rules from the base according to the laws of formal logic. A set of modern methods and algorithms for low-, mid-, and high-level image processing have been used in the AMS structure.

Keywords: breast cancer, histological image, cytological image, fuzzy knowledge base, fuzzy system, automated microscopy system, hybrid intelligent system, convolutional neural networks

1. Introduction

According to the National Cancer Registry of Ukraine, in 2014 there have been 14,908 tumor cases. Moreover, almost 25% of tumors have been diagnosed at a late stage, and 40% of women over 40 years of age have never been properly examined [1].

As a rule, in modern clinical practice, light microscopy is used in diagnosing, which is the area of laboratory diagnostics, where labor-intensive subjective qualitative analysis dominates. For automation of microscopic studies, automated microscopy systems (AMSs) are used. AMSs are software and hardware complexes for digital micro-object processing [2]. The main problem with such systems is image processing quality.

Automated microscopy systems allow conducting microscopic image analysis, selecting objects in manual or automated mode, calculating certain characteristics, and assisting medical diagnosticians to make diagnoses based on these characteristics. Examples of such systems are the following: AxioVision, BioImageXD, ImageJ, MicroManager, MECOS-CH, and others. Some of the mentioned systems have their own hardware (microscopes, photo/video cameras, etc) for research, but most are universal and adaptable to different types of microscopes.

The urgent research area in the application of automated microscopy systems is the development of hybrid intelligent systems that allow automatic image processing to make diagnoses. Such systems are called hybrid because they combine two or more intellectual components.

2. AMS concepts and application areas

Modern automated microscopy systems are characterized by a high cost, high level of complexity, and rigid user interface. In addition, such systems allow image processing only in manual or automated modes, however, they are not able to make diagnoses in automatic mode, which would greatly simplify and improve diagnostic physicians' performance. One more significant issue is a reduction of image processing labor-intensiveness and using modern methods or algorithms of computer vision.

Modern automated microscopy systems for processing the biomedical images of various types, including cytological and histological images, are becoming increasingly popular. Most AMSs consist of hardware (microscope, video camera) and software. The main task of software is to process the input image and identify the objects and features for further diagnosis by an expert or in automatic mode. The most popular systems include: MECOS-C2, TissueFAXS, AnalySISFive, BioVision, VideoTesTMorpho 5.2, BioImageXD, Ariol, ImageJ, analySIS FIVE, MotiImagesAdvanced 3.2, DiMorph, MotiVideoTesTMorpho 5.2, Cell D. Most AMSs are universal, that means that they are not focused on processing of images of the same type, but there are some commercial tools that allow installing separate modules for processing certain types of images, for example, histological ones.

AMSs are widely used in medicine, forensic medicine, and research studies.

2.1. Review of literature on bioimage processing

The problem of AMS development is relevant and the main contribution to its solution has been made by researchers from the United States, France, Germany, Great Britain, China, and Japan. For example, Mitko V. (UK) has developed a technique for segmentation of cells and nuclei on stained histological images [3]. ChenJia-Mei (China) applied the method of support vectors and watershed segmentation of histological images [4]. T. Vrekoussis

researched immunohistochemical samples of breast cancer using the system of automated microscopy ImageJ [5]. Nezved A.M. (Belarus) described the theoretical foundations and methods of image processing necessary for computation of characteristics, which formed the basis of features of medical objects, investigating not only the problems of their analysis but also the problems of recognition [6].

2.2. Problem statement

The main problem in microscopy is automation of the medical diagnostic process. This problem embraces the correct design of biomedical image and signal processing systems, medical expert systems, information-analytical systems, and decision support systems. In the field of histology and cytology, AMSs are used in biomedical image processing.

The purpose of the work is to analyze modern automated microscopy systems, based on the criteria of availability of methods and algorithms at three levels of image processing, and design the AMS structure using modern computer vision algorithms and intelligent data analysis.

3. State of AMS development

3.1. Low-level image processing algorithms

An important stage in image development is image pre-processing. It influences the quality of image and accuracy of output results. Each AMS has its own set of algorithms and image pre-processing methods. The complexity of biomedical microscopic image processing includes identifying contours and desired objects while ignoring unnecessary noises and elements. In **Table 1**, a criteria-based comparison of AMS related to low-level image processing, namely, image pre-processing, is shown. That is why the pre-processing stage is an integral part of automated microscopy system.

Practically all AMSs have a standard set of pre-processing methods, such as changing image contrast, brightness manipulation, image channels, and use of Fourier transform. To choose the best specific filtration algorithm for cytological and histological images, it is necessary to conduct a research on their implementation in well-known automated microscopy systems. The advantage of systems like ImageJ, AxioVisison, MoticImageAdvance is the presence of several algorithms for selecting thresholds in the image, quality of processing depends on both the algorithm, and the image itself. The disadvantage of almost all systems is a limited set of algorithms for filtering and wavelet analysis. Some systems provide such functionality in the form of additional commercial or non-commercial modules.

The digital image is exposed to different types of noise, which are formed at the stage of obtaining an image or its transmission. Typically, noises appear due to the poor quality of photo and video equipment, as well as when transferring images via communication channels. The low image quality may also be caused by a human factor.

There are common filtering algorithms, such as Gaussian, median, averaging, and adaptive. One of the simplest and most natural ways to detect an object is choosing a threshold

Criteria	ImageJ	Axio Visison	BioImageXD	motic	QCapture PRO	Icy	Image Pro Plus	Micro Manager	analySIS FIVE
Contrast	+	+	+	+	+	+	+	+	+
Brightness level change	+	+	+	+	+	+	+	+	+
Low/high frequency filter	+	+	+	+	—	—	+	+	+
Threshold selecting algorithms:	-	+	-	+	+	-	+	+	+
Laplacian	+	-	-	-	-	-	-	+	+
Krish	+	+	+	-	+	+	+	+/-	+
Sobel									
Filters:	+	+	+	+	+	+	+	+	+/-
Gaussian	+	+	+	+	+	-	+	+/-	+/-
Median	+	+	-	-	-	-	-	+/-	+/-
Average									
Fast Fourier Transform	+	+	+	+	+	—	+	+	+
Morphologic operations	+	+	+	+	+	+	+	+	+
Wavelet analysis	-	+	-	-	-	-	+/-	-	-
Haar algorithm	-	-	-	-	-	-	-	-	-
<i>Daubechies algorithm</i>	+	-	-	-	-	-	-	-	-
algorithm “m hat”									

Table 1. Comparative characteristics of AMS low-level image processing.

according to brightness, or thresholding [7]. There are the following common threshold algorithms: Kenny, Sobel’s algorithm, differential selection of thresholds, and refinement of boundaries.

As the result of image filtering algorithms comparison, we can conclude that the median filter showed worse results at the processing speed, however, what is more important, better quality of a final image. Therefore, the final image can be properly processed at the next stages. When choosing AMS, it is necessary to give preferences to those that have this filter in its structure. As the result of comparison of threshold selection algorithms, Canny [8] and Sobel’s [9] algorithms demonstrated the best quality of input image processing. These algorithms are implemented in most AMSs.

A significant advantage of AMS is the presence of image wavelet transform. Wavelet localizes the signal both in space and frequency [10]. A signal can be represented by a set of wave packets (wavelets) formed on the basis of some fundamental function. This set is different in different parts of time interval determination of the signal and is adjusted by factors that have the form of complex time functions.

In signal processing, Fourier transform is usually considered as decomposition of the signal at the frequency and amplitude, that is, the transition from time-space to frequency. By means of Fourier transforms, the frequency domain (spectrum) of the image is received [11]. Fast Fourier Transform (FFT) is a fast algorithm for calculating a discrete Fourier transform [12]. For the direct calculation of the discrete Fourier transform from N data points, $O(N^2)$ arithmetic operations are required, and FFT can calculate the same result using $O(N\log N)$ operations.

3.2. Medium-level processing algorithms

In computer vision systems, image segmentation is one of the most difficult stages. The segmentation stage involves division of the image into areas for which a certain homogeneity criterion is fulfilled, for example, selection of regions of approximately the same brightness in the image. The comparative characteristics of segmentation algorithms are given in **Table 2**.

Table 2. Comparative characteristics of medium-level image processing (+ presence, – absence, +/- availability of additional module).

As the result of the comparison, we can conclude that most systems have in their composition a similar set of elements and segmentation algorithms. Among the above-mentioned AMSs, it is necessary to highlight BioImageXD and AxioVisison, which have the largest set of implemented segmentation algorithms. ImageJ in its composition has only a few algorithms; however, it provides the ability to install additional modules.

The advantage of active contour method is that it divides an image into sub-regions with continuous contours. The contour-based modules use the boundary detector, usually based

AMS	K-means	Intelligent scissors	Snakes	Watershed method	Kruskal algorithm	Grab Cut	Mean shift	Contour coding	RAN SAC	Hough transform
ImageJ	+	+/-	+	+	+/-	+/-	+	—	+	+/-
AxioVision	+	—	+	+	+	+	—	+/-	—	+
BioImage	+	+	+	+	+	—	—	+/-	+	+/-
Motic	+	—	+	—	+	—	—	—	+	—
QCapture	+	+	—	+	—	—	—	+/-	+	+
Image Pro	+	—	+	—	+	+	—	—	—	+
Icy	+	—	—	+	—	—	+	—	—	—
Micro Manager	+	+/-	+	+/-	—	+/-	—	—	+	+

Table 2. Comparative characteristics of medium-level image processing.

on image gradient to find the boundaries of sub-regions and draw contours to the detected boundaries [13].

The “Snake” algorithm is widely used in medical image processing and segmentation. The main disadvantage of the “Snake” algorithm is that the influence of internal energy tends to exaggerate the model excessively generating a straight line [14]. The main advantages of the “Snake” algorithm can be attributed to a relative simplicity of implementation and stability to input data variability.

One of the first interactive segmentation algorithms is Magic Wand algorithm. The algorithm’s action is the following: a user specifies some point of the object, and the algorithm highlights the surrounding pixels with a similar color. Intelligent scissors view the entire image as a graph, each vertex of which corresponds to the pixel of the image. The main limitation is that there are many alternative ways in highly textured areas. The methods of the section of the graph are presented as a weighted non-oriented graph. The pixel or group of pixels is the vertex, and the edges determine the similarity or dissimilarity of the neighboring pixels. Then, the graph (image) is cut in accordance with the criterion created to obtain “good” clusters [15].

The k-mean method is an iterative method used to divide images into K clusters. The principle of the mean shift algorithm is based on finding the maximum probability density of a function that describes the discrete image data. The kernel determines the weight of different points when evaluating the mean [16]. This algorithm is distinguished from others by its processing speed.

The contour analysis (CA) is used to describe, store, compare, and search for objects represented in the form of their external contours [17]. CA allows effectively solving the main problems of image recognition - transfer, zoom, or object rotation. In systems of computer vision, the most popular types of contour encoding are Freeman code, two-dimensional encoding, and polygonal encoding.

The Hough transform is an algorithm that is used to extract elements from an image. This algorithm is used to search for objects belonging to a certain class of figures using the voting procedure [18]. The classic Hough transform algorithm is related to the identification of lines in the image, but later the algorithm was expanded by the possibility of identifying the position of an arbitrary figure, most often of ellipses and circles [19]. RANSAC is an alternative to the Hough transform algorithm [20]. The advantage of the RANSAC algorithm is its ability to give a reliable assessment of model’s parameters. The disadvantage of many AMS is the absence of many encoding methods and algorithms for selection of certain elements, such as lines, circles, or ellipses, which complicates high-level image processing. The complete sets of elements are the following systems: BioImageXD, AxioVision, and analySISFIVE.

3.3. High-level image processing algorithms

The AMS key stage is the stage of selection and recognition of objects in the image, for example, the nuclei or the cytoplasm. The comparative characteristics of AMS high-level image processing are shown in **Table 3**.

Criteria	ImageJ	Axio Vision	BioImage XD	motiC	QCapture PRO	Icy	Image Pro Plus	Micro Manager	analySIS FIVE
Automatic adaptation to image	—	—	+	—	—	—	+	—	—
Object detection	+/-	+/-	—	—	—	—	+/-	+/-	—
Image comparison	—	+	+	+	+	+	+	—	—
Neural network classifiers	-	+	-	-	-	-	-	-	-
SVM	+/-	-	-	-	+/-	-	+	+/-	-
	+/-	+/-	+/-	-	-	-	+/-	-	+/-

Table 3. The comparative characteristics of AMS high-level image processing.

Table 3. The comparative characteristics of AMS high-level image processing (+ presence, – absence, +/- availability of additional module).

Classification is one of the sections of machine learning. Classification of image objects refers to the assignment of an object to the number or class name.

Support vector method is a set of similar teacher-training algorithms used for classification and regression analysis [21]. A special feature of the support vector method is a continuous reduction of empirical classification error.

Convolutional neural network (CNN) combines the selection of elementary image features, the formation of more complex features and recognition [22]. CNN alternates the convolutional layers, sub-sampling layers, and max-pooling layers at the output.

AdaBoost is an algorithm for strengthening the classifiers by combining them into a committee. AdaBoost is adaptive in the sense that each successive committee of classifiers is built on objects that are incorrectly classified by the previous committees [23].

Bayes classifier is a classifier that uses Bayes theorem to determine the probability of belonging to one of the classes. If one can determine which class an object belongs to, the classifier will report that the probability of belonging to this class is equal to 1. In other cases, the classifier will construct a vector whose components are probabilities of belonging to one class or another [24].

4. Image analysis

4.1. Cytological image analysis

Cytological examinations of epithelial cells and structures allow researchers to form suggestions about degrees of epithelial proliferation. The systematization of cytological images with

mastitis and fibroadenoma shows that it is possible to use cytological methods to make a diagnosis [25, 26].

Cytology helps differentiate malignant processes if we find in punctate:

- ductal cells with nuclear enlargement and prominent nucleoli but they are in large sheets with no single cells;
- only a few malignant cells are present;
- malignant cells intermixed with bare bipolar nuclei;
- presence or absence of ductal -foam cells,
- atypical ductal epithelial cells: (Paget's disease, invasive carcinoma).

Cytology defines the characteristics of normal cells (**Figure 1(a)**):

- often scant cellularity (depends on age, hormonal status);
- small groups of ductal cells;
- lobular structures may be seen;
- myoepithelial cells in cell groups (as elongated nuclei) and in the background (ovoid nuclei stripped of the cytoplasm);
- adipose tissue and stroma.

Normal cells: a few small cohesive groups of ductal cells at the most; different from adequacy criteria for FNA;

Cytological structure of the breast cyst (**Figure 1(b)**):

- background of amorphous material;
- degenerate cells and debris;
- foamy macrophages;
- ductal epithelial cells, often apocrine and balling-up;
- myoepithelial cells may not be seen.

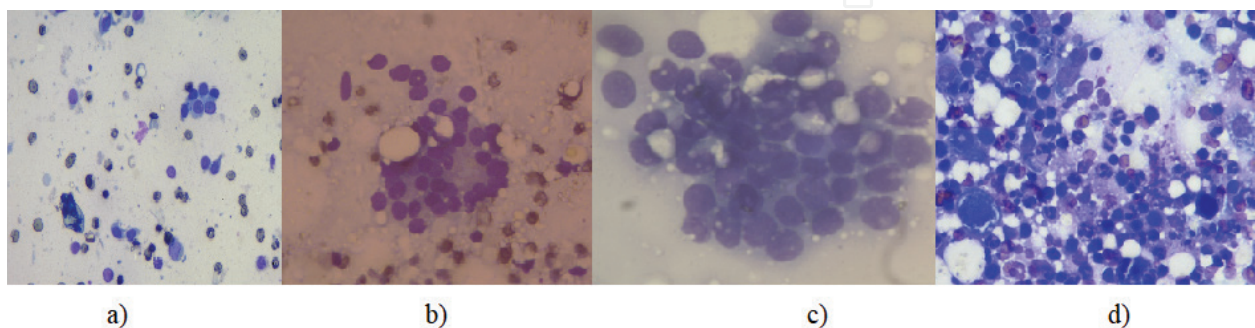


Figure 1. Examples of cytological images.

Cytological structure of fibrocystic change:

- variable numbers of apocrine cells and foam cells;
- variable fat and stroma;
- low to moderate cellularity;
- proteinaceous background;
- cohesive sheets of ductal cells in a honeycomb pattern;
- bare bipolar nuclei dispersed in the background and within or attached to sheets of epithelial cells.

Basic structural elements of Fibroadenoma (**Figure 1(c)**):

- moderate to high cellularity;
- tightly cohesive branching antler-horn or finger-like projections of epithelial cells;
- stromal fragments (metachromatic fibrillary matrix material);
- both ductal and stromal components need to be diagnostic;
- numerous bare bipolar nuclei, bordering and within epithelial clusters;
- may see few foam cells or apocrine cells;
- often mild nuclear atypia with prominent nucleoli, particularly in younger patients.

Basic structural elements of invasive ductal carcinoma, of no special type (**Figure 1(d)**):

- usually very cellular;
- disorganized, loosely cohesive groups;
- single, polygonal, plasmacytoid epithelial cells (which can look deceptively bland);
- absence of bare bipolar nuclei;
- cellular and nuclear pleomorphism (2-4x RBC);
- nuclear border irregularity;
- hyperchromasia;
- nucleoli;

	Papillary cancer	Cystic disease
Nuclear area	13,610,87,392	5581,726
Cytoplasm area	28,833,08333	114,426,8
Nuclear-cytoplasm ratio	0,472,057,524	0,048779884

Table 4. Comparative characteristics of papillary cancer and cystic mastitis.

- there may be mucin vacuole/targetoid inclusion within the cytoplasm;
- mitoses.

Based on experimental studies, we have developed quantitative microscopic features of breast tissue [25]. For example, there is a comparison of mastitis and cystic papillary cancer.

Table 4. Comparative characteristics of papillary cancer and cystic mastitis.

4.2. Histological image analysis

Intraductal cancer is one of the histological cancer types, and it can be solid, cribriform, and papillary. Solid cancer of small and medium ducts is characterized by the formation of solid nests. A peculiar feature of this cancer is the formation of central necrosis. Large regular ducts are fully filled with cancer cells as if they form a cuff around a necrotic core. Cancer cells are atypical and polymorphic; they share irregular mitosis; besides, they differ in polar differentiation (**Figure 2**). Tumor cells are mainly located in three-dimensional structures, some of which form a central lumen reflecting the histological structure of the tumor. There are also scattered isolated epithelial cells, but myoepithelial cells are absent. The background can be clear or hemorrhagic, with no signs of necrosis. Tumor cells are monomorphic, often cylindrical; their nuclei are rounded or oval with a diameter of approximately 1.5 longer than an erythrocyte. Chromatin is fine-grained, condensed near the nuclear membrane. Single small nucleoli can be detected as well.

The beginning of invasive growth is difficult to identify. The absence of basement membrane has no diagnostic value. Invasive growth in stroma is often stimulated by dishormonal proliferation and areas of cancer in situ. But epithelial nests in adipose tissue are pathognomonic as for invasive cancer. With the invasive growth of cells (**Figure 3**), stroma develops in the tissue. Different types of invasive cancer differ not only in cancer cells but also in different correlations between them and their stroma, as well as a stromal nature. The reinforced growth of connective tissue prevents the formation of glandular structures. Growing tumor epithelium is compressed and situated vertically to tissues.

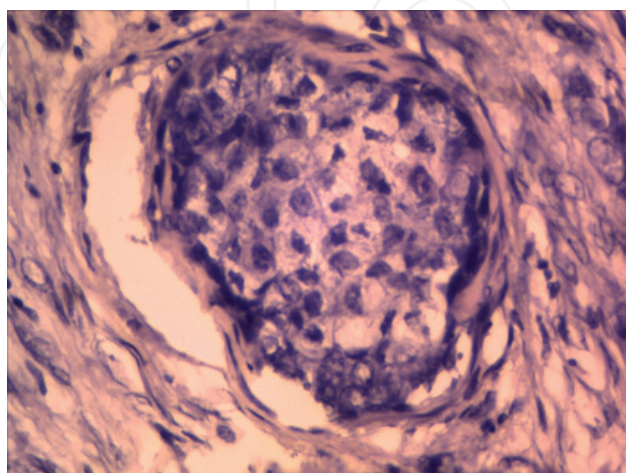


Figure 2. Histological structure of breast cancer intraductal epithelium. Hematoxylin and eosin staining. $\times 200$.

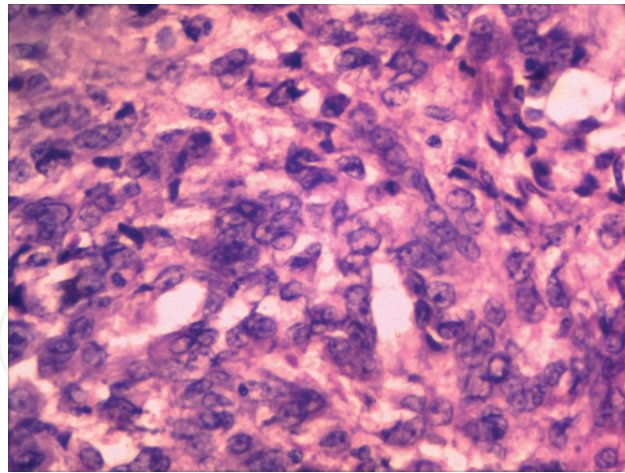


Figure 3. Histological structure of infiltrative breast cancer ductal epithelium Hematoxylin and eosin staining $\times 200$.

Fibroadenoma looks like encapsulated formation with dense consistent fibrous structure. The proliferation of alveoli and intralobular ducts with the growth of connective tissues is microscopically detected. If it surrounds intralobular ducts, this will show pericanalicular fibroadenoma (**Figure 4**). If connecting tissues ingrow in duct wall, they show their pseudo features, and this tumor is called intracanalicular fibroadenoma.

Myoepithelial cells undergo changes. Depending on their functional state, they can be isolated or grouped, elongated and dark, or light. These cells are located between a basement membrane and secreting epithelium of alveoli and small ducts. An exception is a fibrosing adenosis. Basement membrane disappears and proliferative myoepithelial cells penetrate the surrounding connective tissue, where they become looking similar to smooth muscle elements. Microscopic foci appear consisting of clusters and elongated myoepithelial cells including epithelial tubules. Microscopic foci have irregular contours or rounded shapes and clear boundaries. In the latter case, they look like increased or altered lobules. Collagen fibers appear between them and myoepithelial proliferation becomes stiffened (**Figure 5**).

Breast disease (mastopathy) with the dominated cystic component is characterized by cysts clearly separated from surrounding tissues and formed from atrophied lobules and dilated ducts with fibrosing changes of interstitial tissue. Proliferative processes with the development of papillary formations can appear in the epithelium of cysts (**Figure 6**).

Following are the characteristics for diagnosing non-proliferative breast disease (**Figure 7**):

- shallow cysts of alveoli;
- cysts form nests;
- cystic dilated ducts;
- hyalinosis of connective tissue;
- proliferation of connective tissue;
- metaplasia of dark epithelium into white (light);

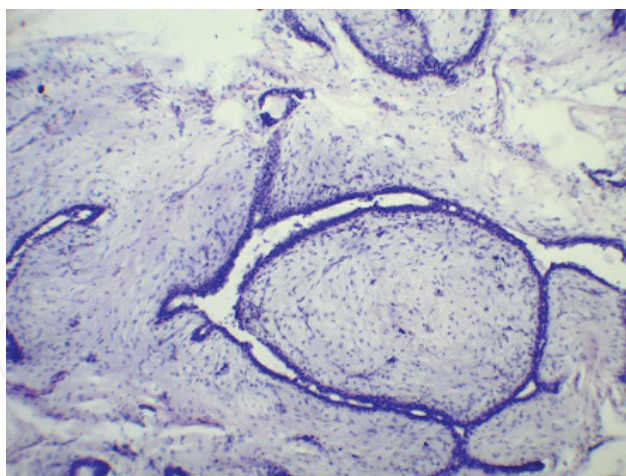


Figure 4. Ingrowth of connective tissue into the duct wall. Hematoxylin and eosin stained. $\times 100$.

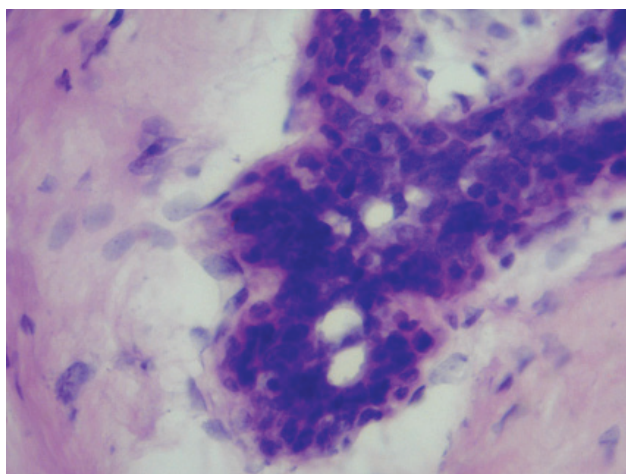


Figure 5. Explicit gland epithelial proliferation. Hematoxylin and eosin staining. $\times 200$.

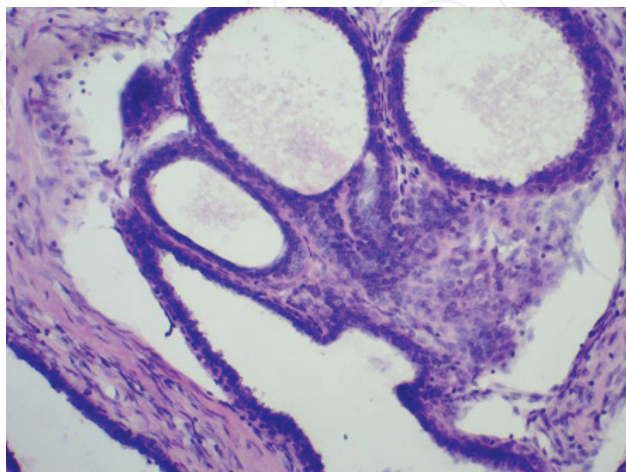


Figure 6. Non-proliferative breast disease. Gland duct expansion.

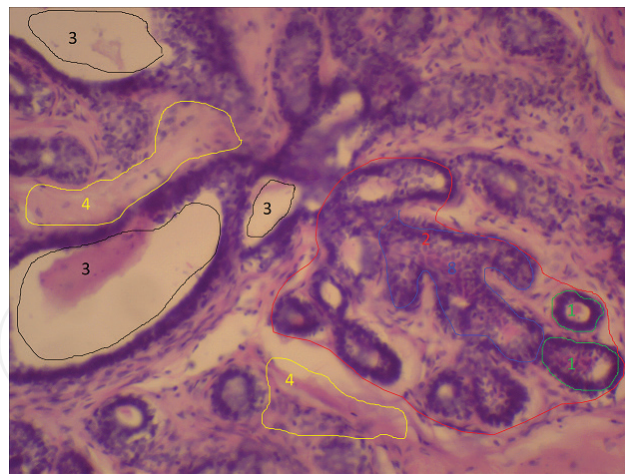


Figure 7. Non-proliferative breast disease. Qualitative features.

- many connective tissues around glands and ducts; pseudo papillary structures;
- atrophy of glandular areas and formation of cysts.

To diagnose non-proliferative breast disease, as it is clearly illustrated in the figure, a doctor needs to see (in histological image) the presence of small cysts of alveoli lobules (1), cysts, which are located in nests (2), cystic dilated ducts (3), hyalinosis of connective tissue, and (4) the formation of pseudo papillary structures (8).

Thus, for automated histological image processing, it is necessary to build a fuzzy knowledge base for diagnosis statement in real time.

5. Fuzzy knowledge base for pathological condition diagnosing

5.1. Fuzzy knowledge base developed on histological image analysis

The main reason for fuzzy logic development was the presence of approximate reasoning in describing processes, systems, and objects by humans [27–29]. This theory is used in various fields of engineering, making it possible to process large amounts of information, solve complex problems in real time without the use of special mathematical and engineering knowledge. Fuzzy logic fundamentals are actively used in medicine for diagnosis confirmation by a doctor.

In most cases, Mamdani fuzzy inference mechanism is applied. The basis for fuzzy logic inference engine is a rule base containing fuzzy “if-then” expressions and membership functions for respective linguistic terms. This should adhere to the following conditions:

- there is at least one rule for each output variable linguistic term;
- for any input variable term, there is at least one rule, in which this term is used as a premise (the left side of the rule).

The results of described above histological image analysis show that there are the following correlations between input image features and diagnosis of breast pathological conditions: the presence of small cysts of alveoli single slices (1), cysts, which are located in nests (2), cystic dilated duct (3), hyalinosis of connective tissue (4), proliferation of connective tissue (5), metaplasia of dark epithelium into white (light) (6); a large amount of connective tissues around glands and ducts (7); pseudo mammilla and atrophy of glandular areas (7); and formation of cysts (8) defines the conclusion that the patient has got a non-proliferative breast disease (mastopathy). It is important to emphasize that all the above-mentioned features are qualitatively described, but the presence of some of them is obligatory for this diagnosis, others can be absent **Figure 6**. There are incompatible features, for example, 6 and 7, 7 and 8, 7 and 3, 8 and 9, 9 and 6. Features 4 and 5 are mandatory.

Example of non-proliferative breast cancer diagnosing rules is the following:

IF 4 AND 5 AND (1 OR 2 OR 3 OR 9), THEN it is a non-proliferative breast disease.

IF 4 AND 5 AND (1 OR 2 OR 3 OR 6 OR 8), THEN it is a non-proliferative breast disease.

IF 4 AND 5 AND (1 OR 2 OR 7 OR 9), THEN it is a non-proliferative breast disease.

IF 4 AND 5 AND (1 OR 2 OR 3 OR 6 OR 8), THEN it is a non-proliferative breast disease.

IF 4 AND 5 AND (1 OR 2 OR 7 OR 9), THEN it is a non-proliferative breast disease.

Applying Fuzzy Logic Toolbox to Matlab and data noted upper, we construct a fuzzy knowledge base for the non-proliferative breast diagnosing, which consists of 26 rules.

The constructed rule base is illustrated in **Figure 8**.

Similarly, for proliferative breast diagnosing, the following characteristics in the histological image are necessary: the proliferation of small ducts myoepithelium and endothelium; interlobular duct dilatation; the proliferation of small ducts and alveoli; small stroma; no basement membrane; myoepithelial proliferating cells move to intralobular connective tissue and become similar to smooth muscle.



Figure 8. Fuzzy knowledge base for non-proliferative breast diagnosis.

However, for this diagnosis, it is necessary to process 62 “if-then” rules.

If the investigated image contains alveoli proliferation, interlobular duct proliferation, porous basophilic connecting tissue; coarse oxyphilic connective tissue; ducts are laid by epithelium and myoepithelium of different functional state; myoepithelium (prolonged dark cells or light cells with spherical inclusions); development of pseudo gland structures; hyalinosis of connective tissue and epithelial atrophy, then histologist diagnoses fibroadenoma. Two hundred and fifty-two rules are necessary to be constructed to process these features.

In addition to the described diagnoses, an expert is also able to diagnose non-infiltrative and infiltrative cancer if there are the following features in the image: polymorphism of cells, sharp increase in size of cells, atypical mitosis, malignant cells accumulation in the lumen of ducts, isolated cell necrosis, invasive growth into the surrounding tissues (adipose tissue), basal membrane abrasion, penetration of tumor cells through the basal membrane, presence of micro alveoli or tubular structures, multiple necrosis, micro-calcification; and cells that do not infiltrate through the basement membrane ducts. **Table 4** demonstrates the qualitative characteristics of these features and their correlations used to construct 511 fuzzy rules.

Based on the described correlations, we can construct a knowledge base of 851 “if-then” rules for diagnosing breast pathological states.

5.2. Fuzzy knowledge base developed on cytological image analysis

The rules of the proposed fuzzy knowledge base are the following:

If there is a small number of hypochromic monomorphic cells and a narrow rim of intensely colored cytoplasm and rounded hyperchromic nuclei, then it is a fibrous non-proliferative breast disease (70%).

If papillary structures and flattened apocrine epithelium are formed, and we observe intense expression of the nucleus, and a narrow rim of intensely colored cytoplasm, and rounded hyperchromatic nuclei, then it is a fibroadenoma (80%) [5]. The proposed fuzzy system of diagnosing breast cancers includes nine inputs and one output (**Figure 9**).

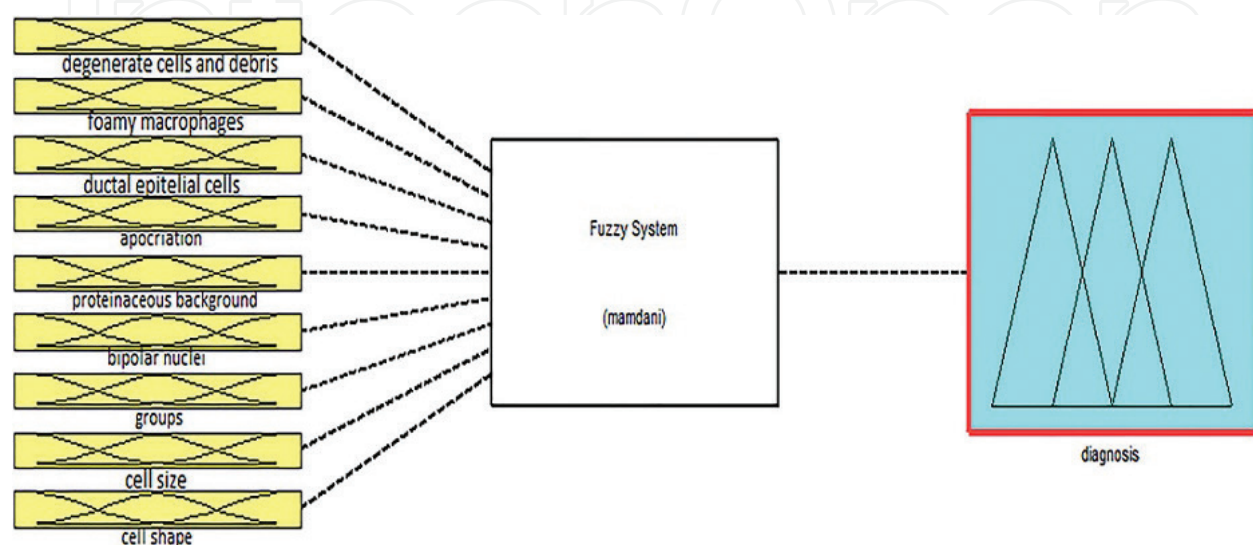


Figure 9. The fuzzy system of diagnosing breast cancers.

Due to the information of part I, the inputs of the proposed fuzzy system are:

- degenerative cells and debris;
- foamy macrophages;
- ductal epithelial cells;
- apocriation;
- proteinaceous background;
- bipolar nuclei;
- groups;
- cell size;
- cell shape.

The output “diagnosis” of the fuzzy system describes the malignant process in the breast:

- breast cyst;
- fibrocystic change;
- fibroadenoma;
- invasive ductal carcinoma.

Most of these inputs are described by quality descriptions. For example, the first input—“degenerate cells and debris” is described by “low”, “medium”, and “high” fuzzy variables showing that these cells are present in the cytological images. Analogically, the inputs of “foamy macrophages”, “ductal epithelial cells”, “apocriation”, “proteinaceous background”, and “bipolar nuclei” can be described by the same variables. The member functions of these inputs can be described by bell function as shown in **Figure 10**.

However, the input “groups” can be described by the variables “cribiform”, “tubular”, “finger-like”, and “cup-shaped”. The member functions of this input are shown in **Figure 11**.

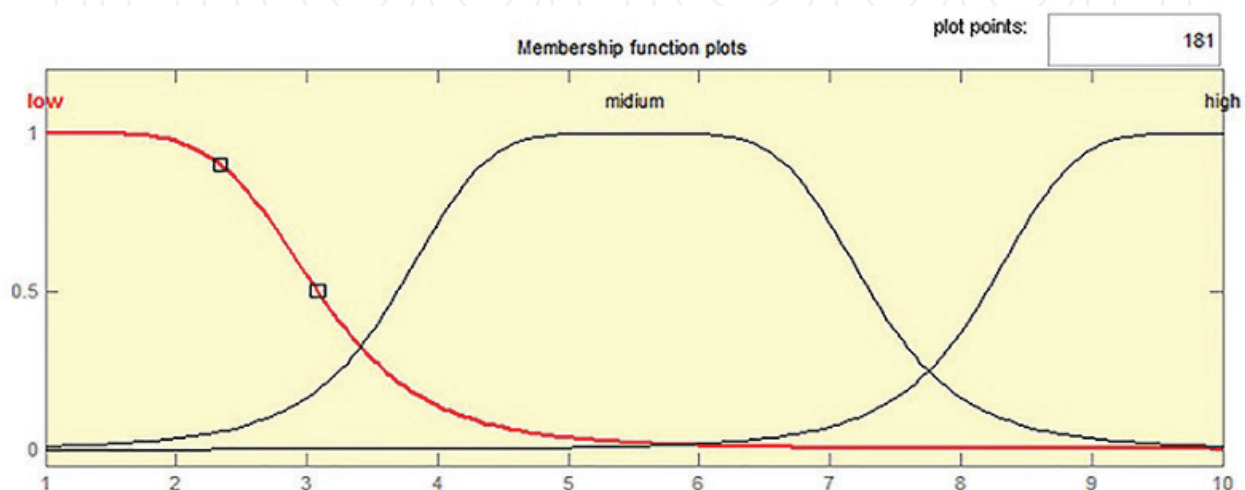


Figure 10. The member functions of input “degenerate cells and debris”.

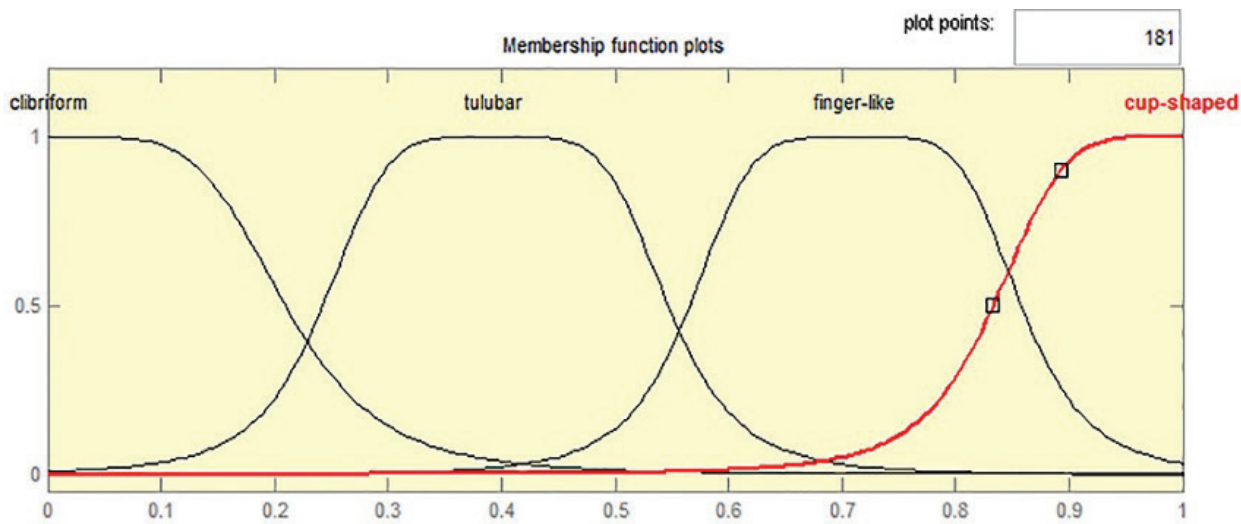


Figure 11. The member functions of input "groups".

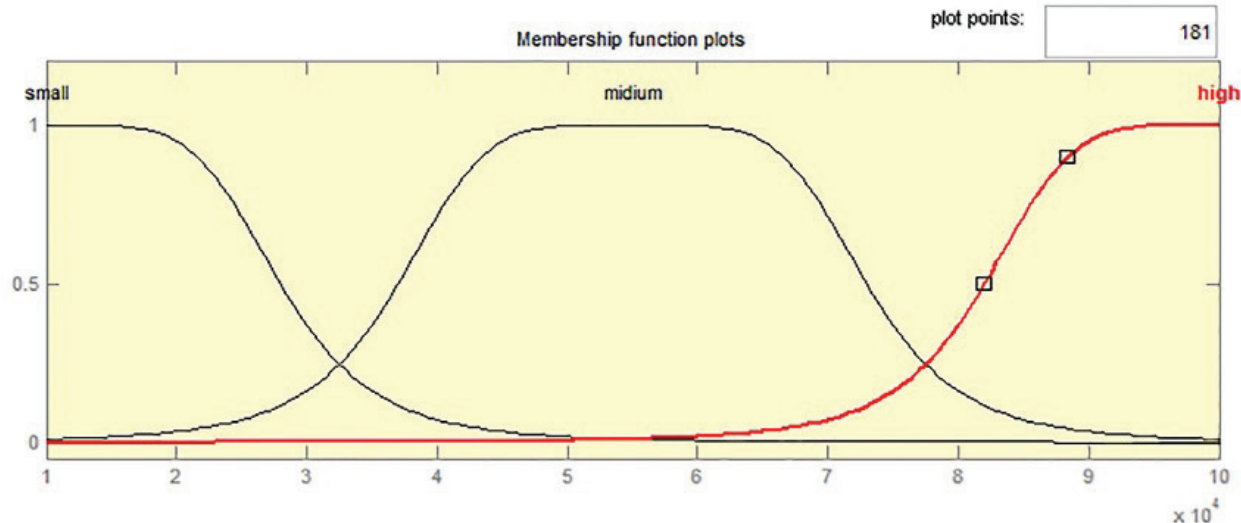


Figure 12. The membership functions of input "cell size".

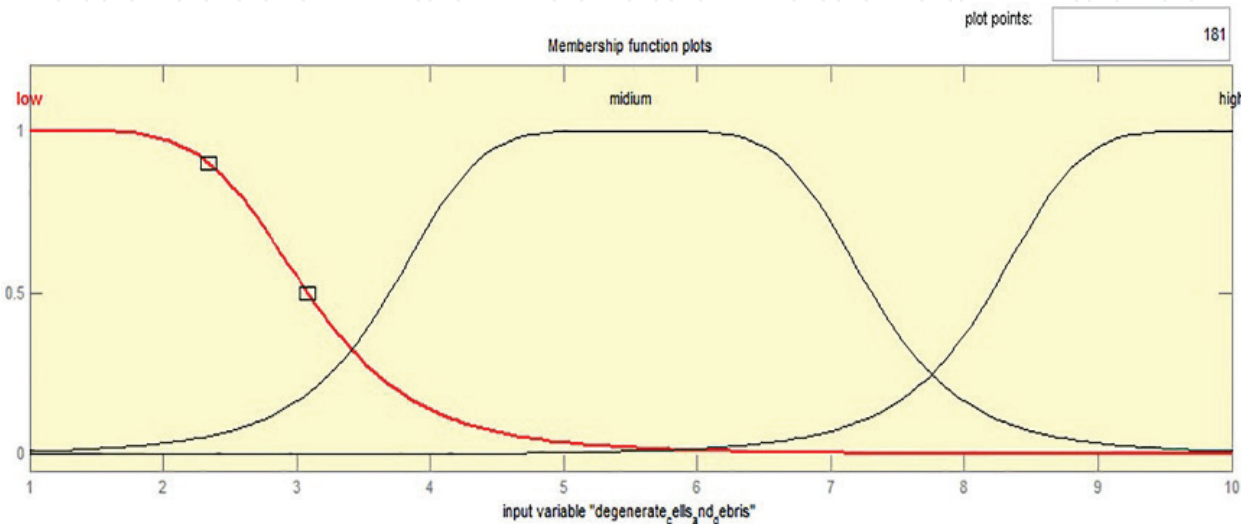


Figure 13. The membership functions of input "cell shape".

The “cell size” and “cell shape” are described by quantitative descriptions, which are shown in **Table 1**. The membership functions of those inputs are shown in **Figures 12** and **13**.

6. Structure and basic modules of the developed intelligent AMS

The intellectual system development is based on the previous research studies of authors [30, 31].

The key characteristic of the developed automated microscopy system in comparison to existing analogues is the presence of adaptive graphical interface for different types of users and, as a result, distribution of access rights to the system. The generalized structure of the developed AMS is presented in **Figure 14**.

We present the basic system modules.

Database. The main groups of system users are treating physician, diagnostic doctor, expert, assistant, and administrator. They communicate using a remote database and a remote FTP server. Currently, in medicine, scientists devote considerable attention to the design of databases for information systems that facilitate the work of physicians. The structure of such relational databases mostly makes it easy to formulate reports and statistical data on patients and their diagnoses. Most of the existing automated microscopy systems for image analysis do not have databases or they have a limited functionality.

DB keeps information about system users, patients’ tests, quantitative and qualitative image characteristics, expert conclusion, etc. Setting master-master or master-slave replication can

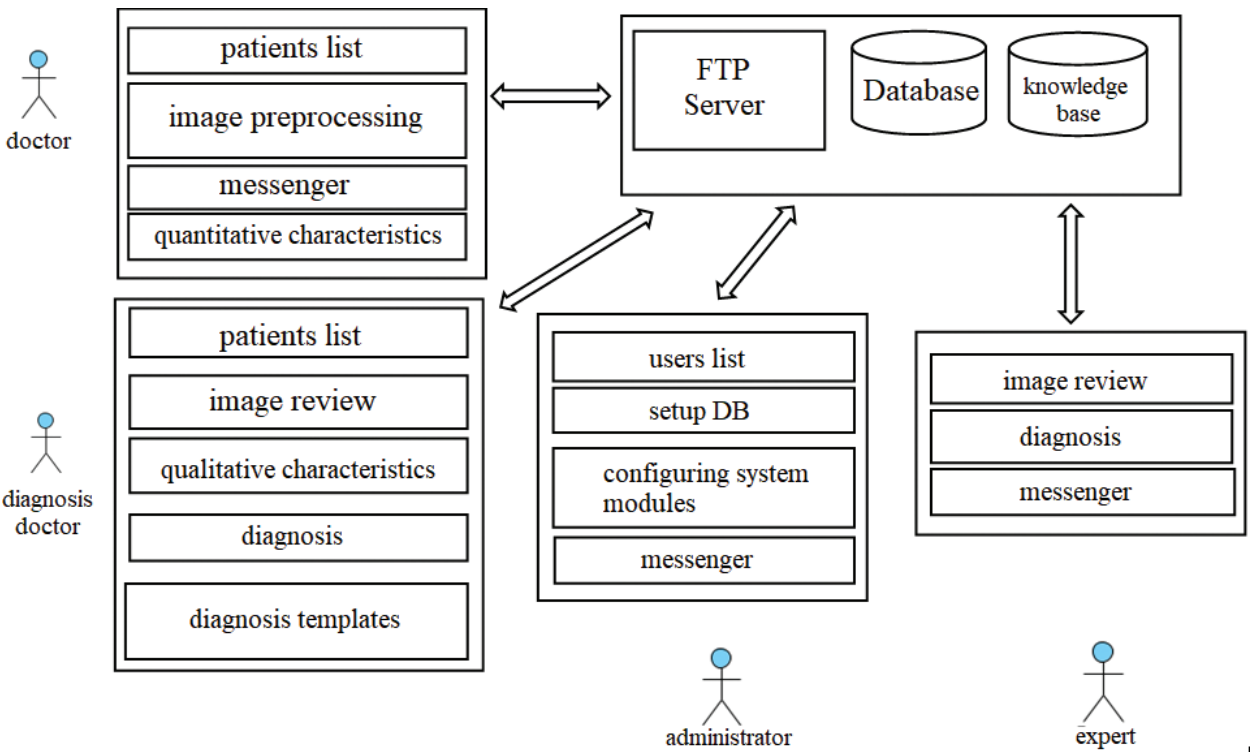


Figure 14. Generalized AMS structure.

greatly improve DB reliability and ensure smooth system performance. The DB datalogical model is illustrated in [23].

In the process of working with patients, an important system element is logging user actions for control. All information about the actions of doctors (adding some information about patients), which is available for viewing by the system administrator, is in the database. The FTP server plays the role of a repository of histological and cytological images. This approach allows implementing a convenient mechanism for image sharing without any extra effort and does not require knowledge of physicians in the field of information technology. To keep the patient's data confidential, all information are encrypted, so the attacker will not be able to identify the image with a definite diagnosis to the particular patient. Images are located in directories with encrypted patient ID and test identifier. The administrator is responsible for configuring and accessing servers.

Patients' registration. Patients' registration module is designed to add, edit, delete, and view information. In a double "click" on the patient's record, a new window with the history of patient's illness appears. The treating physician and diagnostic physician can add information about the research results. This information is stored in the database and allows determining who and when made the diagnosis.

For convenience, mechanisms for sorting patients by alphabet in order of increasing or decreasing and interactive search across all available fields are developed. For example, one can search for a patient's name, article, diagnosis, date of birth, etc.

Messaging. This module is needed to provide communication between doctors. For example, a treating physician can clarify the diagnosis of a particular patient with an expert. The message sender fills in the three main fields in the window: recipient (selected from the system's database), the subject of the message, and text of the message. The "Destination" module has a similar set of attributes and additional "patient identifier".

Image processing. The image processing module is one of the key modules in the developed intelligent AMS. After choosing a patient (by his or her identifier to ensure confidentiality), the user has an option of choosing an experiment for further processing or creating a new experiment. After selecting the image directory, the list of files is displayed in the graphical interface and automatically uploaded to a remote FTP server with the user ID and experiment on the system.

Quantitative and qualitative characteristics. Cytological and histological image processing is characterized by high complexity and requires deep knowledge in this area by AMS users. One of the possible options for automating the biomedical image classification process is to analyze quantitative characteristics of cell nuclei and qualitative characteristics of the entire image.

File containing quantitative characteristics can be exported from AMS for further classification by machine learning algorithms.

Software description. Taking into account the requirements for the developed intelligent automation microscopy system, the design and development of software system play an important role. Therefore, with the increase in system functionality, its complexity also increases.

Any architecture of software system should make the development and maintenance process simpler and more effective. A program with a good architecture is easier to extend, modify, test, and understand. The basis for designing the architecture of the developed AMS is the design template MVC.

7. General structure of convolutional neural networks

In this work, histological and cytological images are transmitted to convolutional neural networks (CNN). The objective of the neural network is to assign the input image to a certain

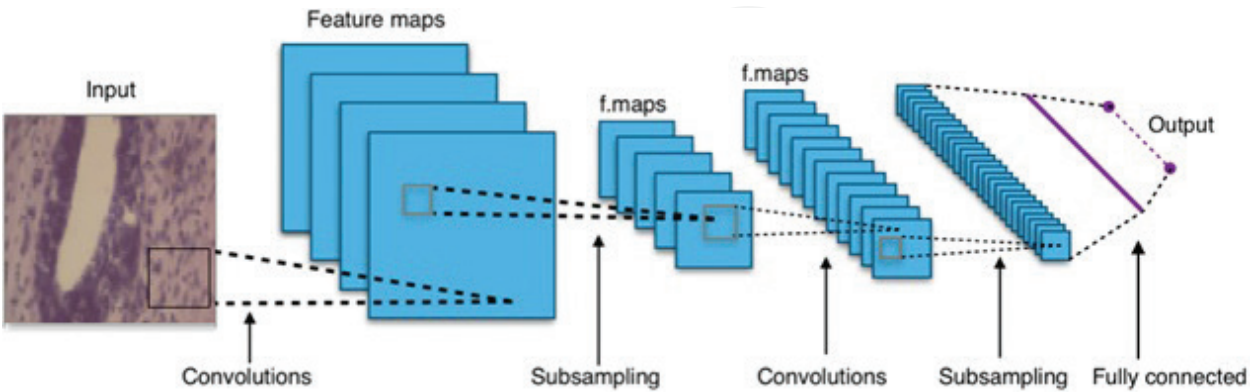


Figure 15. CNN general structure.

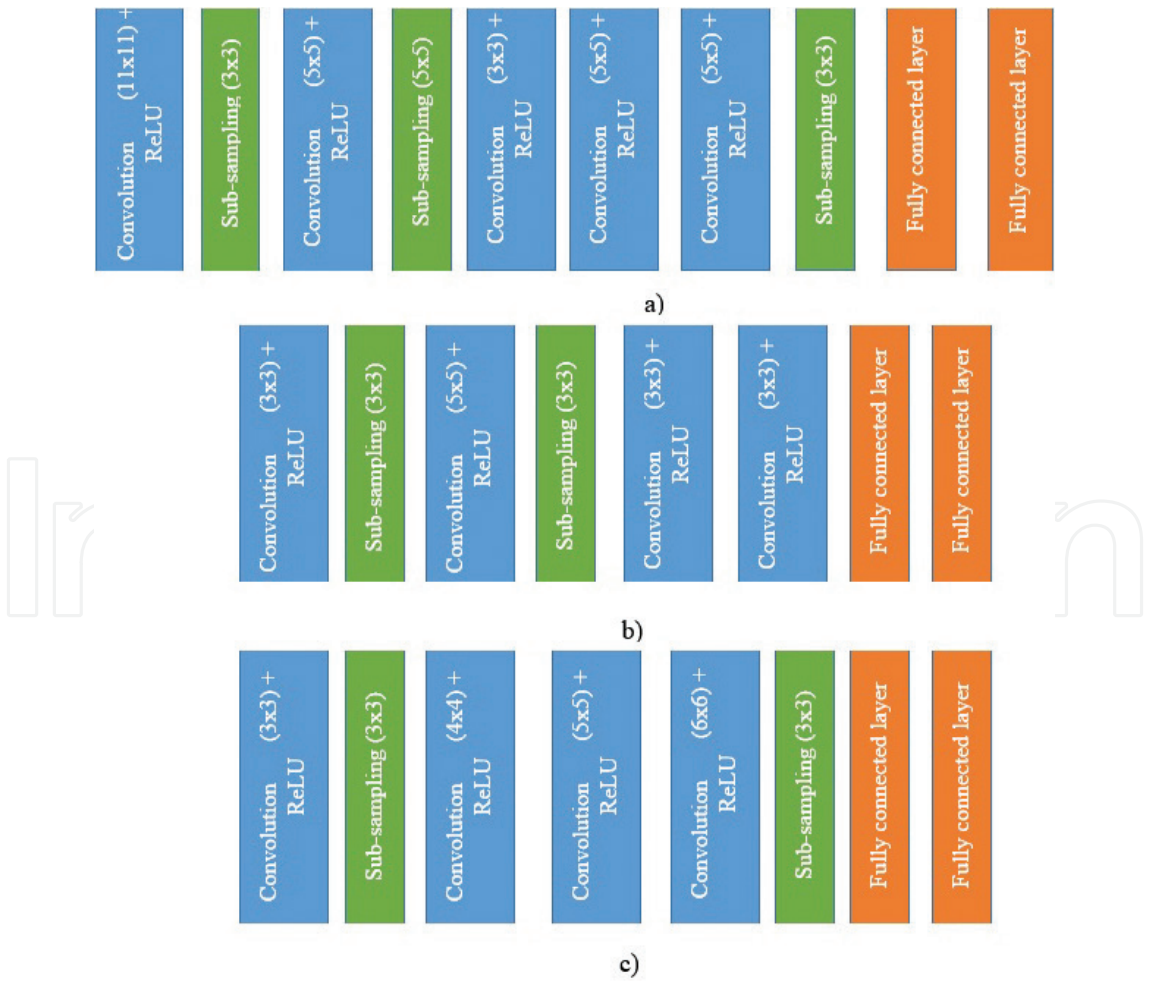


Figure 16. Developed CNN models for cytological and histological image classification.

class. CNN consists of a sequence of convolutional, sub-sampling, and max-pooling layers. The first two types of layers (convolutional and sub-sampling) alternate and form the input vector of features for a multilayer perceptron.

General structure of CNN is shown in **Figure 15**.

The disadvantages of cytological images are a low level of contrast and noise; therefore, several new CNN models are proposed in this paper. Training image sample is divided into the following classes:

- cyto—cystic—mastopathy;
- cyto—mostopathy;
- cyto—non-proliferative—fibro—mastopathy;
- cancer.

We consider the following CNN models for the classification of cytological and histological images (**Figure 16**).

8. Experimental results of cytological and histological image classification and comparative analysis of the developed intelligent AMS with analogues

The comparative characteristics of the existing and developed AMS are shown in **Table 5**.

Table 5. Comparative analysis of AMS (“–” element is absent, “–/+” element is partially present, “+” element is present)

	1	2	3	4	5	6	7	8	9
1									
2									
3									
4				+					
5					+				
6							X		X
7			X					X	
8									X
9									

Table 5. Incompatible features of histological images for non-proliferative breast cancer diagnosis.

Criterion	ImageJ	Image Pro Plus	Dia Morph	Axio Vision	Bio Image XD	QCapture PRO	Micro Manager	Amira	AMS-Diagnosis	Developed AMS
Information protection:	-	-	-	-	+	-	-	+	+	+
Authorization of users	-	-	+	-	+	-	-	-	+	+
Additional Authentication	-	-	-	-	-	-	-	-	-	+
SQL Injection Protection										
Ensuring confidentiality of patient information										
Search by template	—	—	—	—	+	—	—	+	—	+
Patient registration	—	—	—	—	-/+	—	—	+	-/+	+

Table 6. Comparative analysis of AMS.

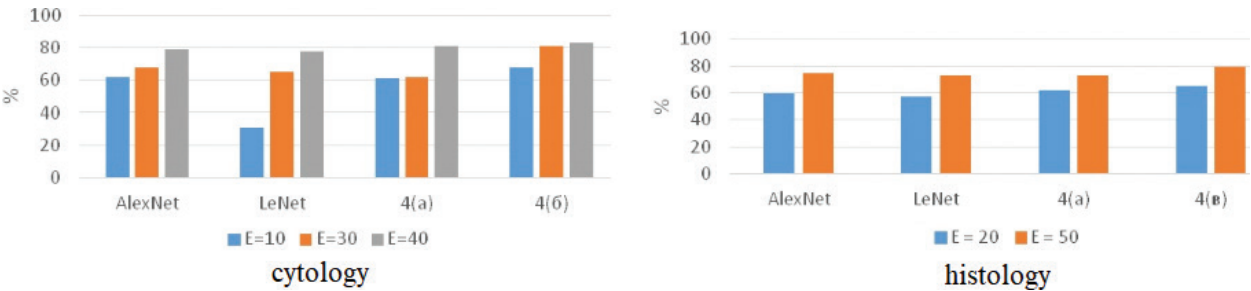


Figure 17. Results of work of CNN models for cytological and histological images classification.

As it can be seen from **Figure 18**, the neural networks showed the best results in comparison with k-nearest neighbors, k-means, and SVM algorithms. In addition, CNN does not require significant image pre-processing, microscopic objects’ selection, and calculation of quantitative characteristics.

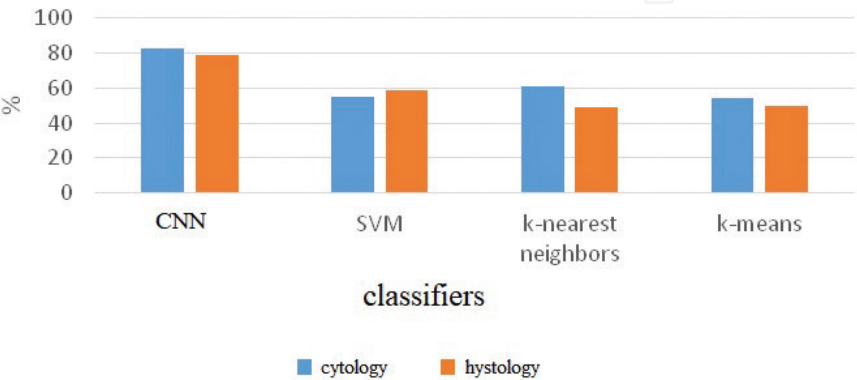


Figure 18. Comparative analysis of cytological image classification.

9. Conclusions

With the application of information technology in all spheres of life, including medicine, there is a need for the development of modern automated microscopy systems. Based on modern methods and algorithms of computer vision and the well-known AMS, comparative characteristics of AMS low-, medium- and high-level image processing were conducted, which allowed highlighting the following basic methods and algorithms: for low level, Gaussian, Adaptive, low/high-frequency filters, Fast Fourier transform, Wavelet analysis; for medium level, feature selection (by contours, areas, angles), segmentation (algorithms “Smart Scissors”, Snakes, Mean-shift, and watershed), object selection (Hough algorithms, RANSAC); for high level, methods of recognition (static, morphological, structural), neural network method, and support vector method.

In this work, analysis of histological and cytological images of breast pre-cancerous conditions has been made, their characteristic features have been determined. This fuzzy system can be used in oncology telemedicine for fast and efficient diagnosing of breast pre-cancerous and cancerous states based on histological image analysis.

The analysis of existing CNN models has been carried out, which made it possible to construct a CNN model for classification of breast pathological conditions. The sequence of convolution layers, sub-sampling, and their input parameters determine the structure of convolutional network model. The CNN results were compared with the known analogues: SVN, k-nearest neighbors, and k-means. Accuracy of classification for cytological images was 83%, and for histological images, it was 79%.

The developed Intelligent AMS, unlike the existing AMS, has an adaptive graphical interface for different user groups, algorithms for automatic image pre-processing and image segmentation, availability of modules for working with remote databases and communication between users, and allows justifying the diagnosis according to the quantitative micro-objects’ characteristics and maximally automating the diagnosing process.

Currently, the main trend in the development of intellectual systems is hybridization, which uses different approaches to artificial intelligence for solving problems.

Acknowledgements

This proposed research was developed during the work on the state budget project “Hybrid intelligent information technology diagnosing precancerous breast cancer based on image analysis” (state registration number 1016 U002500).

Author details

Oleh M. Berezhsky

Address all correspondence to: ob@tneu.edu.ua

Ternopil National Economic University, Ternopil, Ukraine

References

- [1] 14,908 cases of breast cancer (in Ukrainian) [Internet] Available from: http://galinfo.com.ua/articles/torik_v_ukraini_zareiestrovano_14_908_vypadkiv_zahvoryuvannya_na_rak_molochnoi_zalozy_197568.html. [Accessed: 2017-10-01]
- [2] Oleh B, Melnyk G, Batko Y. Modern trends in biomedical image analysis system design. In: Laskovski A, editor. *Biomedical Engineering Trends In Electronics, Communications and Software: InTech*; 2011. p. 461-480. DOI: 10.5772/549
- [3] Veta M, Pluim JP, van Diest PJ, Viergever MA. Breast cancer histopathology image analysis: A review. *Biomedical Engineering, IEEE Transactions*. 2014;**61**:1400-1411. DOI: 10.1109/TBME.2014.2303852
- [4] Chen J-M, Qu A-P, Wang L-W, et al. New breast cancer prognostic factors identified by computer-aided image analysis of HE stained histopathology images. *Scientific Reports*. 2015;**5**:10690. DOI: 10.1038/srep10690
- [5] Vrekoussis T, Chaniotis V, Navrozoglou I, et al. Image analysis of breast cancer immunohistochemistry-stained sections using ImageJ: An RGB-based model. *Anticancer Research*. 2009;**29**:4995-4998
- [6] Nedzved A, Lukashevich P, Belotserkovsky A. A flexible image processing system based on a script-kernel using an intelligent agent. *Iskusstvennyy intellekt*. 2013;**3**:200-208 (in Russian)
- [7] Mehmet S, Bülent S. Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging*. 2009;**13**:146-165. DOI: 10.1117/1.1631315
- [8] Canny J, Computational Approach A. To edge detection, *IEEE trans. Pattern Analysis and Machine Intelligence*. 1986;**8**(6):679-698. DOI: 10.1109/TPAMI.1986.4767851
- [9] Alghurair D. Design of Sobel Operator using field programmable gate Array. In: *Proceedings of International Conference on Technological Advances in Electrical, Electronics and Computer Engineering (TAEECE)*. May 2013. pp. 589-594
- [10] Portilla J, Strela V, Wainwright M, Simoncelli E. Image denoising using scale mixtures of Gaussians in the wavelet domain. *IEEE Transactions on Image Processing*. 2003;**12**:1338-1351. DOI: 10.1109/TIP.2003.818640
- [11] Ren NG. Fourier slice photography. *ACM Transactions on Graphics*. 2005;**24**:735-744. DOI: 10.1145/1186822.1073256
- [12] Castro E, Morandi C. Registration of translated and rotated images using finite Fourier transforms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1987;**PAMI-9**: 700-703. DOI: 10.1109/TPAMI.1987.4767966
- [13] Arbelaez P, Maire M, Fowlkes C, Malik J. Contour detection and hierarchical image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2011;**33**: 898-916. DOI: 10.1109/TPAMI.2010.161
- [14] Amini A, Weymouth T, Jain R. Using dynamic programming for solving variational problems in vision. *IEEE Transactionson Pattern Analysis and Machine Intelligence*. 1990;**12**:855-867. DOI: 10.1109/34.57681

- [15] Felzenszwalb P, Huttenlocher D. Efficient graph-based image segmentation. *International Journal of Computer Vision*. 2004;**59**:167-181. DOI: 10.1023/B:VISI.0000022288.19776.77
- [16] Comaniciu D, Meer P. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2002;**24**:603-619. DOI: 10.1109/34.1000236
- [17] Millasseau SC. Determination of age-related increases in large artery stiffness by digital pulse contour analysis. *Clinical Science*. 2002;**103**:371-377. DOI: 10.1042/cs1030371
- [18] Duda RO. Use of the Hough transform to detect lines and curves in pictures. *Communications of the ACM*. 1972;**15**:11-15. DOI: 10.1145/361237.361242
- [19] Yuen HK, Illingworth J, Kittler J. Detecting partially occluded ellipses using the Hough transform. *Image and Vision Computing*. 1989;**7**:31-37. DOI: 10.1016/0262-8856(89)90017-6
- [20] Schnabel R, Wahl R, Klein R. Efficient RANSAC for point-cloud shape detection. *Computer Graphics Forum*. 2007;**26**:214-226. DOI: 10.1111/j.1467-8659.2007.01016.x
- [21] Cortes C, Vapnik V. Support-vector networks. *Machine Learning*. 1995;**20**:273-297. DOI: 10.1007/BF00994018
- [22] Matusugu M, Mori K, Mitari Y, Kaneda Y. Subject independent facial expression recognition with robust face detection using a convolutional neural network. *Neural Networks*. 2003;**16**:555-559. DOI: 10.1016/S0893-6080(03)00115-1
- [23] Gao W, Zhou ZH. On the doubt about margin explanation of boosting. *Artificial Intelligence*. 2013;**203**:1-18. DOI: 10.1016/j.artint.2013.07.002
- [24] Wan EA. Neural network classification: A Bayesian interpretation. *IEEE Transactions on Neural Networks*. 1990;**1**:303-305. DOI: 10.1109/72.80269
- [25] Berezky OM, Pitsun OY, Verbovyi SO, Datsko TV. Relational database of intelligent automated microscopy system. *Scientific Bulletin of UNFU*. 2017;**27**(5):125-129. DOI: 10.15421/40270525
- [26] Berezsky O, Verbovy S, Dubchak L, Datsko T. Fuzzy system of diagnosing in oncology telemedicine. *Sensors & Transducers*. 2017;**208**:32-38. Available from: http://www.sensorsportal.com/HTML/DIGEST/P_2894.htm Accessed: 2017-10-10
- [27] Alberto d'Onofrio. "Fuzzy oncology": Fuzzy noise in ducedbifurcations and their application to anti-tumorchemotherapy. *Applied Mathematics Letters*. July 2008;**21**(7):662-668. DOI:10.1016/j.aml.2007.05.019
- [28] Seker H, Odetayo MO, Petrovic D, Naguib RNG. A fuzzy logic based-method for prognostic decision making in breast and prostate cancers. In: *Proceedings of the IEEE Transactions on Information Technology in Biomedicine*. June 2003;**7**(2):114-122
- [29] Muhic I. Fuzzy analysis of breast cancer disease using fuzzy c-means and pattern recognition. *Southeast Europe Journal Of Soft Computing [Internet]*. Available from: www.scjournal.com.ba

- [30] Berezsky O, Verbovy S, Datsko T. The intelligent system for diagnosing breast cancers based on image analysis. In: Proceedings of Information Technologies in Innovation Business (ITIB); 7-9 October, 2015, Kharkiv, Ukraine, pp. 27-30. DOI: 10.1109/ITIB.2015.7355067
- [31] Berezsky O, Melnyk G, Datsko T, Verbovy S. An intelligent system for cytological and histological image analysis. In: Proceedings XIII-th of the International Conference on The Experience of Designing and Application of CAD Systems in Microelectronics, CADSM'2015, Polyana-Svalyava (Zakarpatty), Ukraine; 24-27 February 2015. pp. 28-31. DOI: 10.1109/CADSM.2015.7230787

