Research and Analysis of Breast X-Rays Based on Intelligent Technologies

Rimma Tomakova [1[0000-0003-152-4714]], Ilia Egorov [1[0000-0002-1238-3243]], and Aleksandra Brezhneva [2[0000-0001-5226-329X]]

[1] Southwest State University, 50 Let Oktyabrya str., 94, Kursk, 305040, Russia
rtomakova@mail.ru, egorov.ilia1996@mail.ru
[2] Plekhanov Russian University of Economics,
36 Stremyanny lane, Moscow, 115998, Russia
Brezhneva.AN@rea.ru

Abstract. The principles of forming an automated classification system for breast radiographs are considered.
Classification of selected segments in x-ray images is implemented using intelligent technologies based on neural network analysis. For this purpose, 120 images were selected from the MIAS database with morphological data, on the basis of which a training sample was formed. The volume of the training sample is 370 segments, of which 250 segments are characterized by the state of norm and 120 segments contain pathologies. Each analyzed image block corresponds to a predictor described by a three-component vector. The first indicator that evaluates the segment of morphological education is the statistical characteristic of the mode, the second indicator is the mathematical expectation, and the third component of the indicator is the standard deviation.
The software is implemented in the MATLAB2018b environment. The results of the quality classification of the developed software product on control samples are presented. For this purpose, 50 images of mammograms of the breast from the MIAS database were studied: 25 in the normal state and 25 with pathology. The values of positive and negative classification results are established. Diagnostic sensitivity was DH=84%, diagnostic specificity of DS=96%, diagnostic efficiency of DE=90%.

Keywords: breast x-ray, segmentation, neural network classifier, homogeneity criterion, automated system.

1 Introduction

Ecological problems of regions, a wrong way of life and unproductive conduct instrumental in growth of oncologic diseases. Breast cancer is one of the most common cancers currently in the worldwide. Every year in Russia, about 60 thousand

Proceedings of the 10th International Scientific and Practical Conference named after A. I. Kitov "Information Technologies and Mathematical Methods in Economics and Management (IT&MM-2020)", October 15-16, 2020, Moscow, Russia

women diagnosed with breast cancer are admitted to the dispensary, and about 600 thousand patients continue to be monitored by oncologists. The main method for detecting breast tumors is mammography [1]. Since 2017, Russia has been one of the leading manufacturers of high-tech medical equipment and information systems for mammology. The detector "Solo DM-MT", produced by JSC "Medical Technologies Ltd", which can be used in analog and digital mammography, is widely used.

2 Literature Review

As you know, the main developers of applied software for mammogram processing in the world are the companies such as: AccuDetect, The MAMMEX MammoCAD, Syngo Breast Care, Fujifilm's Digital Mammography System [2, 3, 4, 5]. One of the disadvantages of these software products is the closure of program code, which leads to instability and unpredictability in the use of the software. In this regard, in 2018, the Ministry of Digital Development, Communications and Mass Media approved plans for the transition to domestic software developed for the Ministry of Health of Russia. In world practice, recognition methods based on boosting technology and neural network models of classifiers are used for automated image processing [4,5]. Currently, hybrid technologies have become widespread, which allow combining the technologies of trained classifiers and soft computing technologies [6,8]. The recognition problem is most successfully solved using neural network models. The development of methods and algorithms for the identification and classification of images are devoted to the works of A.N. Galushkina, A.N. Gorban, T. Kohonen, F. Wasserman, and J. Hopfield. However, there is no unified methodology for solving applied problems of image classification using artificial neural networks (ANNs). In this regard, it is advisable, in relation to each specific task, to choose not only their architecture, but also the method of forming the space of informative features and the method of teaching ANNs [6,7,8].

Therefore, the purpose of the research is to develop an automated classification system for the analysis of mammograms.

To achieve this purpose, the following tasks were set: highlighting the area of interest that corresponds to the instructions of a mammologist; decomposition of the mammogram into cascading windows for subsequent classification of the selected image areas.

3 Materials and methods

Methods of segmentation of complexly structured images are used to study mammogram images [7,8,9]. The image is divided into segments homogeneous in texture or brightness, their homogeneity index is a priori set. A characteristic feature of the homogeneous segment is the impossibility of separating another segment out it, on the basis of the established criterion of homogeneity [7,8,9,10]. To select segments based on the cascade segmentation method [6,11,12,13], software modules were developed. These modules are combined into an automated system for the classification of images of X-ray images of the breast (ASCIX).

The significant part of the software modules is developed in the MATLAB2018b environment. The choice of the Matlab 2018b environment for development is due to the fact that it has an extensive toolkit for implementing processing procedures and classification images. In addition, the selected software architecture of the automated system for the classification of images of X-ray images of the mammary gland makes it easy to add and / or modify software modules. Structural scheme is shown in Figure 1.

It consists of three main modules: the Cascading Window Formation Module (CWFM), the Cascading Window Combining Module (CWCM), and the Classification and Decision Making Module (CDMM).The purpose of the first of them (CWFM) is that it implements the procedure for segmentation of the breast radiograph using the "top-down" technology described in [6]. In CWFM, the halftone raster image of a breast radiograph is decomposed into segments distributed over hierarchical levels. The criterion of the transition of the segment from one hierarchical level to another (the criterion of indivisibility of the block) is the homogeneity index of brightness of the pixels of the analyzed segment. Wherein, the indivisible segment is assigned a code corresponding to the hierarchical level on which it is located. The segment code is determined by the route the segment moves through the hierarchical levels. The route ends with the procedure for assigning the segment the status of "indivisible". The generated code allows you to determine the relating "indivisible" blocks. The blocks can then be merged if they meet the merge criterion.

The second module (CWCM) is designed to form the final configuration of the segments. The preparation for the subsequent classification of segments takes place in this module. The module (CWCM) is configured to enlarge the segments that meet a certain homogeneity criterion. The homogeneity criterion can be built on the basis of the brightness characteristics of the segments or on the basis of their texture characteristics.

The third module (CDMM) performs a dual-alternative classification of the image segments of the radiograph into two classes: "there is an area of interest", "there is no area of interest". The third module (CDMM) performs a dual-alternative classification of the image segments of the radiograph into two classes: "there is an area of interest", "there is no area of interest". For implementation the classification from attributes of pixels included in the segment, a vector of informative features is formed. Methods and algorithms for creating this vector are described in sufficient detail in [14,15,16,17,18]. In this case, the classification of segment possible that fall into the class "there is an area of interest" into classes "there is a pathology" - " there is no pathology".

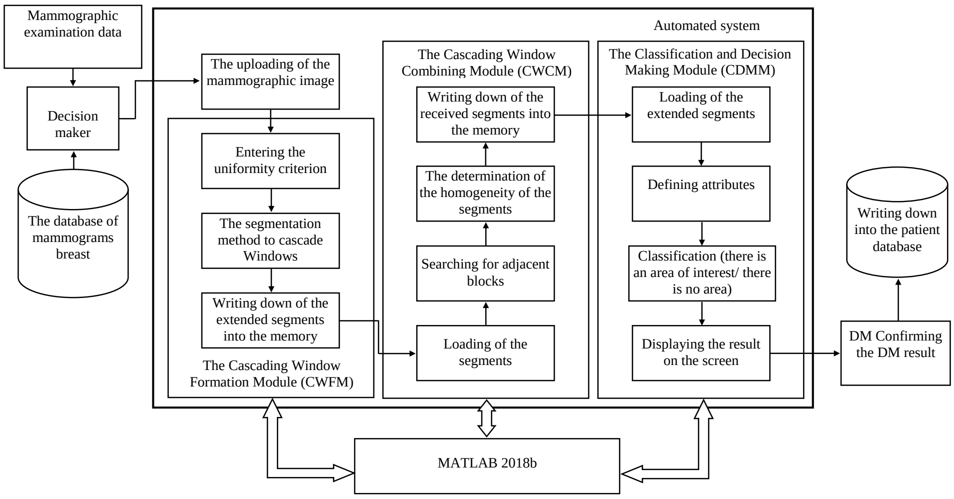


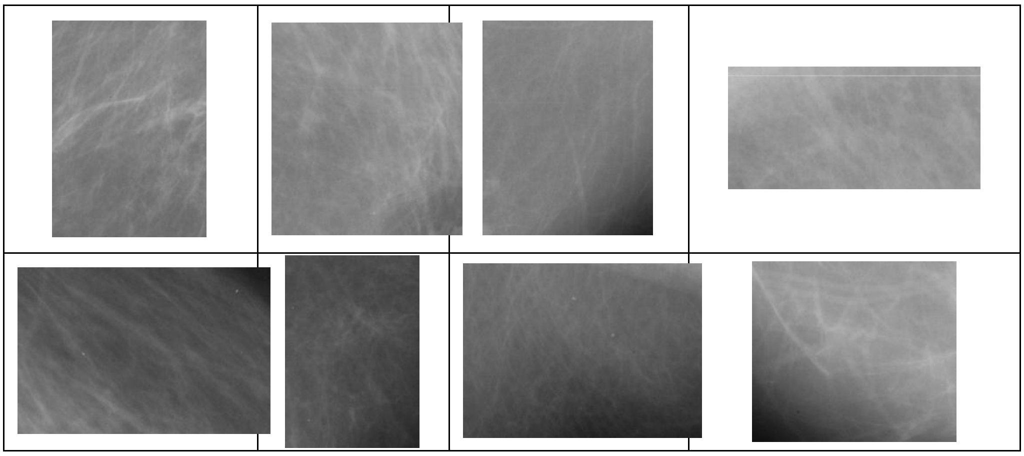
Fig. 1. Structural scheme of an automated mammographic image classification system

To involve the decision-maker (DM) in process of classifications, the classified segments are binarized or presented in the form of a "heat map". The latter way of representing classified segments is expedient when using neural networks with a linear activation function as classifiers. In this case, the proximity of the output of the neural network to one (proximity to the class "pathology") corresponds to the proximity of the pixel shading of the segment to the red color. Such method of data presentation allows the decision-makers both to participate in the process of making diagnostic decisions together with a computer, and to form a database for an automated system.

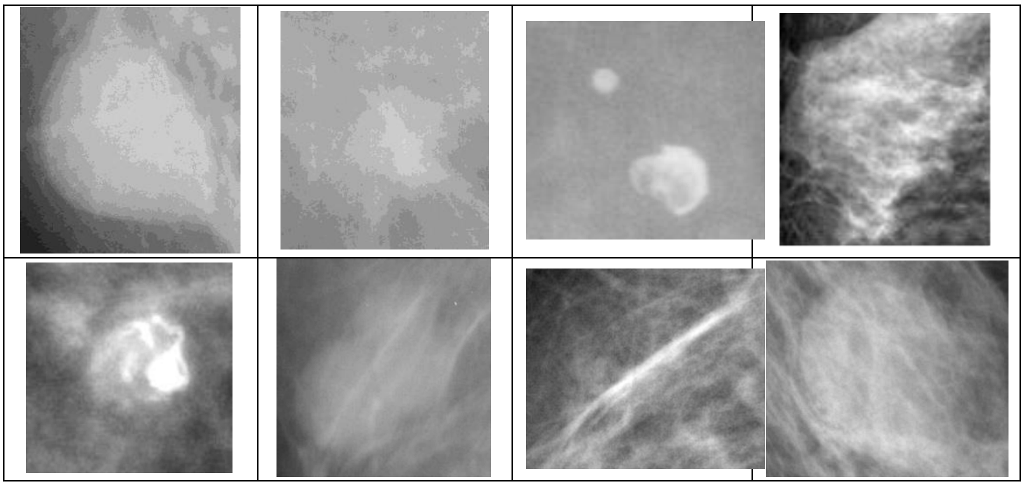
4 Formation of a training sample for the classification of breast diseases based on intelligent technologies

The observational study was carried out to calculate the informative features which are necessary for the functioning of an intelligent system. For this, the test radiographic images of the breast were used from the DDSM database with confirmed diagnoses, which were used as input data in automated system (ASCIX).

The cluster method of sampling was applied in the process of which two classes of analyzed image segments were formed. Class C_1 - contains image segments characterized by the state of the norm. Class C_2 -segments are presented that have morphological formations caused by pathological processes. A case-control study was conducted to develop a dual-alternative classifier. For this purpose, 120 images were selected from the DDSM database with morphological formations. Based on these images, a training sample was formed with a volume of 370 segments. Fragments of test images of reference samples for class C_1 and C_1 are shown in Figure 2.



a)



b)

Fig. 2. Examples of test segments of X-ray images of the breast: a) - norm; b) – pathology

The appointment of the training sample is to establish indicators that characterize the condition of the examined patient. It represents 250 segments of normal breast radiographs and 120 segments contain morphological neoplasms caused by pathological processes. Due to it, two criteria were developed: "norm" - there is no area of interest, "pathology" - there is an area of interest.

Each segment corresponds to the predictor P2, described by the vector of three components. In experimental studies, the statistical characteristic Mo - mode (X1) was chosen as the first indicator evaluating the segment of morphological formation. The second indicator is M - expectation value (X2). The third component of the indicator is SD - mean square deviation (X3).

Table 1 shows fragments of the calculated values: mode, expectation value and mean square deviation for class C_1 segments (there are absent morphological formations with pathology), as well as for C_2 class segments (have morphological neoplasms caused by pathological processes).

Table 1. Fragment of the table of experimental data with informative signs for segments of class C_1 and segments of class C_1

№	informative signs of class C_1			informative signs of class C_2		
	X1	X2	X3	X1	X2	X3
1	2	3	4	2	3	4
1	107	106	16,67	174	186	39,9
2	99	88	29,35	170	184	67,11
3	107	99	19,75	177	208	64,01
4	117	93	25,69	198	168	43,01
5	60	52	25,87	188	201	53,73
6	99	93	22,02	170	156	43,83
7	94	96	20,06	146	160	40,94
8	89	96	17,01	170	205	44,33
9	89	95	19,23	185	182	48,16
10	94	100	19,53	158	150	40,03
11	40	46	31,73	153	173	42,74
12	54	45	29,27	151	145	41,41
13	77	79	16,37	159	187	42,81
14	89	79	29,22	174	212	31,83
15	99	77	30,58	169	182	37,49

The analysis of the experimental data in Table 1 indicates that for the segments in the normal state (class C_1), the mode values do not exceed 143, for the expectation value 153, and for the mean square deviation 36.71.

The averaged values of the mode, mathematical expectation and standard deviation are - 93.50, 88.47, 25.12, respectively.

For segments in which there are morphological formations caused by pathological processes (class C_2), the mode values do not exceed 250, the expectation value is 227, and for the mean square deviation, 67.11. The averaged values of the mode, mathematical expectation and standard deviation are - 176.40, 183.10, 32.15 respectively.

Figure 3 shows the fragments of the studied image segments for the class of norm, as well as its histogram, as well as a fragment of the segment containing the morphological neoplasm and the corresponding histogram.

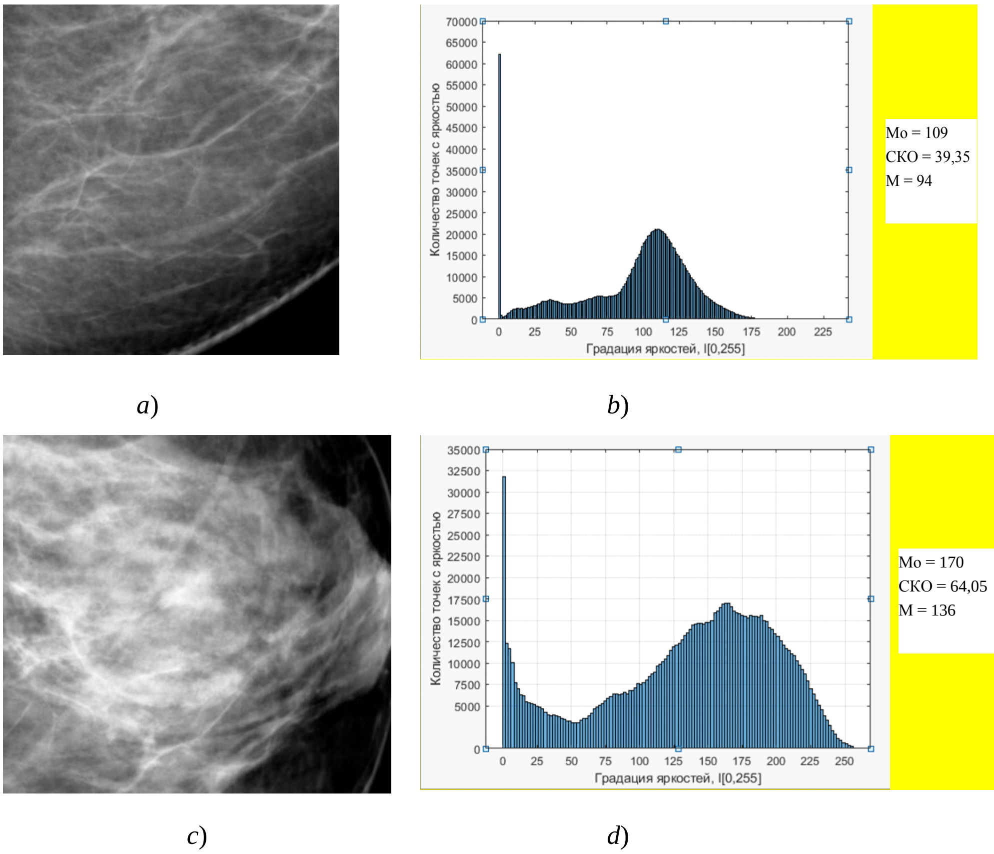
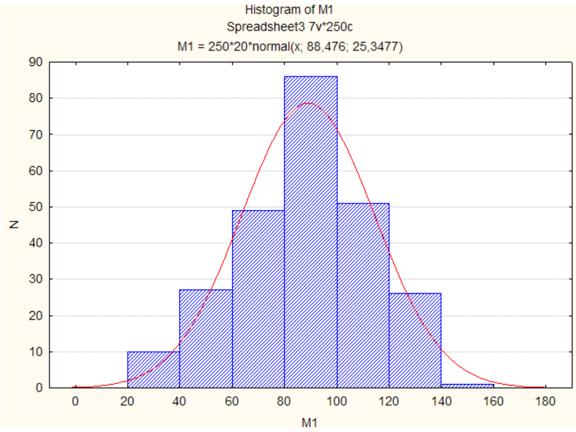


Fig.3. Fragment of the studied segments and their histograms:
a) a fragment of a segment in a normal state; b) the histogram of this segment; c) a fragment of a segment containing a morphological neoplasm; d) histogram of a segment containing a morphological neoplasm

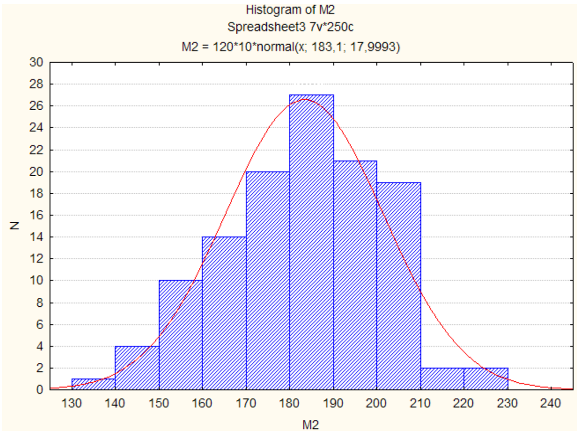
If we accept a priori the law on normality of component distribution of the vector P2, then each class is specified by the three-dimensional normal distribution of the joint probability density of the components of the vector P2. The normal character of such distribution is confirmed by checking its individual components and by the consequence of the central limit theorem [21, 22].

To check the sample for compliance with the normal distribution law, the STATISTICA program was used. With the aim of preliminary analysis, histograms and quantile - quantile of graphics were built.

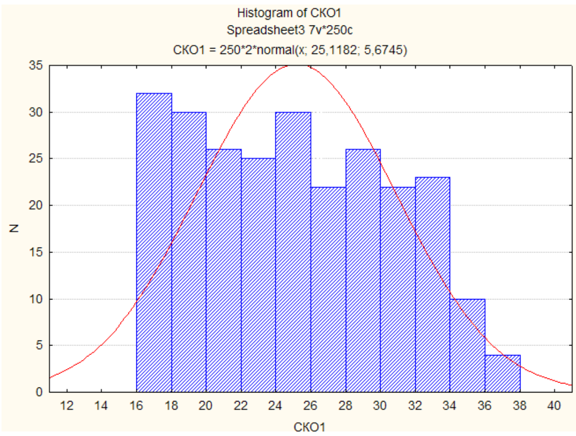
Figure 4 shows the graphs of histograms of indicators: mathematical expectation, standard deviation for classes C_1 and C_2. From the graphs of the histograms it conclusion follows that the obtained data correspond to the Gaussian curve, but there are outliers and anomalous values at the edges of the distribution.



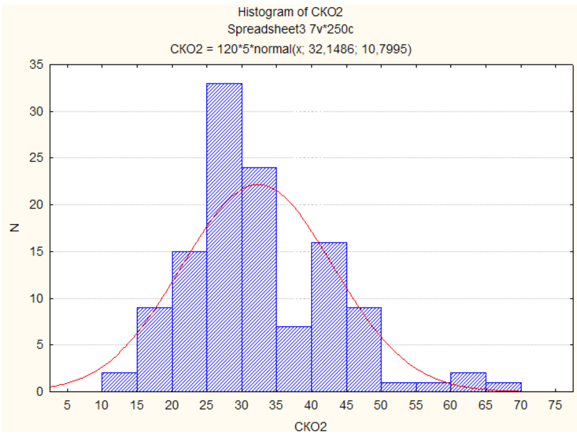
a)



b)



c)



d)

Fig. 4. Histogram of the normal distribution: a) expectation value for class C_1; b) expectation value for class C_2; c) mean square deviation for class C_1; d) mean square deviation for class C_2

Figure 5 depicted the quantile-quantile plots for three indicators: mode, expectation value, mean square deviation for classesC_1 and C_2.

From the quantile-quantile plots, it conclusion follows that all points lie along the line, which does not contradict the normal distribution law.

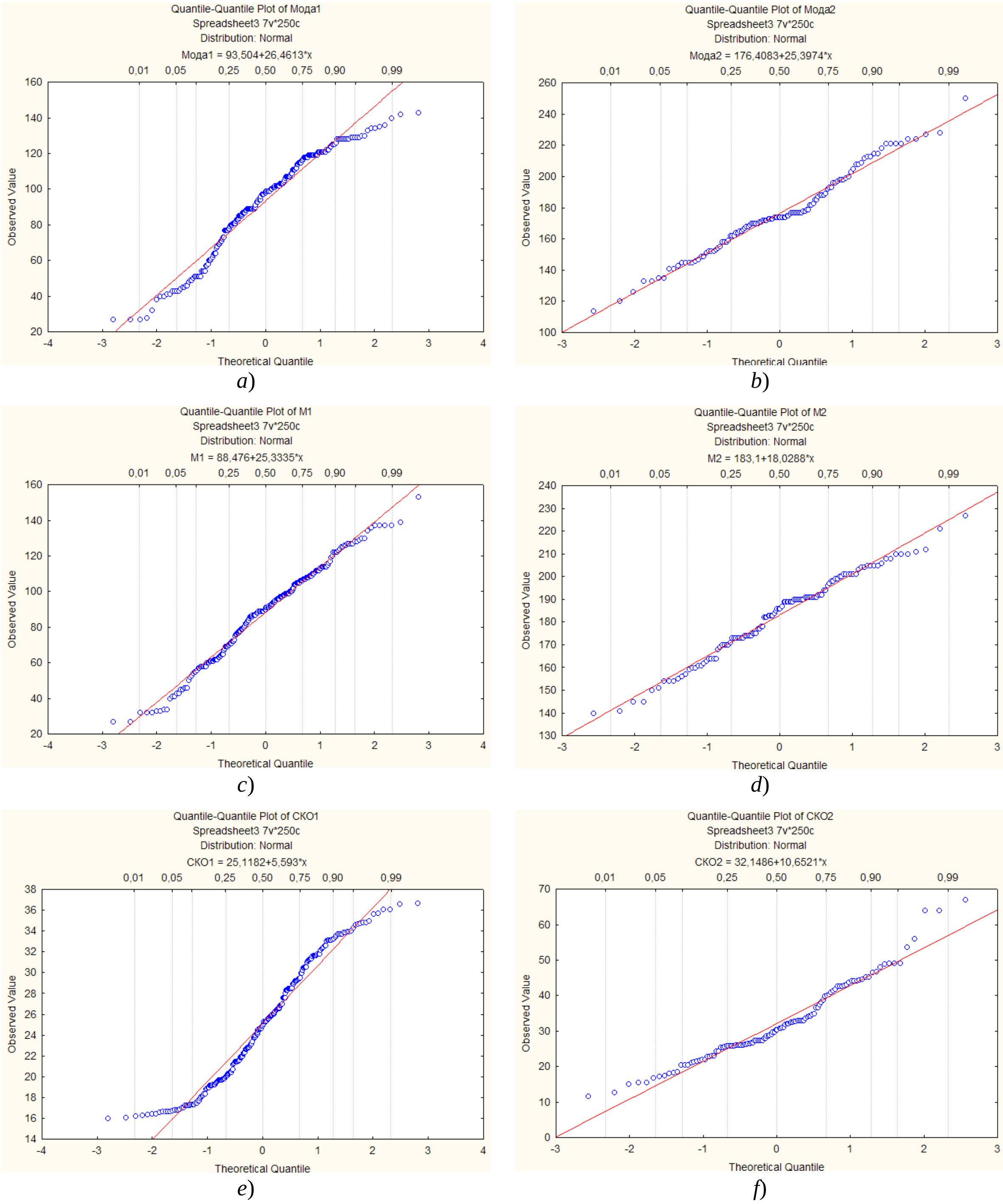


Fig. 5. Quantile-quantile graphs: a) the mode for class C_1 b) the mode for class C_2; c) for the expectation value of class C_1; d) for the expectation value of class C_2; e) for the mean square deviation of class C_1; f) for the mean square deviation of class C_2

For justification compliance with the normal distribution law, an additional check was carried out using the Kolmogorov – Smirnov test [22] with a confidence level of 0.95. Using the STATISTICA program, the Kolmogorov-Smirnov criterion was calculated, the results of which are presented in Table 2.

Table 2. The results of checking three indicators for classes C_1 and C_2 using the STATISTICA program

Variable	Normality tests (spreadsheet 3)		
	N	Max D	Kolmogorov-Smirnov
Mode of class $C1$	250	0,082	p< 0,10
Expectation value of class C_1 (M1)	250	0,081	p< 0,10
Mean square deviation of classC_1 (CKO1)	250	0,079	p< 0,10
Mode of class C_2	120	0,118	p< 0,10
Expectation value of class C_2 (M 2)	120	0,112	p< 0,10
Mean square deviation of class C_2 (CKO 2)	120	0,117	p< 0,10

In the third column of the table shows the calculated sample values of D, and the fourth column shows that all criteria are normally distributed, since the value of the probability p is less than 0.1, which corresponds to a confidence level of 0.95.

To verify that the program is calculated correctly, it is necessary to compare the obtained sampled values of D with the critical value of the Kolmogorov-Smirnov criterion for a confidence level of 0.95. The calculation of the critical value is carried out according to the formula [22]:

$$d = \sqrt{\frac{\ln\frac{2}{\alpha}}{2n} - \frac{1}{6n}}$$

where α is the level of significance of the distribution; n is the sample size.

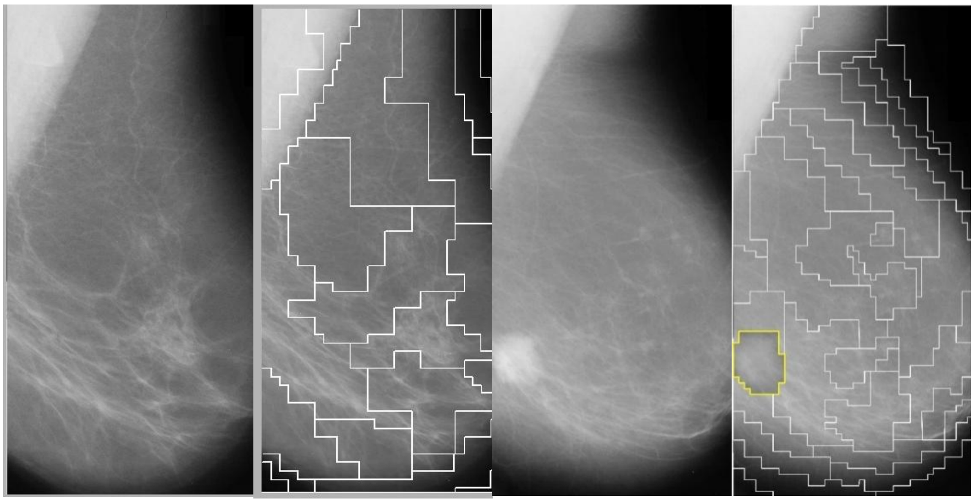
For a sample of 250 values, corresponding to class C1, the critical value is equal $d = 0{,}081921$, and for a sample of 120 values, belonging to class C2, the critical value is $d = 0{,}118243$. If $D \leq d$, then the hypothesis is accepted, if $D > d$, then the hypothesis is rejected [22]. Since all the calculated sample values of D are less than the critical values, then, therefore, the hypothesis is accepted, the samples are normally distributed.

5 Results

To assess the quality of mammograms segmentation by the proposed method, we used collections of X-ray images of the mammary gland from the Digital Database for Screening Mammography –DDSM, created on the basis of the University of South Florida and the University of Washington School of Medicine (available at: http://marathon.csee.usf.edu/ Mammography / Database.html). The database contains both normal cases obtained during the screening examination with the result "normal", and cancer cases confirmed by the results of further research.

50 images of mammograms of the mammary gland were selected from the database, of which the presence of malignant neoplasms was proved in 25 cases by specialists. Computer programs included in the automated system (Fig. 1) provide the selection of segments on the X-ray image that belong to the class “there is an area of interest”. Moreover, the dislocation of these segments should confirm the dislocation of segments with neoplasms, which were identified by a mammologist.

Figure 6 shows the results of classification of X-ray images from the control sample into the classes "there is an area of interest" - "there is no area of interest". If the computer detects an area of interest on the X-ray image, then the patient is assigned additional examinations.



a) *b*) *c*) *d*)

Fig. 6. Results of image segmentation: a) the original image "no area of interest"; b) processed image according to the criterion "no area of interest"; c) the original image “there is an area of interest.”; d) processed image according to the criterion "there is an area of interest."

Table 3 shows the indicators of the quality of diagnostics of X-ray images from the control sample. According to the indicator of diagnostic sensitivity (DH), the intelligent system showed a result of 0.84, and according to the indicator, diagnostic specificity (DS) - 0.96.

Table 3. Results of control tests of the automated system

Condition	Observation result		Total
	Positive	Negative	
there is an area of interest	21	4	25
no area of interest	1	24	25
Total	22	28	50

6 Discussion

Analysis and experimental studies of the known methods of segmentation of halftone raster images have shown that they all have certain drawbacks and cannot be used directly for segmentation of radiographs. Therefore, the segmentation algorithm should be built on a hierarchical principle. The lower hierarchy should contain intelligent agents that provide an improvement in the quality of segmentation by reducing the transition area between segments and increasing the brightness of border pixels. Taking into account D. Hubel's research, it is advisable to use morphological operators for preliminary processing of radiographs. The structure was developed of an intelligent system for the classification of radiographs based on the procedure for forming the selection of an area of interest and were developed uniformity criterions of the selected areas.

7 Conclusion

The approbation of the presented automated system and its software modules showed that out of 50 X-ray images, 45 were classified correctly, which amounted to 90% of the control sample. Moreover, false positive results accounted for 4% of the training sample. According to this indicator, the presented system surpasses the known ones by 2 ... 4%. In addition, there are reserves for reducing this indicator to increase the sensitivity threshold for individual segments. This is due to the fact that in order to send the patient to have an additional examination, it is enough to find only one segment on the image of the breast radiograph that satisfies the given condition, which makes it possible to increase both the diagnostic sensitivity and the diagnostic specificity.

References

1. Boyd H.F., Gio Y., Martin L.J., et. al. Mammographic density and the risk and detection of breast cancer// Engl. JMed. 2007. V. 356. P.227-229.
2. Medical device companies. Scanis // MedWOWGlobal. - K.: Copyright, 2014. - URL: http://ru.medwowglobal.com/company/scanis/91190 (accesed 15.09.2019).
3. Computer-Aided Detection for Digital Mammography syngo MammoCAD // Siemens. – Copyright, 2008. - P. 1-8. - URL: http://www.medical.siemens.com/siemens/en_INT/gg_sps_FBAs/files/brochures/cad/finalpdfMammoCAD.pdf (accesed 29.10.2019).
4. FUJIFILM Digital Mammography CAD // Fujifilm Europe. - M.: Copyright, 2019. - URL: http://www.fujifilm.eu/eu/products/medical-systems/products/p/fujifilm-digital-mammography-cad/ (accesed 15.05.2019).
5. Medical Imaging // Parascript. - M.: Copyright, 2019. - URL: http://www.parascript.com/medical-imaging/ (accesed 10.04.2019).
6. Tomakova R.A., Filist S.A., Pykhtin A.I. Development And Research Of Methods And Algorithms For Intelligent Systems For Complex Structured Images Classification// Journal of Engineering and Applied Sciences. 2017. V. 12. 22. P. 6039-6041.
7. Filist S.A. Dabagov A.R., Malutina I.A. Method for cascade segmentation of radiographs of the breast// [Proceedings of Southwestern state University. Series: Management, computer engineering, computer science. Medical instrumentation.- 2019. - T.9 №1(30). - pp. 49-61.
8. Ledeneva T.M., Podvalny S.L., Stryukov R.K., Degtyarev S.V. Fuzzy modeling of medical expert systems // Biomedical Radioelectronics. 2016. №9. P. 16-24.
9. Filist S.A. Multilayer morphological operators for the segmentation of complexly structured raster halftone images [Text] / S.A. Filist, A.R. Dabagov, I.A. Malutina, etc. // Proceedings of Southwestern state University. Series: Management, computer engineering, computer science. Medical instrumentation.- 2019. - T.9 №3. - pp. 44-63.
10. Rajashekaran, S., Vijayalksmi, G.A. Neural Networks, Fuzzy Logic and Genetic Algorithms: Synthesis and Applications, 2nd Edition. Prentice-Hall, 2017. - 572 p.
11. Yan, Zhicheng, et al. "Hd-cnn: Hierarchical deep convolutional neural network for image classification." International Conference on Computer Vision (ICCV). Vol. 2. 2015. 435-443.
12. Bernikov V.V., Preobrazhensky A.P., Choporov O.N. Possibilities for Parallelizing Image Processing Using OPENCV and PENMP // Modeling, Optimization and Information Technologies. 2019.T.9. №. 2 (25). pp. 110-126.
13. Titov D.V., Bekhtin Y.S., Egoshina I.L., Kaperko A.F. Processing of multi-spectral images for solving the recognition problem // Telecommunications. 2018. №5. pp. 35-38.
14. Korsunov N.I., Toropchin D.A. The method of finding the spam images based on the hash of the key points of the image//International Journal of Computing. 2016. T. 15. № 4. C. 259-264.
15. Filist S.A., Tomakova R.A., Degtyarev S.V., Rybochkin A.F. Hybrid Intelligent Models For Chest X-Ray Image Segmentation//Biomedical Engineering. 2018. V. 51. 5. P. 358-363.
16. Korsunov N.I., Toropchin D.A. Image classification method based on clustering complex objects // Scientific Bulletin of Belgorod State University. Series: Economics. Informatics. 2016. №23 (244). P. 100-103.
17. Tomakova R.A., Filist S.A., Pykhtin A.I. Automatic Fluorography Segmentation Method Based On Histogram Of Brightness Submission In Sliding Window//International Journal of Pharmacy and Technology. 2017. V. 9. 1. P. 28220-28228.

18. Filist S.A., Ali Qaboos, Kuzmin A.A. Formation of an attribute space for classification problems of complexly structured images based on spectral windows and neural network structures. Proceedings of Southwestern state University. Series: Management, computer technology, informatics. Medical instrumentation. 2016. №4 (67). P. 56-68.
19. Korsunov N.I., Toropchin D.A. A method for constructing a spam filter for searching for fuzzy duplicates of images // Information Systems and Technologies. 2017. №1(99). P. 13-20.
20. Korsunov N.I., Ushakova S.N. The structure of the neurocomputer signal classification system // Scientific Bulletin of Belgorod State University. Series: Economics. Informatics. 2019.T. 46. №.3. P. 496-502.
21. Dabagov, A.R. Automated classification system for breast radiographs [Text] / Dabagov A.R., Gorbunov V.A., Filist S.A., Malyutina I.A., Kondrashov D.S. // Medical equipment. - 2019. - №6 (318). - P. 39-41.
22. Dabagov, A.R. Automatic classifier of X-ray images using transparency masks / A.R. Dabagov, A.S. Bugaev // Proceedings of Southwestern state University. Series Management, computer technology, informatics. Medical instrumentation. - 2019. - №.4. - P. 106-125.
23. Dabagov, A.R. A four-stage algorithm for processing X-ray images in the systems of intelligent support for the classification of mammograms [Text] / A.R. Dabagov // System analysis and control in biomedical systems. - 2019. - T. 18, №4. - P. 117 - 127.
24. Zhilyakov E.G., Chernomorets A.A., Bolgova E.V., Kovalenko A.N. Image decomposition on the orthogonal basis of subband matrics eigenvectors//Journal of Engineering and Applied Sciences. 2017. T. 12. № 12. P. 3194-3197.
25. Akimov A.V., Donskikh A.O., Sirota A.A. Models and algorithms for recognizing digital images under the influence of deforming and additive distortions // Bulletin of Voronezh State University. Series: System Analysis and Information Technology. 2018. №1. P. 104-118.
26. Savvin S.V., Sirota A.A. Superpixel segmentation methods and their application for the analysis of images with heterogeneous texture // Bulletin of the Voronezh State University. Series: System Analysis and Information Technology. 2016. № 4. P. 165-172.
27. Zhilyakov E.G., Konstantinov I.S., Chernomorets A.A. Decomposition of images into additive components//International Journal of Imaging and Robotics. 2016. T.16. №1. P. 1-8.
28. Akimov A.V., Sirota A.A. Synthesis and Analysis of Algorithms for Digital Signal Recognition in Conditions of Deforming Distortions and Additive Noise Radioelectronics and Communications Systems. 2017. T.60. №10. P. 458-468.
29. Zhilyakov E.G., Chernomorets A.A., Bolgova E.V. On informational subdomains of spatial frequencies of images // Scientific news of Belgorod State University. Series: Economics. Informatics. 2016. №23 (244). P. 87-92.
30. Egoshina I.L. Complexation of Optical, Ultrasond and X-Ray Images in Intraoperatitive Navigation Systems //Bulletin of the Russian Academy of Sciences: Physics. 2018. T. 82. № 12. P. 1542-1546.
31. Bernikov V.V., Preobrazhensky A.P., Choporov O.N. Analysis of algorithms for detecting moving objects in the video image / Modeling, optimization and information technologies. 2018. T.6. №.3 (22). P. 223-233.
32. Donskikh A.O., Sirota A.A. Training of deep neural networks in a small sample for the classification of biological objects by multiplicative measurements // Bulletin of Voronezh State University. Series: System Analysis and Information Technology. 2019. №4. P. 109-118.
33. Zhilyakov E.G., Chernomorets A.A., Bolgova E.V., Oleynik I.I., Chernomorets D.A. Hidden Data Embedding Method Based on the Image Projections onto the Eigenvectors of

Subinterval Matrices// International Journal of Engineering and Technology(UAE). 2018. T. 7. № 3. P. 72-80.

34. Podvalny S.L., Mugatina V.M., Vasiliev E.M. Faceted neural networks in pattern recognition problems. // Mathematical methods in engineering and technology. - MMT. 2020. T.6. P.90-95.
35. Levenkov K.O., Korovin E.N., Novikova E.I. Neural network modeling of the process of choosing a treatment regimen for patients with chronic pyelonephritis and urolithiasis // Modeling, optimization and information technologies. 2018.T.6, №23 P. 61-71.