Academic Journal of International University of Erbil

Journal Homepage: https://ojs.cihanrtv.com/index.php/public

PRINT-ISSN: 2519-6928

Online-ISSN: 3080-7174

Automated Diagnosis System for Early Breast Cancer Detection

Jamal Othman Ali[1], Alan Anwer Abdulla[2, 3]

[1]Department of Information Technology, Sulaimani Polytechnic University, Sulaimani, Kurdistan, IRAQ

[2]Department of Information Technology, University of Sulaimani, Sulaimani, Kurdistan, IRAQ

[3]Department of Information Technology, University College of Goizha, Sulaimani, Kurdistan, IRAQ

DOI: https://doi.org/10.63841/iue326995

Received 11 Aug 2025; Accepted 09 Sep 2025; Available online 25 Apr 2026

ABSTRACT:

Given that manual disease diagnosis and treatment determination are time-consuming, expensive, and demand skilled professionals, developments in computer technology and medical imaging processing within healthcare environments have improved diagnostic accuracy and early disease detection. Since they overcome the limitations associated with manual diagnostic approaches, automated or computer-assisted diagnostic systems can serve as viable alternatives. These systems are known as computer-aided diagnostic (CAD) systems and are commonly utilised for cancer diagnosis and other medical conditions. Digital image processing plays a vital role in handling and examining medical images for cancer identification and detection. This research introduces a CAD system designed for early breast cancer detection. Following the segmentation of the region of interest (ROI) from mammographic images through thresholding segmentation methods, texture and shape characteristics are extracted. For textural feature extraction, Tamura and Bag of Visual Words (BoVW) techniques are utilised. Additionally, shape-based statistical features, including compactness, sphericity, area, and elongation, are extracted. The images are subsequently categorized as normal or abnormal through k-nearest neighbour (kNN) and support vector machine (SVM) classification algorithms based on the extracted features. Comprehensive experiments conducted on the well-known Mammographic Image Analysis Society (MIAS) mammography image dataset were used to assess the proposed system's performance. The experimental results demonstrate that the proposed system achieves superior performance compared to other existing systems, with an accuracy rate of 99.6%.

Keywords: Breast Cancer, BOVW, Tamura, Descriptor, SVM, KNN.

1 INTRODUCTION

With over 10 million deaths annually, cancer ranks among the primary causes of death globally. Around 2.3 million new cases of breast cancer are identified each year, making it the most common cancer in women globally [1]. The three most significant components in a breast consist of lobules together with ducts and connective tissue; see fig.1. The organs that produce milk have the name 'lobules'. The ducts deliver milk from the nipple through their tubing system [2]. The majority of breast tumours originate from the ducts or lobules. All types of bodily tissue obtain support through connective tissue either to connect with or divide from other tissues. The tissue, along with its fellow tissue types, exists as cells in liquid compartments which are called extracellular matrix (ECM). The cells in connective tissue occupy a distinct arrangement from other tissue types because they reside at low density inside the extracellular matrix [3].

*Corresponding author: author@organization.edu.co

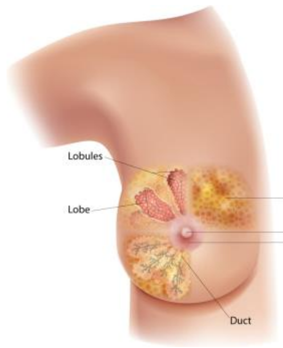

FIGURE 1. Significant Components of Breast [4].

Breast cancer primarily develops in the milk-producing ducts and/or milk-storing lobules, and it is partitioned into several molecular subtypes such as luminal A, luminal B, triple-negative, and HER2-enriched, each with distinct clinical behaviours [5]. As shown in Table 1, the condition is governed by a multitude of determinants that are genetic, hormonal, and lifestyle in nature, with mortality rates of between 5% and 60%, and impacting both males and females, albeit to differing degrees of prevalence.

Table 1. Factors Causing Breast Cancer and Mortality Rate [6] .

Causes	Mortality in %	location
Genetic Factors	5-10%	44,130 (USA)
Hormonal Factors	15-20%	43,700(Global)
Lifestyle Factors	20-60%	685,000(Global)
Age-Related Risk	50-60%	2.3 million (New Cases)
Environmental Exposures	5-10%	Varied Regionally

Global statistics reveal that breast cancer makes up about 25% of all cancer diagnoses among women, with notable differences in incidence and mortality rates depending on the geographical area [7]. Fig. 2 illustrates death rates of breast cancer compared to other common types of cancer, including lung cancer, brain tumours, blood cancer, and skin cancer, highlighting its significant effect on women's health. Early detection is essential for improving patient outcomes, as survival rates can surpass 90% when the disease is caught at localized stages [8].

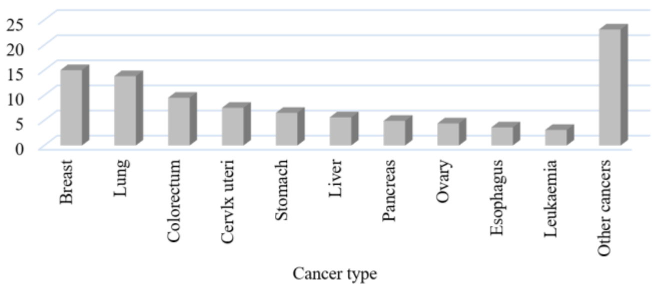

FIGURE 2. The number of deaths of certain common cancers in women around the world in 2018 [9].

A computerized system known as a computer-aided diagnostic (CAD) system uses medical images such as MRIs, X-rays, CT scans, ultrasounds, and microscopic images to help doctors diagnose illnesses, especially cancers [10]. Although mammography has long been the standard screening method, manual interpretation of these images is subject to variability and can be both time-consuming and costly. These limitations have stimulated significant interest in CAD systems capable of assisting radiologists with objective and standardized interpretation of mammographic images [11] [12]. Recent advances in digital image processing and artificial intelligence have enabled more precise extraction of relevant features from medical imaging, allowing automated systems to differentiate between normal and pathological tissues with greater reliability [13]. This paper proposes an automated CAD system for the classification of mammographic images as normal or abnormal in the detection of breast cancer. The proposed system employs a multi-stage strategy comprising sophisticated preprocessing techniques to improve image quality, effective feature extraction to identify significant malignancy indicators, and a dependable classification process.

The remainder of this study is organized as follows. The summary of previous efforts in the literature is given in section 2. Section 3 describes the proposed system. Section 4 summarizes the results of the experiments. Lastly, section 5 provides the conclusion.

2 LITERATURE REVIEW

The CAD system is a computer-based tool used in medical image processing that aids clinicians in making final decisions regarding various diseases, particularly cancers. The goal of the entire procedure is to extract important information from medical images, including ultrasounds, CT scans, and MRIs. Numerous CAD systems have been developed to detect various cancers such as lung cancer, breast cancer, brain tumours, and etc. The focus of this research is detecting breast cancer. Processing and analysing breast mammography images is critical in the early detection of breast cancer. This part highlights the most prominent and relevant contemporary initiatives to detect early breast cancer using digital image processing. Researchers have used a variety of algorithms and analytical techniques to detect breast cancer and classify mammography images into normal or suspicious. Their CAD systems aim to improve early detection capabilities while also providing doctors with more consistent diagnostic support in regular practice. The primary objective in this field of research is to lower the error rate while increasing the accuracy rate of breast cancer diagnosis,

Winsberg et al., in 1967, was the first who introduced a system for detecting the radiographic anomalies in mammogram images. In the pre-processing step, Films of the breasts were produced with the bolus immersion technique, with the breast submerged in either mineral oil or alcohol, resulting in translateral decubitus films. The obtained radiographs were transformed into positive images for scanning. A converted radio-facsimile scanner was utilized with 180 lines per inch resolution and 16 grey-levels. Moreover, the films were split into 64 small rectangles in an 8 × 8 array, standardized relative to the chest wall, nipple, and skin. Four vectors have been calculated for feature extraction, which are the area vector (percentage of sample points at each density level), slice vector (distribution over 16 strips), uniformity vector (density distribution smoothness index), and contiguity vector (density pattern upper/lower half analysis). With regard to the comparison, there are comparison algorithms introduced to compare the matching rectangles on the left and right breasts, and as a result, index numbers are generated to indicate the divergence of the vectors. Finally, visualization of significant differences was done using different line types, with clustering of lines indicating potential lesions. This approach successfully detected lesions in selected cases, particularly in atrophic breasts with solitary lesions, though dense breasts and cystic mastopathy presented greater challenges [14].

Mavroforakis et al., in 2006, developed a CAD system aimed at classifying breast cancer through mammograms, with a focus on analysing the texture of masses. The pre-processing stage involved localized image segmentation and manually marking tumour boundaries. For feature extraction, first-order statistics has been considered. Gray Level Co-occurrence Matrix (GLCM), and Run-Length Matrix (RLM), plus fractal dimension analysis to assess information richness were considered. Several classifiers were employed for the evaluation purposes, including k-nearest neighbour (kNN), support vector machine (SVM), and artificial neural network (ANN). The SVM classifier really shone, boasting an accuracy rate of 83.9%, which is a considerable improvement over linear classifiers like LDA/LSMD, which only managed 69% while ANN achieved 78.2% [15].

Srivastava et al., In 2013, introduced a CAD system for the detection of breast cancer utilising mammography images. The contrast-limited adaptive histogram equalisation (CLAHE) approach is utilised for pre-processing. Thus, the segmentation technique utilises a three-class fuzzy C-means approach. Texture features, geometric/shape features, histogram/statistical features, wavelet-based features, and Gabor features were retrieved for the aim of feature extraction. the feature selection approach termed minimum redundancy maximum relevance has been employed. To categorise the cells as malignant or non-cancerous, three distinct classifiers were utilised: SVM, kNN, and ANN. (SVM) outperformed both (kNN) and (ANN), with an accuracy rate of 85.57% using Mammographic Image Analysis Society (MIAS) dataset [16].

Bhateja et al., in 2018, created a CAD system to classify breast cancer, in the pre-processing step, the sigmoidal transformation was implemented to improve image contrast making it easier to detect any suspicious areas.

For the feature extraction step, 14 Haralick features have been pulled from the GLCM, which captured texture patterns like contrast, correlation, and entropy. Finally, the SVM classifier is implemented to sort mammograms into normal or abnormal. This system reached an accuracy rate of 92.85% on the MIAS dataset [17].

R. Vijayarajeswari et al., in 2019, proposed a CAD system to detect breast cancer utilizing mammogram images. The system began with pre-processing step, with the help of gradient-based thresholding to erase any unnecessary labels. Then, the maximum likelihood estimation is implemented for the segmentation purposes. Regarding to the feature extraction, Hough transform is applied to extract a transform space for intensity feature calculation and intensity features (mean, variance, entropy, and standard-deviation) are calculated from this transformed space. The classification was conducted using SVM achieving an accuracy rate of 94% utilizing MIAS dataset [18].

Fadhil et al., in 2021, introduced a CAD system to spot breast cancer in mammograms. The system started by cleaning up the images using histogram equalization and median filtering to make them clearer. For finding suspicious areas, region growing approach was considered, and then pulled out meaningful patterns using Discrete Wavelet Transform (DWT) to analyse both histogram and texture features. Consequently, four distinct classifiers, Linear Discriminant Analysis (LDA), ANN, Decision Trees, and SVM, were used to make the final call on whether the mammogram image is cancer or not. The SVM worked best and achieved an accuracy rate of 93.6% utilizing MIAS dataset [19].

Chaudhury et al., in 2022, created a CAD system for detecting breast cancer. For the image enhancement purposes, the CLAHE technique was employed. Consequently, the k-means clustering was implemented to segment the (ROI). Eventually, the Fuzzy Support Vector Machine was employed for the classification purposes, which achieving an accuracy rate of 89% on the MIAS dataset [20].

Yan et al., in 2023, designed a CAD system for the detection of breast cancer utilizing mammography images. The suggested system begins with a novel region extraction technique that uses image fusion and filtering to improve contrast and reduce false results. To remove interference, a pectoral muscle removal method with adaptive Otsu thresholding and morphological procedures was used. The automated ROI segmentation technique was then implemented, which made use of dynamic thresholding and region-growing algorithms. Textural and morphological features were extracted from the segmented ROIs before being optimized using a feature weighting technique based on INFO method that used opposition-based learning. To classify mammograms as normal or abnormal, an ensemble model integrating Bagging, kNN, and EigenClass algorithms was used. Experiments on MIAS dataset proved the outperformance of the CAD system, obtaining an accuracy rate of 93.26% [21].

Velumani et al., in 2023, presents a CAD system for detecting breast cancer in mammogram images. The system began with pre-processing using CLAHE to enhance image quality and reduce blurriness. For the segmentation purposes, Particle Swarm Optimization (PSO) was employed to distinguish between nuclei and non-nuclei cells, leveraging parallel computing and global optimization to achieve superior super pixel results. Feature extraction was performed using the GLCM technique, which effectively captured texture patterns to differentiate between various tissue types based on their recognizable characteristics. The extracted features were then classified employing SVM and achieved 98.9% of accuracy rate utilizing MIAS dataset [22].

Prinzi et al., in 2024, presented a CAD system that uses digital mammography images to diagnose and categorize breast cancer. Gray-level discretization with a specified bin width was used for pre-processing in order to normalize intensity levels. Consequently, ITK-SNAP software was used for manual segmentation. For feature extraction, textural features and first-order statistical features were extracted from the GLCM, GLDM, GLSZM, GLRLM, and NGTDM matrices. Variance analysis, correlation filtering, statistical testing, and sequential forward floating selection (SFFS) were used to select features in order to minimize redundancy and enhance relevance. In the end, three classifiers: SVM, Random Forest (RF), and XGBoost— were employed to categorise the ROIs into normal or abnormal. With a 93.3% of accuracy rate, XGBoost emerged as the top performer in the categorization issue, while SVM achieved 86.8% [23].

Hoseini et al., in 2024, presented a CAD system for early breast cancer detection using the MIAS mammography dataset. For pre-processing, ROI was extracted as 100×100 pixel patches, with normal images sampled centrally and abnormal images centred on lesions. Segmentation was implicitly achieved through lesion-centric ROI extraction. For feature extraction, Gabor wavelet transformations were applied to derive 12 textural descriptors, including energy, entropy, inertia, and statistical moments such as kurtosis and skewness. To minimize feature redundancy, feature selection technique Least Significant Difference (LSD) was employed. Finally, kNN classifier was used an accuracy rate of 92% was taken [24].

Kanya et al., in 2024, suggested a system for detecting breast cancer using mammogram images. The pre-processing is started with employing CLAHE technique for enhancing the image contrast. Next, Advanced Gray-Level Co-occurrence Matrix (AGLCM) technique was implemented for extracting texture features. Then, the feature selection technique named Weighted Adaptive Binary Teaching Learning Based Optimization (WA-BTLBO) was employed in order to select most important feature. Finally, XGboost, kNN, Random Forest, ANN, and SVM classifiers were utilized for the classification step. The best outcome was achieved by XGboost which was 98.3% using MIAS dataset, while kNN, ANN, and SVM achieved 96.6%, 97.1%, and 97.8% respectively [25].

Sharma et al., in 2025, designed a CAD system for detecting breast cancer for classifying mammogram masses using the MIAS datasets. Pre-processing involved enhancing the mass region using median filtering, CLAHE, and Gaussian filter to remove noise. The system employs a hybrid feature extraction method applied to the ROI, which fuses texture information from multi-scale multi-orientation (MSMO) Gabor wavelets with statistical features from the GLCM. For the classification purposes, several models were evaluated. The SVM classifier achieved the best accuracy rate of 95.82% [26]. Table 2 presents a summary of accuracies reported by other studies.

Table 2. Reported Accuracies of the Related Works

Reference Number	Accuracy Rate %
[15]	SVM = 83.9%, ANN = 78.2%
[16]	SVM = 85.57%
[17]	SVM = 92.85%
[18]	SVM = 94%
[19]	SVM = 93.6%
[20]	SVM = 89%
[21]	ensemble model = 93.26%
[22]	SVM = 98.9%
[23]	XGBoost = 93.3%, SVM = 86.8%
[24]	kNN=92%
[25]	kNN = 96.6%, ANN = 97.1%, and SVM = 97.8%
[26]	SVM = 95.82%

The rest of this study focuses on the extension and refining of an approach of applying image processing techniques to improve the accuracy rate for early breast cancer detection.

3 PROPOSED CAD APPROACH

CAD systems were created to assist with the diagnosis of a diverse array of maladies, with a particular emphasis on malignancies such as breast cancer, tumours, and lung cancer [29]. This study investigates the potential of mammography images to detect breast cancer at an early stage by implementing a computer-aided diagnosis strategy. The breast is typically depicted in grey and white in mammography; however, the background is frequently black. A tumour may manifest as a dense white region.

The key stage for each CAD system in identifying breast cancer is to isolate the region of interest from the remainder of the image. This section outlines the tactics of the proposed CAD system, starting with the incorporation of several pre-processing techniques to enhance the input image. Consequently, a segmentation technique must be employed to isolate the ROI from the remainder of the input image. Ultimately, from the segmented ROI, several significant features are collected, and classification algorithms are employed to ascertain if the input pictures are normal or anomalous.

3.1 PRE-PROCESSING

Pre-processing is a crucial step in CAD system development since it impacts segments and extracting features. In other words, pre-processing simplifies segmentation and promotes feature extraction from the segmented ROI. In this study, this step includes several pre-processing techniques to improve the quality of mammographic images, such as intensity normalization for equal treatment of different acquisition sources, region props to remove labels, median filter to remove noise without blurring tissue borders, crop the image to the largest contiguous region, CLAHE to improve local contrast particularly for micro-calcifications, and contrast stretching to maximize the use of dynamic range, see figure 3.

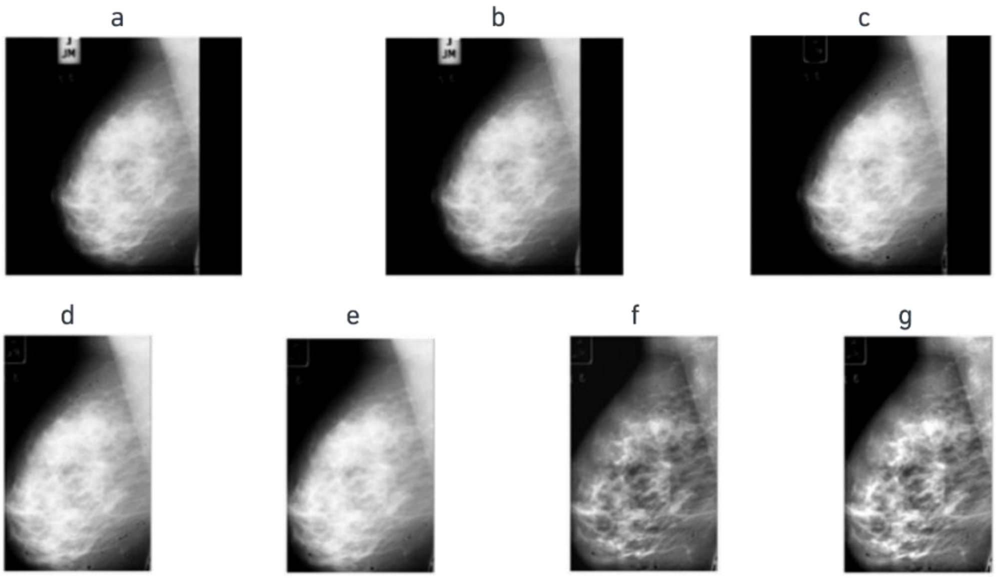

FIGURE 3. Pre-Processing Step: (a) Mammogram Image, (b) Image after implementing Intensity Normalization, (c) Image after removing Label, (d) Cropped Image, (e) Image after implementing Median Filter, (f) Resulted Image after implementing CLAHE and (g) contrast stretch on Image.

3.2 SEGMENTATION

Segmentation is an essential part in the development of a CAD system, since segmented cells are utilised to extract important feature for further classification. In this work, the threshold-based segmentation technique is implemented to separate the ROI from the rest of the input image, image (b) in Fig.4, with the output image being a binary image. The threshold segmentation technique accurately distinguishes lesion boundaries by comparing intensity values. When binary masks made using basic intensity thresholding are applied to the original image, the ROI is properly delineated.

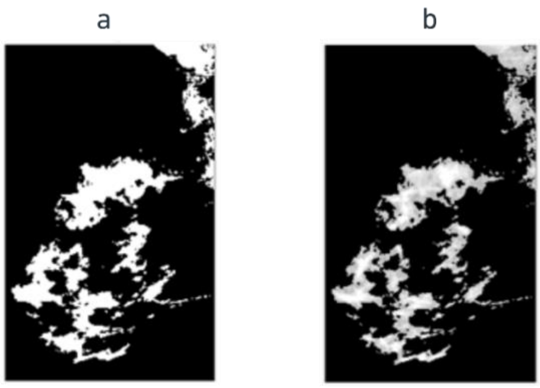

FIGURE 4. Segmentation Step: (a) Binary Image, (b) The ROI is selected in the Original Image based on the Binary Image in (a)

3.3 FEATURE EXTRACTION

The feature extraction process extracts the ROI's most effective and significant features. The retrieved features will be used in the following phase to differentiation between normal and diseased tissue. The extracted features include texture, shape, and visual vocabulary that are taken from the segmented image, which is image (b) in Fig. 4. For the texture feature, Tamura feature extraction technique is employed that computed features such as directionality, coarseness, contrast, regularity, line-likeness, and roughness. Regarding to the shape feature, certain statistical-shape based features including compactness, sphericity, area, and elongation are extracted. Eventually, for the visual vocabulary features, visual descriptors are extracted using a Bag of Visual Words (BOVW) technique to represent margins, textures, calcifications, and density variations. Additionally, all the 60 extracted features are concatenated for the purposes of classification. Table 3 presents the designated feature name.

Table 3. Extracted Features

Feature Type	Feature Name
Tamura	Coarseness, Contrast, Directionality, Line-likeness, Regularity, Roughness
Shape	Compactness, Sphericity, Area, Elongation
BOVW	50 visual word frequencies

3.4 CLASSIFICATION

Finally, the features obtained are incorporated into classifier algorithms, which is produce a model capable of distinguishing between abnormal and normal. The input images are categorized as abnormal or normal employing SVM and kNN classifier techniques, and k-fold cross validation is implemented by all classifiers with k values of 5, 10, 15, and 20. Fig. 5. depicts the proposed CAD block diagram.

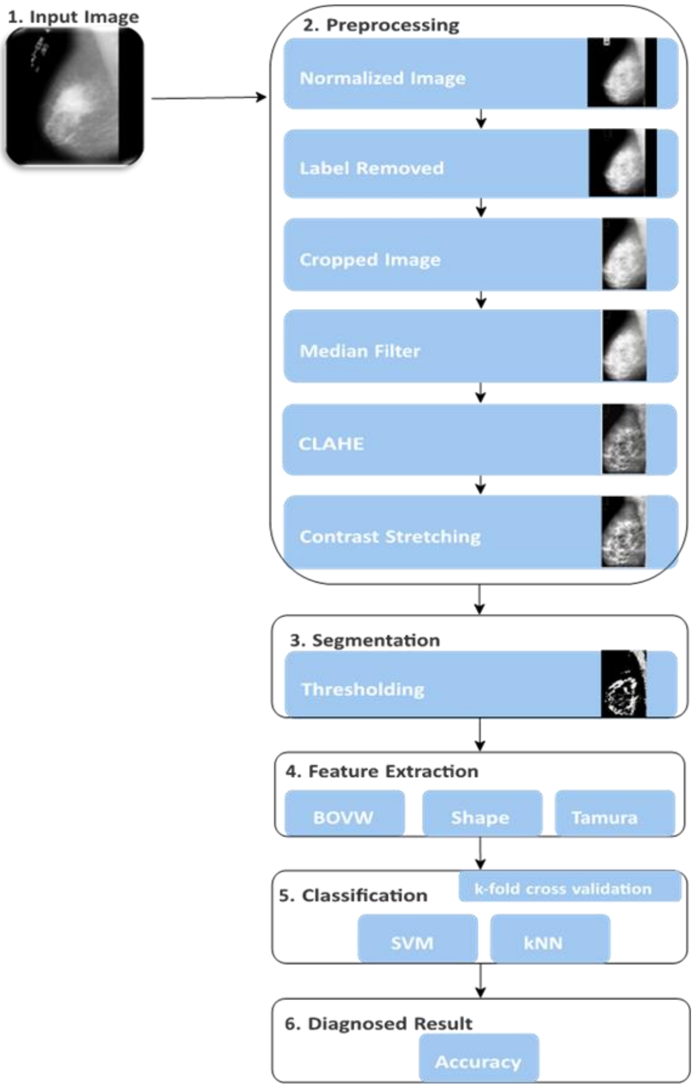

FIGURE 5. Illustrates the block diagram of the Proposed System.

5 PROPOSSED EXPERIMENTAL OUTCOMES

The primary objective of the suggested CAD methodology is to differentiate breast cells as either normal or anomalous. This section includes thorough assessments that evaluate the effectiveness of the proposed method in terms of accuracy rate. The proposed technique is also compared to prior Systems.

5.1 DATASET

To evaluate the efficacy of the proposed CAD system, tests are conducted using the publicly accessible and renowned Mammographic Image Analysis Society (MIAS) collection of pictures. The MIAS collection comprises 322 Images, including 208 normal and 114 abnormal specimens [27] All images possess a same resolution of 1024 by 1024 pixels.

5.2 RESULT

The MATLAB programming language was employed to construct the proposed CAD system. Two classifiers—KNN and SVM—are implemented to assess the accuracy of each extracted feature. The accuracy rates for the extracted features using SVM and KNN are illustrated in Tables 4 and 5, respectively. Furthermore, each evaluation test was conducted with considering a variety of k-fold values. K-fold cross-validation is the process of evaluating the model's performance against new data by dividing a dataset into k folds. The data sample is partitioned into k categories. The accuracy of classifiers with k values of 5, 10, 15, and 20 is assessed using k-fold cross-validation in this work. In addition, the accuracy rate was determined by employing the following formula [28]:

$$TP + TN/(TP + TN + FP + FN) \quad (1)$$

Where: TP, FP, TN, and FN refer to true positive, false positive, true negative, and false negative respectively.

Table 4. Accuracy Rate for the Extracted Features using SVM.

Feature Extraction Techniques	5 k-fold	10 k-fold	15 k-fold	20 k-fold	Average
Tamura	97.51	97.5	97.51	97.53	97.51
Shape	87.88	87.59	88.57	88.52	88.14
BOVW	99.38	99.37	99.39	99.37	99.38
Combined	99.69	99.68	99.67	99.68	99.68

Table 5. Accuracy Rate of Extracted Features By using kNN.

Feature Extraction Techniques	5 k-fold	10 k-fold	15 k-fold	20 k-fold	Average
Tamura	97.82	98.12	98.15	98.14	98.06
Shape	87.57	86.62	87.66	87.27	87.28
BOVW	99.37	99.37	99.38	99.37	99.37
Combined	99.38	99.38	99.38	99.37	99.38

It is evident from tables 4 and 5 that combining the extracted features instead of taking them separately yielded the proper results.

In tables 4 and 5, that the number of visual words extracted using the BOVW feature extraction technique equals 100. Further investigation is being conducted into the count for testing the various numbers of visual words. For this purpose, several numbers of visual words are considered, beginning with 25 and increasing by 25 until reaching 200 visual words. The results show that the best performance was obtained when 25 visual terms were used, see tables 6 and 7.

Table 6. The Accuracy Rate for the Different Numbers of Extracted Visual Words using SVM.

Number of Visual Words	5 k-fold	10 k-fold	15 k-fold	20 k-fold	Average
25	99.69	99.69	99.69	99.68	99.69
50	99.68	99.69	99.68	99.68	99.68
75	99.69	99.68	99.69	99.11	99.54
100	99.69	99.68	99.67	99.68	99.68
125	99.30	99.68	99.41	99.21	99.40
150	99.70	99.68	99.69	99.30	99.59
175	99.37	99.68	99.69	99.68	99.60
200	99.68	99.69	99.68	99.37	99.61

Table 7. The Accuracy Rate for the Different Numbers of Extracted Visual Words using kNN.

Number of Visual Words	5 k-fold	10 k-fold	15 k-fold	20 k-fold	Average
25	99.68	99.68	99.68	99.68	99.68
50	99.68	99.68	99.38	99.68	99.60
75	99.38	99.38	99.38	99.37	99.37
100	99.38	99.38	99.38	99.37	99.38
125	99.37	99.37	99.36	99.37	99.37
150	99.38	99.38	99.39	99.37	99.38

175	99.37	99.68	99.69	99.37	99.53
200	99.38	99.38	99.38	99.37	99.38

The findings in tables 6 and 7 reveal that the BOVW feature extraction technique yielded the best results when 25 visual words were used. Furthermore, the proposed approach yields the best results when employing the SVM classifier technique rather than kNN, with an accuracy rate of 99.69%. Focusing on the best result obtained using the BoVW approach, in which 25 different evaluation metrics were tested, Table 8 provides the corresponding results for metrics such as accuracy, recall, precision and f1-score.

Table 8. Performance Evaluation of the Proposed Approach using Difference Measurements.

		5 k-fold	10 k-fold	15 k-fold	20 k-fold	Average
SVM	Accuracy	99.69	99.69	99.69	99.68	99.69
	Precision	99.52	99.13	99.52	99.12	99.32
	Recall	99.02	99.02	99.04	99	99.02
	F1-score	99.26	99.04	99.25	99.01	99.14
kNN	Accuracy	99.68	99.68	99.68	99.68	99.68
	Precision	96.3	95.48	95.04	95.81	95.66
	Recall	99.02	99.52	99.52	99.54	99.40
	F1-score	97.61	97.42	97.17	97.52	97.43

5.3 STATE-OF-THE-ART COMPARISON

More trials have been conducted to evaluate the suggested strategy with current efforts; see tables 9 and 10. Three of the works are used the SVM see table 9, while the other two are used kNN see table 10.

Table 9: Accuracy Rates of the Evaluated CAD Approaches Implementing SVM.

Systems	Accuracy Rate (%)
Velumani et al., in 2023 [22]	98.9%
Sharma et al., in 2025 [26]	95.82%
Kanya et al., in 2024 [25]	93.8%
Proposed	99.69%

Table 10: Accuracy Rates of the Evaluated CAD Approaches Utilising KNN.

Systems	Accuracy Rate (%)
Hoseini et al., in 2024 [24]	92%
Kanya et al., in 2024 [25]	92.7%
Proposed	99.68%

The findings shown in tables 9 and 10 indicate that the suggested technique surpassed the current state-of-the-art.

CONCLUSION

The advancement of medical image processing applications in the healthcare sector has enhanced the quality and precision of illness diagnosis and early detection, as manual disease or cancer detection and treatment identification is expensive, time-intensive, and necessitates skilled professionals. Algorithms for medical image processing can more reliably, rapidly, and cost-effectively identify specific diseases. Breast cancer is one of the foremost causes of mortality among women across all cancer types. Therefore, preventing fatalities from breast cancer requires early detection. As a result, present machine learning algorithms can predict the early identification of breast cancer cells. The primary aim of creating a computer-aided diagnosis (CAD) system for mammography pictures is to assist physicians and diagnostic specialists by providing an alternate viewpoint, hence enhancing confidence in the diagnostic process. This study aimed to create an efficient CAD system for the early detection of breast cancer. Based on the testing results, the proposed CAD method surpassed the existing techniques with an accuracy of 99.69%, though this evaluation was conducted on a limited dataset and may require validation across diverse populations and imaging conditions. While these findings are promising, the clinical implementation remains subject to regulatory approval processes and integration challenges within existing healthcare workflows. Future work should focus on prospective clinical trials to validate the system's performance in real-world healthcare settings.

FUNDING

No funding was received for this work

CONFLICTS OF INTEREST

The author declares no conflict of interest.

DATA AVAILABILITY

The dataset used in this study is publicly available. The MIAS (Mammographic Image Analysis Society) breast cancer dataset can be accessed at: https://www.kaggle.com/datasets/kmader/mias-mammography

REFERENCES

[1] Z. Momenimovahed and H. Salehiniya, “Epidemiological characteristics of and risk factors for breast cancer in the world,” 2019, Dove Medical Press Ltd. doi: 10.2147/BCTT.S176070.

[2] S. K. Biswas, S. Banerjee, G. W. Baker, C. Y. Kuo, and I. Chowdhury, “The Mammary Gland: Basic Structure and Molecular Signaling during Development,” Apr. 01, 2022, MDPI. doi: 10.3390/ijms23073883.

[3] T. J. McKee, G. Perlman, M. Morris, and S. V. Komarova, “Extracellular matrix composition of connective tissues: a systematic review and meta-analysis,” Scientific Reports, vol. 9, no. 1, art. no. 10542, Jul. 22, 2019, doi: 10.1038/s41598-019-46896-0.

[4] M. H. Aziz and Alan Anwer Abdulla, “Computer-Aided Diagnosis for the Early Breast Cancer Detection,” UHD Journal of Science and Technology, vol. 7, no. 1, pp. 7–14, Jan. 2023, doi: 10.21928/uhdjst.v7n1y2023.pp7-14.

[5] N. Harbeck, “Breast cancer is a systemic disease optimally treated by a multidisciplinary team,” Dec. 01, 2020, Nature Research. doi: 10.1038/s41572-020-0167-z.

[6] Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al., "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries," A Cancer Journal for Clinicians, vol. 71, no. 3, pp. 209–249, May 2021, doi: 10.3322/caac.21660.

[7] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, “Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries,” A Cancer Journal for Clinicians, vol. 68, no. 6, pp. 394–424, Nov. 2018, doi: 10.3322/caac.21492.

[8] Islami, F., Gansler, T., et al., "American Cancer Society’s report on the status of cancer disparities in the United States, 2023," A Cancer Journal for Clinicians, vol. 74, no. 2, pp. 136–166, Mar. 2024, doi: 10.3322/caac.21812.

[9] M. A. Elsadig, A. Altigani, and H. T. Elshoush, “Breast cancer detection using machine learning approaches: a comparative study,” International Journal of Electrical and Computer Engineering, vol. 13, no. 1, pp. 736–745, Feb. 2023, doi: 10.11591/ijece.v13i1.pp736-745.

[10] H. Arimura, T. Magome, Y. Yamashita, and D. Yamamoto, “Computer-aided diagnosis systems for brain diseases in magnetic resonance images,” Sep. 2009. doi: 10.3390/a2030925.

[11] Y. W. Chang et al., “Artificial intelligence for breast cancer screening in mammography (AI-STREAM): preliminary analysis of a prospective multicenter cohort study,” Nature Communications, vol. 16, no. 1, Dec. 2025, doi: 10.1038/s41467-025-57469-3.

[12] M. L. Giger, N. Karssemeijer, and J. A. Schnabel, “Breast image analysis for risk assessment, detection, diagnosis, and treatment of cancer,” Jul. 2013. doi: 10.1146/annurev-bioeng-071812-152416.

[13] A. Hamidinekoo, E. Denton, A. Rampun, K. Honnor, and R. Zwiggelaar, “Deep learning in mammography and breast histology, an overview and future trends,” Medical Image Analysis, vol. 47, pp. 45–67, Jul. 2018, doi: 10.1016/j.media.2018.03.006.

[14] F. Winsberg, M. Elkin, J. Macy, V. Bordaz, and W. Weymouth, “Detection of Radiographic Abnormalities in Mammograms by Means of Optical Scanning and Computer Analysis!” Radiology, vol. 89, no. 2, pp. 211–215, May 1967, doi: 10.1148/89.2.211.

[15] M. E. Mavroforakis, H. V. Georgiou, N. Dimitropoulos, D. Cavouras, and S. Theodoridis, “Mammographic masses characterization based on localized texture and dataset fractal analysis using linear, neural and support vector machine classifiers,” Artificial Intelligence in Medicine, vol. 37, no. 2, pp. 145–162, Jun. 2006, doi: 10.1016/j.artmed.2006.03.002.

[16] S. Srivastava, N. Sharma, S. K. Singh, and R. Srivastava, “Design, analysis and classifier evaluation for a CAD tool for breast cancer detection from digital mammograms,” International Journal of Biomedical Engineering and Technology, vol. 13, no. 3, pp. 270–300, Jan. 2013, doi: 10.1504/IJBET.2013.058447.

[17] V. Bhateja, S. C. Satapathy, S. S. Das, et al., "Haralick features-based classification of mammograms using SVM," in Advances in Intelligent Systems and Computing, Springer Verlag, 2018, pp. 787–795. doi: 10.1007/978-981-10-7512-4_77.

[18] R. Vijayarajeswari, P. Parthasarathy, S. Vivekanandan, and A. A. Basha, “Classification of mammogram for early detection of breast cancer using SVM classifier and Hough transform,” Measurement (Lond), vol. 146, pp. 800–805, Nov. 2019, doi: 10.1016/j.measurement.2019.05.083.

[19] A. F. Fadhil and H. Kahramanli Ornek, “A Computer-Aided Detection System for Breast Cancer Detection and Classification,” Selcuk University Journal of Engineering Sciences, vol. 20, no. 01, pp. 23–31, 2021.

[20] S. Chaudhury, T. Shelke, K. Sau, P. Prathar, S. S. Roy, et al., "Effective Image Processing and Segmentation-Based Machine Learning Techniques for Diagnosis of Breast Cancer," Computational and Mathematical Methods in Medicine, vol. 2022, 2022, doi: 10.1155/2022/6841334.

[21] F. Yan, H. Huang, W. Pedrycz, and K. Hirota, “Automated breast cancer detection in mammography using ensemble classifier and feature weighting algorithms,” Expert Syst Appl, vol. 227, p. 120282, Oct. 2023, doi: 10.1016/J.ESWA.2023.120282.

[22] J. Velumani, A. A. Issa, G. R. Kumar, A. H. A. Hussein, and A. M. Shakir, “Enhancing Breast Cancer Screening: Mammogram Image Analysis with Multiclass Support Vector Machine,” in 3rd IEEE International Conference on Mobile Networks and Wireless Communications, ICMNWC 2023, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/ICMNWC60182.2023.10435830.

[23] F. Prinzi, A. Orlando, S. Gaglio, and S. Vitabile, “Interpretable Radiomic Signature for Breast Microcalcification Detection and Classification,” Journal of Imaging Informatics in Medicine, vol. 37, no. 3, pp. 1038–1053, Feb. 2024, doi: 10.1007/s10278-024-01012-1.

[24] F. Hoseini, H. Sepehrzadeh, and M. Kheyri, "Early Detection of Breast Cancer by LSD Analysis and KNN Classification on MIAS Mammography Database," Quarterly Scientific Journal of Technical and Vocational University, vol. 1, no. 1, pp. 13–24, Spring 2024, doi: 10.48301/JSSC.2024.190496.

[25] L. Kanya Kumari and B. Naga Jagadesh, “An adaptive teaching learning based optimization technique for feature selection to classify mammogram medical images in breast cancer detection,” International Journal of System Assurance Engineering and Management, vol. 15, no. 1, pp. 35–48, Jan. 2024, doi: 10.1007/s13198-021-01598-7.

[26] S. Sharma, T. Choudhury, and Y. Singh, “Advanced feature extraction for mammogram mass classification: a multi-scale multi-orientation framework,” International Journal on Smart Sensing and Intelligent Systems, vol. 18, no. 1, Jan. 2025, doi: 10.2478/ijssis-2025-0022.

[27] “The Mammographic Image Analysis Society Digital Mammogram Database – ScienceOpen.” Accessed: Jun. 14, 2025. [Online]. Available: https://www.scienceopen.com/book?vid=119a594e-fa40-4bf8-9eac-950ceddb7ed1

[28] R.G. Congalton, R. G. Oderwald, and R. A. Mead, “Assessing Landsat Classification Accuracy Using Discrete Multivariate Analysis Statistical Techniques.” Photogrammetric Engineering and Remote Sensing., vol. 49, no. 12, pp. 1671–1678, Dec. 1983.

[29] E. Altobelli, P. M. Angeletti, M. Ciancaglini, and R. Petrocelli, “The Future of Breast Cancer Organized Screening Program Through Artificial Intelligence: A Scoping Review,” Feb. 01, 2025, Multidisciplinary Digital Publishing Institute (MDPI). doi: 10.3390/healthcare13040378.