Applied Medical Informatics
Research Letter
Vol. 45, No. 1/2023, pp: 6-17

Classification of Breast Cancer Tumors using a Random Forest on Mammogram Images

Saiful BUKHORI*, Suci Dwi MAYSAROHA and Januar ADI

Department of Computer Science, University of Jember, Jalan Kalimantan No. 37 Jember, East Java, 68121, Indonesia
E-mail(*): saiful.ilkom@unej.ac.id

* Author to whom correspondence should be addressed;

Received: February 20, 2023 /Accepted: March 28, 2023/ Published online: March 30, 2023

Abstract
Mammography can detect lumps in early-stage breast cancer when the tumors are small and cannot be felt when touched. Mammography is an X-ray of breast tissue and the diagnosis has limitations because it has low contrast and noise. We proposed a computer vision method to classify mammogram images to reduce the visual limitations of images and doctor's subjectivity. The classification process is carried out by recognizing and processing mammographic images and analyzing them using the random forest method to obtain appropriate knowledge. Before being classified, the image is enhanced by changing the format, rotated, cropped, enlarged the contrast using contrast stretching, and removed noise using a Gaussian filter. The enhanced image was extracted using the Gray Level Co-occurrence Matrix (GLCM) method. Classification of breast cancer based on mammogram images with the random forest algorithm gave an accuracy of 70.8%, a precision of 85.7%, and a recall equal to 25%.

Keywords: Mammography; Breast Cancer; Computer Vision; Random Forest; Gaussian Filter

Introduction

Breast cancer forms in the cells of breast tissue and grows very fast and is difficult to control [1]. The types of breast cancer are determined by the location of the cells in the breast that turns into cancer [2]. Breast cancer can start from three main parts of the breast: lobules (glands that produce milk), ducts (tubes that carry milk to the nipples), and connective tissue (the fibrous and fatty tissue that surrounds and holds the parts together as a whole) [3]. Most breast cancers start in the ducts or lobules [4]. The rapid spread of breast cancer causes it to be the second leading cause of death from cancer in women in the worldwide after lung cancer [5].

The magnitude of death by cancer is because it was found at an advanced stage. More than 80% of breast cancer is found at an advanced stage [6]. The advanced stage is the severity of cancer that occurs when cancer cells have spread to other organs making it difficult to treat. The more advanced the stage, the lower the survival rate. Only 22% of stage IV breast cancer patients survive over 5 years [7]. Based on the cases of death caused by these cancers, it is necessary to identify breast cancer early. Early identification can be made by screening the patient's breasts to classify them as normal or abnormal. There are several ways to check for breast cancer, including Breast Self-Exams (BSEs), Clinical Breast Exams (CBE), mammography, breast ultrasound, breast Magnetic Resonance Imaging (MRI), and biopsy [8].

Breast Self-Exam is a screening done to determine whether there is a lump or physical oddity in the breast as a symptom of breast cancer. It is recommended to have a BSE once a month so that changes can be identified. However, this screening is certainly not enough to decide whether or not cancer is present, a clinical breast cancer test must be carried out. Before examining with the help of medical devices, the doctor will examine the breasts with bare hands, Clinical Breast Exams (CBE). The CBE aims to determine the breast's shape, size, color, and texture to detect the possibility of cancer [9]. The doctor or nurse will examine the breast in a circular motion to detect the location of a lump or tumor in the breast. Apart from around the breasts, the doctor will also examine the lymph nodes in the armpits and the top of the collarbone. If there is swelling or a lump, the doctor will carry out a more specific follow-up examination.

Mammography is an examination to confirm the diagnosis of breast cancer, both in symptomatic and asymptomatic women [10]. Mammography can detect lumps at an early stage of breast cancer that are still small or cannot be felt when touched. Mammography is an X-ray photo of breast tissue [11]. However, the results of this mammography image have limitations because it has low contrast and high noise [12]. Breast ultrasound is a breast cancer test using sound waves to display images on a computer screen. This screening can detect changes in the breast, such as lumps or tissue changes. In addition, breast ultrasound can also distinguish whether the lump contains fluid, which means a breast cyst or a solid lump. Breast MRI is a breast cancer test using magnets and radio waves [13]. The combination of magnets and radio waves will produce images of all parts of the breast and show soft tissue very clearly. An MRI examination is a follow-up examination when a person is diagnosed with breast cancer to know the size of cancer and look for possible tumors on the breast. A breast biopsy is done when a physical exam, mammography, or other breast cancer imaging test shows cancer cells to be present. This test procedure is carried out by taking a tissue sample suspected of cancer cells in it. This tissue sample is then examined in the laboratory with a microscope.

This research proposes a computer vision method to classify mammogram images to reduce images' visual limitations and doctors' subjectivity. Several previous research has examined this computer vision method for detecting breast cancer [14, 15, 16], some of them used machine learning [17, 18, 19] while other used deep learning [20, 21, 22]. However, this research has not yielded optimal results because image enhancement and noise removal have not been carried out. The aim of this research was to classify the breast cancer based on mammogram image using the random forest method applied to enhanced images by changing the format, rotated, cropped, increased contrast using contrast stretching, and removing noise using a Gaussian filter. Features are extracted from the transformed image using the GLCM and used as input for classification using a random forest. To get the best classification model, testing was carried out at the stages of contrast stretching, gaussian filter and classification with random forest.

Material and Method

This research used mammogram data obtained from the Mammographic Image Analysis Society (MIAS) dataset [23], one of the oldest data sets compiled by British researchers. The dataset was downloaded for free from the Cambridge University repositor on January 2, 2023. The dataset consists of 322 mammogram data with details of 207 normal breast images and 115 abnormal breast images. The mammogram view is taken using the Medio Lateral Oblique (MLO) or side view. Mammograms in this dataset have four image sizes: small (1600×4320 pixels), medium (2048×4320 pixels), large (2600×4320 pixels), and extra-large (5200×4320 pixels). Mammograms are as*.pgm files with a gray level resolution of 8 bits per pixel. The model proposed in the research classification of breast cancer based on mammogram image using random forest is shown in Figure 1. The experiment was carried out with a computer hardware processor 2.2 Ghz, Intel Core i7 6-core 2.2 GHz, Turbo Boost up to 4.1 GHz, with shared L3 cache of 9 MB, 2400 MHz memory storage media of 16 GB and SSD 512 GB. The software used is jupyter notebook 6.5.2, python 3.

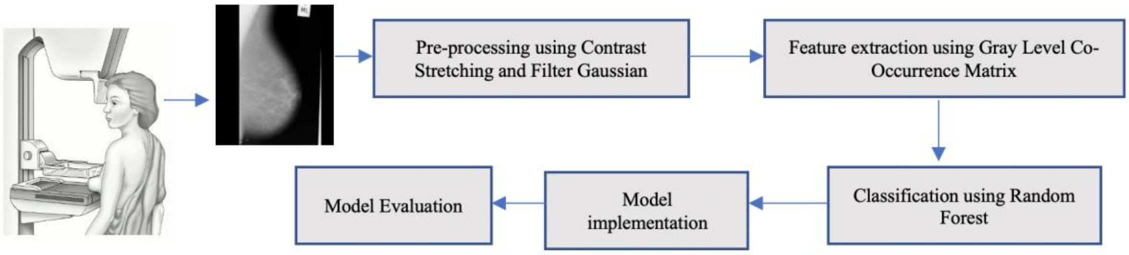

Figure 1. The proposed model

At the preprocessing stage, image quality improvement is carried out to produce images that are ready for analysis. A special application that supports the extension (.pgm) is required to open images in the dataset. To make it easy to implement in a classification system, the dataset format which originally had the extension (.pgm) was converted to a mammogram image in (.png) format. This format supports the process of data compression without losing the information in it. After converting the image format, the image is rotated and cropped manually to get clear images and specific part of the object under study. Image rotation is done so that side-view images can be seen more clearly. Meanwhile, cropping is done to remove useless parts of the scanned image and focus on the breasts. Cropping is also done so that the image size becomes uniform with a ratio of 7:5. Screened and cropped mammogram images have increased contrast values using contrast stretching (see Figure 2).

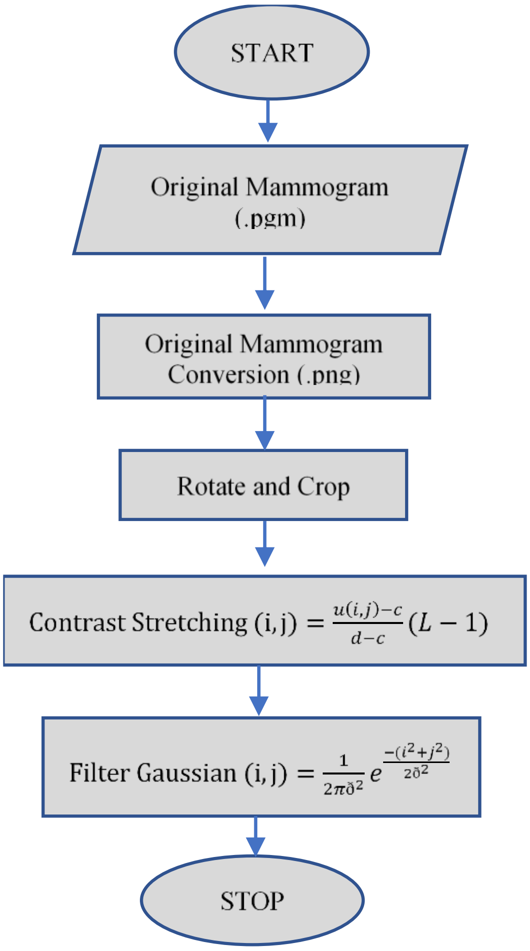

Figure 2. Preprocessing flow

The implementation of the contrast stretching method is used to increase the image contrast by flattening the spread of pixel values so that the image has better contrast. To determine the level of image contrast, minimum and maximum percentile values range from 0 - 100, 1 - 99, 2 - 98, 3 - 97, 4 - 96, 5 - 95, and 6 – 94 are used. After the contrast is increased, a mammogram usually has a little noise. To reduce noise, we uses Gaussian filters with various sigma values, namely 1, 2, 3, and 4, to make mammogram images smoother.

Mammograms with improved image quality at the preprocessing stage were then extracted using the Gray Level Co-occurrence Matrix (GLCM) method [24] (Figure 3). The GLCM matrix calculation uses 4 angles (0, 45, 90, and 135 degrees) and spatial distance d = 1. The GLCM matrix that has been formed is used as a reference in calculating five textures, namely energy, contrast, correlation, homogeneity, and entropy. Based on the angles and textures used, the total feature values generated from mammogram images are 20 features.

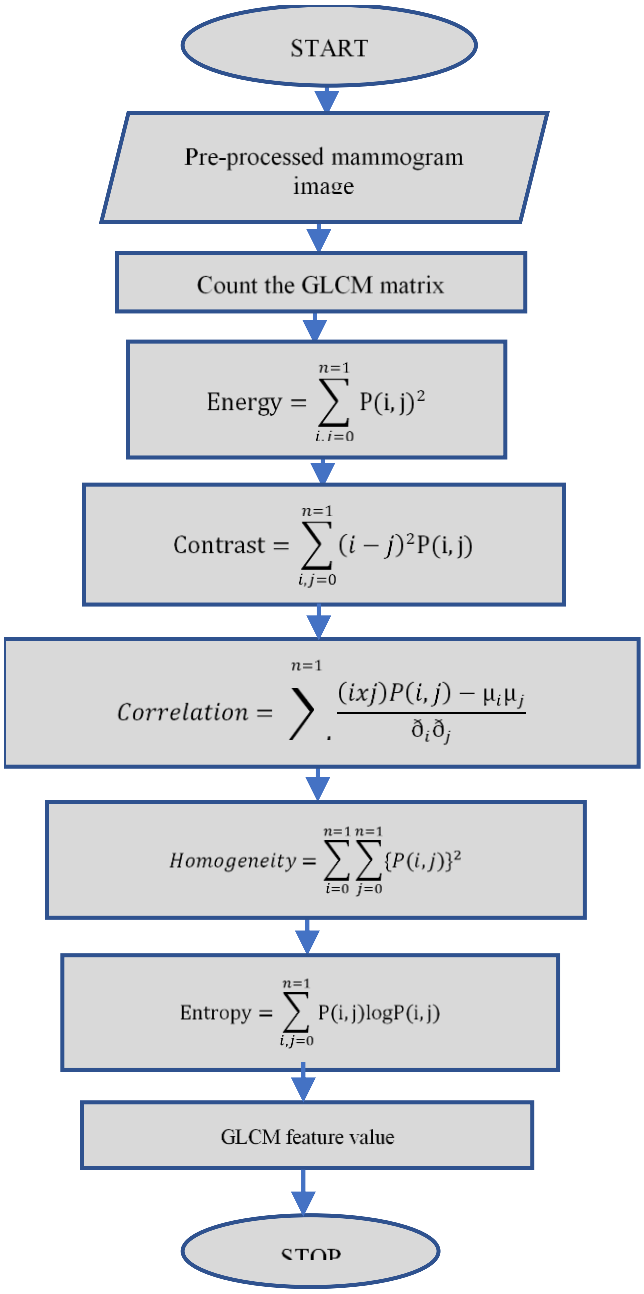

Figure 3. Extracted using the GLCM

The features that have been extracted are then classified using a random forest (Figure 4). The GLCM feature values dataset is divided into two parts randomly regardless of the presence or absence of disease: training and testing sets. The ratio between training data and test data used varies with a ratio of 50% : 50%, 60% : 40%, 70% : 30%, and 80% : 20%. This ratio results in the performance of different models. In addition to ratio data, the number of trees used also varies, starting with 50 trees, 100 trees, 200 trees, 300 trees, 400 trees, 500 trees, 600 trees, 700 trees, 800 trees, 900 trees, and 1000 trees. The more training data and trees used do not necessarily provide better performance. Computation costs will increase as training data and the number of trees used are increased.

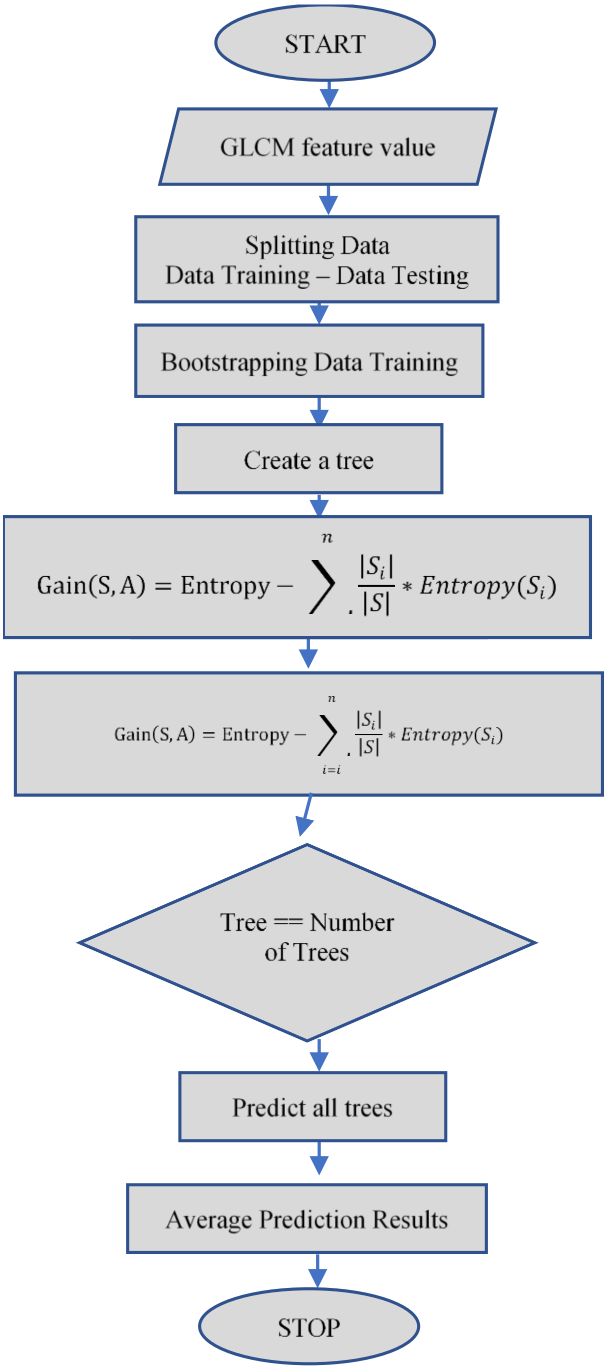

Figure 4. Classification uses a random forest

Bootstrapping is then carried out by forming new training data based on a simple random sampling of existing training data with returns. The amount of new training data generated is in accordance with the number of trees used. The new training data set will be formed into a tree by determining the nodes from the root to the terminal (p) using entropy (S) calculations and information acquisition. The Information Gain (S,A) value is affected by the attributes as nodes in the tree formed. The formation of the tree is carried out on each new training data that is generated. After all the trees are formed, then each tree makes predictions using test data. The final decision in the form of classification is determined by the prediction with the highest value in the test data. The end result of the classification is an indication of normal or abnormal breast cancer. The entropy and information gain values formula uses are presented in Eq (1) and (2).

$$Entropy\ (S) = \sum_{i=1}^{n} -p_i log_2 p_i \tag{1}$$

$$Gain(S,A) = Entropy(S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} * Entropy(S_i) \tag{2}$$

The final result of classification is an indication of normal or abnormal breast cancer. The design is implemented into a model. After that, an evaluation of the model is carried out. Model evaluation aims to measure the performance of the random forest classification algorithm by comparing the predicted results with the actual classification results. Evaluation of the model uses the confusion matrix with the performance measures used in precision, recall, and accuracy, as shown in Figure 5. The calculation of the performance matrix is carried out in each experiment so that a comparison of training data and testing data and the optimal number of trees can be found.

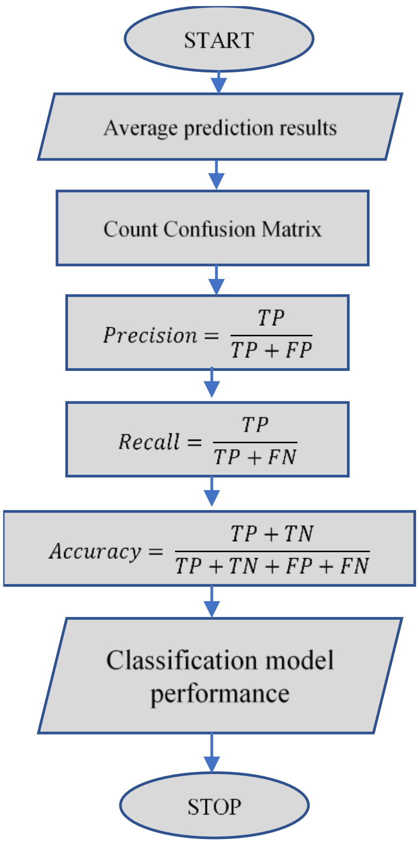

Figure 5. Evaluation of the model (TP = true positive, TN = true negative, FP = false positive, FN = false negative)

To find the best Random Forest model, the researchers conducted several experiments. The first experiment focused on the contrast stretching stage, while the second experiment focused on the gaussian filter stage [25,26]. In this research, the first experiment used varying minimum and maximum percentile values, while the second experiment used varying sigma values.

Results and Discussion

Structure Percentile Values

The first experiment was to change the percentile value at the contrast stretching stage. This experiment aims to find the best contrast enhancement in mammogram images. The percentile values tested were 0-100, 1-99, 2-98, 3-97, 4-96, 5-95, and 6-94. Variation in the value of this percentile was a feature extraction to test the increase in image contrast directly without going through the Gaussian filter stage. The percentile experiment will produce 7 different datasets. Each dataset will be classified and evaluated to determine the best model. Figure 6 presents the results of the percentile experiment.

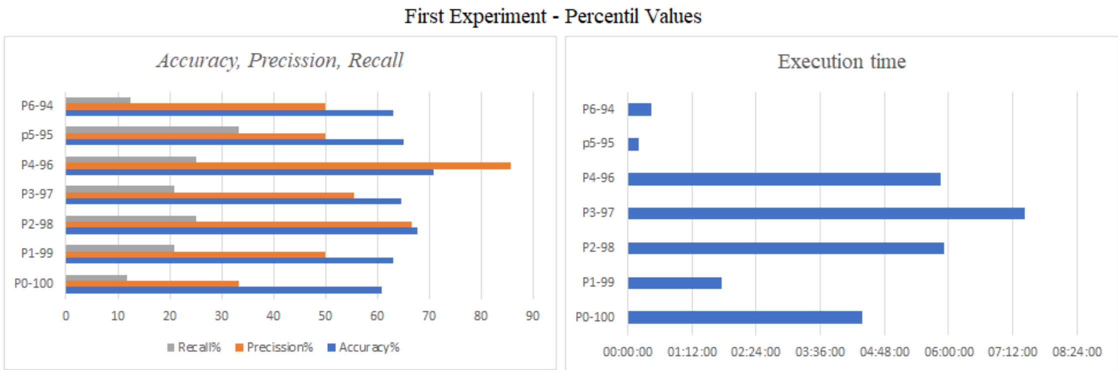

Figure 6. Graph of percentile value experiment results

The resulting accuracy, precision, and recall values are 70.8%, 85.7%, and 25%, with an execution time of 5 hours 51 minutes 6 seconds. The best model uses a 80:20 training data : testing data, and 600 trees. Table 1 shows the model evaluation results at the 4-96th percentile.

Sigma Values

The second experiment was carried out at the stage of removing noise with a gaussian filter. The noise removal power depends on the sigma value used. The higher the sigma, the more noise will be removed and the image will be blurry. The sigma experiment was carried out using sigma = 1, sigma = 2, sigma = 3, and sigma = 4. The image with the noise removed is the image from contrast stretching with the best percentile, namely 4-96. After feature extraction, each dataset will be classified and evaluated to determine the best model. Figure 7 presents the results of the sigma experiment.

Based on the experimental graph, it is known that the sigma value that produces the best classification model is sigma = 3. The resulting accuracy, precision, and recall values are 65.9%, 53.8%, and 15.6% with an execution time of 10 minutes 28 seconds. The best model uses a comparison of training data: testing data 60 : 40 and the number of trees is 100. The following Table 2 shows the details of the evaluation results of the model at sigma = 3.

Table 1. Evaluation results of the 4-96 percentile value model

Training Data : Testing Data	Number of Trees	Accuracy%	Precision%	Recall%	Execution Time
50:50	50	57.8	31.8	15.8	7m3s
	100	60.2	37	17.5	12m34s
	200	59	33.3	15.8	25m
	300	57.8	31	15.8	34m57s
	400	60.2	37.9	19.3	50m9s
	500	59	33.3	15.8	58m9s
	600	59.6	35.7	17.5	1h8m37s
	700	58.4	32.1	15.8	1h28m7s
	800	59	34.5	17.5	1h25m38s
	900	59.6	35.7	17.5	1h44m41s
	1000	59.6	35.7	17.5	1h55m4s
40:60	50	58.9	37.5	26.7	9m5s
	100	66.7	54.2	28.9	21m43s
	200	65.1	50	31.1.	44m29s
	300	62.8	44.4	26.7	59m45s
	400	62.8	44.4	26.7	1h26m59s
	500	62	43.3	28.9	1h40m17s
	600	61.2	40.7	24.4	1h55m54s
	700	62.8	44.4	26.7	2h12m17s
	800	64.3	48	26.7	2h42m30s
	900	63.6	45.8	24.4	2h57m41s
	1000	64.3	48	26.7	3h13m20s
30:70	50	66	52.6	29.4	21m15s
	100	63.9	47.1	23.5	34m10s
	200	57.7	31.6	17.6	1h3m2s
	300	61.9	42.1	23.5	1h30m5s
	400	58.8	33.3	17.6	2h28m2s
	500	62.9	44.4	23.5	3h5m34s
	600	64.9	50	29.4	3h47m15s
	700	62.9	45	26.5	4h5m12s
	800	59.8	35.3	17.6	4h25m26s
	900	62.9	44.4	23.5	4h54m40s
	1000	62.9	43.8	20.6	5h25m37s
20:80	50	67.7	63.6	29.2	26m41s
	100	67.7	71.4	20.8	1h39m27s
	200	67.7	66.7	25	2h43m12s
	300	69.2	70	29.2	3h15m0s
	400	69.2	75	25	3h49m53s
	500	69.2	75	25	4h2m8s
	600	70.8	85.7	25	5h51m6s
	700	66.2	60	25	5h52m5s
	800	66.2	60	25	8h6m51s
	900	66.2	60	25	7h56m7s
	1000	67.7	63.6	29.2	8h33m58s

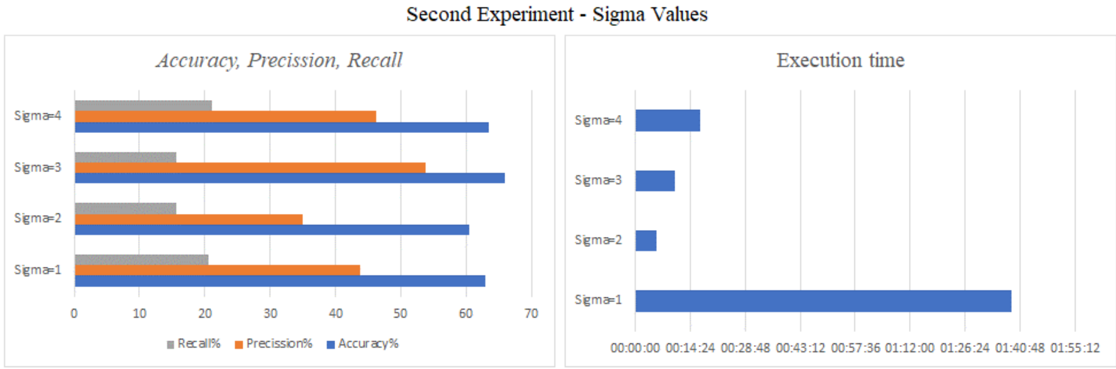

Figure 7. Graph of sigma value experiment results

After conducting the first experiment using various percentile values and the second experiment using various sigma values, it can be seen that the first experiment resulted in a better classification model than the second experiment. The first experiment resulted in the best accuracy, precision, and recall values of 70.8%, 85.7%, and 25% with an execution time of 5 hours 51 minutes 6 seconds, while the second experiment resulted in the best accuracy, precision, and recall values of 65.9%, 53.8%, and 15.6% with an execution time of 10m28s. The best classification model is obtained by doing contrast stretching on the mammogram image using the minimum-maximum percentile of 4-96, then extracting the features using the Gray Level Co-occurrence Matrix, and classifying it in the random forest algorithm using a comparison of training data: 80 : 20 testing data and the number of trees 600. The gaussian filter method to remove noise from the mammogram image was not used because based on the second experiment, this filter could not improve the classification results. The use of the higher sigma value or standard deviation on the gaussian filter will affect the number of areas identified so that there are differences in texture in the mammogram image.

The testing using various Percentile Values to find the best contrast value in the mammogram image followed by Sigma Values testing to remove noise with a Gaussian filter will find a random forest model with the most optimal accuracy. Image enhancement which consists of the process of finding the best contrast value and removing the best noise is a purely subjective processing technique [25], therefore, experiments are needed for each object and for each specific purpose in order to obtain the most optimal model. This model needs to be developed for other objects and other purposes by increasing the number of experiments using the same process.

Conclusions

The classification of breast cancer based on mammogram images with the random forest algorithm gave the accuracy, precision, and recall results, namely 70.8%, 85.7%, and 25%. These results were obtained after the mammogram image was contrasted using contrast stretching with the minimum and maximum percentile values used were 4 and 96. The corrected mammogram image extracted its texture features using the Gray Level Co-occurrence Matrix method. The selection of methods and variables used can affect the random forest algorithm classification results. Based on the experimental results, it is known that the number of trees and the optimal comparison of training-testing data for classifying mammogram images with the random forest algorithm are 600 and 80:20 with an execution time of 5 hours 51 minutes 6 seconds.The use of the Gaussian filter method to remove noise from mammogram images is also known not to be able to improve the classification results of the random forest algorithm.

Table 2. Results of model evaluation with sigma value = 3

Training Data : Testing Data	Number of Trees	Accuracy%	Precision%	Recall%	Execution Time
50:50	50	59	23.5	7	3m10s
	100	60.2	36	15.8	7m45s
	200	61.5	35.3	10.5	12m42s
	300	61.5	36.8	12.3	20m4s
	400	62.7	42.9	15.8	26m56s
	500	62.7	42.1	14	31m50s
	600	60.9	36.4	14	40m26s
	700	62.7	42.1	14	44m28s
	800	62.7	42.9	15.8	49m10s
	900	62.7	41.2	12.3	56m25s
	1000	62.1	37.5	10.5	1h4m19s
40:60	50	62.8	41.2	15.6	5m14s
	100	65.9	53.8	15.6	10m28s
	200	62.8	38.5	11.1	19m1s
	300	61.2	30.8	8.9	30m12s
	400	63.6	41.7	11.1	42m10s
	500	62.8	30.4	8.9	55m16s
	600	63.6	37.5	6.7	1h1m6s
	700	64.3	44.4	8.9	1h14m28s
	800	64.3	44.4	8.9	1h23m14s
	900	64.3	45.5	11.1	1h36m53s
	1000	62.8	36.4	8.9	1h43m35s
30:70	50	58.8	25	8.8	8m52s
	100	63.9	44.4	11.8	19m25s
	200	62.9	37.5	8.8	34m59s
	300	62.9	37.5	8.8	51m30s
	400	62.9	37.5	8.8	1h12m50s
	500	61.9	33.3	8.8	1h25m42s
	600	62.9	37.5	8.8	1h47m10s
	700	62.9	37.5	8.8	2h1m43s
	800	62.9	37.5	8.8	2h23m54s
	900	62.9	37.5	8.8	2h25m41s
	1000	61.9	33.3	8.8	2h59m58s
20:80	50	56.9	16.7	4.2	10m59s
	100	60	33.3	8.3	26m58s
	200	61.5	42.9	12.5	52m52s
	300	63.1	50	8.3	1h15m44s
	400	63.1	50	8.3	1h46m8s
	500	63.1	50	8.3	2h4m22s
	600	61.5	40	8.3	2h43m6s
	700	62.8	36.4	8.9	1h18m19s
	800	64.3	44.4	8.9	1h20m
	900	61.5	40	8.3	3h34m55s
	1000	63.1	50	8.3	4h20m44s

References

1. Feng Y, Spezia M, Huang S, Yuan C, Zeng Z, Zhang L, et al. Breast cancer development and progression: Risk factors, cancer stem cells, signaling pathways, genomics, and molecular pathogenesis. Genes Dis. 2018;5(2):77-106. doi: 10.1016/j.gendis.2018.05.001
2. Beňačka R, Szabóová D, Guľašová Z, Hertelyová Z, Radoňák J. Classic and New Markers in

Diagnostics and Classification of Breast Cancer. Cancers. 2022;14(21):5444. doi:10.3390/cancers14215444
3. Christgen M, Cserni G, Floris G, Marchio C, Djerroudi L, Kreipe H, et al. Lobular Breast Cancer: Histomorphology and Different Concepts of a Special Spectrum of Tumors. Cancers (Basel). 2021;13(15):3695. doi: 10.3390/cancers13153695.
4. Biswas SK, Banerjee S, Baker GW, Kuo CY, Chowdhury I. The Mammary Gland: Basic Structure and Molecular Signaling during Development. Int J Mol Sci. 2022;23(7):3883. doi: 10.3390/ijms23073883.
5. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. CA Cancer J Clin. 2022;72(1):7-33. doi: 10.3322/caac.21708.
6. Nelson DR, Brown J, Morikawa A, Method M. Breast cancer-specific mortality in early breast cancer as defined by high-risk clinical and pathologic characteristics. PLoS One. 2022;17(2):e0264637. doi: 10.1371/journal.pone.0264637.
7. Pascual MC, Montaño JJ, Franch P, Sánchez-Contador C, Ramos M. Survival of Breast Cancer by Stage, Grade and Molecular Groups in Mallorca, Spain. J Clin Med. 2022;11(19):5708. doi: 10.3390/jcm11195708.
8. Luo C, Wang L, Zhang Y, Lu M, Lu B, Cai J, et al. Advances in breast cancer screening modalities and status of global screening programs. Chronic Dis Transl Med. 2022;8(2):112-123. doi: 10.1002/cdt3.21.
9. Sarker R, Islam MS, Moonajilin MS, Rahman M, Gesesew HA, Ward PR. Knowledge of breast cancer and breast self-examination practices and its barriers among university female students in Bangladesh: Findings from a cross-sectional study. PLoS One. 2022;17(6):e0270417. doi: 10.1371/journal.pone.0270417.
10. Barlow WE, Lehman CD, Zheng Y, Ballard-Barbash R, Yankaskas BC, Cutter GR, et al. Performance of diagnostic mammography for women with signs or symptoms of breast cancer. J Natl Cancer Inst. 2002;94(15):1151-9. doi: 10.1093/jnci/94.15.1151.
11. Lin C-H, Lai H-Y, Chen P-Y, Wu J-X, Pai C-C, Su C-M, et al. Breast Lesions Screening of Mammographic Images with 2D Spatial and 1D Convolutional Neural Network-Based Classifier. Applied Sciences. 2022;12(15):7516.
12. Almalki YE, Soomro TA, Irfan M, Alduraibi SK, Ali A. Computerized Analysis of Mammogram Images for Early Detection of Breast Cancer. Healthcare (Basel). 2022;10(5):801. doi: 10.3390/healthcare10050801.
13. Mann RM, Cho N, Moy L. Breast MRI: State of the Art. Radiology. 2019;292(3):520-536. doi: 10.1148/radiol.2019182947.
14. Sadoughi F, Kazemy Z, Hamedan F, Owji L, Rahmanikatigari M, Azadboni TT. Artificial intelligence methods for the diagnosis of breast cancer by image processing: A review. Breast Cancer Targets Ther. 2018;10:219-30. doi: 10.2147/BCTT.S175311.
15. Basurto-Hurtado JA, Cruz-Albarran IA, Toledano-Ayala M, Ibarra-Manzano MA, Morales-Hernandez LA, Perez-Ramirez CA. Diagnostic Strategies for Breast Cancer Detection: From Image Generation to Classification Strategies Using Artificial Intelligence Algorithms. Cancers (Basel). 2022;14(14):3442. doi: 10.3390/cancers14143442.
16. Hassan NM, Hamad S, Mahar K. Mammogram breast cancer CAD systems for mass detection and classification: a review. Multimedia Tools and Applications 2022;81(14): 20043-75. doi: 10.1007/s11042-022-12332-1.
17. Naji MA, El Filali S, Aarika K, Benlahmar EH, Abdelouhahid RA, Debauche O. Machine Learning Algorithms for Breast Cancer Prediction and Diagnosis. Procedia Comput. Sci. 2021;191:487-92. doi: 10.1016/j.procs.2021.07.062.
18. Singh S, Prasad J, Prasad S, Naik S. Breast Cancer Prediction using Supervised Machine Learning Techniques. Proc. 2021 IEEE 18th India Counc. Int. Conf. INDICON 2021;4:348-5. doi: 10.1109/INDICON52576.2021.9691609.
19. Islam MM, Haque MR, Iqbal H, Hasan MM, Hasan M, Kabir MN. Breast Cancer Prediction: A Comparative Study Using Machine Learning Techniques. SN Comput Sci. 2020;1(5):1–14. doi: 10.1007/s42979-020-00305-w.
20. Mahoro E, Akhloufi MA. Applying Deep Learning for Breast Cancer Detection in Radiology.

Curr Oncol. 2022 Nov 16;29(11):8767-8793. doi: 10.3390/curroncol29110690.
21. Abunasser BS, Al-Hiealy MRJ, Zaqout IS, Abu-Naser SS. Breast Cancer Detection and Classification using Deep Learning Xception Algorithm. Int J Adv Comput Sci Appl. 2022;13(7):223-8. doi: 10.14569/IJACSA.2022.0130729.
22. Ahn SH, Kim E, Kim C, Cheon W, Kim M, Lee SB, et al. Deep learning method for prediction of patient-specific dose distribution in breast cancer. Radiat Oncol. 2021 Aug 17;16(1):154. doi: 10.1186/s13014-021-01864-9.
23. Suckling J, Parker J, Dance D, Astley S, Hutt I, Boggis C, et al. Mammographic Image Analysis Society (MIAS) database v1.21 [Dataset]. 2015. Available from: https://www.repository.cam.ac.uk/handle/1810/250394.
24. Simanjuntak TI, Suwilo S, Sembiring RW. Analysis of Detection of Drow and Entire Co-Occurrence Matrix GLCM Method on the Classification of Image. J Phys Conf Ser. 2019;1361(1):012028. doi: 10.1088/1742-6596/1361/1/012028.
25. Negi SS, Bhandari YS. A hybrid approach to Image Enhancement using Contrast Stretching on Image Sharpening and the analysis of various cases arising using histogram. International Conference on Recent Advances and Innovations in Engineering, ICRAIE 2014, 2014, pp. 1–6. doi: 10.1109/ICRAIE.2014.6909232.
26. Shaheed K, Szczuko P, Abbas Q, Hussain A, Albathan M. Computer-Aided Diagnosis of COVID-19 from Chest X-ray Images Using Hybrid-Features and Random Forest Classifier. Healthcare. 2023;11(6):837. doi:10.3390/healthcare11060837.