

Nataliya Shakhovska,
Ivan Izonin,
Sergio Montenegro,
Yannick Estève,
Jaime Campos,
Natalia Kryvinska
(Eds.)



Proceedings of the 2nd International Workshop on
Informatics & Data-Driven Medicine (IDDM 2019): Volume 1

Lviv, Ukraine
November, 2019

Shakhovska N., Izonin I., Montenegro S., Estève Y., Campos J., Kryvinska N., (Eds.):
The 2nd International Workshop on Informatics & Data-Driven Medicine (*IDDM 2019*):
Volume 1. Lviv, Ukraine, November 11-13, 2019, CEUR-WS.org, online

This volume represents the proceedings of the Workshop, with Posters and Demonstrations track, of the 2nd International Workshop on Informatics & Data-Driven Medicine, held in Lviv, Ukraine, in November, 2019. It comprises 33 contributed papers that were carefully peer-reviewed and selected from 67 submissions.

Copyright © 2019 for the individual papers by the papers' authors.
Copyright © 2019 for the volume as a collection by its editors. This
volume and its papers are published under the Creative Commons
License Attribution 4.0 International (CC BY 4.0).

IDDM 2019

Informatics & Data-Driven Medicine

Proceedings of the 2nd International Workshop on Informatics & Data-Driven Medicine (IDDM 2019)

Lviv, Ukraine, November 11-13, 2019.

Edited by

Nataliya Shakhovska *

Ivan Izonin **

Sergio Montenegro ***

Yannick Estève ****

Jaime Campos *****

Natalia Kryvinska *****

* Lviv Polytechnic National University, Ukraine

** Lviv Polytechnic National University, Ukraine

*** Julius-Maximilians-Universität Würzburg, Germany

**** University of Le Mans, France

***** Linnaeus University, Sweden

***** University of Vienna, Austria

Table of Contents

- Cover and Preface
- Organization and Committees

Key note speeches

- Online Hybrid Probabilistic-Fuzzy Clustering in Medical Data Mining Tasks (invited paper) 1-15
Yevgeniy Bodyanskiy, Anastasiia Deineko, Iryna Pliss, Olha Chala
- Cytological Image Classification Using Data Reduction (invited paper) 16-29
Oleh Berezsky, Oleh Pitsun, Taras Dolynyuk, Lesia Dubchak, Nadiya Savka, Grygoriy Melnyk, Vasyl Teslyuk
- On the Mathematical Model of Nonlinear Vibrations of a Biologically Active Rod with Consideration of the Rheological Factor (invited paper) 30-42
Petro Pukach, Volodymyr Il'kiv, Myroslava Vovk, Olha Slyusarchuk, Yulia Pukach, Yuriy Mylyan, Winfried Auzinger
- Information System for Visual Analyzer Disease Diagnostics (invited paper) 43-56
Ihor Bodnarchuk, Nataliia Kunanets, Serhii Martsenko, Oleksandr Matsiuk, Anastasiia Matsiuk, Volodymyr Pasichnyk, Roman Tkachuk, Hryhorii Shymchuk
- Concept of Intelligent Decision Support System in the Legal Regulation of the Surrogate Motherhood (invited paper) 57-68
Tetiana Hovorushchenko, Alla Herts, Yelyzaveta Hnatchuk

Session 1. Artificial intelligence

- Fuzzy System For Breast Disease Diagnosing Based On Image Analysis
Oleg Berezsky, Lesia Dubchak, Natalia Batryn, Tamara Datsko, Kateryna Berezska, Oleh Pitsun, Yuri Batko 69-83
- A Method for Extracting a Breast Image from a Mammogram Based on Binarization, Scaling and Segmentation 84-98
Eugene Fedorov, Tetyana Utkina, Kostiantyn Rudakov, Andriy Lukashenko, Serhii Mitsenko, Maryna Chychuzhko, Valentyna Lukashenko
- Identification of Biometric Images Using Latent Elements 99-108
Mariya Nazarkevych, Mykola Logoyda, Serhii Dmytruk, Yaroslav Voznyi, Olga Smotr
- Implementation of Selective Pressure Mechanism to Optimize Memory Consumption in the Synthesis of Neuromodels for Medical Diagnostics
Serhii Leoshchenko, Andrii Oliinyk, Sergey Subbotin, Tetiana Zaiko, Nataliia Gorobii 109-120
- Mathematical Modeling of Heat Transfer in Anisotropic Biophysical Materials, Taking Into Account the Phase Transition Boundary 121-132
Yaroslav Sokolovskyy, Iryna Boretska, Bogdana Gayvas, Igor Kroshnyy, Volodymyr Shymansky, Michal Gregus
- Mathematical Models and Analysis of Deformation Processes in Biomaterials with Fractal Structure 133-144
Yaroslav Sokolovskyy, Maryana Levkovich, Olha Mokrytska, Svitlana Yatsyshyn, Yaroslav Kaspryshyn, Christine Strauss

- Statistical Modeling of Diffusion Processes with a Fractal Structure 145-154
Ivan Sokolovskyy, Natalia Shakhovska
- Applying Regression Analysis to Study the Interdependence of Thyroid, Adrenal Glands, Liver, and Body Weight in Hypothyroidism and Hyperthyroidism 155-164
Olha Ryabukha, Ivanna Dronyuk
- Technique Medical Image Compression by Linear Algebra Methods 165-174
Dmytro Kucherov, Galina Rosinska, Natalia Khalimon, Ludmila Onikienko
- An Overview of Ateb-Theory Mathematical Apparatus for Data Confidentiality in Medical Computer Networks 175-184
Ivanna Dronyuk, Olga Fedevych, Danylo Lizanets, Natalia Kryvinska
- Application Artificial Intelligence for Real-Time Monitoring, Diagnostics, and Correction Human State 185-194
Tetiana Shmelova, Olexandr Sechkok
- Technique of Gene Regulatory Networks Reconstruction Based on ARACNE Inference Algorithm 195-207
Sergii Babichev, Bohdan Durnyak, Vsevolod Senkivskyy, Oleksandr Sorochnytskyi, Mykhailo Kliap, Orest Khamula
- Automated Document Analysis for Quick Personal Health Record Creation
Nataliya Boyko, Olena Pylypiv, Yulia Peleshchak, Yuriy Kryvenchuk, Jaime Campos 208-221
- Using a Bayesian Network to Assess the Atmospheric Pollution Influence on Immunological Parameters 222-233
Volodymyr Lytvynenko, Mariia Voronenko, Serge Sitalo, Oleg Boskin, Iryna Lurie, Nataliia Savina, Yaroslav Tanasiichuk, Nataliia Krugla

Session 2. Information systems in medicine

- Classification and Recognition of Medical Images Based on the SGTm Neuroparadigm 234-245
Viktor Khavalko, Ivan Tsmots, Anastasija Kostyniuk, Christine Strauss
- Information-Analytical Support to Medical Industry 246-257
Ivan Tsmots, Vasyl Teslyuk, Anatolyi Batyuk, Viktor Khavalko, Andreas Mladenow
- Analysis of Medical Systems to Control the Driver's Condition to Improve Traffic Safety 258-267
Mariya Nazarkevych, Vasyl Brytkovskyy, Michal Gregus
- Exploratory Analysis of Neuroblastoma Data Genes Expressions Based on Bioconductor Package Tools 268-279
Sergii Babichev, Bohdan Durnyak, Vsevolod Senkivskyy, Oleksandr Sorochnytskyi, Mykhailo Kliap, Orest Khamula
- Investigation of Mathematical Model of Acoustic Wave Propagation through Relax Environment in Ultrasound Diagnostics Problems 280-289
Zinovii Nytrebych, Volodymyr Il'kiv, Oksana Malanchuk, Winfried Auzinger
- Intellectual Analysis of Making Decisions Tree in Information Systems of Screening Observation for Immunological Patients 281-296
Lyubomyr Chyrun, Eugene Leshchynskyy, Vasyl Lytvyn, Antonii Rzhеuskyi, Victoria Vysotska, Yuriy Borzov

- Synthesis Control System Physiological State of a Soldier on the Battlefield
 " 297-306
Yurii Kryvenchuk, Ihor Helzynskyy, Tetiana Helzhynska, Nataliya Boyko, Roman Danel
- Fractal Distribution of Medical Data in Neural Network 307-318
Nataliya Boyko, Maxim Kuba, Lesia Mochurad, Sergio Montenegro
- A Robust Model for Integration of Artificial Intelligence Methods in Patient Care 319-329
Serge Dolgikh
- "Analysis of the Early Posttraumatic Period Pathophysiology in Case of the Severe Combined Thoracic Trauma Using Multivariate Logistic Regression" 330-339
Myroslav Stupnytskyi, Viktor Zhukov, Tatyana Gorbach, Oleksii Biletskii, Hakan Kutucu
- Development of Genetic Methods for Predicting the Incidence of Volumes of Emissions of Pollutants in Air 340-353
Andrii Oliinyk, Ievgen Fedorchenko, Alexander Stepanenko, Alexander Katschan, Yuliia Fedorchenko, Anastasia Kharchenko, Dmytro Goncharenko
- Methods and Models of Intellectual Processing of Texts for Building Ontologies of Software for Medical Terms Identification in Content Classification 354-368
Vasyl Lytvyn, Yevhen Burov, Petro Kravets, Victoria Vysotska, Andriy Demchuk, Andrii Berko, Yuriy Ryshkovets, Serhii Shcherbak, Oleh Naum
- Medical News Aggregation and Ranking of Taking into Account the User Needs 369-382
Nataliya Antonyuk, Lyubomyr Chyrun, Vasyl Andrunyk, Andrii Vasevych, Sofia Chyrun, Aleksandr Gozhij, Irina Kalinina, Yuriy Borzov
- Determining the Probability of Heart Disease using Data Mining Methods 383-394
Kseniia Bazilevych, Ievgen Menailov, Kirill Fedulov, Sergey Goranina, Dmytro Chumachenko, Pavlo Pyrohov

2019-11-03: submitted by Ivan Izonin, metadata incl. bibliographic data published under [Creative Commons CC0](#)

2019-11-03: published on [CEUR-WS.org](#) [valid HTML5]

A Method for Extracting a Breast Image from a Mammogram Based on Binarization, Scaling and Segmentation

Eugene Fedorov¹[0000-0003-3841-7373], Tetyana Utkina¹[0000-0002-6614-4133],
Kostiantyn Rudakov¹[0000-0003-0000-6077], Andriy Lukashenko²[0000-0002-6016-1899],
Serhii Mitsenko¹[0000-0002-9582-7486], Maryna Chychuzhko¹[0000-0001-5329-7897],
Valentyna Lukashenko¹[0000-0002-6749-9040]

¹ Cherkasy State Technological University, Cherkasy, Shevchenko blvd., 460, 18006, Ukraine
{t.utkina, ckc, k.rudakov, s.mitsenko,
m.chychuzhko}@chdtu.edu.ua, fedorovee75@ukr.net

² E. O. Paton Electric Welding Institute, Kyiv, Bozhenko str., 11, 03680, Ukraine
ineks-kiev@ukr.net

Abstract. The paper proposes a method for extracting a breast image from a mammogram. For this, mammogram binarization, scaling and segmentation of a binary scaled mammogram with the subsequent selection of the maximum connected component that corresponds to a breast image have been suggested. The proposed binarization uses uniform quantization that simplifies the selection of the threshold value for different mammograms. The proposed binary mammogram scaling uses a fast wavelet transform and an arithmetic mean filter with threshold processing which accelerates further segmentation. The proposed segmentation of binary scaled mammograms uses density clustering to extract connected components that can more accurately extract the breast image. The proposed method for processing a mammogram based on binarization, scaling, and segmentation can be used in various intelligent medical diagnostic systems.

Keywords: breast image extraction from a mammogram, binarization, quantization, threshold processing, scaling, fast wavelet transform, arithmetic mean filter, segmentation, density clustering.

1 Introduction

Currently, the methods of automatic and automated diagnostics of microcalcifications of mammary glands [1-2], nodes in the lungs [3]; polyps [4], pulmonary embolism [5]; brain tumors [6] et al. based on artificial intelligence approaches and applied to digital images are widely used.

For automatic and automated medical diagnostics, the extracting of a breast image from the mammogram, for which the methods of binarization, scaling and segmentation of images can be used, plays an important role.

For binarization of images, usually use such approaches as:

- automatic selection of a single-level global threshold (for example, Otsu's method) [7, 8];
- automatic selection of a single-level local threshold (for example, the methods of Bernsen, Eikvil, Niblack, Sauvola, Christian) [9, 10].

These methods have one or more of the following limitations:

- they do not perform binarization accurately;
- require a laborious procedure for determining the threshold value;
- require a laborious procedure for determining additional parameters.

In this regard, it is relevant to create a method of mammogram binarization, which will eliminate these limitations.

For image segmentation, usually use such approaches as:

- determination of the boundaries of the regions (as the boundaries of the regions, pixels with a large intensity gradient, as well as differing in color are selected) [11];
- definition of regions (regional growth, separation and merging of regions, watershed) [12];
- taxonomic [13];
- histogram [14];
- based on partial differential equations [15].
- variational [16];
- graph [17];
- based on Markov random field [18].

Taxonomic approach is the most popular of them.

Traditional methods of taxonomic approach are the following:

1. Methods based on partition (partition-based, partitioning-based) or center (center-based) (for example, k-means [19], PAM (k-medoids) [19], FCM [20], ISODATA methods [21]).
2. Methods of a model mixture or based on distribution (distribution-based) or model (model-based) (for example, EM [22]).
3. Density-based methods (for example, DBSCAN [23], OPTICS methods [24]).
4. Hierarchical methods:
 - agglomerative or ascending (bottom up) (for example, centroid communication, Ward's, single connection, full connection, group average methods) [25, 26];
 - divisive or top down (for example, DIANA, DISMEA methods) [27, 28].

Methods of taxonomic approach can also be based on metaheuristics [29, 30] and artificial neural networks [31-33].

These methods have one or more of the following limitations:

- have high computational complexity;
- do not allow to emit noise and random emissions;
- clusters cannot have different shapes and sizes;

- require the setting of the number of clusters;
- require the determination of parameter values.

In this regard, it is relevant to create a method of mammogram segmentation, which will eliminate these limitations.

Pre-scaling, which helps to reduce image size, is one of the ways to speed up segmentation.

For scaling images, usually use such approaches as:

- the method of the nearest neighbor [34];
- filtration (bilinear, bicubic, Lanczos and other filters) [35];
- supersampling (oversampling, mip-card) [36];
- spectral transformations [9].

These methods have one or more of the following limitations:

- have high computational complexity;
- provide low quality of images;
- require the determination of parameter values.

In this regard, it is relevant to create a method for mammogram scaling, which will eliminate these limitations.

The purpose of the work is to create a method for extracting a breast image from a mammogram based on binarization, scaling and segmentation. To achieve the goal, the following tasks have been set and solved:

1. To create a technique for mammogram binarization based on quantization and threshold processing.
2. To develop a technique for binary mammogram scaling.
3. To create a technique for binary scaled mammogram segmentation based on density clustering.
4. To develop a technique for determining the maximum connected component of a binary scaled mammogram, which corresponds to a breast image.
5. To create a technique for transforming the initial mammogram based on the maximum connected component of a binary scaled mammogram.
6. To conduct a numerical study.

2 Mammogram binarization based on quantization and threshold processing

In the paper a uniform scalar quantizer, which is optimal (the root-mean-square error of quantization is minimal) is used, and the quantization step is constant.

The proposed mammogram binarization includes the following steps:

1. Set the 8-bit mammogram $s(n_1, n_2)$, $n_1 \in \overline{1, N_1}$, $n_2 \in \overline{1, N_2}$. Set the number of quantization levels of pixel values L that is a multiple of the power of two. Set the min-

imum and maximum pixel values s_{\min} (in the case of the 8-bit image $s_{\min} = 1$), and s_{\max} (in the case of the 8-bit image $s_{\max} = 256$), respectively. Set the threshold value T .

2. Calculate the boundaries of the quantized pixel values for the corresponding quantization levels

$$d_i = s_{\min} + \frac{(s_{\max} - s_{\min})i}{L}, \quad i \in \overline{0, L}.$$

3. Calculate the quantized pixel values for the corresponding quantization levels

$$r_i = \frac{d_i + d_{i-1}}{2}, \quad i \in \overline{1, L}.$$

4. Quantize the 8-bit mammogram in the form

$$Q(s(n_1, n_2)) = \begin{cases} r_1, & d_0 < s(n_1, n_2) \leq d_1 \\ \dots & \dots \\ r_L, & d_{L-1} < s(n_1, n_2) \leq d_L \end{cases}, \quad n_1 \in \overline{1, N_1}, n_2 \in \overline{1, N_2}.$$

5. Perform the threshold processing in the form

$$b(n_1, n_2) = \begin{cases} 1, & Q(s(n_1, n_2)) > T \\ 0, & Q(s(n_1, n_2)) \leq T \end{cases}, \quad n_1 \in \overline{1, N_1}, n_2 \in \overline{1, N_2}.$$

As a result, the binary mammogram is formed.

3 Mammogram scaling

To scale the mammogram, the methods based on filtering and two-dimensional fast wavelet transform are offered in the paper.

3.1 Mammogram scaling based on arithmetic mean filter and threshold processing

The proposed mammogram scaling based on arithmetic mean filter and threshold processing includes the following steps:

1. Set the binary mammogram $b(n_1, n_2)$, $n_1 \in \overline{1, N_1}$, $n_2 \in \overline{1, N_2}$. Set the scaling parameter P , which determines the length of the square window as 2^P . Set the threshold value T .
2. Set the line number of the binary scaled mammogram $n_1 = 1$.
3. Set the column number of the binary scaled mammogram $n_2 = 1$.
4. Calculate the average pixel value in a window of $2^P \times 2^P$ size

$$\mu(n_1, n_2) = \frac{1}{2^p 2^p} \sum_{l_1, l_2} b(l_1, l_2),$$

$$l_1 \in \overline{(n_1 - 1)2^p + 1, (n_1 - 1)2^p + 2^p}, l_2 \in \overline{(n_2 - 1)2^p + 1, (n_2 - 1)2^p + 2^p}.$$

5. Convert the binary mammogram to

$$\hat{b}(n_1, n_2) = \begin{cases} 1, & \mu(n_1, n_2) > T \\ 0, & \mu(n_1, n_2) \leq T \end{cases}.$$

6. If it is not the end of the current line of the binary scaled mammogram, i.e. $n_2 < N_2 / 2^p$, then increase the column number of the current line of the binary scaled mammogram, i.e. $n_2 = n_2 + 1$. Go to step 4.
7. If it is not the last line of the binary scaled mammogram, i.e. $n_1 < N_1 / 2^p$, then increase the line number of the binary scaled mammogram, i.e. $n_1 = n_1 + 1$. Go to step 3.

As a result, the binary scaled mammogram is formed.

3.2 Mammogram scaling based on two-dimensional fast wavelet transform

The proposed mammogram scaling based on two-dimensional fast wavelet transform includes the following steps:

1. Set the binary mammogram $b(n_1, n_2)$, $n_1 \in \overline{1, N_1}$, $n_2 \in \overline{1, N_2}$. Set the scaling parameter P that determines the number of decomposition levels.
2. Set the decomposition level number $i = 1$.
3. For each line x , $x \in \overline{0, N_1 / 2^{i-1} - 1}$, at the current i -th decomposition level, a convolution of this line with transition functions FIR-HPF and FIR-LPF, respectively, is performed:

$$\tilde{d}_i(x, m) = \sqrt{2} \sum_{k=0}^{N_2/2^{i-1}-1} c_{i-1}(x, k) g(k + 2m), \quad m \in \overline{0, N_2 / 2^i - 1},$$

$$\tilde{c}_i(x, m) = \sqrt{2} \sum_{k=0}^{N_2/2^{i-1}-1} c_{i-1}(x, k) h(k + 2m), \quad m \in \overline{0, N_2 / 2^i - 1},$$

where $c_0(x-1, y-1) = b(x, y)$.

4. For each column y , $y \in \overline{0, N_2 / 2^i - 1}$, at the current i -th decomposition level, a convolution of this column with transition functions FIR-HPF and FIR-LPF, respectively, is performed:

$$d_i^{(d)}(m, y) = \sqrt{2} \sum_{k=0}^{N_1/2^{i-1}-1} \tilde{d}_i(k, y)g(k+2m), \quad m \in \overline{0, N_1/2^i-1},$$

$$d_i^{(v)}(m, y) = \sqrt{2} \sum_{k=0}^{N_1/2^{i-1}-1} \tilde{d}_i(k, y)h(k+2m), \quad m \in \overline{0, N_1/2^i-1},$$

$$d_i^{(h)}(m, y) = \sqrt{2} \sum_{k=0}^{N_1/2^{i-1}-1} \tilde{c}_i(k, y)g(k+2m), \quad m \in \overline{0, N_1/2^i-1},$$

$$c_i(m, y) = \sqrt{2} \sum_{k=0}^{N_1/2^{i-1}-1} \tilde{c}_i(k, y)h(k+2m), \quad m \in \overline{0, N_1/2^i-1}.$$

5. If $i < P$, then $i = i + 1$. Go to step 1.

6. Convert the values of approximating coefficients to the range of values $\{0,1\}$

$$c_{\min} = \min_{x,y} c_p(x, y), \quad x \in \overline{0, N_1/2^p-1}, \quad y \in \overline{0, N_2/2^p-1},$$

$$c_{\max} = \max_{x,y} c_p(x, y), \quad x \in \overline{0, N_1/2^p-1}, \quad y \in \overline{0, N_2/2^p-1},$$

$$\hat{b}(x+1, y+1) = \text{round} \left(\frac{c_p(x, y) - c_{\min}}{c_{\max} - c_{\min}} \right), \quad x \in \overline{0, N_1/2^p-1}, \quad y \in \overline{0, N_2/2^p-1},$$

where $\text{round}(x)$ is the x rounding.

As a result, the binary scaled mammogram is formed.

4 Binary scaled mammogram segmentation based on density clustering

The proposed segmentation of binary scaled mammogram includes the following steps:

1. Set the binary scaled mammogram $\hat{b}(n_1, n_2)$, $n_1 \in \overline{1, \hat{N}_1}$, $n_2 \in \overline{1, \hat{N}_2}$, where $\hat{N}_1 = N_1/2^p$, $\hat{N}_2 = N_2/2^p$. Set the size of the pixel neighborhood D (in the case of Moore neighborhood $D=9$). Initialize the pixel marking matrix $g(n_1, n_2) = 0$, $n_1 \in \overline{1, \hat{N}_1}$, $n_2 \in \overline{1, \hat{N}_2}$. Initialize the counter of the current number of connected components $c = 0$.
2. Set the mammogram line number $n_1 = 1$.
3. Set the mammogram column number $n_2 = 1$.
4. Calculate the current pixel number $i = (n_1 - 1)\hat{N}_2 + n_2$.
5. If the i -th pixel is already marked, i.e. $g(n_1, n_2) \neq 0$, then go to step 20.

6. Determine the neighborhood of the i -th pixel

$$U_{i,\varepsilon} = \{e \mid \widehat{b}(l_1 + n_1, l_2 + n_2) = 1\}, \quad e = (l_1 + n_1 - 1)\widehat{N}_2 + l_2 + n_2,$$

$$l_1 \in \{-1, 0, 1\}, l_2 \in \{-1, 0, 1\}.$$

7. If not all neighbors of the i -th pixel fall into its neighborhood, i.e. $|U_{i,\varepsilon}| < D$, then mark the i -th pixel as noise or random emission, i.e. $g(n_1, n_2) = -1$. Go to step 20.

8. Increase the counter of the current number of connected components $c = c + 1$.

9. Mark the i -th pixel as the c -th cluster, i.e. $g(n_1, n_2) = c$.

10. Create a multitude $S = U_{i,\varepsilon}$.

11. Extract from the set S the first element, i.e. $v = s_1$, and remove it from the set S , i.e. $S = S \setminus \{v\}$.

12. Calculate the coordinates of the v -th pixel in the mammogram

$$m_2 = v \bmod \widehat{N}_2, \quad m_1 = [(v - m_2) / \widehat{N}_2],$$

where $[\cdot]$ – taking the integer part of the number, \bmod – division modulo.

13. If the v -th pixel has been marked as noise or accidental release, i.e. $g(m_1, m_2) = -1$, then mark it as the c -th cluster, i.e. $g(m_1, m_2) = c$.

14. If the v -th pixel is already marked, i.e. $g(m_1, m_2) \neq 0$, then go to step 19.

15. Mark the v -th pixel, i.e. $g(m_1, m_2) = c$.

16. Determine the neighborhood of the v -th pixel

$$U_{v,\varepsilon} = \{e \mid \widehat{b}(l_1 + m_1, l_2 + m_2) = 1\}, \quad e = (l_1 + m_1 - 1)\widehat{N}_2 + l_2 + m_2,$$

$$l_1 \in \{-1, 0, 1\}, l_2 \in \{-1, 0, 1\}.$$

17. If not all neighbors of the v -th pixel fall into its neighborhood, i.e. $|U_{v,\varepsilon}| < D$, then go to step 19.

18. Combine the set S with the neighborhood of the v -th pixel, i.e. $S = S \cup U_{v,\varepsilon}$.

19. If the set S still contains pixels, i.e. $|S| > 0$, then go to step 11.

20. If it is not the end of the mammogram current line, i.e. $n_2 < \widehat{N}_2$, then increase the column number of the mammogram current line, i.e. $n_2 = n_2 + 1$. Go to step 4.

21. If it is not the last line of the image, i.e. $n_1 < \widehat{N}_1$, then increase the mammogram line number, i.e. $n_2 = n_2 + 1$. Go to step 3.

As a result, the pixel marking matrix of the segmented binary scaled mammogram is formed.

5 Determination of the maximum connected component of the binary scaled mammogram, which corresponds to a breast image

The proposed definition of the maximum connected component of the binary scaled mammogram includes the following steps:

1. Define the pixel marking matrix $g(n_1, n_2)$, $n_1 \in \overline{1, \widehat{N}_1}$, $n_2 \in \overline{1, \widehat{N}_2}$, where $\widehat{N}_1 = N_1 / 2^p$, $\widehat{N}_2 = N_2 / 2^p$. Set the number of the connected components c . Initialize the vector of the counters of the connected components dimensions $z(n) = 0$, $n \in \overline{1, c}$.
2. Set the line number of the pixel marking matrix $n_1 = 1$.
3. Set the column number of the pixel marking matrix $n_2 = 1$.
4. If the binary pixel refers to the connected component, i.e. $g(n_1, n_2) > 0$, then increase the size counter of the connected components, i.e. $z(g(n_1, n_2)) = z(g(n_1, n_2)) + 1$.
5. If it is not the end of the current line of the pixel marking matrix, i.e. $n_2 < \widehat{N}_2$, then increase the column number of the current line of the pixel marking matrix, i.e. $n_2 = n_2 + 1$. Go to step 4.
6. If it is not the last line of the pixel marking matrix, i.e. $n_1 < \widehat{N}_1$, then increase the line number of the pixel marking matrix, i.e. $n_1 = n_1 + 1$. Go to step 3.
7. Determine the number of the maximum connected component

$$c^* = \arg \max_n z(n), n \in \overline{1, c}.$$

As a result, the number of the maximum connected component of the binary scaled mammogram, which corresponds to a breast image, is determined.

6 Transformation of the original mammogram based on the maximum connected component of the binary scaled mammogram

1. Set the pixel marking matrix $g(n_1, n_2)$, $n_1 \in \overline{1, \widehat{N}_1}$, $n_2 \in \overline{1, \widehat{N}_2}$, $\widehat{N}_1 = N_1 / 2^p$, $\widehat{N}_2 = N_2 / 2^p$. Set the number of the maximum connected component c^* . Set the 8-bit mammogram $s(l_1, l_2)$, $l_1 \in \overline{1, N_1}$, $l_2 \in \overline{1, N_2}$. Set the scaling parameter P .
2. Set the line number of the pixel marking matrix $n_1 = 1$.
3. Set the column number of the pixel marking matrix $n_2 = 1$.
4. Convert the original mammogram to

$$s(l_1, l_2) = \begin{cases} s(l_1, l_2), & g(n_1, n_2) = c^* \\ 0, & g(n_1, n_2) \neq c^* \end{cases}$$

$$l_1 \in \overline{(n_1 - 1)2^P + 1, (n_1 - 1)2^P + 2^P}, \quad l_2 \in \overline{(n_2 - 1)2^P + 1, (n_2 - 1)2^P + 2^P}.$$

5. If it is not the end of the current line of the pixel marking matrix, i.e. $n_2 < \widehat{N}_2$, then increase the column number of the current line of the pixel marking matrix, i.e. $n_2 = n_2 + 1$. Go to step 4.
6. If it is not the last line of the pixel marking matrix, i.e. $n_1 < \widehat{N}_1$, then increase the line number of the pixel marking matrix, i.e. $n_1 = n_1 + 1$. Go to step 3.

As a result, the mammogram containing only a breast image is formed.

7 Numerical research

In the paper, the proposed method for extracting a breast image from a mammogram is investigated.

The quantization level of 16 and the threshold value of 0.5 are selected.

Fig. 1,a shows the original 8-bit mdb274 mammogram from the standard mini-MIAS mammogram database. Image size is 1024x1024 pixels.

Fig. 1,b shows the resulting 8-bit mdb274 mammogram that does not use scaling ($P=0$).

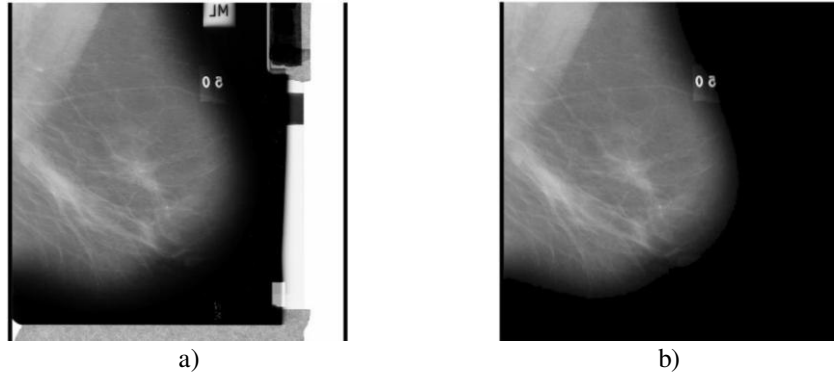


Fig. 1. 8-bit mdb274 mammogram:
a) the original, b) the resulting with $P=0$.

According to the experiments for mammograms from the standard mini-MIAS database, which are presented in Fig. 2, for scaling using the arithmetic mean filter with threshold processing, use the scaling parameter value $P=3$.

Such a value of the scaling parameter P , on the one hand, does not lead to significant changes in the shape of a breast image (this is typical for values 4, 5, 6), which

impair the visual perception, and, on the other hand, does not lead to a significant slowdown in segmentation (this is typical for values 1, 2).

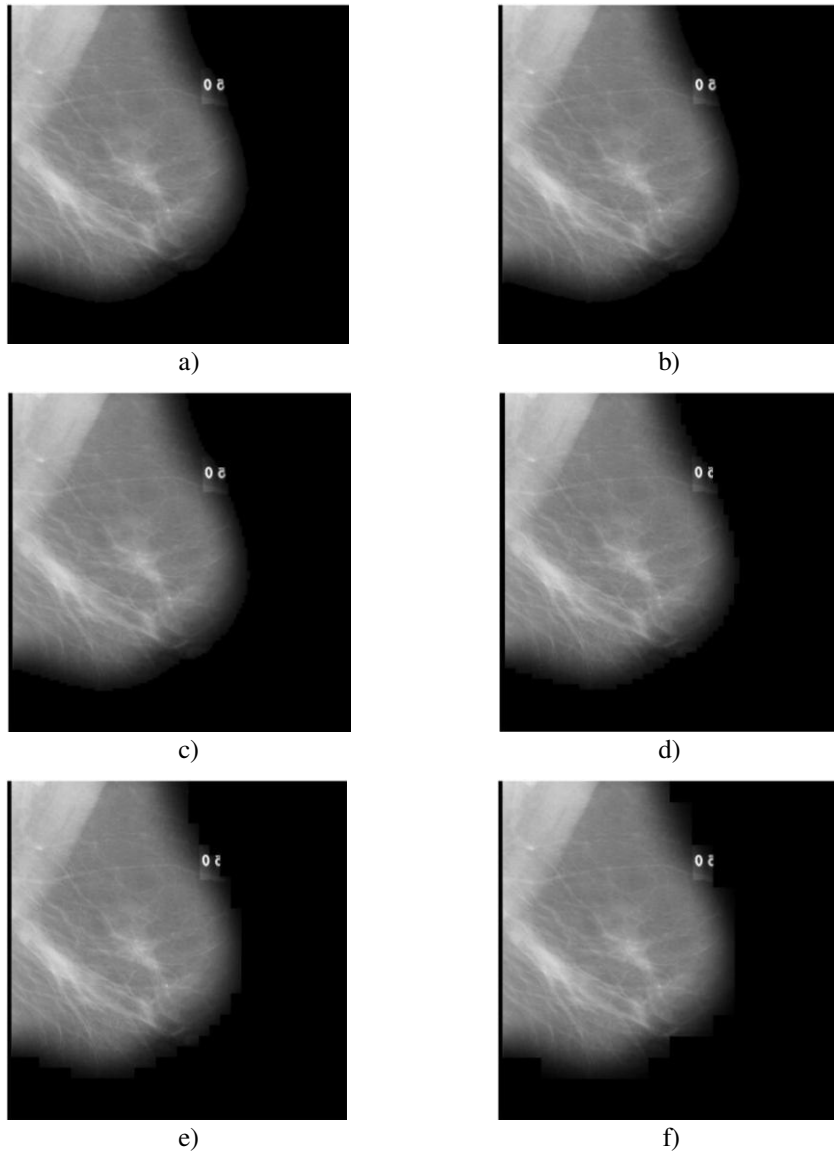


Fig. 2. 8-bit mdb274 mammogram in the case of scaling using the arithmetic mean filter with threshold processing: a) the resulting with $P = 1$; b) the resulting with $P = 2$; c) the resulting with $P = 3$; d) the resulting with $P = 4$; e) the resulting with $P = 5$; f) the resulting with $P = 6$.

According to the experiments for mammograms from the standard mini-MIAS database, which are shown in Fig. 3, for scaling using the fast wavelet transform by

means of Daubechies wavelet of length 8 (denoted as db4), use the scaling parameter value $P = 2$.

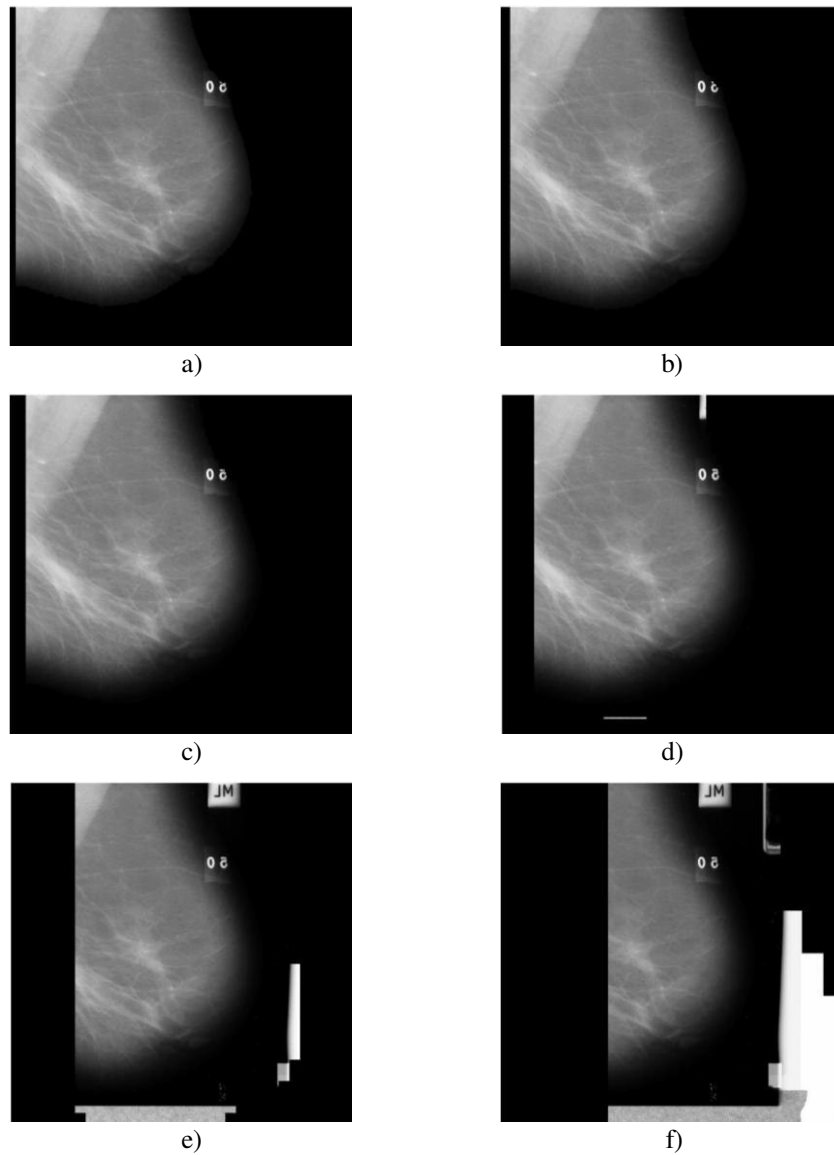


Fig. 3. 8-bit mdb274 mammogram in the case of scaling using the fast wavelet transform: a) the resulting with $P = 1$; b) the resulting with $P = 2$; c) the resulting with $P = 3$; d) the resulting with $P = 4$; e) the resulting with $P = 5$; f) the resulting with $P = 6$.

Such a value of the scaling parameter P , on the one hand, does not lead to significant changes in the shape of a breast image (this is typical for values 4, 5, 6), which

impair visual perception, and, on the other hand, does not lead to a significant slow-down in segmentation (this is typical for values 1, 2).

According to Fig. 2-3, the fast wavelet transform gives smoother edges than the arithmetic mean filter with threshold processing, but truncates a breast image, and also requires the choice of the wavelet and its length, and has greater computational complexity, while the arithmetic mean filter with threshold processing does not require the setting of additional parameters.

The dependence of the segmentation time of the binary mdb274 mammogram on the scaling parameter P is shown in Fig. 4. The experiments have been conducted on a computer with an Intel Pentium Quad-Core processor with a clock frequency of 2.58 GHz.

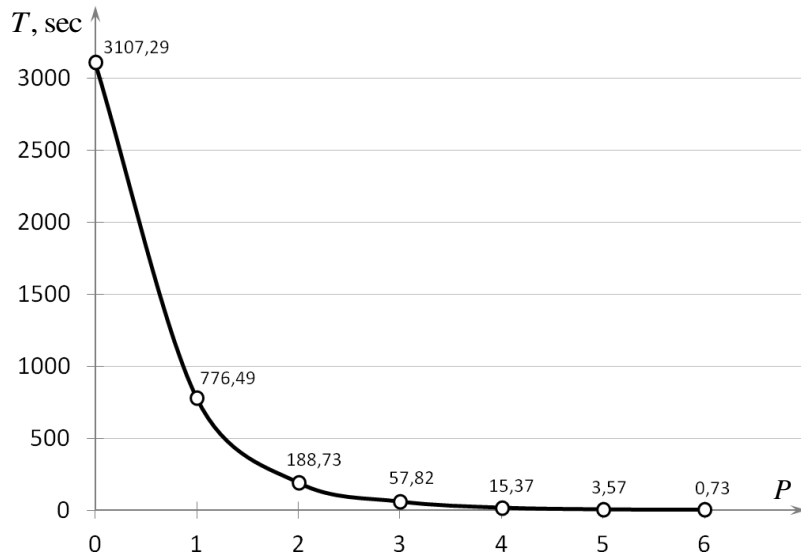


Fig. 4. Dependence of the segmentation time of mdb274 mammogram on scaling parameter.

According to Fig. 4, the dependence of the segmentation time on the scaling parameter is close to exponential one and shows that, starting from $P = 3$, the segmentation time changes only slightly.

8 Conclusions

1. To solve the problem of improving the quality of medical diagnostics, the corresponding methods of digital image processing have been investigated. The research data show that today, the methods of binarization, scaling and segmentation are actively used to extract a breast image from a mammogram.
2. The created method for extracting a breast image from a mammogram sequentially performs:

- binarization based on uniform quantization and threshold processing, which allows to more accurately separate the background and mammogram objects, simplifies the selection of the threshold value for different mammograms;
 - binary mammogram scaling based on the arithmetic mean filter with threshold processing and fast wavelet transform, which accelerates further segmentation. To increase the accuracy and speed of scaling experimentally, the scaling parameter is determined;
 - binary scaled mammogram segmentation based on density clustering for extraction of connected components, which allows to more accurately extract a breast image, not take noise and accidental releases into account, extract connected components of different shapes and sizes, not indicate the number of connected components;
 - selection of the maximum connected component that matches a breast image;
 - transformation of the initial mammogram based on the maximum connected component of the binary scaled mammogram.
3. The proposed method for extracting a breast image from a mammogram based on binarization, scaling and segmentation can be used in various intelligent medical diagnostic systems.

References

1. Zhang, W., Doi, K., Giger, M.L., Wu, Y., Nishikawa, R.M., Schmidt, R.A.: Computerized detection of clustered microcalcifications in digital mammograms using a shift-invariant artificial neural network. *Medical Physics*. 21, 517–524 (1994). doi: 10.1118/1.597177
2. Chan, H.-P., Lo, S.-C.B., Sahiner, B., Lam, K.L., Helvie, M.A.: Computer-aided detection of mammographic microcalcifications: Pattern recognition with an artificial neural network. *Medical Physics*. 22, 1555–1567 (1995). doi: 10.1118/1.597428
3. Lo, S.-C., Lou, S.-L., Lin, J.-S., Freedman, M., Chien, M., Mun, S.: Artificial convolution neural network techniques and applications for lung nodule detection. *IEEE Transactions on Medical Imaging*. 14, 711–718 (1995). doi: 10.1109/42.476112
4. Tajbakhsh, N., Gurudu, S.R., Liang, J.: A Comprehensive Computer-Aided Polyp Detection System for Colonoscopy Videos. *Lecture Notes in Computer Science Information Processing in Medical Imaging*, 327–338 (2015). doi: 10.1007/978-3-319-19992-4_25
5. Tajbakhsh, N., Gotway, M.B., Liang, J.: Computer-Aided Pulmonary Embolism Detection Using a Novel Vessel-Aligned Multi-planar Image Representation and Convolutional Neural Networks. *Lecture Notes in Computer Science Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2015*. 62–69 (2015). doi: 10.1007/978-3-319-24571-3_8
6. Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P.-M., Larochelle, H.: Brain tumor segmentation with Deep Neural Networks. *Medical Image Analysis*. 35, 18–31 (2017). doi: 10.1016/j.media.2016.05.004
7. Bow, S.-T.: *Pattern recognition and image preprocessing*. Marcel Dekker, New York (2002).
8. Pratt, W.K.: *Digital Image Processing*. John Wiley & Sons (2016).
9. Gonzalez, R.C., Woods, R.E.: *Digital image processing*. Pearson, New York, NY (2018).

10. Ritter, G.X., Wilson, J.N.: Handbook of computer vision algorithms in image algebra. CRC Press, Boca Raton (2001).
11. Martin, D., Fowlkes, C., Malik, J.: Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 26, 530–549 (2004). doi: 10.1109/TPAMI.2004.1273918
12. Ballard, D.H., Brown, C.M.: Computer vision. Prentice-Hall, Englewood Cliffs (1982).
13. Comaniciu, D., Meer, P.: Mean shift: a robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 24, 603–619 (2002). doi: 10.1109/34.1000236
14. Shapiro, L.G., Stockman, G.C.: Computer vision. Prentice-Hall, Upper Saddle River, NJ (2001).
15. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. *International Journal of Computer Vision*. 1, 321–331 (1988). doi: 10.1007/BF00133570
16. Chan, T., Vese, L.: Active contours without edges. *IEEE Transactions on Image Processing*. 10, 266–277 (2001). doi: 10.1109/83.902291
17. Wu, Z., Leahy, R.: An optimal graph theoretic approach to data clustering: theory and its application to image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 15, 1101–1113 (1993). doi: 10.1109/34.244673
18. Geman, S., Geman, D.: Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. PAMI-6, 721–741 (1984). doi: 10.1109/TPAMI.1984.4767596
19. Brusco, M.J., Shireman, E., Steinley, D.: A comparison of latent class, K-means, and K-medial methods for clustering dichotomous data. *Psychological Methods*. 22, 563–580 (2017). doi: 10.1037/met0000095
20. Bezdek, J.C.: Pattern recognition with fuzzy objective function algorithms. Plenum, New York (1987).
21. Ball, G.H., Hall, D.J.: Isodata, a novel method of data analysis and pattern classification. Stanford Research Institute, Menlo Park, CA (1965).
22. Fu, Z., Wang, L.: Color Image Segmentation Using Gaussian Mixture Model and EM Algorithm. *Multimedia and Signal Processing Communications in Computer and Information Science*. 61–66 (2012). doi: doi.org/10.1007/978-3-642-35286-7_9
23. Ester, M., Kriegel, H.-P., Sander, J., Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise . In: Simoudis, E., Han, J. (eds.) *KDD'96 Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, pp. 226–231. AAAI Press, Portland, Oregon (1996).
24. Ankerst, M.M., Breunig, M.M., Kriegel, H.-P.M., Sander, J.M.: Optics: Ordering points to identify the clustering structure. *Proceedings of the 1999 ACM SIGMOD international conference on Management of data - SIGMOD 99*. 28, 49–60 (1999). doi: 10.1145/304182.304187
25. Kaufman, L., Rousseeuw, P.J.: Finding groups in data: an introduction to cluster analysis. Wiley-Interscience, Hoboken, NJ (2005).
26. Mirkin, B.G.: Clustering for data mining a data recovery approach. Chapman & Hall/CRC, Boca Raton, FL (2012).
27. Aggarwal, C.C., Reddy, C.K.: Data Clustering: Algorithms and Applications. Chapman and Hall/CRC, Boca Raton, FL (2018).
28. Gan, G., Ma, C., Wu, J.: Data clustering: theory, algorithms, and applications. SIAM, Society for Industrial and Applied Mathematics, Philadelphia, PA (2007).

29. Subbotin, S., Oliinyk, A., Levashenko, V., Zaitseva, E.: Diagnostic Rule Mining Based on Artificial Immune System for a Case of Uneven Distribution of Classes in Sample. *Communications - Scientific Letters of the University of Zilina*. 18(3), 3–11 (2016).
30. Fedorov, E., Lukashenko, V., Utkina, T., Lukashenko, A., Rudakov, K.: Method for Parametric Identification of Gaussian Mixture Model Based on Clonal Selection Algorithm. *Computer Modeling and Intelligent Systems*. 2353, 41–55 (2019).
31. Fedorov, E., Lukashenko, V., Patrushev, V., Lukashenko, A., Rudakov, K., Mitsenko, S.: The Method of Intelligent Image Processing Based on a Three-Channel Purely Convolutional Neural Network. *Informatics & DataDriven Medicine*. 2255, 336–351 (2018).
32. Tkachenko, R., Izonin, I.: Model and Principles for the Implementation of Neural-Like Structures Based on Geometric Data Transformations. *Advances in Intelligent Systems and Computing Advances in Computer Science for Engineering and Education*. 578–587 (2018). doi: 10.1007/978-3-319-91008-6_58
33. Larin, V.J., Fedorov, E.E.: Combination of PNN network and DTW method for identification of reserved words, used in aviation during radio negotiation. *Radioelectronics and Communications Systems*. 57, 362–368 (2014). doi: 10.3103/S0735272714080044
34. Bannore, V.: *Iterative-Interpolation Super-Resolution Image Reconstruction a Computationally Efficient Technique*. Springer Berlin Heidelberg, Berlin, Heidelberg (2009).
35. Friedland, G.: *Multimedia computing*. Cambridge University Press, Cambridge (2014).
36. O’Gorman, L., Sammon, M.J., Seul, M.: *Practical algorithms for image analysis: description, examples, programs, and projects*. Cambridge University Press, Cambridge (2009).