INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

OPEN ACCESS

Received: 19-03-2024

Accepted: 17-05-2024

Published: 08-06-2024

Citation: Sridevi V (2024) Breast Cancer Examination in Digitized Mammograms using Integrated K-Means Clustering with Garbor Filter and Shrunk Kernel KNN Method . Indian Journal of Science and Technology 17(23): 2444-2454. https://doi.org/10.17485/IJST/v17i23.736

* **Corresponding author**.

vissridevi@gmail.com

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Breast Cancer Examination in Digitized Mammograms using Integrated K-Means Clustering with Garbor Filter and Shrunk Kernel KNN Method

V Sridevi[1]*

1 Assistant Professor, Department of Computer Science, PSG College of Arts & Science, Coimbatore, Tamil Nadu, India

Abstract

Objectives: To suggest an intelligent classification system for efficient breast cancer diagnosis that distinguishes between benign and malignant breast cancer. The goal of the research is to develop a unique CAD system for the detection & classification of breast cancer using novel K-Means clustering (KMC) with Gabor Filter (GF) and Shrunk Kernel K-Nearest Neighbor (KNN) classifier. **Methods:** Two different sorts of perspectives, such as Craniocaudal (CC) and Mediolateral oblique (MLO) mammograms are employed to improve diagnostic effectiveness. Utilizing an adaptive K-means clustering technique to segment the tumor. The Gabor filter is used in conjunction with the k-means clustering method to extract the features of the CC and MLO perspectives. The mammography image is finally classified into benign and malignant using a unique Shrunk Kernel K-Nearest Neighbor (SKKNN) classifier. The biopsy-proven annotated mammograms from the CBIS-DDSM dataset are used in this study. There were 6156 occurrences in the dataset with MLO and CC view of 1331 normal, 858 benign and 889 malignant mammograms. **Findings:** The experimental findings showed that the suggested model KMC-GF and SKKNN can accurately detect breast cancer at an early stage. The accuracy, sensitivity , specificity, AUC, precision, F1-measure for SKKNN was 92.56%, 93.8%, 92.75%, 95.2%, 93.93%, and 94.5% which are higher comparing single view features. **Novelty:** This technique could be employed in the medical field to diagnose breast cancer and also produce few false positive results. This method reduces the workload for radiologists while still providing reliable diagnostics without the need for expensive procedures or a lot of equipment.

Keywords: SKKNN; Adaptive K-means segmentation; Gabor filter; MLO and CC Mammogram; KMC-GF

1 Introduction

Worldwide, breast cancer is the most frequent cancer among women. In the 1990s, it was the fourth most prevalent cancer in India; today, it is the top cancer. According to epidemiological research, by 2030, there will be over 2 million cases of BC worldwide[(1)].

Accurate breast cancer diagnosis requires computer-aided detection (CAD) systems that employ machine learning techniques. Early breast cancer detection can be facilitated by these CAD systems. The survival rate of breast cancer rises when it is discovered early enough to allow for improved therapy. There is need for a computer aided detection (CAD) systems which uses machine learning approach to provide accurate diagnosis of breast cancer. These CAD systems can aid in detecting breast cancer at an early stage. When, breast cancer is detected early enough, the survival rate increases because better treatment can be provided. There are various existing research techniques available for the detection of Breast Cancer. However, the existing research works are observed to have more issues which might degrade the overall performance of the method. A hybrid combination of K-means and Gaussian Mixture Model are proposed for breast cancer segmentation and detection[2]. They proved that an artificial neural network is the best candidate and gives better detection. However, this accuracy is so far approximately 91% and can lead more false negative detection. In the study[3], a mass detection approach based on CNN deep features and unsupervised extreme learning machine (ELM) clustering was proposed. The study examined a breast CAD method based on feature fusion using CNN deep features. In order to distinguish between benign and malignant breast tumours, a feature set that combines deep, morphological, texture, and density features is constructed here. An ELM classifier is then created utilising the fused feature set. However, the size of high resolution digital mammography pictures and the abundance of nuclei make automated nucleus detection difficult.

Two X-ray projection views Craniocaudal (CC) and Mediolateral oblique (MLO) are obtained for each breast during mammography screening procedure[4,5]. When diagnosing breast cancer, radiologists frequently use all available views. The results from the same views of the two breasts are combined in the bilateral analysis. Numerous earlier methods for detecting mammogram lesions only looked at one view, making it impossible to gather as much rich data from single view analysis. As a result, this study adopts two-view strategy for the mass detection and malignancy classification of mammograms. The classification of mammographic lesions using Pixel N-gram features and a variety of classifiers like MLP, KNN and SVM are done in[6,7], and it was discovered that the performance showed the effect of improving N. The obtained results were evaluated and it was observed that performance attained with MLP classifier excels that achieved with KNN or SVM or classifiers.

Deep learning approach are used to detect breast cancer by combining LASSO regression features with deep convolutional neural networks (CNNs)surpassed multiview CNN without feature fusion with an accuracy of 99.2 in[8].A Radial Basis Function Network (RBF)-based system was introduced by[9]. Using RBF network classifiers, the decision-making system aids in the categorization of malignancies. The MLP method obtained 54.1667% accuracy overall, whereas the RBF neural network achieved a classification accuracy of 79.166%, demonstrating the efficacy of the RBF neural network in categorizing the mammography images. The back propagation neural network (BPNN) classification model was used by[10]. The method correctly recognises the tumour in its early stages with 99.0% classification accuracy on MIAS and DDSM datasets. For the identification of mammography cancer, fine-tuned transfer learning network model VGG16 in conjunction with two-view LASSO regression feature fusion was proposed by[11]. The 95.24 accuracy rate proved its effectiveness and improved clinical decision making.

A study[12] discussed about an inventive scheme that includes pre-processing stage and K-mean clustering feature extraction for Speed-Up Robust Features (SURF) selection. According to the findings, a decision tree model is outperformed by the automated DL strategy that is suggested, which applies K-mean clustering with MSVM. The results of the studies demonstrate that the suggested methodologies achieved average accuracy (ACC) rates of 95%, 94%, and 98%. When SVM is utilized, the increased sensitivity rate is realized at 3%, specificity is at 2%, and ROC area is at 0.99. A modified version of SVM classification for an automated CAD System for mammogram classification are used in[13]. The technology could prove to be a valuable tool for radiologists to support their decision-making when interpreting mammograms, according to the findings of performance analysis. The effectiveness of the radiologists' diagnostic proficiency could be increased by using this to make a precise and proactive judgement. The article[14] proposed effective technique for the quick identification of breast cancer. This method employed a stack made up of the three algorithms decision tree, SVM and KNN, and adopts the CRISP-DM procedure to develop a collaborative model. The meta classifier performance is compared using the three collaboration model to that of the individual works of DT, SVM, and KNN. Chi-square analysis is used to evaluate the top characteristics. The study[15] proposed the three steps of feature extraction, multiple view feature fusion and classification to identify the breast cancer. They achieved 98.4% of accuracy to classify breast cancer using hybrid feature with CNN classifier.

An innovative Back Propagation Boosting Recurrent Widening Model (BPBRW) with a Hybrid Krill Herd African Buffalo Optimization (HKH-ABO) technique was generated in this research[16] that uses breast MRI data. The initial training of the system is achieved with the help of MRI breast images. Python is helpful in the model simulation. It is shown that the accuracy rate of this model is about 99.6%.

The model[17] introduced that depends on the concept of transfer learning. To avoid fitting problem and provide constant results through the enlargement of numerous mammography images, various augmentation methods were used, including rotation, scale, and shifting. On the MIAS database, ResNet50 achieved 89.5% accuracy and NASNet-Mobile 70%. The efficiency and effectiveness of Pre-trained categorization networks are quite remarkable, rendering them highly desirable for diagnostic performance. The study[18] recommended a machine learning-based algorithm for classifying cancer. It was found from this research that multivarious conditions can influence the outcomes, which was not considered after the examination. This research emphasizes the usage of the optimized SVM or Nave Bayes that yields an accuracy of 100%. Deep Residual learning model is combined with a Decision Tree Machine Learning mechanism to accomplish breast cancer prediction with efficiency in[19]. Hence, the proposed study was referred to be RDT model. Recall, accuracy, accuracy, and specificity are used to develop and verify the RDT model. HDL model offers better prediction accuracy in terms of the traditional mechanisms. Two views of MLO and CC mammogram to enhance diagnostic efficiency were used in[20]. In feature extraction, integration of conventional K-means clustering and Gabor filter was used to extract both texture and shape features. The research[21] introduced the first order procedure of feature determination which is baseline for the classification using KNN. The highest accuracy levels were obtained with K = 5 for classification with cross-validation and K = 15 for classification without cross-validation, both of which produced accuracy values of 91.8%. Deep learning and machine methods are used for classification[22–24] and diagnosis using mammograms where the accuracy is achieved 91%. Few limitations found in terms of feature extractions and deep spot analysis of mammograms.

The examination of single views has served as the foundation for the majority of CAD system development. The development of CAD methods that make use of information from several perspectives such as bilateral views of the same breast is of great interest to reduce the amount of false positives and increase consistency. In order to find abnormal asymmetry densities, radiologists are trained to compare the left and right breasts. Earlier screening views are utilized to spot developing density. Screening with two mammograms mediolateral oblique MLO and craniocaudal (CC) is also recommended which increases the detection accuracy of breast abnormalities. Two projections may show lesions hidden by glandular tissue in one projection. Numerous earlier methods for detecting mammogram lesions only looked at one view, making it impossible to gather as much rich data from single view analysis. As a result, this study adopts two-view strategy for the mass detection and malignancy classification of mammograms.

2 Methodology

The CBIS-DDSM[25] dataset was used in this study which contains the biopsy-proven annotated mammograms. Pre-processing, segmentation, and feature extraction were performed when the training images are supplied to the computer. The suspicious regions are segmented from the mammography region using the adaptive K-Means segmentation technique. The segmented regions are passed to the feature extraction step, where the KMC-GF algorithm is used. The KMC-GF procedure consists of two steps: the first involves applying the adaptive K-Means clustering approach to a segmented region to cluster it, and the second involves using the Gabor filter to extract features from the cluster region. In order to classify mammography area, the SKKNN classifier is used. Figure 1 describes the proposed architecture using SKKNN in BC Detection.

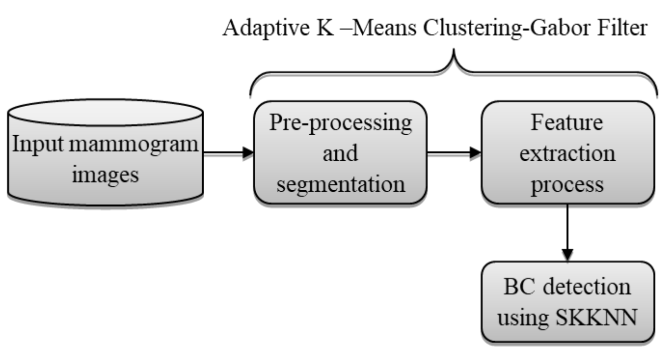

Fig 1. Proposed Architecture using SKKNN in BC Detection

2.1 Data Set

In this research, we make use of the biopsy-proven annotated mammograms from the CBIS-DDSM dataset. The collection includes bilateral breast scans taken from CC and MLO perspectives. Based on the diagnosis of the radiologists, we extract

1331 normal, 858 benign and 889 malignant mammograms from the database. So total of 6156 mammograms are used. The partitioning the dataset is shown in Table 1.

Table 1. Partitioning Specification of Dataset

Data / Type	Normal	Benign	Malignant	Total
Train	1120	647	678	2445
Test	211	211	211	633
Total	1331	858	889	3078

2.2 Image Pre-Processing

The quality of the image is increased through image pre-processing. Background noise, tape artifacts, high-intensity rectangle labels, edge shadowing effects, and low-intensity labels in mammography images are the types of noise that have been noticed. Due to the complexity and variability of breast tissue, the mammographic image has low contrast and therefore the doctor only extracts a limited amount of information from the image. Misdiagnoses occur as a consequence of even skilled doctors being unable to identify hidden MC. We used several preprocessing methods in this work to smooth, brighten, denoise, and identify edges in breast pictures to enhance their look. Sharp edges were preserved while noise was reduced using adaptive median, Gaussian, and bilateral filtering. Artifact and pectoral muscle were eliminated by denoising. Tissue density in the non-breast region was substantially correlated, which might affect future mammography analyses.

To increase the image quality and smoothness, this research used image enhancement techniques based on wavelet analysis, CLAHE, and adaptive unsharp masking, as illustrated in Figure 2. The suggested method reduces superfluous background data, highlights the image's weak borders and calcification spots, and accentuates tiny calcification sites. In order to accurately assess the impact of image edge enhancement and to measure the image's denoising to confirm the efficacy of the procedure, this research employed the contrast improvement index (CII). Additionally, in order to assess the improved performance of mammography images, we assessed the CII and PSNR. A greater denoising effect is shown by higher PSNR and CII values

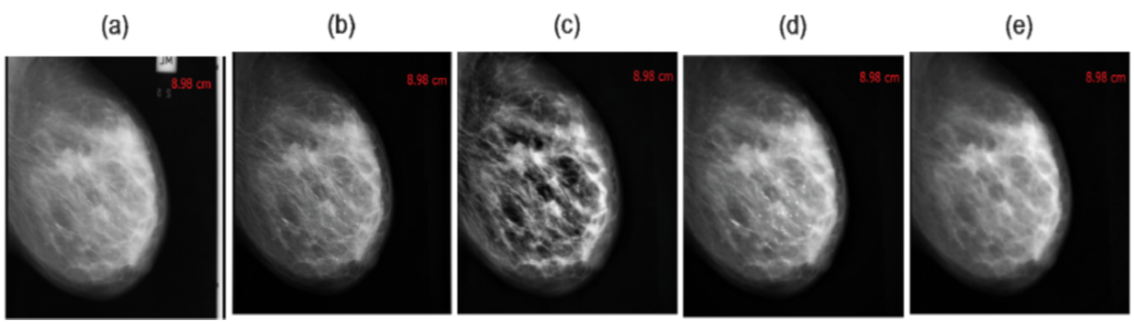

Fig 2. Pre-processing - (a) original Mammogram Image, (b) Adaptive Unsharp Masking, (c) Image Enhancement (CLAHE), (d) Dilation, (e) Erosion

2.3 Adaptive K Means Segmentation

Starting with a selection of k inputs from the provided database, the adaptive k means clustering method is run. The K randomly chosen elements are used to create the clusters. Each K element that makes up an element has a set of qualities that combine to generate the cluster properties. Figure 3 shows the flowchart of the adaptive K-Means segmentation procedures.

Based on the aforementioned algorithm, the distance between the provided element and the clusters is calculated. The distance should be taken into consideration depending on the qualities, which is a crucial factor that is also normalized. Consequently, none of the qualities either dominate or are neglected by the result. The most common use of the Euclidean distance is given by Equation (1).

$$Euclidean\ Distance = \sqrt{(E_{11}-E_{12})^2+(E_{12}-E_{22})^2+\cdots+(E_{1n}-E_{2n})^2} \tag{1}$$

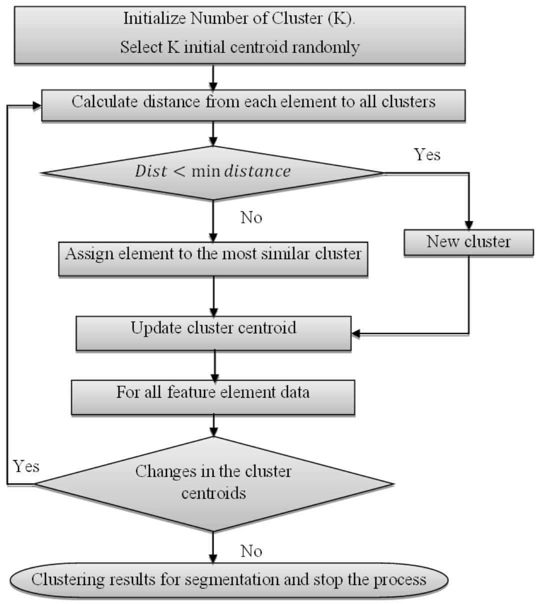

Fig 3. Flowchart of AKM Segmentation

For the purpose of dropping the square root function, the derived distance function must be adjusted. To compare properties in this method, several weights are needed for each property. It is determined and stored as a triangular matrix how far apart each cluster is from the others. For every element that isn't in a cluster, the distance is determined. The obtained k values are given in Gabor filter for the extraction of features from mammogram image.

2.4 Proposed Gabor Filter for Feature Extraction

Gabor filters have been used for image coding, image representation, texture segmentation, target recognition, edge detection, retina identification, and more. The Gabor family of filters has gained popularity in recent years because they can imitate the properties of some cells in the visual brain of animals. The primary visual cortex is thought to undertake identical orientation and Fourier space decomposition tasks, according to biological studies, making them appear reasonable for a technological vision system. Additionally, it has been shown that these 2D band-pass filters offer the best localization capabilities in the spatial and frequency domains, making them ideal for extracting image edges or features that are oriented in a particular frequency range.

As a sinusoidal plane modulated by a GE, a Gabor filter may be thought of as having a certain frequency and direction. Equation (2) is a possible representation.

$$GE(x,y) = e^{\frac{1}{2}\left[\frac{x_2}{\Sigma_x^2} + \frac{y_2}{\Sigma_y^2}\right]} e^{-j2\Phi(u_0x+v_0y)} \tag{2}$$

Two 2D Gaussian functions GF from Equation (3) make up the filter's response in the Fourier frequency domain.

$$GF(u,v) = GE_1 + GE_2 \tag{3}$$

where $\Sigma_u = \frac{1}{2\pi\Sigma_x}$ and $\Sigma_v = \frac{1}{2\pi\Sigma_y}$ assuming that the Fourier transform's origin has been centered and are the standard deviation along two orthogonal directions. To create the edge histogram descriptors that will serve as the classification features, a collection of edge histogram descriptors for each alarm segment is produced with its n counterparts present in n Gabor-filtered images. After clustering the EHD features with Adaptive k-means clustering method, a SKKNN classifier is used to reduce the number of false alarms[5]. The histogram that is produced by EHD indicates the local frequency of four distinct kinds of edges at each band. These corners have angles of 90º vertical, 0 degrees horizontal, 45º diagonal, and 135º diagonal. The vertical histogram frequency for a particular segment at each band is defined as the proportion of pixels in the vertical edge-extracted image with the highest intensity values in comparison to the pixel values in the other three directional edge-extracted images (horizontal, 45° diagonals, and 135° diagonal). Four directional frequencies may be combined to generate a four-dimensional EHD signature, and the remaining three directional frequencies can be calculated in a similar manner. The EHD characteristics are expressed as in **Equation 4**.

$$EHD\,(m,n) = \frac{No.\,of\ max\ intensity\ pixels\ at\ direction\ n}{Alarm\ segment\ area} \tag{4}$$

$$m = 1,\, 2,\, 3,\, 4,\, 5;\ n = 1,\, 2,\, 3,\, 4$$

The EHD computation is identical to counting the quantities of pixels with maximum intensity in each band's direction. For instance, if band 1's EHD signature has the biggest vertical frequency, band 1 will be dominated by vertical edges. Such a cancer segment EHD characteristic displays information on both directional edges and frequency scales from low to high frequencies. Regardless of the absolute intensity settings, the statistical properties of the EHD are consistent and dependable. Both the MLO and CC views of the mammography are used to extract features. So we can have two feature vectors.

The GLCM matrix is created by applying a Gabor filter on the cropped ROI after it has been trimmed. Co-occurrence matrix characteristics such as contrast, energy, entropy, mean, standard deviation, homogeneity, correlation, entropy, cluster shade, and cluster prominence are among the second order statistical features that may be derived from this information and computed with the use of characteristics mentioned in Equations (5), (6), (7), (8), (9), (10), (11) and (12).

$$Mean = \frac{1}{M*N}\sum_{i}^{M}\sum_{j}^{N} I\,[\,i\,,\,j\,] \tag{5}$$

$$Homogeneity = \sum_{i}^{M}\sum_{j}^{N}\frac{I\ [i,j]}{1+[i+j]^2} \tag{6}$$

$$Correlation = \sum_{i}^{M}\sum_{j}^{N}\frac{(i,\mu)(j,\mu)}{\sigma}\,(i,j] \tag{7}$$

$$Inverse\ Difference\ Moment = \sum_{i}^{M}\sum_{j}^{N}\frac{I[i,j]}{[i,j]^2}\ where\ i\ \neq j \tag{8}$$

$$Contrast = \sum_{i}^{M}\sum_{j}^{N}[i,j]^2 I^2\,(i,j] \tag{9}$$

$$STD = \sqrt{I\,(i,j] - I\,(\bar{i},j]^2} \tag{10}$$

$$Cluster\ Shade = \sum_{i}^{M}\sum_{j}^{N}(\,i\,M_x\ + j\,M_y)^3\, I\,(i,j] \tag{11}$$

$$Cluster\ Prominance = \sum_{i}^{M}\sum_{j}^{N}(i\,M_x + j\,M_y)^4\ I\,(i,j] \tag{12}$$

Where $M_x = \sum_{i}^{M}\sum_{j}^{N} i\,I[i,j]$ and $M_y = \sum_{i}^{M}\sum_{j}^{N} j\,I[i,j]$ μ - Mean σ – Variance of co-occurrence Matrix

The following explanations provide the previous attribute's physical significance.

Contrast: The intensity contrast between a pixel and its neighbour over the source image.

Correlation: Measures the degree to which pixels in the whole image are statistically associated with one another; range = [-1 1]. A fully positively or negatively correlated image has a correlation of either 1 or -1.

Energy: Squared element GLCM summation; range = [0 1]. For an unchanging image, energy equals 1.

Homogeneity: Range = [0 1] indicates how nearly the GLCM element distribution follows the GLCM diagonal. For a diagonal GLCM, homogeneity is 1.

Entropy: It is a randomness measure which is used to describe the texture of the input image.

Cluster shade and cluster prominence: Indicators of the skewness, or lack of symmetry, of the matrix.

After the features are determined for each view, Feature fusion is applied. To create a single feature vector that is more discriminative than either of the input feature vectors, two feature vectors are combined in a process known as feature fusion. Feature level fusion is done by DCAFUSE employing a Discriminant Correlation Analysis (DCA)-based methodology. It gathers the train and test data matrices and accompanying class labels from two modalities X and Y and merges them into a single feature set Z.

2.5 Proposed SKKNN for BC Detection

Although KNN is effective at classifying data, when the training set grows large, the computational cost often prevents its use in practical projects. This section introduces an effective and quick model based on the KNN baseline model for enhancing classification performance and improving training effectiveness. In this part, the kernel method is presented together with the KNN model. It makes use of the kernel method's characteristic to increase the dimension of features and improve classification performance in the KNN baseline model. The Shrunk Kernel technique is then introduced, which primarily aims to cut down on kernel computation.

The fundamental principle of the reduced kernel technique is to compute the kernel matrix using just a portion of the training data from each class. According to this research, the training samples are used to construct the kernel matrix. This kind of feature representation is equivalent to or superior to the traditional kernel matrix.

The following Equation (13) illustrates how the Shrunk kernel mathematical works.

$$\bar{SK}_{tr} = G\left(U_s, V, \sigma\right) \tag{13}$$

The low-dimensional characteristics are converted to high-dimensional features through the kernel approach. Use the Gaussian Kernel technique to process the characteristics in this research. The following Equation (14) displays its mathematical formula:

$$G\left(U_s, V, \sigma\right) = e^{\left(-\frac{\left(U-\mu_U\right)^2+\left(V-\mu_V\right)^2}{2\sigma^2}\right)} \tag{14}$$

where $\mu_U$and $\mu_v$are the average of the input data U and V, respectively, and σ denotes the user-defined kernel parameter. $\bar{SK}$is the reduced kernel matrix, s is the chosen kernel matrix computation percentage, and U_s denotes the chosen s input data sample units. Following Equations (15) and (16) may be used to generate the Shrunk kernel matrix of training and testing data.

$$\bar{SK}_{tr} = G\left(U_{tr_s}, U_{tr}, \sigma\right) \tag{15}$$

$$\bar{SK}_{te} = G\left(U_{te_s}, U_{te}, \sigma\right) \tag{16}$$

where U_{tr_s} indicates the training observation class percentages U_{tr}. Choosing a certain proportion of training samples, along with speeding up the calculation of the kernel matrix, the reduced kernel methodology maintains the high-dimensional features of the original kernel method. The pseudocode of SKKNN is depicted in algorithm 2.1.

Algorithm 1. Pseudocode of SKKNN

Input: Training matrix U_tr; Testing matrix U_te; Number of training data L; Parameter of SKKNN SK; Kernel parameter σ; The percentage of Shrunk kernel matrix s;

Step 1: Choose s percentage samples from the training data for each class as Xtr P;

Step 2: Using equation (SK) ̅tr, determine the reduced kernel matrix for the training features.

Step 3: Create the reduced kernel matrix using equation (SK) ̅te to test features.

Step 4: Loop: for i∈1,…,L do

Find the distance between training and testing samples
end for
Step 5: Sort the distance in the ascending order;
Step 6: Selecting the top K vectors from the sorted collection to serve as an index
Step 7: Set the forecasting the class label Y based on the most frequent class of processed index.
Step 8: end for return Y

3 Results and Discussion

The working platform of MATLAB was used to construct the suggested breast cancer detection system. The confusion matrix is used to estimate the proposed framework's performance indicators. For both perspectives, the performance metrics are calculated. Performance metrics include the following terms: TP (True Positive): The abnormal region is appropriately classified as abnormal by the classifiers. False Positive (FP): The classifiers misclassify a region that is normally occurring as abnormal. TN (True Negative): The normal zone is appropriately classified as normal by the classifiers. FN (False Negative): The aberrant region is misclassified as normal by the classifiers. To evaluate the learning models, the following metric's detail and mathematical expression are used:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

$$Sensitivity\ or\ Recall = \frac{TP}{TP+FN}$$

$$Specificity = \frac{TN}{TN+FP}$$

$$AUC = \frac{Sensitivity\ +\ Specificity}{2}$$

$$Precision = \frac{TP}{TP+FP}$$

$$Prediction\ Error\ = \frac{FP\ + FN}{FP\ +\ FN\ +\ TP\ +\ TN}$$

$$F1\ Score = 2\frac{Precision * Recall}{Precision + Recall}$$

$$FPR = \frac{FP}{FP+TN}$$

$$FNR = \frac{FN}{FN+TP}$$

$$Younden's\, index\ =\ Sensitivity - (1 - Specificity)$$

Using these performance metrics, the score of all proposed classifiershave been measured for single-view feature and two-view features, shown in Table 2. Table 2 ensures that the single-view features give the highest accuracy with 81.9%. However, two-view feature reaches 97.50% with 0.015% error. After observing these tables this research ensure that two-view feature fusion enhances some accuracy in terms of detection.

Table 2. Numerical Results of BC Detection

SKKNN	Accuracy (%)	Sensitivity	Specificity	AUC	Precision	Prediction Error	F1 Score	FPR	FNR	Youden's Index
CC View Features	81.8	71.2	81.09	89.9	76.7	12.0	84.2	18.9	17.6	179.8
MLO View Features	81.9	81.3	76.72	82.5	72.7	18.	88.9	23.2	11.6	165.1
Two View Feature Fusion	97.5	96.8	92.75	95.2	93.9	4.5	94.5	6.44	3.1	190.4

The accuracy, senstitivity , specificity, AUC, precision, F1-measure for SKKNN was 97.56%, 96.8%, 92.75%, 95.2%, 93.93%, and 94.5% which are higher comparing single view features. The FPR, FNR and prediction error for SKKNN was 6.44, 3.15 and 4.5 which is lower comparing single view features. The Youden's Index for SKKNN was 190.41 which is higher comparing singe view features. In spite of the fact that the execution of the fundamental show is marginally distinctive on size of information, the proposed system has moved forward the execution of the essential show.

The comparison graph of accuracy, sensitivity, specificity, AUC, precision and f1-score for proposed model SKKNN with Linear Binary Pattern and SVM is appeared in Figure 4.

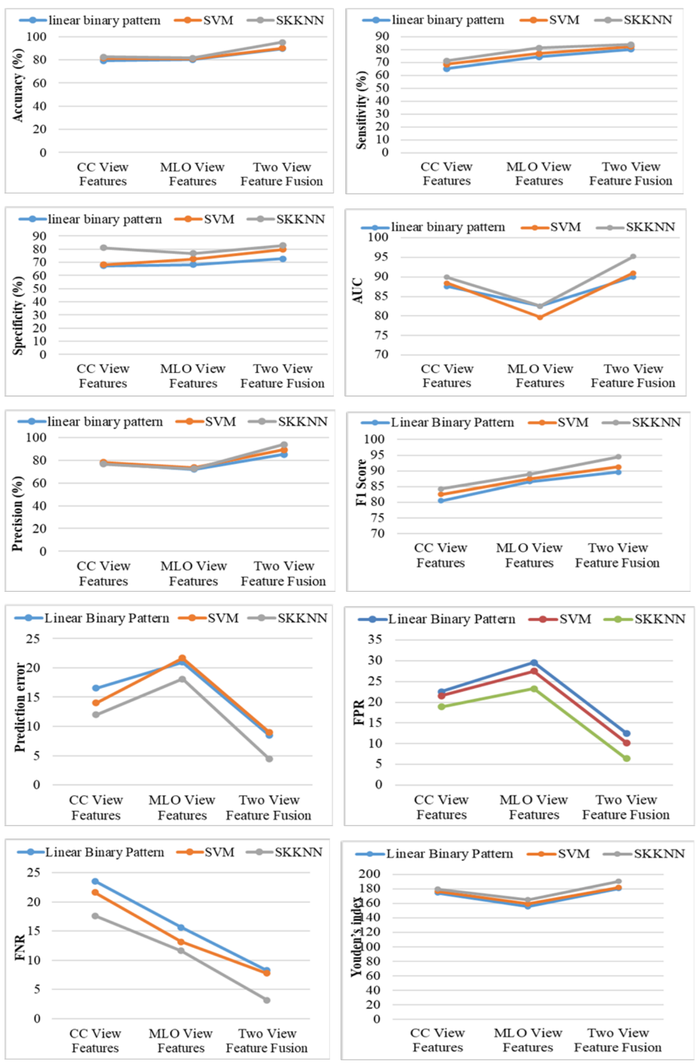

Fig 4. Chart Comparison of Linear Binary Pattern, SVM and SKKNN

3.1 Comparison with other works

However, a number of academics have recently focused on breast cancer and have put forth several approaches and methodologies that demonstrate the varied outcomes of this study. Table 3 presents a comparison analysis demonstrating that our proposed methodology has a greater level of accuracy than any other research work stated in this table. This study solely took into account the DDSM dataset that was comparable to it and the machine learning techniques that other scientists had used. Our suggested model SKKNN obtained substantial outcomes in terms of feature optimization and the usage of important characteristics, even though some of the studies showed slightly greater accuracy than ours.

Table 3. Comparison of our work with the most related works

Author	Method	Accuracy
Ibrahim et al. 2020 [9]	RF	96.59
Arooj et al. 2022 [14]	XGboost	97.1
Diaz et al. 2021 [21]	KNN	94.35
Vijayarajeswari et al., 2019 [10]	SVM	96.72
Alshammari et al. 2021 [18]	MLP	97.9
Proposed	SKKNN	97.5

4 Conclusion

In comparison to single view, using a double view yields superior results. Breast cancer detection in this chapter is done using MLO and CC view mammography images. The segmentation of regions is done using the Adaptive K-means approach. In this study, the KMC-GF feature extraction approach is used. The benefit of the KMC-GF approach is that it gives shape and orientation characteristics while also extracting fine texture features from the clustered area. KMC-GF characteristics provide a good result for breast cancer early diagnosis when compared to GLCM features. The view of mammography images is then classified as normal or malignant using the SKKNN classifier. As a result, this technique could be employed in the medical field to diagnose breast cancer and also produce few false positive results. The main benefit of this strategy is that it lightens the burden for radiologists. Mammography screening lowers mortality and aids in early detection, although it is not a perfect screening method. The suggested can be applied in the future to imaging modalities like ultrasound and MRI that are used to identify breast cancer. Modern transfer learning architectures can also be used to identify and categorize breast cancer as benign or malignant.

References

1) Lei S, Zheng R, Zhang S, Wang S, Chen R, Sun K, et al. Global patterns of breast cancer incidence and mortality: A population-based cancer registry data analysis from 2000 to 2020. *Cancer Communications*. 2021;41(11):1183–1194. Available from: https://doi.org/10.1002/cac2.12207.
2) Jebarani PE, Umadevi N, Dang H, Pomplun M. A Novel Hybrid K-Means and GMM Machine Learning Model for Breast Cancer Detection. *IEEE Access*. 2021;9:146153–146162. Available from: https://doi.org/10.1109/ACCESS.2021.3123425.
3) Wang Z, Li M, Wang H, Jiang H, Yao Y, Zhang H, et al. Breast cancer detection using extreme learning machine based on feature fusion with CNN deep features. *IEEE Access*. 2019;7:105146–105158. Available from: https://doi.org/10.1109/ACCESS.2019.2892795.
4) Li Y, Chen H, Zhang L, Cheng L. Mammographic mass detection based on convolution neural network. In: 24th International Conference on Pattern Recognition (ICPR). IEEE. 2018;p. 3850–3855. Available from: https://doi.org/10.1109/ICPR.2018.8545557.
5) Li H, Chen D, Nailon WH, Davies ME, Laurensonv D. A deep dual-path network for improved mammogram image processing. In: IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE. 2019;p. 1224–1228. Available from: https://doi.org/10.1109/ICASSP.2019.8682496.
6) Kulkarni P. Fine Grained Classification of Mammographic Lesions using Pixel N-grams. *Asian Journal For Convergence In Technology (AJCT) 2019*. 2019;5(1):1–5. Available from: https://asianssr.org/index.php/ajct/article/view/734.
7) Yang Z, Cao Z, Zhang Y, Tang Y, Lin X, Ouyang R, et al. MommiNet-v2: Mammographic multi-view mass identification networks. *Medical Image Analysis*. 2021;73:1–11. Available from: https://doi.org/10.1016/j.media.2021.102204.
8) Sridevi V, Samath JA. A combined deep CNN-lasso regression feature fusion and classification of MLO and CC view mammogram image. *International Journal of System Assurance Engineering and Management*. 2024;15:553–563. Available from: https://doi.org/10.1007/s13198-023-01871-x.
9) Ibrahim AO, Ahmed A, Abdu A, Abd-Alaziz R, Alobeed MA, Saleh AY, et al. Classification of mammogram images using radial basis function neural network. In: Emerging Trends in Intelligent Computing and Informatics - Data Science, Intelligent Information Systems and Smart Computing;vol. 1073 of Advances in Intelligent Systems and Computing. Springer. ;p. 311–320. Available from: https://doi.org/10.1007/978-3-030-33582-3_30.
10) Vijayarajeswari R, Parthasarathy P, Vivekanandan S, Basha AA. Classification of mammogram for early detection of breast cancer using SVM classifier and Hough transform. *Measurement*. 2019;146:800–805. Available from: https://doi.org/10.1016/j.measurement.2019.05.083.
11) Sridevi V, Samath JA. Design and implementation of transfer learned deep CNN with feature fusion for automated mammogram classification. *International Journal of Health Sciences*. 2022;6(S6):3033–3047. Available from: https://doi.org/10.53730/ijhs.v6nS6.10267.

12) Kaur P, Singh G, Kaur P. Intellectual detection and validation of automated mammogram breast cancer images by multi-class SVM using deep learning classification. *Informatics in Medicine Unlocked*. 2019;16:1–15. Available from: https://doi.org/10.1016/j.imu.2019.100239.
13) Kayode AA, Akande NO, Adegun AA, Adebiy MO. An automated mammogram classification system using modified support vector machine. *Medical Devices: Evidence and Research* . 2019;12:275–284. Available from: https://doi.org/10.2147/MDER.S206973.
14) Arooj S, Atta-Ur-Rahman, Zubair M, Khan MF, Alissa K, Khan MA, et al. Breast Cancer Detection and Classification Empowered With Transfer Learning. *Frontiers in Public Health*. 2022;10:1–18. Available from: https://doi.org/10.3389/fpubh.2022.924432.
15) Sridevi V, Samath JA. MLO and CC view of feature fusion and mammogram classification using deep convolution neural network. *International Journal of Health Sciences*. 2022;6(S7):5196–5207. Available from: https://doi.org/10.53730/ijhs.v6nS7.13106.
16) Dewangan KK, Dewangan DK, Sahu SP, Janghel R. Breast cancer diagnosis in an early stage using novel deep learning with hybrid optimization technique. *Multimedia Tools and Applications*. 2022;81:13935–13960. Available from: https://doi.org/10.1007/s11042-022-12385-2.
17) Alruwaili M, Gouda W. Automated Breast Cancer Detection Models Based on Transfer learning. *Sensors*. 2022;22(3):1–16. Available from: https://doi.org/10.3390/s22030876.
18) Alshammari MM, Almuhanna A, Alhiyafi J. Mammography Image-Based Diagnosis of Breast Cancer Using Machine Learning: A Pilot Study. *Sensors* . 2022;22(1):1–20. Available from: https://doi.org/10.3390/s22010203.
19) Jayandhi G, Jasmine JL, Seetharaman R, Joans SM. Breast Cancer Prediction Based on Mammographic data by Hybrid Resnet and Decision Tree. In: International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), 2022. IEEE. IEEE. 2022;p. 234–238. Available from: https://doi.org/10.1109/ICSCDS53736.2022.9760754.
20) Sridevi V, Samath JA. Advancement on Breast Cancer Detection Using Medio-Lateral-Oblique (MLO) and Cranio-Caudal (CC) Features. 2020. Available from: https://ssrn.com/abstract=3664619.
21) Diaz RAN, Swandewi NNT, Novianti KDP. Malignancy Determination Breast Cancer Based on Mammogram Image With K-Nearest Neighbor. In: 1st International Conference on Cybernetics and Intelligent System (ICORIS). IEEE. 2019. Available from: https://doi.org/10.1109/ICORIS.2019.8874873.
22) Sahu B, Panigrahi A. Efficient role of Machine learning classifiers in prediction and detection of breast cancer. In: 5th International Conference on Next Generation Computing Technologies (NGCT-2019). 2020;p. 1–9. Available from: https://dx.doi.org/10.2139/ssrn.3545096.
23) Mohapatra S, Muduly S, Mohanty S, Ravindra JVR, Mohanty SN. Evaluation of Deep Learning Models for Detecting Breast Cancer using Mammograms. *Sustainable Operations and Computers*. 2022;3:296–302. Available from: https://doi.org/10.1016/j.susoc.2022.06.001.
24) Sahu B, Mohanty SN, Rout SK. A Hybrid Approach for Breast Cancer Classification and Diagnosis. *EAI Endorsed Transactions on Scalable Information Systems*. 2019;6(20):1–8. Available from: https://dx.doi.org/10.4108/eai.19-12-2018.156086.
25) Guan Q, Huang Y, Zhong Z, Zheng Z, Zheng L, Yang Y. Thorax disease classification with attention guided convolutional neural network. *Pattern Recognition Letters*. 2020;131:38–45. Available from: https://dx.doi.org/10.1016/j.patrec.2019.11.040.