BI-RADS CATEGORIES AND BREAST LESIONS CLASSIFICATION OF MAMMOGRAPHIC IMAGES USING ARTIFICIAL INTELLIGENCE DIAGNOSTIC MODELS

F. Turk[*], *E. Akkur*[†], *O. Eroğul*[‡]

Abstract: According to BI-RADS criteria, radiologists evaluate mammography images, and breast lesions are classified as malignant or benign. In this retrospective study, an evaluation was made on 264 mammogram images of 139 patients. First, data augmentation was applied, and then the total number of images was increased to 565. Two computer-aided models were then designed to classify breast lesions and BI-RADS categories. The first of these models is the support vector machine (SVM) based model, and the second is the convolutional neural network (CNN) based model. The SVM-based model could classify BI-RADS categories and malignant-benign discrimination with an accuracy rate of 86.42% and 92.59%, respectively. On the other hand, the CNN-based model showed 79.01% and 83.95% accuracy for BI-RADS categories and malignant benign discrimination, respectively. These results showed that a well-designed machine learning-based classification model can give better results than a deep learning model. Additionally, it can be used as a secondary system for radiologists to differentiate breast lesions and BI-RADS lesion categories.

Key words: *breast cancer, mammography, BI-RADS, convolutional neural network, support vector machines*

Received: April 4, 2023 **DOI:** 10.14311/NNW.2023.33.023
Revised and accepted: December 8, 2023

1. Introduction

Nowadays, breast cancer (BC) poses a severe threat to the health of the female population [1]. Early diagnosis of breast cancer is an essential step in increasing survival rates [2]. Mammography is the most recommended imaging tool for breast screening that uses low-dose X-ray for early diagnosis [3, 5]. Radiologists

[*]Fuat Turk – Corresponding author; Department of Computer Engineering, Kırıkkale University, Kırıkkale, Turkey, E-mail: fturk@kku.edu.tr

[†]Erkan Akkur; Turkish Medicines and Medical Devices Agency, Ankara, Turkey

[‡]Osman Eroğul; Department of Biomedical Engineering, TOBB ETU University of Economics & Technology, Ankara, Turkey

examine the mammogram images to identify breast lesions as malignant or benign. The Breast Imaging Reporting and Data Systems (BI-RADS) is a comprehensive guide radiologists use to identify breast lesions. The BI-RADS terminology provides standardization of breast imaging, report organization, assessment categories, and classification of breast lesions. Based on the level of suspicion, breast lesions can be categorized into six BI-RADS categories: incomplete (category 0), negative (category 1), benign (category 2); probably benign (category 3), suspicious (category 4), high probability of malignancy (category 5); and proved cancer (category 6) [4]. A radiologist usually analyzes many mammogram images manually, which is repetitive and prone to human error. Therefore, the process of deciding BI-RADS categories and classifying breast lesions can be a challenging task for radiologists. This decoding process can also lead to misinterpreted mammograms and unnecessary biopsy rates. Computer-aided diagnostic (CAD) models can reduce these difficulties and assist radiologists [5]. CAD systems play an important role in improving BC's diagnostic performance. The main goal of such systems is to minimize the interpretation error by diminishing the number of false positives that lead to unnecessary biopsies. CAD systems mainly consist of machine learning (ML) and deep learning (DL) approaches. ML approaches depend on handcrafted feature extraction, feature selection, and classification processes. Conversely, DL approaches extract the discriminating features automatically by a learning process from the dataset during the training [6]. Some CAD models focus on classifying breast lesions, while others are designed to organize BI-RADS categories [7–15]. However, as a result of the examination of the studies on the subject in the literature, it has been observed that very few studies automatically classify both breast lesion and BI-RADS categories. In this study, two different CAD models based on ML and DL that automatically classify breast lesions and BI-RADS categories (2, 3, 4 and 5) are proposed.

Several CAD systems have been designed to classify BI-RADS categories and breast lesions as malignant or benign. Chokri and Farida [7] presented an ML-based CAD model for classifying BI-RADS and mammographic breast lesions. Twenty-three handcrafted features and multiplayer perception (MLP) are used. Their approach showed 88.02% and 83.85% accuracy for breast lesions and BI-RADS categories. Boumaraf et al. [8] introduced a machine-learning CAD model for classifying BI-RADS categories (2, 3, 4, 5). The experiments were tested on the DDSM database. They extracted 130 handcrafted features. The modified genetic algorithm was used for the feature selection process A backpropagation neural network (BPNN) is utilized for the classification process. Their suggested model achieved 84.5% accuracy, 84.4% positive predictive value, and 94.8% negative predictive value. Tsai et al. [9] used a deep neural network (DNN) model to classify BI-RADS categories. A public dataset was utilized for experiments. The suggested model achieved 94.22% accuracy. Domingues et al. [10] presented a CNN model for classifying BI-RADS categories. The experimental results showed 83.4% accuracy. Punitha et al. [11] used an optimized region-growing technique to segment malignant and benign breast masses. Gray level co-occurrence matrix (GLCM) and gray level run matrix texture features were used. The Digital Database for Screening Mammography (DDSM) database was preferred. Feed-forward neural network (FFNN) was used as the classification algorithm. The proposed model

achieved 98.1% sensitivity and 97.8% specificity values. Ting et al. [12] suggested a CNN-based model for breast lesion classifications. The presented model reached 90.5% of accuracy. Ketabi et al. [13] proposed a CAD model for differentiating breast lesions. Histogram and GLCM methods were used for the feature extraction (FE) process. The genetic algorithm was utilized to select the optimal feature subset. Finally, they used the support vector machine approach for the classification process. Their classification results were 90% accuracy. Li et al. [14] suggested an ML-based approach to predict mammographic microcalcifications' pathological status. A data set of 260 patients was used. 837 textures and nine geometric features were extracted from the mammographic microcalcifications. The recursive feature elimination technique was used for the feature selection process. After the feature selection process, ten features were selected. Logistic regression (LR), support vector machine (SVM), k-nearest neighbor (k-NN) and naive Bayes (NB) algorithms were utilized for the classification process. The SVM algorithm achieved 80% accuracy compared to other algorithms. Stelzer et al. [15] suggested a model using texture features and ML algorithms to detect calcifications in mammography. A data set of 226 patients was used in the study. Two hundred forty-nine image features are derived from the grey-value histogram, grey-level co-occurrence and run-length matrices. The principal component analysis (PCA) was used to select the most discriminative features. The multiplayer perceptron (MLP) was chosen as the classification algorithm. The model proposed in the study demonstrated the potential to reduce 37.1–45.7% of unnecessary biopsies for one false negative per reader. Although there are separate studies on the classification of breast lesions and BI-RADS categories in the literature, there are very few studies that automatically classify both breast lesion and their BI-RADS categories together.

The study is organized as follows: Some existing studies on the subjects are summarized in Section 1. The materials and methodology are explained in Section 2. Experimental results are described in Section 3. Section 4 summarizes the discussion and conclusion of the study.

2. Materials and methodology

This study proposes two CAD models to classify BI-RADS categories and breast lesions. The architecture of the proposed CAD models is presented in Fig. 1.

2.1 Study population

This study used a mammography dataset retrieved from Ankara Training and Research Hospital, Department of Radiology. The Institutional Ethics Committee of Ankara Training and Research Hospital approved this retrospective dataset. All patients who underwent digital mammography between April 2015 and April 2020 were retrieved from the picture archiving and communication systems (PACS). Patient consent was obtained on the condition that all data were anonymized. The inclusion criteria are (1) patients who had suspected breast lesions and accepted digital mammography and (2) patients who were confirmed with benign and malign breast lesions by histopathologic examinations or the ones who were confirmed with benign lesions as a result of two years of radiological periodic follow-up. Exclusion

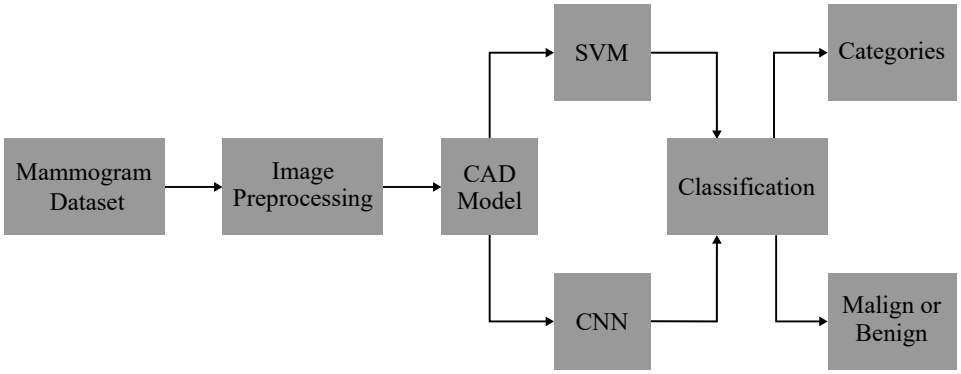

Fig. 1 *The framework of the suggested CAD models.*

criteria were: 1) any treatment history for breast cancer before mammography imaging and 2) poor quality of mammographic images. 3) breast lesions are not fully displayed in craniocaudal (CC) and/or mediolateral oblique (MLO) views.

The data set consists of 264 mammogram images of 139 (mean age: 58.26 ±11.96) patients in total. Data augmentation is applied to increase the number of mammogram images to avoid overfitting [16]. During the data augmentation process, the images were not shifted (no matter vertically or horizontally). Because it was thought that the shifting process would affect the tumour tissue positionally and cause errors. Only 60° angle rotation and zoom operations are performed. After the data augmentation process, the dataset consists of 565 images. Tab. I shows the number of benign, malignant, and BI-RADS category samples before and after the data augmentation.

Samples	Before augmentation	After augmentation
Benign	103	256
Malign	161	309
Total	264	565
2	30	80
3	21	73
4	76	215
5	137	197
Total	264	565

Tab. I *The number of breast lesions and BI-RADS categories.*

The dataset is randomly divided into training, validation, and test data, illustrated in Tab. II.

Samples	Training	Validation	Testing
Benign	194	30	32
Malign	212	48	49
Total	406	78	81
2	57	11	12
3	53	10	10
4	149	32	34
5	147	25	25
Total	406	78	81

Tab. II *The training, validation, and testing samples.*

2.2 Image preprocessing

Mammogram images contain many artefacts, pectoral texture, and inference noise that affect the clarity of the images and cause false positive rates. Image preprocessing is a crucial step to improve the performance of CAD models [17]. First, unwanted areas are cut off, and the image is cropped (this step is performed with the img.crop parameter). During the image preprocessing phase, image.enhancement functions in the Python OpenCV library were used. Brightness and contrast parameters have been improved by 1.5, respectively. Then, while the size of the original mammogram images was 3584×2816, unnecessary areas (black areas on the image that did not contain tissue samples) were removed. The image was then cropped by setting the resolution to 224×224. Thus, faster training aimed to reduce the workload in the deep learning phase. Some sample images of the datasets before and after image processing are shown in Fig. 2.

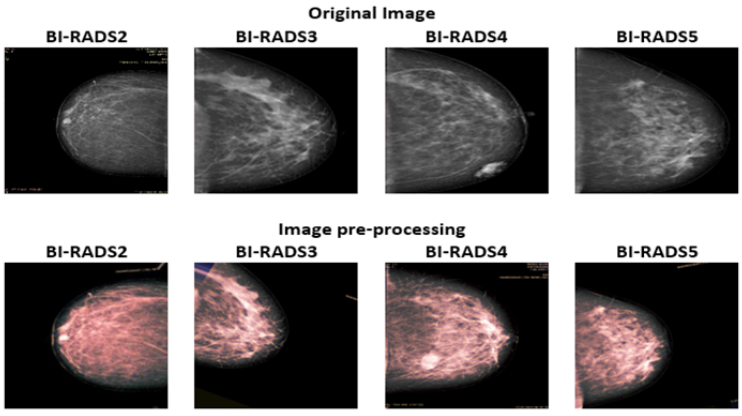

Fig. 2 *The framework of the suggested CAD models.*

The original images are presented in gray color on the official website. We took them as they are. Then, we applied image processing steps such as sharpness, blurriness and brightness with ready-made functions in the OPENCV library in Python. Finally, we colored the gray image with the "cvtColor" function in the OPENCV library so that the texture and lesion differences on the image could be seen better.

2.3 SVM CAD model

The flowchart diagram of the SVM-based CAD model, which is the first model to classify BI-RADS categories and breast lesions, is shown in Fig. 3. The model starts with the segmentation process applied to determine the region of interest (ROI). The first step of the segmentation process is cropping the region where the breast lesion is located and removing unnecessary areas. Depends on mammography images and their corresponding ROI integrated with three categories. A total of 127 features, including 16 geometric and texture features, which are 15 histograms, 52 gray level co-occurrence matrix (GLCM) and 44 gray level run matrix (GLRM), were extracted for each ROI [18–20]. GLCM and GLRLM features were computed for 0°, 45°, 90° and 135° directions for each ROI [16]. The extracted features are stated in Tab. III.

The next stage is the feature selection process. The least absolute shrinkage and selection operator (LASSO) regression is used to select the most discriminant features in the training dataset. 10-fold cross-validation and λ (standard error of the minimum mean-square error criteria) are preferred to choose optimal features for LASSO. Then, corresponding λ values are computed. The features with zero coefficient are eliminated, and the remaining values of λ (non-zero coefficient) are determined useful [21,22]. After the feature selection process, the ML classification model starts. As an ML, the SVM algorithm is used for regression and classification

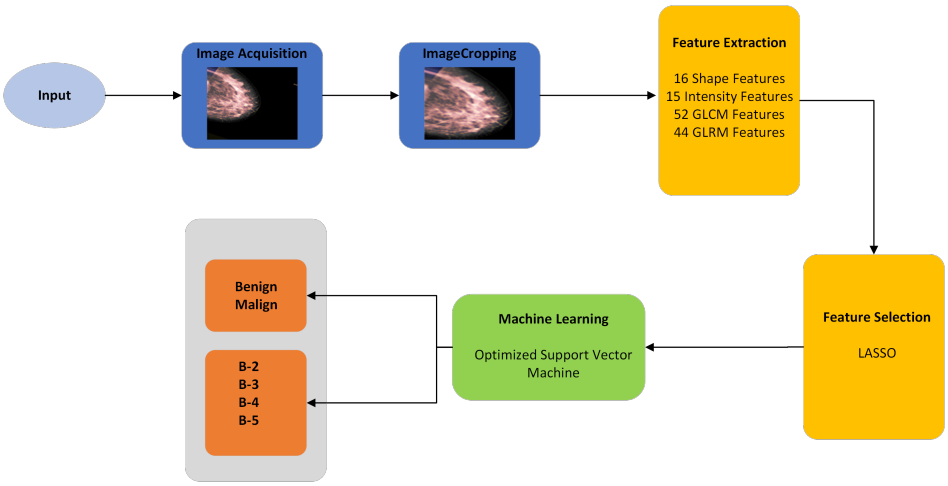

Fig. 3 *The suggested machine learning-based CAD model architecture.*

Methods	Features	Dimensions
Geometric	Area, perimeter, max. radius, min. radius, euler number, eccentricity, solidity, entropy, equivalent diameter, elongatedness, circulation 1, circulation 2, compactness, dispersion, thinness ratio, shape index	16
Histogram	Mean, standard deviation, variance, smoothness, mean absolute deviation, skewness, kurtosis, minimum, maximum, 10th percentile, 90th percentile, interquartile range, range, root mean square, median	15
GLCM	Contrast, correlation, energy, entropy, homogeneity, sum of square and mean, sum of variance and entropy, difference variance and entropy, information measure of correlation 1 and 2	52
	13 values were obtained for each corresponding to the four directions (0°, 45°, 90°, 135°)	
GLRM	Short run and long run emphasis, gray-level and run length nonuniformity, run percentage, short run low and short run high gray-level run emphasis, long run low and long run high gray-level run emphasis, low gray-level and high gray-level run emphasis	44
	11 values were obtained for each corresponding to the four directions (0°, 45°, 90°, 135°)	
Total		127

Tab. III *The extracted features.*

processes. The algorithm classifies the data by finding the best hyperplane that separates one class from the other classes [23]. The SVM algorithm includes hyperparameters such as kernel function (KF), box constraint level (BCL), and kernel scale (KS). Hyperparameter optimization is used for the algorithm to achieve the best performance for the data. Bayesian optimization method is used in this study. Bayesian optimization automatically sets the hyperparameters of SVM. This algorithm is a practical approach for parameter search and is a black-box optimization technique. The algorithm builds a probabilistic model by selecting a prior probability distribution over the optimized function. Then, it combines with sample information to obtain a posterior function. The algorithm's work depends on an iterative Gaussian process [24]. The suggested optimized SVM algorithm is implemented in MATLAB using the Statistical and Machine Learning Toolbox [25]. The setting of hyperparameters for SVM is shown in Tab. IV.

Algorithm	Hyperparameters	Search range
SVM	KF	Gaussian, linear, quadratic, cubic
	KS	[0.001–1000]
	BCL	[0.001–1000]

Tab. IV *The setting of hyperparameters of SVM.*

2.4 CNN-based CAD model

The second suggested CNN-based CAD model is a convolutional neural network that classifies BI-RADS categories and breast lesions. CNN is one of the popular deep neural networks used in classifying images, object detection and image recognition. It comprises three layers: convolution, pooling, and fully connected. FE has been performed in the convolutional layer by applying suitable filters. A pooling layer is used to reduce the dimensionality of feature spaces. A pooling layer is used to reduce the dimensionality of feature spaces. The higher-level features are extracted in fully connected (FC) layers with a corresponding weight. These achievable features are processed to classify according to output categories corresponding to the original input [26]. Fig. 4 illustrates the architecture of the suggested CNN model. The input image is given to the models as 224×224 and goes through 4 convolutional layers. Convolution is performed with 5×5 filters in the first two layers and 3×3 filters in the next layers. From this stage onwards,

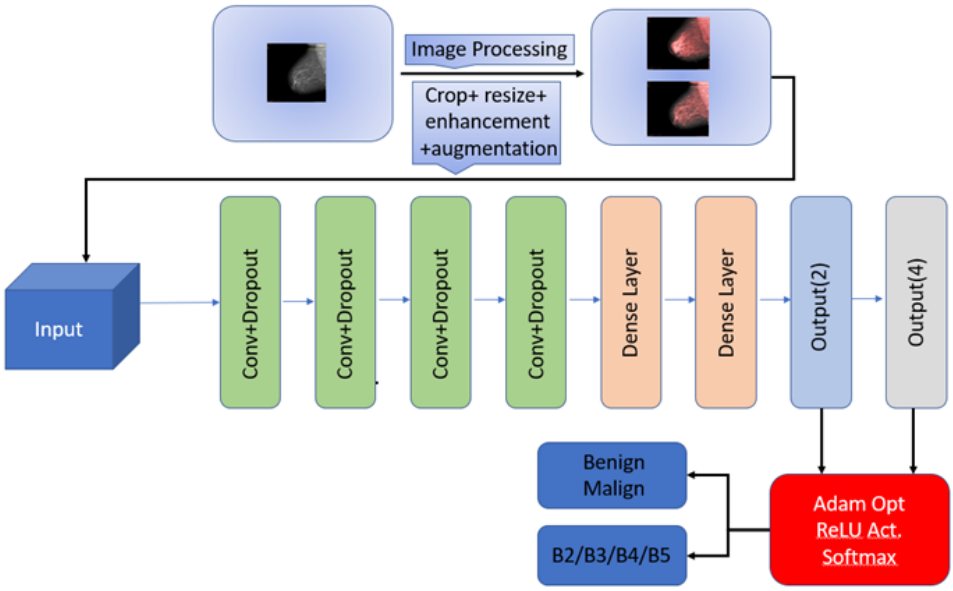

Fig. 4 *The suggested deep learning-based CAD model architecture – CNN-based model structures for BI-RADS and breast lesion classification.*

max-pooling and dropout layers are used while moving from each layer to the next. After the four convolutional layers, two dense layers were added. The soft-max function is used for classification. An activation function is utilized to reduce the feature dimension, and ReLU was used as an activation function. Adam Optimizer was used for optimization.

Tab. V shows the layer structures, output sizes, and parameter numbers for breast lesions and BI-RADS categories classifications. Output shape can have 2 or 4 classes (benign-malignant, or multi-class). The param expression shows the mathematical values of the row, column and depth expressions of the matrices created according to the size of the images through the CNN model. Since the input size of the first image is not important on the model, the input size can be taken as "none". Both architectural structures start with a 2D convolutional layer and continue with a 2×2 filter, max pooling and 0.1 dropout layer, respectively. This process was carried out with four block layers (conv2D, max pooling, dropout). Then, the architecture is finalized by flattening with a flattened layer (conversion to a one-dimensional matrix) and returning the desired output with a dense layer. The last dropout layer applies a 0.2% simplification. It should be noted that only the number of classes in the previous drizzle layer of the architectures is different. This situation has been planned in this way to facilitate the application.

Layer type	Output shape (2/4 class)	Number of parameters
Conv2D_1	(None, 224, 224, 32)	896
Maxpooling_1	(None, 112, 112, 32)	0
Dropout_1	(None, 112, 112, 32)	0
Conv2D_2	(None, 112, 112, 64)	18496
Maxpooling_2	(None, 56, 56, 64)	0
Dropout_2	(None, 56, 56, 64)	0
Conv2D_3	(None, 56, 56, 128)	73856
Maxpooling_3	(None, 18, 18, 128)	0
Dropout_3	(None, 18, 18, 128)	0
Conv2D_4	(None, 18, 18, 256)	295168
Maxpooling_4	(None, 6, 6, 256)	0
Dropout_4	(None, 6, 6, 256)	0
Flatten_1	(None, 9216)	0
Dense_1	(None, 64)	589888
Dropout_5	(None, 64)	0
Dense_2	(None, 2)	130
Dense_4	(None, 4)	130

Tab. V *The layer structures, output sizes, and parameter numbers for breast lesions and BI-RADS categories classifications.*

3. Experimental results

The suggested CAD models are examined according to the two different scenarios using different evaluation measures. The presentation and analysis of experimental results are shown in this section. Four evaluation metrics—accuracy, precision, recall and F-score were utilized.

3.1 First scenario — BI-RADS classification

The BI-RADS categories are classified as SVM and CNN-based models in the first scheme. The classification of results is shown in this section. The results are shown in terms of the test dataset.

3.1.1 SVM model results

Tab. VI shows the selection features after using the LASSO feature selection process. After the feature selection process, 15 discriminant features (6 geometric, 3 histograms, 3 GLCM and 2 GLRM) remained for the classification of BI-RADS categories.

Feature	Feature type
Eccentricity	Geometric
Entropy	Geometric
Compactness	Geometric
Dispersion	Geometric
Thinness ratio	Geometric
Shape index	Geometric
Mean	Histogram
Std. deviation	Histogram
Skewness	Histogram
Sum of square (0° degree)	GLCM
Sum of square (45° degree)	GLCM
Difference variance (0° degree)	GLCM
Run length nonuniformity (0° degree)	GLRM
Run percentage (0° degree)	GLRM
Run percentage (90° degree)	GLRM

Tab. VI *The most discriminant features for BI-RADS classification.*

The confusion matrix and the performance matrix of the SVM algorithm for BI-RADS classification are presented in Tab. VII and Fig. 5. 70 of 81 breast lesions are correctly classified in their BI-RADS classes. The overall accuracy is 86.42%, and the classification results for each BI-RADS category (2, 3, 4, 5) are 95.06%, 91.36%, 91.36% and 95.06%.

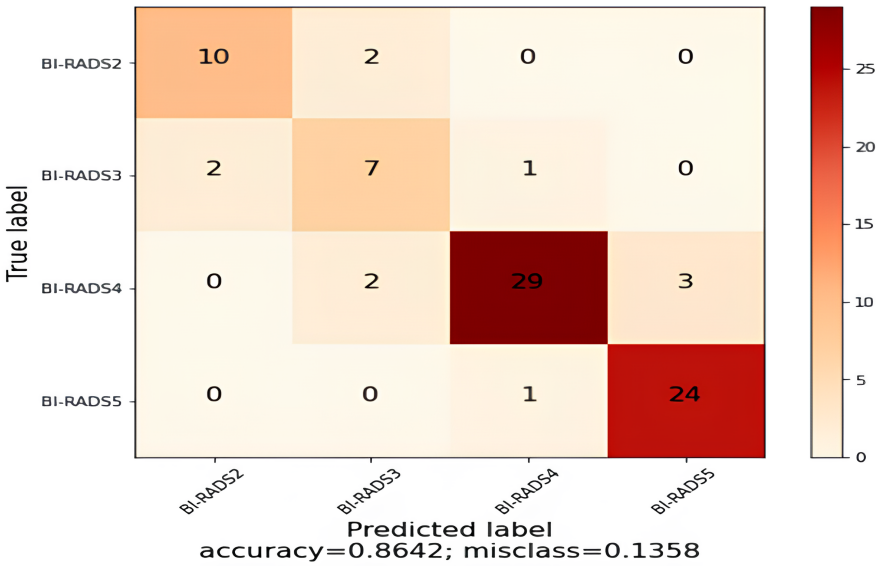

Fig. 5 *The confusion matrix of the SVM algorithm.*

Class	Accuracy [%]	Precision [%]	Recall [%]	F1-score [%]
BI-RADS-2	95.06	83.30	83.30	83.30
BI-RADS-3	91.36	70.00	63.60	66.70
BI-RADS-4	91.36	85.30	93.50	89.20
BI-RADS-5	95.06	96.00	88.90	92.30
Overall acc.	86.42			

Tab. VII *Performance metrics of the SVM algorithm.*

3.1.2 CNN-based model results

The confusion matrix and the performance matrix of the CNN algorithm for BI-RADS classification are presented in Tab. VIII, and Fig. 6. 64 of 81 breast lesions are correctly classified in their BI-RADS classes. The overall accuracy is 79.01%, and the classification accuracy results for each BI-RADS category (2, 3, 4, 5) are 88.9%, 91.36%, 84% and 93.08%.

3.2 Second scenario — breast lesion classification

For breast lesion classification as malignant or benign, SVM and CNN-based models are performed, respectively. The results are demonstrated in terms of the testing dataset.

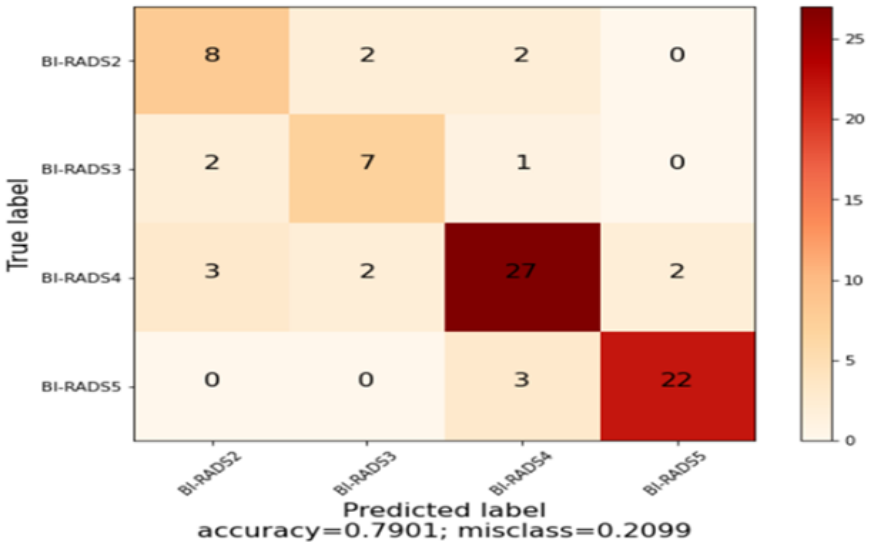

Fig. 6 *The confusion matrix of the CNN algorithm.*

Class	Accuracy [%]	Precision [%]	Recall [%]	F1-score [%]
BI-RADS-2	88.90	66.70	61.50	64.00
BI-RADS-3	91.36	70.00	63.60	66.70
BI-RADS-4	84.00	79.40	81.80	80.60
BI-RADS-5	93.80	88.00	91.70	89.80
Overall acc.	79.01			

Tab. VIII *The confusion matrix and the performance matrix of the CNN algorithm.*

3.2.1 SVM model results

Tab. IX shows the most discriminant features after using the LASSO feature selection process. After the feature selection process, 11 discriminant features (2 geometric, 4 histograms, 2 GLCM and 3 GLRM) remained for breast lesion classification.

The confusion matrix and the performance matrix of the SVM algorithm for breast lesion classification are presented in Tab. X and Fig. 7. 75 of 81 breast lesions are classified correctly, and 3 misclassifications are performed for each malignant and benign lesion. The SVM algorithm showed an overall accuracy rate of 92.59%.

3.2.2 CNN-based model results

The confusion matrix and the performance matrix of the CNN algorithm for breast lesion classification are presented in Tab. XI and Fig. 8. 68 of 81 breast lesions are

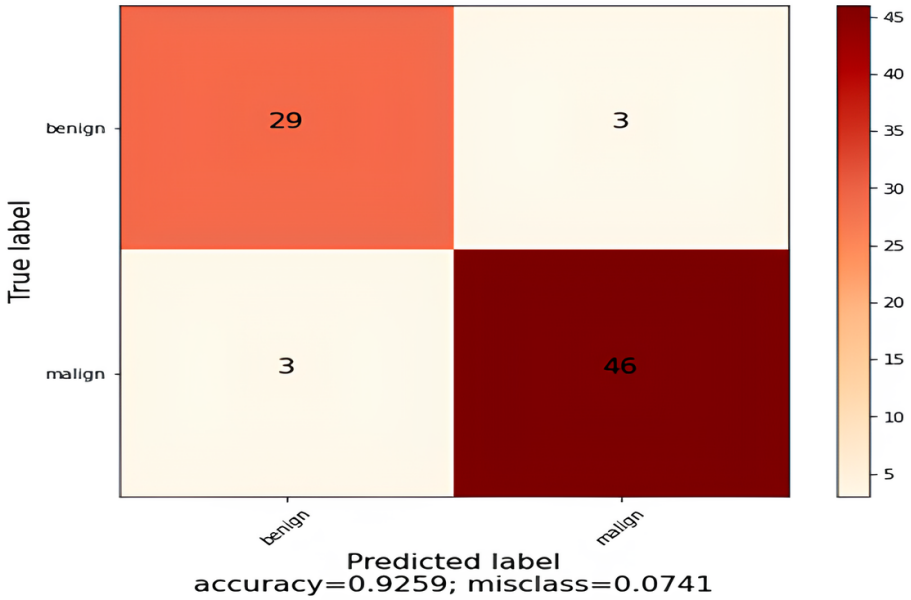

Fig. 7 *The confusion matrix of the SVM algorithm.*

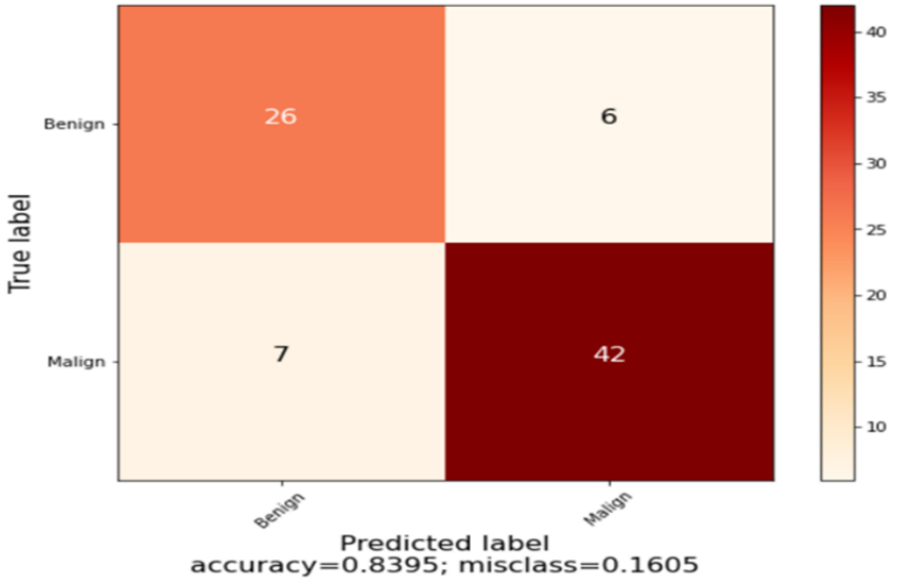

Fig. 8 *The confusion matrix of the CNN algorithm.*

Feature	Feature type
Compactness	Geometric
Thinness ratio	Geometric
Mean	Histogram
Std. Deviation	Histogram
Skewness	Histogram
Kurtosis	Histogram
Sum of square (0° degree)	GLCM
Difference entropy(0° degree)	GLCM
Run length nonuniformity (0° degree)	GLRM
Run percentage (0° degree)	GLRM
Run percentage (90° degree)	GLRM

Tab. IX *The most discriminant features for breast lesion classification.*

Class	Accuracy [%]	Precision [%]	Recall [%]	F1-score [%]
Benign	92.59	90.60	90.60	90.60
Malign	92.59	93.90	93.90	93.90
Overall acc.	92.59			

Tab. X *The performance matrix of the SVM algorithm.*

Class	Accuracy [%]	Precision [%]	Recall [%]	F1-score [%]
Benign	83.95	78.80	81.30	80.00
Malign	83.95	87.50	85.70	86.60
Overall acc.	83.95			

Tab. XI *The performance matrix of the CNN algorithm.*

classified correctly. 6 benign cases and 7 malignant cases were incorrectly classified. The overall accuracy obtained using the CNN model is 83.95%.

3.3 Comparison of the results of SVM and CNN-based CAD models for BI-RADS categories and breast lesions

Fig. 9 presents the classification accuracy rates for each BI-RADS category obtained using the SVM and CNN models. Except for the BI-RADS 3 category, the SVM model gives better results than the CNN model for all other categories (2, 4, 5). SVM and CNN models show the same accuracy value for the BI-RADS 3 category.

The overall accuracy rates for all BI-RADS categories using CNN and SVM models are shown in Fig. 10. With an 86.42% accuracy rate, the SVM model offers a better classification rate than the CNN model with 79.01% accuracy.

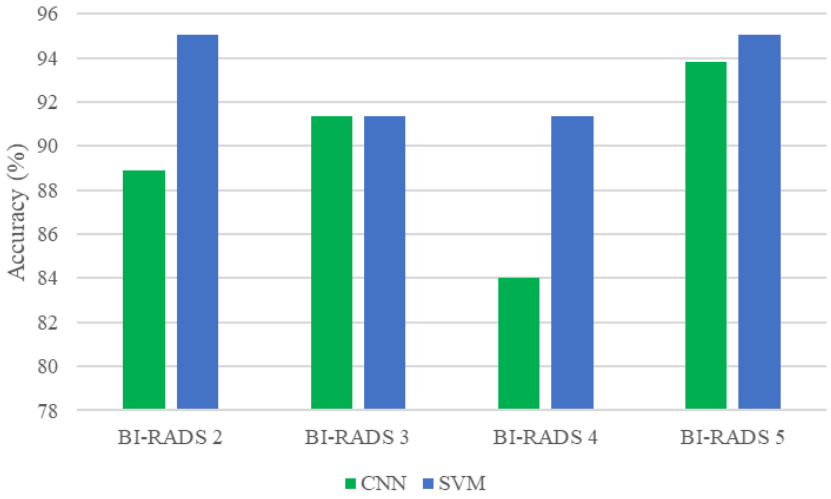

Fig. 9 *The classification accuracy rates for each BI-RADS category were obtained using the SVM and CNN models.*

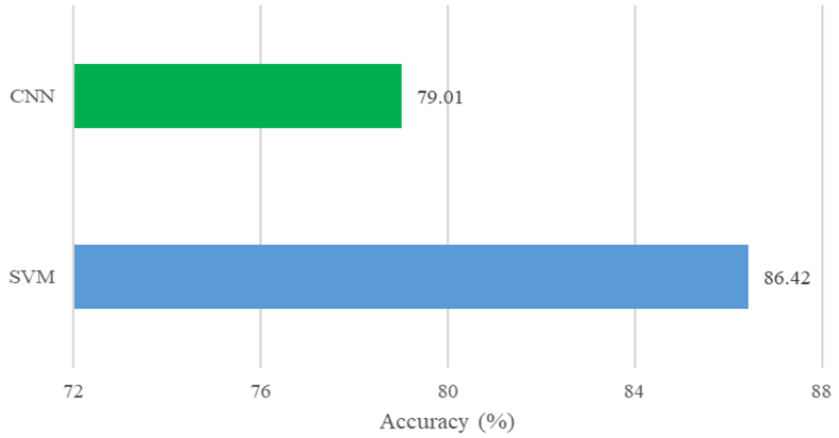

Fig. 10 *The overall accuracy rates for all BI-RADS categories using CNN and SVM models.*

The results of classifying breast lesions as malignant or benign using CNN and SVM models are shown in Fig. 11. When the results of SVM and CNN were compared, the SVM model with 92.59% accuracy showed a better classification rate.

Fig. 12 shows the ROC analysis of the test results of the proposed CNN model and the SVM algorithm. It is possible to conclude that the models are successful since they are beyond the green line area. However, the SVM based model outperformed.

Fig. 13 shows mammogram images misclassified in differentiating malignant and benign breast lesions and classifying breast lesions in the BI-RADS categories.

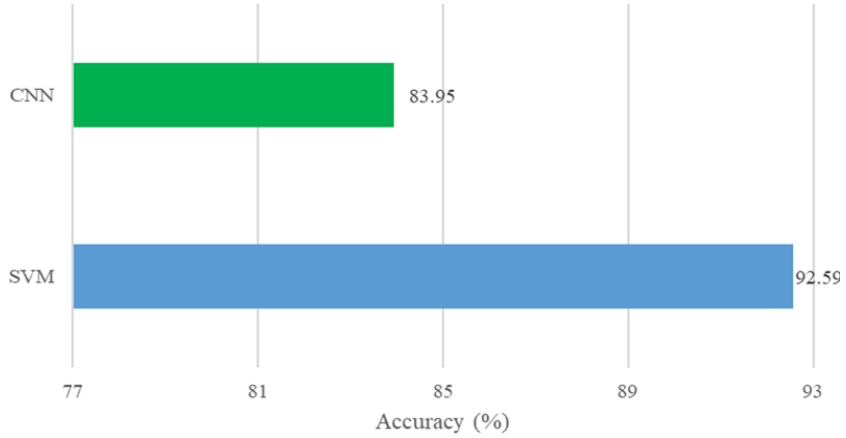

Fig. 11 *The results of classifying breast lesions as malignant or benign using CNN and SVM models.*

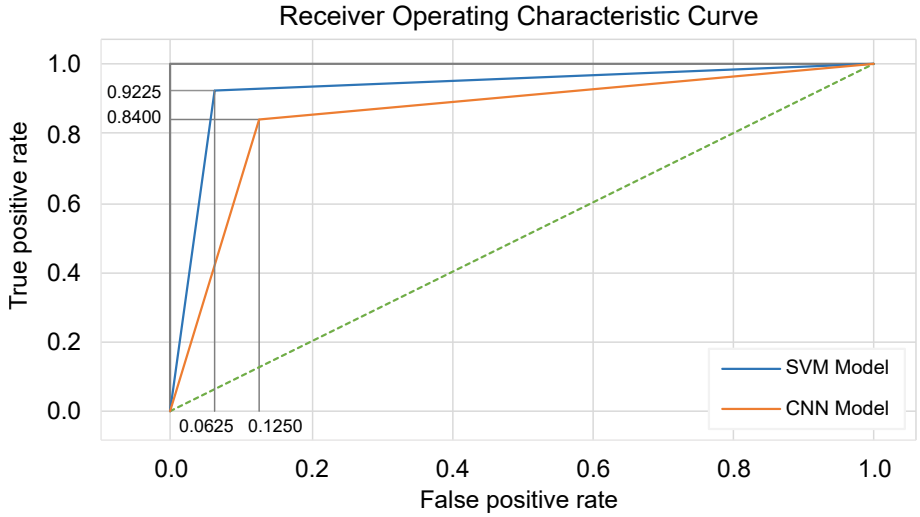

Fig. 12 *Comparison of ROC analysis results.*

Here, true class refers to the data belonging to classes that are previously known to be true, and predicted class refers to the results of the classes that are predicted by our model.

4. Discussion and conclusion

According to the BI-RADS guidelines, BI-RADS 2 shows a high probability of benign lesions, BI-RADS 3 defines the probability of benign lesions, BI-RADS 4 illustrates suspicious lesions, and BI-RADS 5 describes a high probability of ma-

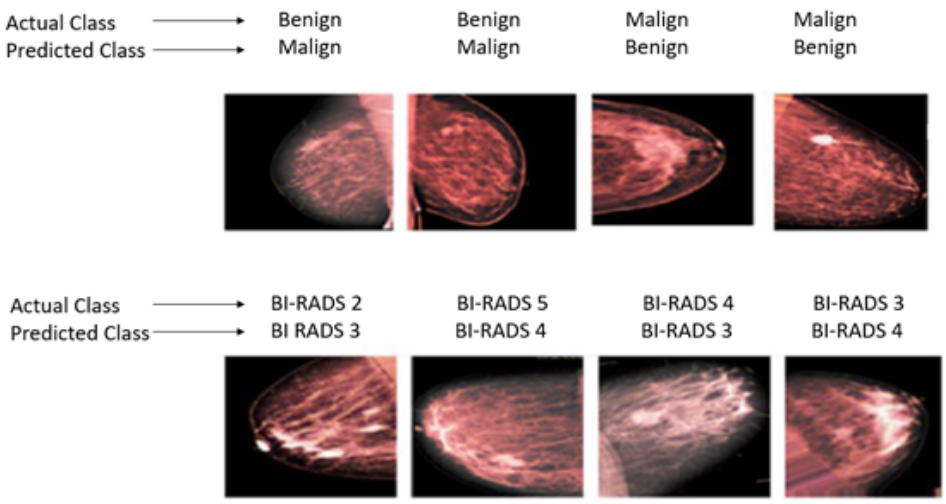

Fig. 13 *Misclassified mammogram images (SVM testing result).*

lignant lesions. In clinical practice, differentiating between malignant and benign breast lesions and classifying breast lesions of BI-RADS categories are complicated tasks for radiologists due to individual differences in breast density and daily workload. When the images are examined, it is seen that it is difficult to predict the BI-RADS categories and malignant/benign classification of breast lesions with the naked human eye. Therefore, using CAD systems is very useful to assist radiologists in decision-making.

This study aims to provide a practical approach to BI-RADS categories and breast mass malignancy classification. For this purpose, SVM-based and CNN-based CAD systems are designed. Both the CNN and SVM models are uniquely designed in terms of their architectural structures and principles of use. These proposed models may provide a new systematic perspective in breast cancer diagnosis. Comparing the classification rates obtained by the two proposed systems for the classification of BI-RADS categories and malignant-benign discrimination, it can be seen that the SVM-based system has a better classification rate. This study demonstrates that ML methods are still a popular choice for classification tasks and can be utilized as an alternative to deep learning, particularly in the medical field.

Tab. XII compares the model presented in this study and similar studies in the literature. There are separate studies in the literature that classify BI-RADS and mass malignancy. However, there are very few studies that automatically classify both breast lesion and BI-RADS categories together. In addition, studies in the literature often use publicly available datasets rather than clinical datasets. Therefore, this study is valuable because it uses clinical data. When compared to studies in the literature, the model presented in this study achieved 86.42% and 92.59% accuracy rates with the SVM model and 79.01% and 83.95% accuracy rates with the CNN model in both BI-RADS category classification and mass malignancy classification.

Ref	Task	Dataset	Methods	Accuracy [%]
[7]	BI-RADS classification	DDSM	MLP	88.02
	Mass malignancy classification			83.85
[8]	BI-RADS classification	DDSM	BPNN	84.50
[9]	BI-RADS classification	A public dataset	DNN	94.22
[10]	BI-RADS classification	InBreast	CNN	83.40
[11]	Mass malignancy classification	DDSM	FFNN	98.10 (sensitivity)
[12]	Mass malignancy classification	A public dataset	CNN	90.50
[13]	Mass malignancy classification	DDSM	SVM	90.00
[14]	Microcalcification classification	A public dataset	SVM	80.00
[15]	Microcalcification classification	A public dataset	MLP	82.00
Our	BI-RADS classification	A public dataset	SVM	86.42
study	Mass malignancy classification			92.59

Tab. XII *Classification results within the present study are compared with the related studies.*

It is expected that deep learning algorithms will have a higher classification rate than machine learning algorithms in general. Due to the small-scale dataset used in our study, the SVM-based model outperformed the CNN-based model in terms of classification rate. Based on this information, it is possible to conclude that SVM or other machine learning methods can yield better results than deep learning algorithms in small-scale data sets. Larger volume datasets are needed for deep learning algorithms.

It is clear from the experimental results that the suggested SVM-based model can be utilized to classify breast lesions and their BI-RADS categories. In this way, it is thought that the proposed SVM-based model will contribute to the clinical process in the hospital. Designing a diagnostic tool for BC is a significant task since BC threatens many women. Therefore, we believe that the proposed SVM-based CAD model can be considered an auxiliary system for radiologists in classifying breast lesions with high classification rates.

References

[1] SUNG H., FERLAY J., SIEGEL R.L., LAVERSANNE M., SOERJOMATARAM I., JEMAL A., BRAY F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, 2021, *71*(3), pp. 209–249, doi: 10.3322/caac.21660.

[2] WANG L. Early diagnosis of breast cancer. *Sensors*, 2017, *17*(7), 1572.

[3] WELCH H.G., PROROK P.C., O'MALLEY A.J., KRAMER B.S. Breast-cancer tumor size, overdiagnosis, and mammography screening effectiveness. *New England Journal of Medicine*, 2016, 375(15), pp. 1438–1447, doi: 10.1056/nejmoa1600249.

[4] SPAK D.A., PLAXCO J.S., SANTIAGO L., DRYDEN M.J., DOGAN B.E. BI-RADS®fifth edition: A summary of changes. *Diagnostic and interventional imaging*, 2017, *98*(3), pp. 179–190, doi: 10.1016/j.diii.2017.01.001.

[5] JALALIAN A., MASHOHOR S., MAHMUD R., KARASFI B., SARIPAN M.I.B., RAMLI A.R.B. Foundation and methodologies in computer-aided diagnosis systems for breast cancer detection. *EXCLI journal*, 2017, *16*, 113, doi: 10.1016/j.clinimag.2012.09.024.

[6] HASSAN N.M., HAMAD S., MAHAR K. Mammogram breast cancer CAD systems for mass detection and classification: a review. *Multimedia Tools and Applications*, 2022, pp. 1–33, doi: 10.1007/s11042-022-12332-1.

[7] CHOKRI F., HAYET FARIDA M. Mammographic mass classification according to Bi-RADS lexicon. *IET Computer Vision*, 2017, 11(3), pp. 189–198, doi: 10.1049/iet-cvi.2016.0244.

[8] BOUMARAF S., LIU X., FERKOUS C., MA X. A new computer-aided diagnosis system with modified genetic feature selection for bi-RADS classification of breast masses in mammograms. *BioMed Research International*, 2020, 2020, doi: 10.1155/2020/7695207.

[9] TSAI K.J., CHO M.C., LI H.M., LIU S.T., HSU J.H., YEH W.C., HWANG S.H. A High-Performance Deep Neural Network Model for BI-RADS Classification of Screening Mammography. *Sensors*, 2022, *22*(3), 1160, doi: 10.3390/s22031160.

[10] DOMINGUES I., ABREU P.H., SANTOS J. Bi-rads classification of breast cancer: a new pre-processing pipeline for deep models training. In: *2018 25th IEEE international conference on image processing (ICIP)* (pp. 1378–1382). IEEE, 2018, doi: 10.1109/icip.2018.8451510.

[11] PUNITHA S., AMUTHAN A., JOSEPH K.S. Benign and malignant breast cancer segmentation using optimized region growing technique. *Future Computing and Informatics Journal*, 2018, 3(2), pp. 348–358, doi: 10.1016/j.fcij.2018.10.005.

[12] TING F.F., TAN Y.J., SIM K.S. Convolutional neural network improvement for breast cancer classification. *Expert Systems with Applications*, 2019, 120, pp. 103–115, doi: 10.1016/j.eswa.2018.11.008.

[13] KETABI H., EKHLASI A., AHMADI H. A computer-aided approach for automatic detection of breast masses in digital mammogram via spectral clustering and support vector machine. *Physical and Engineering Sciences in Medicine,* 2021, 44(1), pp. 277–290, doi: 10.1007/s13246-021-00977-5.

[14] LI M., ZHU L., ZHOU G., HE J., JIANG Y., CHEN Y. Predicting the pathological status of mammographic microcalcifications through a radiomics approach. *Intelligent Medicine*, 2021, 1(03), pp. 95–103, doi: 10.1016/j.imed.2021.05.003.

[15] STELZER P.D., STEDING O., RAUDNER M.W., EULLER G., CLAUSER P., BALTZER P.A.T. Combined texture analysis and machine learning in suspicious calcifications detected by mammography: Potential to avoid unnecessary stereotactical biopsies. *European Journal of Radiology*, 2020, 132, 109309, doi: 10.1016/j.ejrad.2020.109309.

[16] SHORTEN C., KHOSHGOFTAAR T.M. A survey on image data augmentation for deep learning. *Journal of Big Data*, 2019, 6(1), pp. 1–48, doi: 10.1186/s40537-019-0197-0.

[17] MAITRA I.K., BANDYOPADHYAY S.K. CAD Based Method for Detection of Breast Cancer. *Oriental Journal of Computer Science and Technology*, 2018, 11(3), pp. 154–168, doi: 10.13005/ojcst11.03.04.

[18] FONTI V., BELITSER E. Feature selection using lasso. VU Amsterdam research paper in business analytics, 2017, 30, pp. 1–25, doi: 10.5220/0005827003810386.

[19] VADIVEL A., SURENDIRAN B. A fuzzy rule-based approach for characterization of mammogram masses into BI-RADS shape categories. *Computers in Biology and Medicine*, 2013, 43(4), pp. 259–267, doi: 10.1016/j.compbiomed.2013.01.004.

[20] HARALICK R.M., SHANMUGAM K., DINSTEIN I.H. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 1973, (6), pp. pp. 610–621, doi: 10.1109/tsmc.1973.4309314.

[21] GALLOWAY M.M. Texture analysis using gray level run lengths. *Computer graphics and image processing,* 1975, 4(2), pp. 172–179, doi: 10.1016/s0146-664x(75)80008-6.

[22] TIBSHIRANI R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society*: *Series B (Methodological)*, 2019, 58(1), pp. 267–288, doi: 10.1111/j.2517-6161.1996.tb02080.x.

[23] KARAMIZADEH S., ABDULLAH S.M., HALIMI M., SHAYAN J., JAVAD RAJABI M. Advantage and drawback of support vector machine functionality. *In 2014 international conference on computer, communications, and control technology (I4CT) (pp. 63–65). IEEE*, 2014, doi: 10.1109/i4ct.2014.6914146.

[24] SNOEK J., LAROCHELLE H., ADAMS R.P. Practical Bayesian optimization of machine learning algorithms. *Advances in neural information processing systems*, 2012 , 25, doi: 10.21203/rs.3.rs-3312122/v1.

[25] MATLAB and Statistics Toolbox Release 2020a, The MathWorks, Inc., Natick, Massachusetts, United States.

[26] MOHAPATRA S., MUDULY S., MOHANTY S., RAVINDRA J.V.R., MOHANTY S.N. Evaluation of deep learning models for detecting breast cancer using histopathological mammograms Images. *Sustainable Operations and Computers,* 2022, 3, pp. 296–302, doi: 10.1016/j.susoc.2022.06.001.