

Web-Based System for Remote Mammogram Processing

Marina Milošević, Dejan Vujičić, Đorđe Damnjanović, Marija Vujović

Department of Computer and Software Engineering
University of Kragujevac, Faculty of Technical Sciences
Čačak, Serbia

marina.milosevic@ftn.kg.ac.rs, dejan.vujicic@ftn.kg.ac.rs, djordje.damnjanovic@ftn.kg.ac.rs, marija.vujovic@ftn.kg.ac.rs

Abstract: This paper presents a web-based system for the remote processing of mammographic images using Matlab Web Service. The Matlab Web Service is a toolbox for accessing Matlab applications from the internet. Our system for mammogram processing is a Matlab application stored on a local workstation. By using a web browser users can upload their data and view the results obtained after processing. The presented computer system for mammogram processing is based on feature extraction techniques for detecting abnormalities in digital mammograms. A total of 20 texture features based on gray-level co-occurrence matrices are used to evaluate the effectiveness of textural information possessed by mass regions. To investigate the ability of feature set in differentiating abnormal from normal tissue, we employed three different classifiers: the support vector machine, the Naive Bayes classifier and the K-Nearest Neighbor classifier.

Keywords-*Breast cancer; Image processing; Mammography; Matlab; Web service;*

I. INTRODUCTION

Breast cancer is the most common malignancy in women. It is often characterized by a lack of early symptoms, which results in late detection of the disease. Breast cancer screening with mammography has been proven very effective in preventing breast cancer death. Nonetheless, due to the very complex anatomy of the breast and small differences in the density of various breast tissues, the analysis of mammographic images (mammograms) is significantly hindered. Thus, the software support for the mammographic detection of tumors (Computer-Aided Diagnosis - CAD) has become very important.

There are many commercial and non-commercial CAD systems based on image processing to assist in diagnosis and therapy. However, most of these systems are available only as local applications that run on high-performance workstations. In that case, the processing power of the application is not available to other workstations. Using web-based tools, users can access the CAD system functionalities wherever the internet is available, without taking into account operating system and processing power limitations.

There are many web-based medical software tools for accessing DICOM (Digital Image Communication in Medicine) images through the web [1]-[3]. Presented web

applications are usually picture archiving and communication systems (PACS) or Medical Informatics Systems (MIS). Various studies have focused on medical image processing web-based tools available for research and diagnosis purposes. Mahmoudi with the group of authors in [4] presented a web-based, interactive, extendable, 2D and 3D medical image processing and visualization application. Medical image preprocessing, registration, and segmentation methods are implemented. They used a four-layered software design consisting of an algorithm layer, web-user-interface layer, server communication layer and wrapper layer. Jiménez with the group of authors in [5] presented a web platform for computing and analyzing the 3D fractal dimension of magnetic resonance images (MRIs) of the brain. The presented web platform is successfully applied in a study where an analysis and characterization of groups of 3D MR images are performed for three neurodegenerative diseases: Multiple Sclerosis, Intrauterine Growth Restriction and Alzheimer's disease. Paper [6] presents a method for integration of image processing and display into a web-based image distribution server to enable users to access, view and manipulate PACS DICOM images using Web browsers. A Web-based platform that allows registered users to upload, visualize, process and comment on medical images is developed by Millan and Yunda [7]. The system contains a social network module, a video conference and webinar module and an image visualization, annotation and processing module. Cai with the group of authors developed a system that provides efficient and flexible digital medical image data access and manipulation over the Internet [8]. The presented system has the potential to integrate and associate data from multiple vendors' systems and multiple imaging modalities at a regional level. The authors of [9] reviewed medical image processing, cloud computing and cloud computing benefits due to medical image processing. They also presented tools and methods for medical image processing using the cloud. Yuan with the group of authors in [10] proposed a web-based platform Rayplus for medical image processing. The system allows access to applications written in C++ through the internet.

A computer system for tumor diagnosis in mammograms described in this paper includes three modules: image

preprocessing, feature extraction, and classification. The preprocessing phase consists of the image de-noising, region of interest (ROI) extraction and image contrast enhancement. In the second phase, we analyzed the texture measures based on gray-level co-occurrence matrices (GLCM) for the classification of preprocessed mammograms. Support vector machine (SVM) classifier, Naive Bayes classifier, and K-Nearest Neighbor (k-NN) classifier were employed to estimate the classification performances of the extracted textural features in the final stage of mammogram processing.

The significance of the realized CAD system increases if it is available to a larger number of users at any time. Matlab web service application enables access to CAD system for the detection of breast cancer in mammograms from remote locations at every moment.

CAD system for breast cancer detection is available as a Matlab application, stored on a local workstation. User can upload the mammographic image and enter the necessary data that is sent to a server for processing. After executing the Matlab application on the server, user receives a notification of the classifier's estimates about the breast cancer existence.

The rest of this paper is structured as follows. Section 2 describes modules of the CAD system for breast cancer detection: image preprocessing, feature extraction, and classification, as well as classification performance. Section 3 gives a general description of the Matlab web service. The user interface of the Matlab web service application is described in Section 4. Conclusions are given in Section 5. The last section contains a literature review.

II. CAD SYSTEM FOR DETECTION OF TUMOR IN DIGITIZED MAMMOGRAMS

A. Image preprocessing

Preprocessing of mammograms aims to provide better conditions for the extraction of desired objects. The preprocessing phase of the presented CAD system consists of three segments: image de-noising, region of interest extraction, and enhancement of image contrast.

Image noise removal is an important part of many CAD systems based on image processing. A 2D adaptive median filter [11] is used to remove noise from mammograms. Each pixel in the image compares with its surrounding neighbor pixels. A pixel that is different from the majority of its neighbors is considered as noise and replaced by the neighborhood median value. The size of the neighborhood is adjustable.

The region of interest extraction in mammogram processing involves background and pectoral muscle removal. It is a necessary step of preprocessing phase, especially in the case of scanned analog mammograms that often contain undesirable objects. ROI extraction procedure provides extraction of relevant characteristics of mammograms. We extracted the region of interest using the algorithm presented in [12]. The

mentioned algorithm includes skin line detection, detection of MLO view type, extraction of the breast region and pectoral muscle removal procedure. Original mammograms from the Mini-MIAS database [13] after de-noising and output of the ROI extraction procedure are depicted in the Fig.1 (a) and the Fig.1 (b), respectively.

In the third part of the preprocessing stage, an adjustment of pixel intensity values is performed. Values between low and high intensities of the input image are mapped to values between low and high intensities of the output image. Values below the low intensity of the input image are mapped to the low intensity of the output image and those above the high intensity of the input image are mapped to the high intensity of the output image. The result of the contrast enhancement operation is shown in the Fig.1 (c).

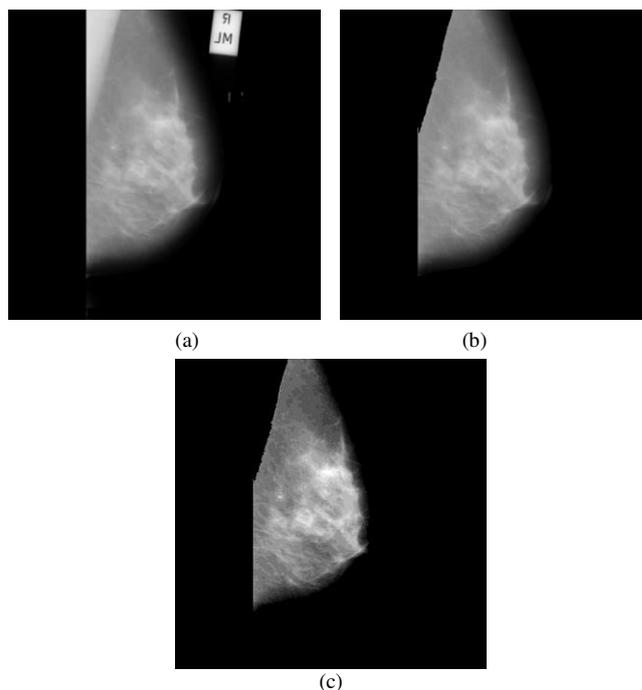


Figure 1. Mammogram preprocessing: (a) Mammogram mdb266 from mini-MIAS database after de-noising; (b) Detected region of interest; (c) Contrast enhancement operation result.

B. Features extraction and classification

Architectural distortion in the breast tissues arises as a consequence of the presence of tumor masses. As a result, textural information of mammograms that contain masses and mammograms without masses are significantly different. In the present study, the classification of ROIs as normal or abnormal is based on the GLCM texture measures.

The GLCM is a method based on calculation how often two pixels with specific values and in a specified spatial relationship occur in an image. The spatial relationship of the pixels is defined by orientation θ and distance d . Each element in the GLCM is the sum of the number of times that the pixel with one specific value occurred in the specified spatial relationship to a pixel with another specific value in the input

image. GLCM can be calculated in four different angles θ (0° , 45° , 90° , 135°) and for different values of distance d . The size of the GLCM is determined by the number of gray levels in the image.

In this study, 4 matrices corresponding to 4 directions $\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ$ and pixel distance $d = 1$ were computed for each ROI. A large number of co-occurrences derived from ROI is achieved using distance $d = 1$. Four values corresponding to the four matrices were obtained for each feature. Texture feature is represented by an average value.

We used a total of 20 GLCM features: Angular Second Moment (Energy), Contrast, Correlation, Variance, Inverse Difference Moment (Homogeneity), Sum Average, Sum Variance, Sum Entropy, Entropy, Difference Variance, Difference Entropy, Information Measure of Correlation 1 and Information Measure of Correlation 2 proposed by Haralick [14], Autocorrelation, Dissimilarity, Cluster Shade, Cluster Prominence and Maximum Probability proposed by Soh [15], and features Inverse Difference Normalized and Inverse Difference Moment Normalized proposed by Clausi [16].

The classification process consists of two phases: the training and the testing phase. The classifier is trained using the known data. Then, a trained classifier performed classification of the unknown data. Support Vector Machine [17], K-Nearest Neighbor classifier [18] and Naive Bayes classifier with diagonal covariance matrix estimate [19] were employed to classify the normal and abnormal cases.

C. Classification performance

Mammogram classification results presented in [12] confirm the effectiveness of the applied method for two-class mammogram classification. A five-fold cross-validation method, confusion matrix, and receiver operating characteristic analysis were performed to evaluate the classification performance. The data set used in this study consists of 300 randomly selected mammograms from the local database, 150 abnormal and 150 normal patterns. Masses of varying size and location are present in abnormal mammograms. The classification performances are given in the Table I.

TABLE I. CLASSIFICATION PERFORMANCE ACHIEVED USING THE SVM CLASSIFIER, K-NN CLASSIFIER, AND NAIVE BAYES CLASSIFIER IN THE TEST WITH THE MAMMOGRAMS FROM THE LOCAL DATABASE

Performance Measures	Classification schemes		
	SVM	k-NN	Naive Bayes
TP	121	84	114
FP	20	71	32
FN	29	66	36
TN	130	79	118
Accuracy	83.7%	54.3%	77.3%
Sensitivity	80.7%	56%	76%
Specificity	86.7%	52.7%	78.7%

TP=true positives; FP=false positives; FN=false negatives; TN=true negatives

An accuracy ratio of 83.7% indicates that a SVM classifier can be trained to effectively classify mammograms from the local database. Results in Table I also indicate that there is no significant difference in classification accuracy between the SVM classifier (83.7%) and the Naive Bayes classifier (77.3%). Classification results obtained using the SVM classifier and Naive Bayes classifier are much better than the classification results obtained using the k-NN classifier, with an accuracy ratio of 83.7% and 77.3% according to 54.3%.

III. MATLAB WEB SERVICE

Matlab Web Service (MWS) is a toolbox for the MATLAB software package that allows users to run Matlab applications remotely over the internet using standard web technologies. MWS is not part of the standard Matlab installation. It is installed subsequently as a separate toolbox. In this paper, the Modelit Matlab Webservice Toolbox [20] was used. In the simplest MWS configuration, the user enters the input values using the web browser and then runs the Matlab applications. Web browser runs on a client computer while Matlab application runs on another (server) machine. After executing the Matlab application, the server computer sends the output of the application to the user-designed HTML (HyperText Markup Language) document in the web browser. In this way, users can access Matlab applications over the internet, from any remote location in an easy manner at the lowest possible cost and without software installation.

The MWS application is a combination of Matlab executable files (m-files), HTML files, and graphics. Knowledge of Matlab programming and HTML programming are the only requirements to design a Matlab application that runs on a Matlab Web Service. The realization of the MWS application can be divided into the following steps:

- Creating a user interface in the form of HTML documents for the collection of input data from the user and display of results obtained after the execution of the Matlab application,
- Writing the Matlab application as an executable m-file that:
 - Receives the data entered by the client into an input HTML document;
 - Processes the entered data and generates the required output data (numerical values, text, images, etc.);
 - Forwards the output data to the output HTML document.
- Specifying the name of the main executable m-file and associated configuration data in the configuration file.

Fig. 2 shows the interaction between a client application (client) and Matlab over the internet. The client application (usually a web browser) opens the input HTML document (web page) to upload the input parameters, both numeric and graphical (images). After entering the required input values in the input HTML document, the client request is forwarded to the *matweb* component via the HTTP service. The *matweb* component is a TCP/IP (Transmission Control Protocol/Internet Protocol) client that communicates with the

Matlab server. It uses the Common Gateway Interface (CGI) to extract data from the input HTML document and transfer them to the Matlab server.

The Matlab server is a multithreaded TCP/IP server that manages communication between the client application and Matlab. It is configured to listen the requests on the TCP/IP ports that are defined in the *matlabserver.conf* file. The Matlab server loads the requested m-file into Matlab. After execution has been completed the Matlab server gives the received output data to the *matweb* component. After that, the *matweb* sends them to the client in an output HTML document through HTTP service. During the execution of m-file, images or any other data files can be created and sent to the output HTML document. The web browser will download them from the HTTP service.

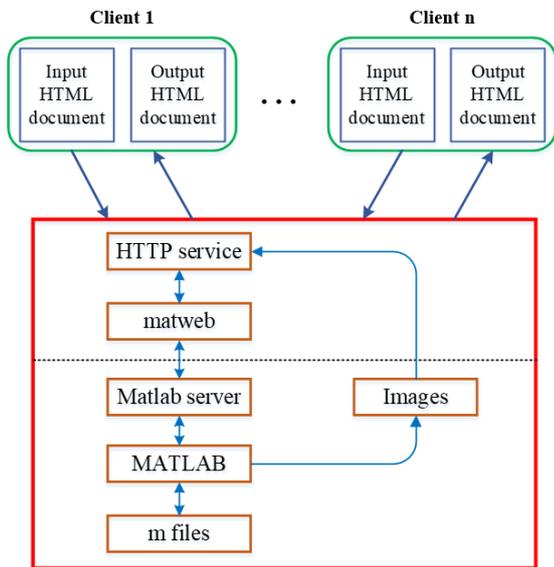


Figure 2. Communication between clients and Matlab over the internet.

IV. RESULTS

The user interface of the MWS application consists of two basic HTML documents: one for uploading mammograms and entering values of the input parameters and the other for displaying results obtained after executing the Matlab code. Every internet user can use the MWS application on their computer with the appropriate web browser.

The user interface for uploading mammograms and entering values of the input parameters is shown in the Fig. 3. Button “Browse” gives users the ability to choose a mammogram to be processed. After a successful mammogram upload and its display that confirms that the mammogram is correctly uploaded, it is necessary to specify values of the four input parameters: contrast-correction value, binary image enhancement value, the threshold for pectoral muscle removal, and length of the pectoral muscle. The parameter binary image enhancement value can be selected from a dropdown menu. The user can enter different values of the input parameters until he finds the optimum values that provide a satisfactory result. By clicking the “Submit” button, the uploaded image

and the entered values are sent to the application for processing. The “Reset” button allows the user to clear all entered values, including the selected mammogram.

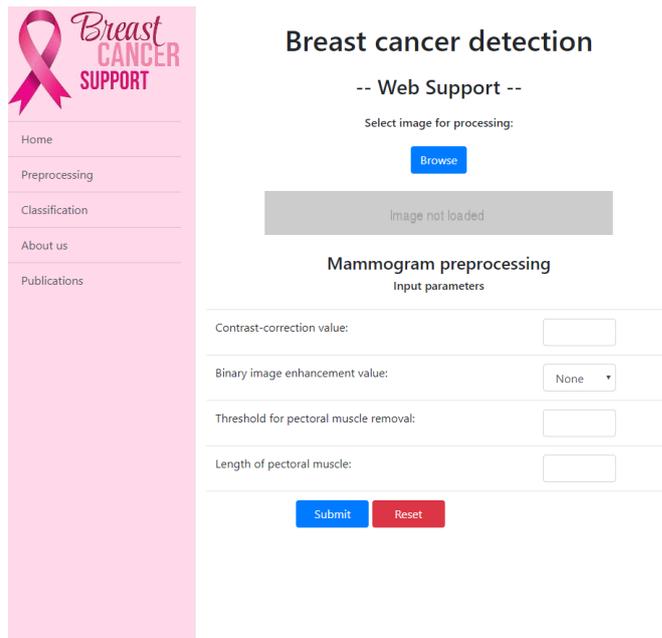


Figure 3. Web page for uploading mammograms and entering values of input parameters.

After processing the input data, a new web page is opened showing the detected region of interest - a mammogram with removed background and pectoral muscle, and a notification about estimations of all three classifiers (Fig. 4). Each classifier estimates whether the mammogram is positive (there's a sign of cancer) or negative (there's no sign of cancer).

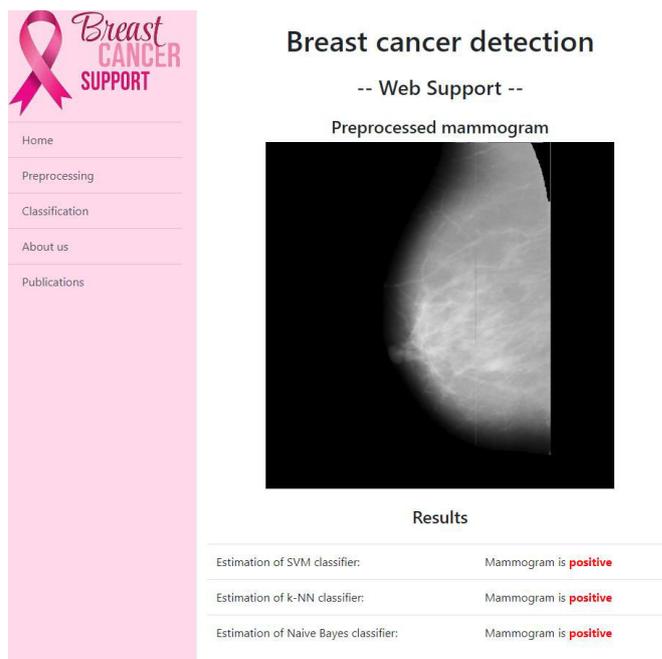


Figure 4. Web page for displaying mammogram processing results.

On the left side of the HTML documents shown in the Fig. 3 and Fig. 4 are links that provide information for easier use of the application. This information is especially useful when the user first encounters an application. A description of the mammogram preprocessing procedure is provided on the web page referenced by the “Preprocessing” link. Information about image de-noising, region of interest extraction, and enhancement of image contrast is displayed on this web page. The method for mammogram ROI detection, composed of several steps, is described in detail. To set optimal values of the input parameters, users must read this informative text. A link “Classification” leads the users to information about texture feature extraction and classification of patterns. Basic information about the authors of the application is given on the link “About us”. The authors of the presented application are authors of several published scientific papers whose topic is the mammogram classification. A list of mentioned scientific papers is provided on the link “Publications”. Papers [21], [22] and [23] are part of this list.

V. CONCLUSION

The paper describes a mammogram processing system that is available through the internet using Matlab Web Services. The user interface consists of web pages for data entry and display of results. Therefore, in addition to a web browser, users do not need additional software to use this web-based application.

Users of the MWS application for breast cancer detection can upload their own mammogram and enter different values of the input parameters until they find the optimum values that provide a properly determined region of interest. After processing the input data, users get a notification about estimations of three different classifiers. The notification contains information on whether the ROI extracted from their mammogram is positive or negative.

The most important advantage of MWS is that users can use applications written in Matlab without installing Matlab software, whose license price is quite high. Besides, users don't need to have any special knowledge about programming in Matlab to run applications. To execute the MWS application, it is enough to have a suitable web browser and be familiar with the purpose of the Matlab application, to set the correct values of the input parameters. The system can be used on almost every main web browsers available today. We have tested the system on Mozilla Firefox, Google Chrome, Microsoft Edge, Opera, Vivaldi, and Safari web browsers.

MWS applications are available through the standard HTTP protocol. This allows users to connect with Matlab through any application that implements this protocol, such as Java applets.

The advantages of the proposed CAD system for breast cancer detection in mammographic images are the good classification results and very short processing time (less than 10 seconds in our testing cases, from image upload to obtaining results). The accuracy of the detection algorithm is good but not satisfactory. Therefore, it is not possible to rely

on their diagnosis. In other words, the result of a detection algorithm should be used only as help in deciding, confirming the diagnosis, or to remark on the oversight.

In addition to numerous advantages, the presented MWS application has some disadvantages. One of the disadvantages is the fixed structure of the Matlab application. In some cases, the ability to change the application structure would be useful.

Future work will focus on the further improvement of classification performance by reducing the number of input texture features and using the other classification methods, such as neural networks. In addition, the focus will be on the discrimination between benign and malignant masses. Also, while the current system is served on local computers and is not publicly offered, the idea is to publish the system online. Of course, this step has to be done carefully, and after many tests of the system's stability.

ACKNOWLEDGMENT

This study was supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia, and these results are parts of Grant No. 451-03-47/2023-01/200132 with the University of Kragujevac - Faculty of Technical Sciences Čačak.

REFERENCES

- [1] D. Klimes, M. Kubasek, R. Smid and L. Dusek, “Internet-based system for anti-tumor chemotherapy evaluation,” *Computer Methods and Programs in Biomedicine*, vol. 93, no. 3, pp. 292–296, 2009.
- [2] D. N. Kennedy et al., “WebParc: a tool for analysis of the topography and volume of stroke from MRI,” *Medical & Biological Engineering & Computing*, vol. 48, no. 3, pp. 215–228, 2010.
- [3] H. Shen et al., “MIAPS: a web-based system for remotely accessing and presenting medical images,” *Computer Methods and Programs in Biomedicine*, vol. 113, no. 1, pp. 266–283, 2014.
- [4] S. E. Mahmoudi et al., “Web-based interactive 2D/3D medical image processing and visualization software,” *Computer Methods and Programs in Biomedicine*, vol. 98, no. 2, pp. 172–182, 2010.
- [5] J. Jiménez et al., “A Web platform for the interactive visualization and analysis of the 3D fractal dimension of MRI data,” *Journal of Biomedical Informatics*, vol. 51, pp. 176–190, 2014.
- [6] J. Zhang, J. Sun and J. N. Stahl, “PACS and Web-based image distribution and display,” *Computerized Medical Imaging and Graphics*, vol. 27, no. 2–3, pp. 197–206, 2003.
- [7] J. Millan and L. Yunda, “An Open-Access Web-based Medical Image Atlas for Collaborative Medical Image Sharing, Processing, Web Semantic Searching and Analysis with Uses in Medical Training, Research and Second Opinion of Cases,” *Nova*, vol.12, no.22, pp. 143–150, 2014.
- [8] W. Cai, D. Feng and R. Fulton, “Web-based digital medical images,” *IEEE Computer Graphics and Applications*, vol. 21, no. 1, pp. 44 - 47, 2001.
- [9] A. Mirarab, N. G. Fard and M. Shamsi, “A cloud solution for medical image processing,” *International Journal of Engineering Research and Applications*, vol. 4, no. 7, pp. 74–82, 2014.
- [10] R. Yuan, M. Luo, Z. Sun, S. Shi, P. Xiao and Q. Xie, “RayPlus: a Web-Based Platform for Medical Image Processing,” *Journal of Digital Imaging*, vol. 30, no. 2, pp. 197–203, 2017.
- [11] H. Hwang and R. A. Haddad, “Adaptive median filters: New algorithms and results,” *IEEE Transactions on Image Processing*, vol. 4, no. 4, pp. 499–502, 1995.

- [12] M. Milosevic, D. Jankovic and A. Peulic, "Comparative Analysis of Breast Cancer Detection in Mammograms and Thermograms," *Biomedical Engineering/ Biomedizinische Technik*, vol. 60, no. 1, pp. 49-56, 2015, DOI: 10.1515/bmt-2014-0047.
- [13] J. Suckling et al., "The mammographic images analysis society digital mammogram database," *Exerpta Medica*, vol. 1069, pp. 375-378, 1994.
- [14] R. M. Haralick, K. Shanmugam and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 3, no. 6, pp. 610-621, 1973.
- [15] L. K. Soh and C. Tsatsoulis, "Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices," *IEEE Transactions on geoscience and remote sensing*, vol. 37, no. 2, pp. 780-795, 1999.
- [16] D. A. Clausi, "An analysis of co-occurrence texture statistics as a function of grey level quantization," *Canadian Journal of Remote Sensing*, vol. 28, no. 1, pp. 45-62, 2002.
- [17] B. E. Boser, I. M. Guyon and V. N. Vapnik, "A training algorithm for optimal margin classifiers," *Proceedings of the Fifth Annual ACM Workshop on Computational Learning Theory*, pp. 144-152, 1992.
- [18] N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," *The American Statistician*, vol. 46, no. 3, pp. 175-185, 1992, DOI: 10.1080/00031305.1992.10475879.
- [19] I. Rish, "An Empirical Study of the Naive Bayes Classifier," *Proceedings of the IJCAI-2001 Workshop on Empirical Methods in Artificial Intelligence*, vol. 3, no. 22, pp. 41-46, 2001.
- [20] Modelit Webserver Toolbox for Matlab, available at: <https://www.modelit.nl/index.php/webserver-toolbox-for-matlab>
- [21] M. Radovic, M. Milosevic, S. Ninkovic, N. Filipovic, A. Peulic, "Parameter optimization of a computer-aided diagnosis system for detection of masses on digitized mammograms," *Technology and Health Care*, vol. 23, no. 6, pp. 757-774, 2015, DOI: 10.3233/THC-151034.
- [22] M. Radovic, M. Djokovic, A. Peulic, N. Filipovic, "Application of Data Mining Algorithms for Detection of Masses on Digitalized Mammograms," *Proceedings of the 5th International Conference on Information Society and Technology (ICIST 2015)*, Kopaonik, Serbia, March 8-11, pp. 13-18, 2015.
- [23] M. Radović, M. Đoković, A. Peulić, N. Filipović, "Application of data mining algorithms for mammogram classification," *Proceedings of the 13th International IEEE Conference on Bioinformatics and Bioengineering (BIBE)*, Chania, November 10-13, pp. 1-4, 2013.