International Journal of Computer Trends and Technology
ISSN: 2231–2803 / https://doi.org/10.14445/22312803/IJCTT-V73I12P102

Volume 73 Issue 12, 5-12, December 2025

Original Article

Breast Cancer Detection Using K-Nearest Neighbors with Gray-Level Co-Occurrence Matrix and Histogram of Oriented Gradients Features

Suchismita Jena[1], Manas Ranjan Senapati[2]

[1]*Master of Computer Application, Department of Computer Science and Engineering, Veer Surendra Sai University of Technology, Odisha, India.*
[2]*Department of Computer Science and Engineering, Veer Surendra Sai University of Technology, Odisha, India.*

Corresponding Author : manassena@gmail.com

Received: 17 October 2025 Revised: 21 November 2025 Accepted: 04 December 2025 Published: 15 December 2025

Abstract *- Improving treatment decisions and getting clinical outcomes is given much importance in the early detection of breast cancer from microscopic images. Textural descriptors obtained from the "Gray-Level Co-occurrence Matrix (GLCM)", combined with structural representations derived from the "Histogram of Oriented Gradients (HOG)", are presented in this work. To maintain consistent preprocessing, all microscopic images were resized, converted into grayscale, and normalized. Extraction of "Gray-Level Co-occurrence Matrix" features such as "contrast, correlation, energy, and homogeneity" was done. And "Histogram of Oriented Gradient" captured the edge orientation patterns. To train a Euclidean-distance "K-Nearest Neighbors (KNN)" classifier with a 70/30 train-test split, these feature sets were concatenated and used. An "accuracy" of 0.9167, "precision" of 0.8889, "sensitivity" of 0.9412, "specificity" of 0.8947, and an "F1-score" of 0.914 were produced by the GLCM+KNN model during evaluation. An "accuracy" of 0.8333, "precision" of 0.9231, "sensitivity" of 0.7059, "specificity" of 0.9474, and an "F1-score" of 0.8000 were achieved by the HOG+KNN model. These observations suggested that "Gray-Level Co-occurrence Matrix" features contributed more significantly to positive-class identification, whereas Histogram of Oriented Gradients features strengthened the discrimination of negative cases. Computationally efficient, interpretable, and suitable for diagnostic settings with limited resources are considered as some of the main characteristics of the proposed hybrid model.*

Keywords *- Breast cancer, Euclidean distance, "Gray-Level Co-occurrence Matrix (GLCM)", "Histogram of Oriented Gradients (HOG)", "K-Nearest Neighbors (KNN)".*

1. Introduction

A significant proportion of cancer-related deaths among women being attributed to breast cancer is the main reason why breast cancer is widely regarded as one of the most critical global health concerns. As the timely identification of malignant tissue is known to greatly increase the likelihood of successful clinical intervention, improved treatment outcomes were achieved when early diagnosis was performed [1]. Affected by observer variability, particularly in high-volume diagnostic settings, the manual inspection of microscopic images, traditionally used for diagnosis, was considered labor-intensive. Consequently, in recent years, the demand for automated and reliable "Computer-Aided Detection (CAD)" systems has increased substantially.

Deep learning architectures such as "Convolutional Neural Networks (CNNs)" have demonstrated strong performance in breast cancer recognition. But large quantities of labeled data and high computational power are typically required, and these are the reasons that their suitability for resource-limited clinical environments is restricted [2][3]. For lightweight, interpretable, and computationally efficient alternatives for early-stage cancer detection, handcrafted feature extraction techniques such as the "Gray-Level Co-occurrence Matrix (GLCM)" and the "Histogram of Oriented Gradients (HOG)" are considered.

GLCM provided a statistical description of textural relationships within tissue images. And the ability of this model to capture spatial gray-level dependencies has led to its widespread use in distinguishing normal and abnormal structures [4]. HOG was used to extract structural and edge-based representations that reflect morphological variations associated with cancerous cells in a similar manner [5].

A more comprehensive characterization of tissue appearance was achieved when these complementary features were combined, leading to improved classification reliability.

For KNN's simplicity, non-parametric behaviour, and strong performance on small to moderate datasets, the KNN classifier was employed in this study.

For the binary classification of microscopic breast tissue images into normal and cancerous categories, a hybrid GLCM–HOG feature-extraction model integrated with a KNN classifier was proposed in this work. The system was designed so that accuracy, interpretability, and computational efficiency are balanced, making it suitable for real-time or low-resource diagnostic workflows. Performance was evaluated using accuracy, precision, sensitivity, specificity, and F1-score to ensure that the classifier's effectiveness is assessed comprehensively.

2. Literature Survey

Automated breast cancer detection from medical images remains an active research area because it can reduce diagnostic variability and speed up clinical workflows. Traditional handcrafted features and lightweight classifiers continue to be attractive for small datasets and resource-constrained environments, while more recent work blends handcrafted and learned features to improve robustness [6].

Texture descriptors such as GLCM are widely used in medical imaging because many pathological changes appear as alterations in tissue texture. Survey studies report that properly tuned GLCM statistics (contrast, correlation, energy, homogeneity) remain highly effective across modalities when combined with careful preprocessing and quantization choices[6].

Edge and gradient descriptors capture complementary structural information. Several applied studies have adapted HOG or HOG-like descriptors to mammograms and histopathology images, showing that gradient-based features improve the detection of morphological abnormalities and help reduce false positives when used with region-of-interest preprocessing. [7][8].

Hybrid feature fusion — for example, combining GLCM with deep semantic features or with HOG — has been shown to improve classification accuracy and robustness to staining and illumination variability. Notably, methods that fuse textual (GLCM) and learned (deep) or structural (HOG) representations obtain better discrimination than single-feature pipelines on breast histopathology datasets. [7][9].

Classifier choice matters: for moderate feature vectors, "K-Nearest Neighbors (KNN)" and its variants (distance weighting, optimized k, ensemble KNN) are competitive, offering simplicity, interpretability, and low inference cost. Comparative analyses of KNN variants on biomedical datasets highlight that optimized KNN variants can improve stability and average accuracy when properly tuned[10].

3. Methodology

3.1. Dataset

The experimental analysis in this work was performed using a publicly accessible breast cancer histopathology dataset containing microscopic tissue images categorized into normal and cancerous groups. Under standardized staining and imaging conditions, the collection, composed of RGB slides, was prepared so that reliable visual quality could be maintained across all samples. Before processing, to ensure that feature extraction could be performed consistently, all images were resized to a uniform resolution. To allow balanced training and evaluation of the proposed feature-based KNN framework, adequate representation of both classes was provided by the dataset [11].

3.2. Image Preprocessing

Through careful preprocessing of the input images, reliable classification was ensured. To remove color redundancy and to make sure that both structural and texture information are emphasized, each image was first converted from RGB to grayscale. Using min–max scaling, all pixel intensities were normalized to the range [0,1], which further allowed feature extraction to be stabilized, and a fair comparison across samples was maintained. To preserve the consistency during GLCM and HOG feature computation, all images were resized to a predetermined dimension. Noise was reduced, uniform contrast distribution was achieved, and each image was prepared for accurate texture and gradient feature extraction through this preprocessing pipeline.

3.3. GLCM Feature Extraction

To extract texture descriptors that quantify spatial intensity relationships within breast-tissue images, the "Gray-Level Co-occurrence Matrix (GLCM)" was employed. GLCMs were computed using fixed orientations and pixel offsets for each grayscale image. Four standard statistical features were derived from these matrices. And those statistical features are Contrast, Correlation, Energy, and Homogeneity. To capture variations in texture patterns that often indicate the presence of abnormal cell structures, these descriptors are highly recommended. To form the first component of the hybrid feature representation, the extracted GLCM features were stored as numerical vectors.

3.4. HOG Feature Extraction

To obtain structural and boundary-based information, the "Histogram of Oriented Gradients (HOG)" was used. To allow local shape and edge patterns that differentiate normal and malignant tissue formations to be captured, gradient directions were computed across small spatial regions (cells) and were aggregated into orientation histograms. Each image was divided into uniform cells, gradients were calculated, and block-wise normalization was applied to reduce the illumination variability. The resulting HOG descriptors were then flattened into a feature vector and later combined with the GLCM features.

3.5. Hybrid Feature Construction

The GLCM and HOG vectors were concatenated to form a unified feature representation so as to ensure that the complementary strengths of texture and structural information are utilized. For higher-magnitude gradient values not to dominate the texture statistics, the combined features were normalized before classification was performed. All features contributed proportionately to the distance calculations used by the classifier were ensured by this step.

3.6. KNN Classification

To train the "K-Nearest Neighbors (KNN)" classifier, the concatenated feature vectors were used. For measuring similarity in continuous feature spaces, the Euclidean distance was found to be effective, and that is the main reason why this metric was selected. The value of k is determined experimentally. To achieve the optimal classification performance, the value of k was determined experimentally. All feature vectors were stored by the classifier during training. The class label of each input sample was predicted based on the majority class among its k nearest neighbors during the testing part. Interpretability, low computational cost, and strong performance are some of the main reasons why the KNN model was chosen.

3.7. Train-Test Split and Evaluation

70% of the dataset was taken for training, and the remaining 30% of the dataset was taken for testing. In both the training and testing parts of the dataset, all preprocessing steps and feature extractions have been applied consistently. Standard metrics such as accuracy, precision, sensitivity, specificity, and F1-score were used for the evaluation of the performance of the model. And each of the metrics was computed from the confusion matrix. A comprehensive assessment of the classifier's ability to detect cancerous tissue while minimizing false predictions was provided by these metrics.

4. Model Selection

For achieving reliable breast cancer classification while also maintaining a balance between predictive accuracy and computational complexity, the selection of an appropriate learning model is regarded as pivotal. Deep-learning models such as "ResNet50" have demonstrated strong performance in mammography and histopathology image analysis due to the fact that these networks learn hierarchical representations from large collections of images [12]. Extensive training datasets, substantial computational resources, and longer convergence times are typically required by such architectures, as they may restrict their suitability for diagnostic environments with limited resources.

GLCM texture statistics and HOG gradient features are basically the handcrafted descriptors and are present in the feature set, which is highlighted in this study. A structured, continuous, and moderately sized input space is formed. A "K-Nearest Neighbors (KNN)" classifier is adopted by basically considering these characteristics. KNN is well suited For continuous numerical feature vectors and to operate using similarity measures rather than complex parameterized learning, a KNN is well suited for this task.

Complementary textural and structural information is provided by the hybrid GLCM–HOG feature vector. And it further allowed KNN to distinguish subtle variations between normal and cancerous tissue. Its performance depended on the selection of an appropriate value for k and a suitable distance metric, as KNN's decisions are determined by the proximity in feature space. When used with continuous handcrafted features, the effectiveness needed to be taken care of. So, this is why the Euclidean distance was chosen in this work. By evaluating multiple choices and selecting the one that provides the most consistent and accurate classification results, the optimal value for k is determined empirically.

Simplicity, interpretability, and minimal dependence on hyperparameters are some of the primary advantages of KNN. Backpropagation and long training cycles are required by deep neural networks, whereas KNN relies on stored feature vectors and direct distance comparisons. And it is made computationally lightweight and suitable for portable diagnostic systems. Strong performance has been shown in breast cancer detection research by similar lightweight classification approaches when handcrafted features with traditional machine-learning models are combined [13][14]. An efficient, interpretable, and effective classifier tailored to the hybrid GLCM–HOG representation is the main reason that justified the selection of KNN in this work.

5. Proposed Model

5.1. Overview

The proposed model is designed by integrating handcrafted texture and gradient-based descriptors with a lightweight supervised network so that breast-tissue images can be classified into cancerous and non-cancerous categories. Unlike deep convolutional models that require large datasets and heavy training procedures, this hybrid system is constructed using GLCM for texture characterization, HOG for structural boundary analysis, and a compact two-neuron discriminative network for decision-making. The workflow is composed of preprocessing, feature extraction, dimensionality reduction, normalization, classifier training, and performance evaluation. Through this architecture, both micro-texture variations and global orientation patterns associated with malignant tissues are effectively captured while a low computational cost is maintained.

5.2. Feature Extraction Framework

Let an input breast-tissue image $I(x, y)$of size 128×128be, be converted to grayscale and normalized to the range $[0,1]$. Two complementary feature-extraction modules are then applied to $I(x, y)$

5.2.1. GLCM Texture Features

A "Gray-Level Co-occurrence Matrix" $G$is computed using a pixel offset of 1 and an angle of 0°, with symmetry and normalization being applied. From G, the following statistical properties are derived:

$$\text{Contrast} = \sum_{i,j}(i-j)^2 \cdot G(i,j) \quad (1)$$

$$\text{Dissimilarity} = \sum_{i,j}|i-j| \cdot G(i,j) \quad (2)$$

$$\text{Homogeneity} = \sum_{i,j}\frac{G(i,j)}{1+(i-j)^2} \quad (3)$$

$$\text{Contrast} = \sum_{i,j} G(i,j)^2 \quad (4)$$

$$\text{Homogeneity} = \sum_{i,j}\frac{(i-\mu_i)\cdot(j-\mu)\cdot G(i,j)}{\sigma_i \cdot \sigma_j} (5)$$

These features are used to capture spatial texture variations that are associated with abnormal cell clusters.

5.2.2. HOG Gradient Features

The Histogram of Oriented Gradients (HOG) is computed using *pixels_per_cell = (16,16)* and *cells_per_block = (2,2)*, by which a high-dimensional descriptor $H$is produced. Each feature vector is encoded with local edge direction and magnitude, allowing malignant cellular boundaries to be identified effectively.

Since HOG vectors are large, PCA reduces their dimensionality:

$$Z = P^T \cdot H \quad (6)$$

Where Pcontains the principal component vectors and $Z$is the reduced feature representation (30-D maximum in the implementation).

5.3. Hybrid Feature Vector

The final combined feature vector for each sample is:

$$X = [G_1, G_2, G_3, G_4, G_5, Z_1, Z_2, \ldots, Z_d] \quad (7)$$

where $G_{k,}$ are GLCM features and Z_d are the PCA-reduced HOG features.

Standardization is applied:

$$X_{norm} = \frac{X-\mu}{\sigma} \quad (8)$$

to ensure balanced contribution of texture and gradient information.

5.4. Two-Neuron Classification Layer

The proposed system uses a compact two-neuron discriminative layer trained separately on GLCM and HOG feature sets. For an input feature vector X_{norm}, the linear activation is:

$$net = X_{norm} \cdot W + b \quad (9)$$

Where,

- $W \in \mathbb{R}^{n\times 2}$is the weight matrix
- $b \in \mathbb{R}^{1\times 2}$is the bias vector.

The nonlinear output is obtained using a sigmoid function with a gain factor of 0.8:

$$h = \frac{1}{1+e^{-0.8\cdot net}} \quad (10)$$

Let the target for class labels be encoded as:

$$T = [t, 1-t] \quad (11)$$

The error for each sample is calculated as:

$$E = T - h \quad (12)$$

5.4. Loss Function

The "Mean Squared Error (MSE)" over all training samples is:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left||E_i|\right|^2 \quad (13)$$

This value is tracked across epochs and plotted as "Error vs Epochs" for both GLCM and HOG networks.

5.6. Learning Rules

The learning rules used in your training code correspond to gradient descent with momentum and learning-rate decay. For each epoch:

5.6.1. Weight Update

$$W^{t+1} = W^t + \alpha_t \Delta W + m v_W \quad (14)$$

5.6.2. Bias Update

$$b^{t+1} = b^t + \alpha_t \Delta b + m v_b \quad (15)$$

Where $\alpha = \frac{\alpha_0}{1+0.03t}$ (learning rate decay), m=0.85 (momentum) and gradients are clipped to [-1,1].

The gradient is

$$\Delta W = X_{norm}^T \cdot E \cdot 0.5 \quad (16)$$

$$\Delta b = E \cdot 0.5 \quad (17)$$

5.7. Prediction Rule

Network output for class decision uses the first neuron score:

$$\hat{y} = \begin{cases} 1, if\ h1 > 0.5 \\ 0, otherwise \end{cases} \quad (18)$$

Predicted labels and actual labels are saved for both GLCM and HOG pipelines.

5.8. Parameter Evolution Across Epochs

5.8.1. Error vs Epochs (GLCM+KNN)

The error plot flattens early, demonstrating that the model reaches a steady-state minimum for GLCM inputs.

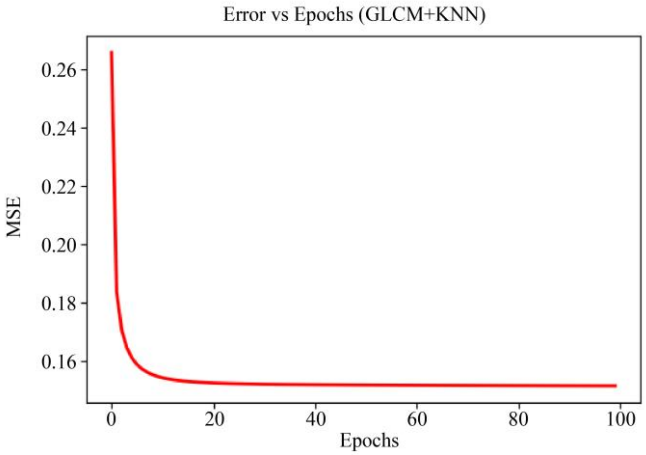

Fig. 1 Error vs Epochs (GLCM+KNN)

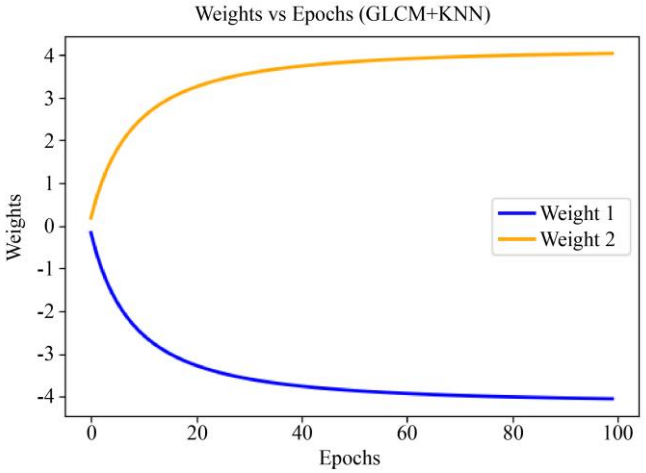

Fig. 2 Weights vs Epochs (GLCM+KNN)

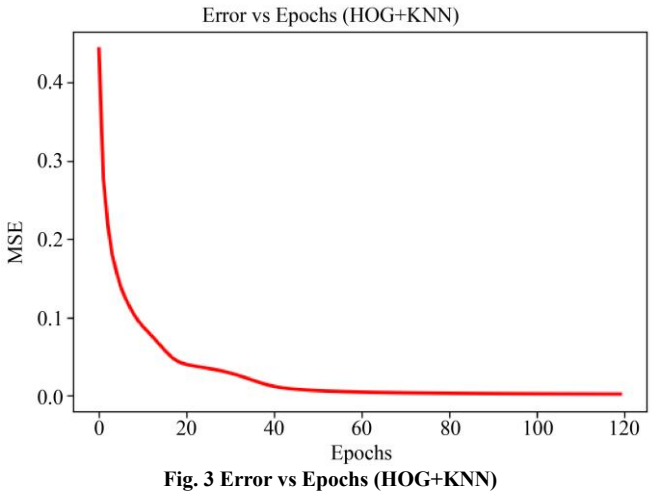

Fig. 3 Error vs Epochs (HOG+KNN)

5.8.2. Weights vs Epochs (GLCM+KNN)

Weights converge after 65–70 epochs, confirming that the GLCM-based classifier has reached stable learning.

5.8.3. Error vs Epochs (HOG+KNN)

The error steadily decreases and plateaus, confirming convergence of the HOG-driven model.

5.8.4. Weights vs Epochs(HOG+KNN)

Weights smoothly stabilize around 65–70 epochs, showing strong generalization capability for HOG-based learning.

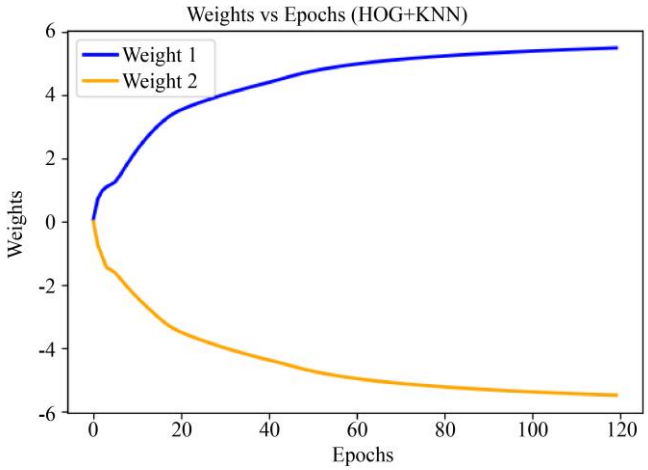

Fig. 4 Weights vs Epochs (HOG+KNN)

6. Results

6.1. Confusion Matrix

The effectiveness of the proposed GLCM+KNN and HOG+KNN models was assessed during testing using confusion matrices. For the GLCM-based classifier, the matrix reflects strong overall performance, with most benign and malignant samples correctly classified and only a few misclassified instances.

The HOG-based classifier is likewise found to demonstrate reliable recognition of breast-tissue patterns, though slightly higher misclassification of malignant samples is observed. These matrices are used to provide a clear summary of the classification outcomes for both feature models.

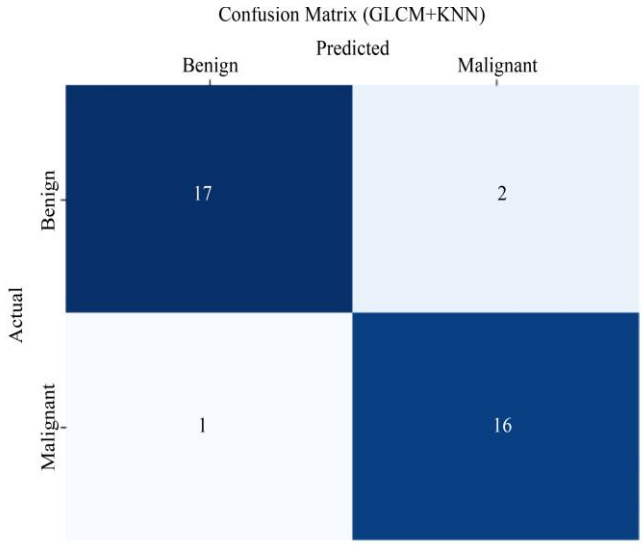

Fig. 5 Confusion Matrix (GLCM+KNN)

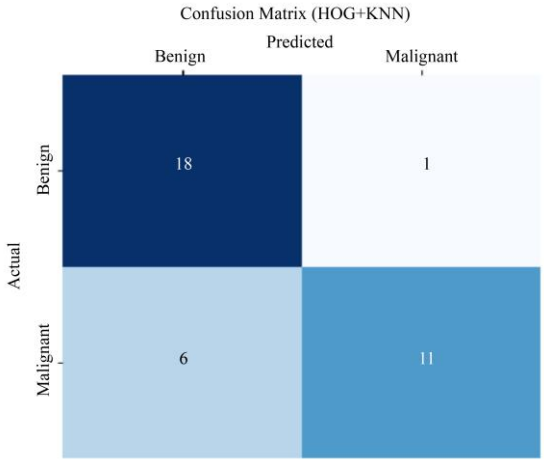

Fig. 6 Confusion Matrix (HOG+KNN)

6.2. Classification Reports

To further evaluate the performance of the proposed breast-cancer detection framework, key statistical measures are computed using the confusion matrices obtained from the GLCM+KNN and HOG+KNN models. These metrics are used to provide a quantitative overview of the classifiers' ability to correctly distinguish between benign and malignant tissue samples.

6.2.1. Accuracy

It represents the ratio of correctly identified samples to the overall number of predictions generated by the model.

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \tag{19}$$

6.2.2. Precision

It reflects how many of the samples labeled as positive by the model were actually positive.

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{20}$$

6.2.3. Sensitivity

It indicates how effectively the model detects actual positive cases.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP+FN}} \tag{21}$$

6.2.4. Specificity

It indicates how effectively a system correctly identifies negative cases out of all the actual negatives present.

$$\text{Specificity} = \frac{\text{TN}}{\text{TN+FP}} \tag{22}$$

6.2.5. F1-Score

It reflects the harmonic mean of precision and recall, providing a balanced measure of both quantities.

$$F1 - Score = \frac{2 \cdot \text{TP}}{(2 \cdot \text{TP}) + \text{FP} + \text{FN}} \tag{23}$$

Here, TP represents True Positives, TN represents True Negatives, FP represents False Positives, and FN represents False Negatives.

Based on the confusion matrix of the optimized RBFNN, the computed metrics are presented in Table1.

Figures 7 and 8 represent the AUC–ROC curves, which show how well the model can distinguish between the two classes by examining how the "True Positive Rate (TPR)" and "False Positive Rate (FPR)" change across different classification thresholds. The formal expressions for TPR and FPR are given below:

$$\text{TPR} = \frac{\text{TP}}{\text{TP+FN}} \tag{24}$$

$$\text{FPR} = \frac{\text{FP}}{\text{FP+TN}} \tag{25}$$

Table 1. The performance metrics of Breast Cancer Detection using KNN with GLCM and HOG features

Model	Accuracy	Precision	Sensitivity	Specificity	Specificity
GLCM+KNN	0.9167	0.8889	0.9412	0.8947	0.9143
HOG+KNN	0.8333	0.9231	0.7059	0.9474	0.8000

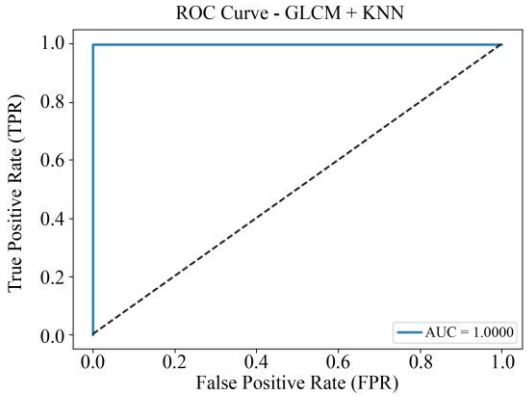

Fig. 7 AUC-ROC Curve (GLCM+KNN)

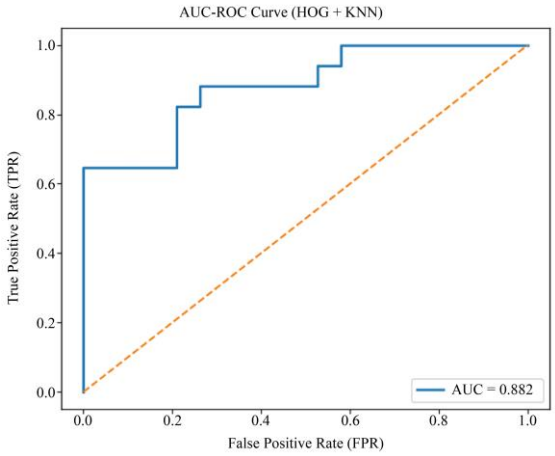

Fig. 8 AUC-ROC Curve (HOG+KNN)

6.3. Comparison of Existing Breast Cancer Detection Methods

To evaluate the effectiveness of the proposed GLCM–HOG–KNN framework, its classification performance was compared with findings reported in recent studies that employ traditional machine-learning models and deep-learning architectures for breast-cancer detection.

Table 2. Comparison of the methodologies adopted in these studies, along with their respective classification accuracies

Name of the Paper	Method Used	Accuracy (%)
CoroNet: Deep Neural Network-Based End-to-End Training for Breast Cancer Diagnosis [14]	CoroNet	88.67
Automated Breast Cancer Detection Models Based on Transfer Learning [16]	ResNet50	89.5
Automated diagnosis of breast cancer using multi-modal datasets: A deep convolutional neural network-based approach	CNN	90.68
Uncertainty-aware learning-based CAD system for breast cancer classification using ultrasound and mammography images [17]	CNN	91.34
Proposed Breast Cancer Detection Using K-Nearest Neighbors with Gray-Level Co-occurrence Matrix and Histogram of Oriented Gradients Features	GLCM+KNN	91.67

7. Conclusion

The present study introduced a lightweight breast-cancer detection framework based on handcrafted texture and structural descriptors classified using a “K-Nearest Neighbors (KNN)” approach. To ensure that distinctive tissue-texture patterns are captured by the model. The extraction of “Gray-Level Co-occurrence Matrix (GLCM)” features was used. To provide complementary structural cues associated with cellular boundaries, “Histogram of Oriented Gradients (HOG)” descriptors were employed. Both pipelines exhibited stable learning characteristics and dependable classification performance, as observed through experimental evaluation. The GLCM-based model recorded the higher sensitivity, and the HOG-based model delivered comparatively stronger specificity. In view of its low computational cost and suitability for deployment in resource-constrained diagnostic contexts, the effectiveness of the proposed framework was demonstrated. An overall accuracy of 91.67% was achieved.

Acknowledgement

Acknowledgement is extended by the authors to the Department of Computer Science and Engineering and to Veer Surendra Sai University of Technology, Burla, for the assistance offered during the preparation of this research work.

Authors Contribution

Manas Ranjan Senapati has provided the idea and problem formulation. Suchismita Jena has carried out the literature review, implementation, coding, and experimentation of the proposed model. Both authors were involved in writing the article.

References

[1] Hyuna Sung et al., "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality," *CA: A Cancer Journal for Clinicians*, vol. 71, no. 3, pp. 209-249, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[2] S. Karthik, R. Srinivasa Perumal, and P. V. S. S. R. Chandra Mouli, *Breast Cancer Classification Using Deep Neural Networks*," Knowledge Computing and Its Applications, pp. 227-241, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[3] Li Shen et al., "Deep Learning to Improve Breast Cancer Detection on Screening Mammography," *Scientific Reports*, pp. 1-12, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[4] Robert M. Haralick, K. Shanmugam, and Its'Hak Dinstein, "Textural Features for Image Classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 3, no. 6, pp. 610-621, 1973. [CrossRef] [Google Scholar] [Publisher Link]

[5] N. Dalal, and B. Triggs, "Histograms of Oriented Gradients for Human Detection," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Diego, CA, USA, vol. 1, pp. 886-893, 2005. [CrossRef] [Google Scholar] [Publisher Link]

[6] Faeze Kiani, "Texture Features in Medical Image Analysis: A Survey," *arXiv preprint*, pp. 1-7, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[7] Yan Hao et al., "Breast Cancer Histopathological Images Recognition Based on Low Dimensional Three-Channel Features," *Frontiers in Oncology*, vol. 11, pp. 1-15, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[8] Amjad Rehman Khan et al., "Identification of Anomalies in Mammograms through Internet of Medical Things Diagnosis System," *Computational Intelligence and Neuroscience*, vol. 2022, no. 1, pp. 1-12, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[9] Yan Hao et al., "Breast Cancer Histopathological Images Classification Based on Deep Semantic Features and Gray Level Co-Occurrence Matrix (GLCM) Features," *PLOS ONE*, vol. 17, no. 5, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[10] Shahadat Uddin et al., "Comparative Performance Analysis of K-Nearest Neighbour (KNN) Algorithm and Its Different Variants for Disease Prediction," *Scientific Reports*, pp. 1-11, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[11] Breast Histopathology Images. [Online]. Available: https://www.kaggle.com/datasets/paultimothymooney/breast-histopathology-images

[12] Harmandeep Singh, Vipul Sharma, and Damanpreet Singh, "Comparative Analysis of Proficiencies of Various Textures and Geometric Features in Breast Mass Classification Using K-Nearest Neighbour," *Visual Computing for Industry, Biomedicine and Art*, vol. 5, no. 3, pp. 1-19, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[13] Reem Jalloul, H.K. Chethan, and Ramez Alkhatib, "A Review of Machine Learning Techniques for the Classification and Detection of Breast Cancer from Medical Images," *Diagnostics*, vol. 13, no. 14, pp. 1-24, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[14] Nada Mobark, Safwat Hamad, and S.Z. Rida, "CoroNet: Deep Neural Network-Based End-to-End Training for Breast Cancer Diagnosis," *Applied Sciences*, vol. 12, no. 14, pp. 1-12, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[15] Madallah Alruwaili, and Walaa Gouda, "Automated Breast Cancer Detection Models Based on Transfer Learning," *Sensors*, vol. 22, no. 3, pp. 1-16, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[16] Debendra Muduli, Ratnakar Dash, and Banshidhar Majhi, "Automated Diagnosis of Breast Cancer Using Multi-Modal Datasets: A Deep Convolutional Neural Network Based Approach," *Biomedical Signal Processing and Control*, vol. 71, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[17] Chegini M. (Mohaddeseh Chegini) — "Uncertainty-Aware Deep Learning-Based CAD System for Breast Cancer Classification Using Ultrasound and Mammography Images," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 12, no. 1, pp. 1-17, 2023. [CrossRef] [Google Scholar] [Publisher Link]