IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 15, No. 4, August 2026, pp. 3556~3567
ISSN: 2252-8938, DOI: 10.11591/ijai.v15.i4.pp3556-3567

A machine learning framework for skin cancer classification using texture descriptors

Shikha Malik[1], Vaibhav V. Dixit[2]
[1]Department of Electronics and Communication Engineering, G. H. Raisoni College of Engineering and Management, Savitribai Phule Pune University, Pune, India
[2]Department of Electronics and Telecommunication, Rasiklal M. Dhariwal Sinhgad Technical Institutes Campus, Savitribai Phule Pune University, Pune, India

Article Info

Article history:

Received Nov 9, 2025
Revised Jun 23, 2026
Accepted Jul 9, 2026

Keywords:

Geometric transformation augmentation
Gray level co-occurrence matrix
Skin cancer classification
Support vector machine
Texture features

ABSTRACT

Skin cancer remains a critical health and economic concern worldwide. Timely and accurate diagnosis is crucial for improving mortality rate of patients. Although automated machine learning (ML) models assist doctors in clinical assessment of skin cancer from dermoscopic images, their performance often suffers from extreme class imbalance, as benign image samples greatly exceed malignant ones. This leads to false detection and delayed diagnosis of skin cancer. To overcome this issue, the study proposes an efficient and lightweight geometric transformation (GT)–augmented support vector machine (SVM) framework for early diagnosis of skin cancer. It effectively addresses the class imbalance issues present in the datasets and improves the detection of positive cases. The novel pipeline integrates preprocessing, morphology preserving GT, rotation-invariant texture feature extraction, feature validation, and feature scaling for performing binary classification using an optimized SVM framework. This framework provides balanced and accurate lesion classification even if image samples are insufficient. Experimental results have successfully achieved a true positive rate (TPR) of 93% on PH^2 and 81.1% on International Skin Imaging Collaboration 2016 (ISIC-2016) dataset, which is better than conventional ML models. These findings prove that proposed GT-SVM framework is a lightweight, interpretable, and computationally efficient approach for early skin cancer diagnosis. Future developments shall explore multi-class lesion classification, validation across diverse clinical datasets, and hybrid feature fusion.

Corresponding Author:

Shikha Malik
Department of Electronics and Telecommunication
G. H. Raisoni College of Engineering and Management, Savitribai Phule Pune University
New Gate No. 1200, Domkhel Road, Wagholi, Pune 412207, Maharashtra, India
Email: shikha.malik.phdetc@ghrcem.raisoni.net

1. INTRODUCTION

The rising skin cancer incidence is a critical health and economic concern worldwide. The disease primarily stems from the epidermis (outermost) stratum of the skin. This occurs due to damage to cellular deoxyribonucleic acid (DNA) that leads to genetic mutations and unregulated cell multiplication. Skin cancer exists as acute malignant melanoma (MEL) and the more frequent non-melanoma varieties such as basal cell carcinoma (BCC) and squamous cell carcinoma (SCC). Global Cancer Statistics reported nearly 19.3 million new patients and over 10 million cancer fatalities in 2020 [1]. Melanoma skin cancer has the highest mortality rate because it metastasizes and severely affects surrounding tissues [2]. Epidemiological data

highlight that melanoma incidences are higher in men than in women. The forecasts indicate that if the same current trends exist, then melanoma cases could reach 510,000 with a growth rate of 68% by 2040 [3].

In a clinical scenario, the diagnosis of skin cancer is made by dermatologists by performing visual inspection of the lesions along with the individual's history, age, location, and morphological features of the lesions. The accuracy of clinical diagnosis by this method, however, becomes low due to broad morphological variations among the lesions. Artificial intelligence (AI) and image processing tools can be valuable options for assisting dermatologists in better, more reliable, and timely diagnosis of skin cancer [4], [5]. Even with these advances, class imbalance in dermoscopic skin cancer datasets remains the key limitation during automated diagnosis, as the benign lesion samples greatly exceed malignant ones. This imbalance exists in several public repositories such as International Skin Imaging Collaboration 2016 (ISIC-2016) and PH^2, which causes AI models to prefer majority benign classes and produce low true positive rates (TPR) for minority malignant lesions [6]. As a result, the rising number of false negative cases is delaying timely treatment and adversely affecting the survival of patients. Researchers explored several machine learning (ML) methods as an automated approach to skin cancer detection. This includes i) clinically driven methods, ii) handcrafted feature-based classification, and iii) deep learning approaches. The following section reviews key advances in these areas, highlighting their shortcomings and identifying the research gaps that motivated the development of the proposed geometric transformation (GT)–augmented support vector machine (SVM) framework for cancer classification.

During clinical diagnosis of skin cancer, the ABCD rule is followed, which measures values of four main parameters: asymmetry (A) in pigmented lesions, irregularities in borders (B), variations in lesion colors (C), and diameter (D) of lesions. The malignant lesions commonly have irregular shapes, uneven borders, non-uniform color spread, and a minimum 6 mm diameter. These diagnosis criteria have been commonly implemented in studies involving manual methods of skin cancer detection. An ABCD clinical prototype system proposed in [7] achieved 92.22% specificity, 90% accuracy, and 85% sensitivity on the PH^2 dataset. Another similar research [8] involved the ABCD parameters and total dermoscopy score (TDS) calculations, achieving 84% accuracy and 60.5% sensitivity for melanoma detection. Both approaches did not employ ML methods and utilized only classical image processing techniques that provided only acceptable accuracy. This makes it difficult for the model to work effectively on a wide scale with complex datasets. These clinical approaches do not include automatic feature learning and lack the ability to handle class imbalance or heterogeneous data samples. Consequently, recent studies have focused on advanced approaches such as ML and deep learning to strengthen clinical performance and improve generalization on new data.

Feature engineering techniques have been employed in several works for the extraction of color, shape, and texture attributes. Warsi *et al*. [9] developed a three-dimensional analysis tool that extracts color and texture information from the PH^2 dataset by utilizing red, green, and blue (RGB) color channels. They classified features using a proposed multilayer perceptron (MLP) algorithm. The results of the simulation show that the 3D tool offered improved performance compared with existing benchmark models. However, researchers did not employ alternative classification methods for comparison and evaluated them on a single dataset. Thus, the model failed to generalize across changing image conditions and populations. Tumpa and Kabir [10] proposed a hybrid texture-feature extraction from ISIC and PH^2 datasets. The model attained 97.7% accuracy with an artificial neural network (ANN) classifier. Although the model outperformed several existing classifiers and extracted a rich set of texture descriptors, its performance was not compared against baseline ML algorithms. An ensemble framework proposed by Alizadeh and Mahloojifar [11] integrated convolutional neural network (CNN), Haralick, and local binary pattern (LBP) texture features. This hybrid feature model performed better on PH^2 and ISIC-2019 datasets but attained insufficient sensitivity of 52% on ISIC-2016. This low sensitivity on ISIC-2016 is due to severe imbalance between benign and malignant lesions. Similarly, Yu *et al*. [12] combined features from multiple CNN architectures using fisher vector encoding to enhance lesion classification. Although GT and color normalization were used to minimize class imbalance, the model still demonstrated reduced sensitivity because of high complexity and inadequate feature selection. Haider *et al*. [13] proposed framework involved active contour region of interest (ROI) extraction, LBP texture descriptor, k-nearest neighbor (KNN), and support vector machine (SVM) decision models for detection of pigmented lesions in PH^2 images. The system was superior to manual lesion extraction but did not perform well on automatic ROI extraction. They assessed the model only on the PH^2 dataset without considering its class imbalance issues, therefore, it is not capable of deployment in healthcare. Feature-based classification has the potential to be applied across multiple application domains [14], [15]. It can improve the outcomes provided meaningful features are selected and validated precisely.

In recent years, deep learning architectures have been adopted widely as an automated method to extract informative and complex patterns from dermoscopic images. Memon *et al*. [16] employed the Xception algorithm for the prediction of skin cancer and attained the highest accuracy score of 93% on the

HAM10000 dataset. However, they worked only on a small set of the HAM10000 database; therefore, its application is limited in actual clinical settings. Tang *et al*. [17] implemented a dual-network architecture called global-part convolutional neural network with data-transformed ensemble learning (GP-CNN-DTEL) that completely captures image information along with precise lesion details. Ensemble learning was applied to normalized color images to improve robustness. The model achieved 99.7% specificity and 86.3% accuracy, but it has low sensitivity (32%), which indicates its limited ability to detect malignant lesions. This is caused by class imbalance and strict reduction of false positive cases.

Another study by Wu *et al*. [18] proposed the attention and residual learning densely connected convolutional network (ARDT-DenseNet) framework consisting of complex connections, attention mechanisms, and residual learning for reliable classification of melanoma. Even after employing imbalance mitigation techniques, the model still underperforms due to the overfitting problem and deep-layer architecture. Similar work by Balaha and Hassan [19] implemented Attention U-Net for ROI extraction, pre-trained CNN classifiers, and a sparrow search optimization (SSO) algorithm on the PH^2 dataset for melanoma identification. The attention U-Net with DenseNet-201 yielded 95% accuracy, 86% sensitivity, 93% specificity, and 92% F1-score. However, the model was highly complex due to dense architectures and complex optimization layers. The high computational cost made it less operative for practical clinical applications.

The examination of prior research shows that most work has used imbalanced datasets such as ISIC and PH^2 for training their models. This leads to model bias and hence reduces its ability to detect malignant cases. Furthermore, many studies have utilized only a single dataset with insufficient samples for testing their model. Thus, their model performance cannot be validated across diverse clinical scenarios. Another common limitation is that the majority of research is focused on improving overall accuracy, which is an inadequate measure for imbalanced classes. There is very limited attention to the TPR in the literature, an important evaluation metric for reliable melanoma detection. The deep learning approaches implemented to improve performance are computationally complex and less interpretable. This makes them unsuitable for regular clinical practice. To address the above-described challenges, this study proposes an efficient automated approach for classifying skin cancer that is reliable, interpretable, and computationally efficient, designed to optimize the TPR and can be effectively deployed in clinical environments.

This study's main research contributions are described as follows:

i) The proposed framework applies morphology preserving GT to the minority class samples present in the ISIC-2016 and PH^2 datasets. This effectively tackles the class imbalance issues and enhances the TPR for malignant minority cases.
ii) The framework effectively captures distinct image features through the extraction of rotation-invariant Haralick texture and statistical descriptors from the gray level co-occurrence matrix (GLCM) approach.
iii) It implements hypothesis testing for feature validation and assesses the statistical significance of extracted features. This makes the model more interpretable and ensures reliable feature representation. There is very limited attention on this statistical aspect in earlier studies.
iv) The proposed GT-SVM framework is a lightweight and interpretable framework that presents an effective alternative to complex deep models for binary classification of skin lesions under limited data conditions.
v) The proposed framework is thoroughly evaluated and compared with existing ML approaches using ISIC-2016 and PH^2 datasets.

The structure of the remaining paper is as follows. Section 2 provides a detailed explanation of the methodology and implementation steps of the proposed framework. This is to support the reproduction of the experimental results for further extending research work. The results of the proposed model, key findings, and the framework's implications and limitations are discussed in section 3. Section 4 concludes the study by briefly presenting the clinical usefulness of the proposed framework and directions for future research.

2. METHOD

The research aims to improve the early detection of malignant lesions by proposing an efficient approach for classifying skin cancer from dermoscopic images. To achieve this goal, the GT-SVM framework is suggested, which enhances the minority classes, captures distinct image features, performs feature validation, and classifies lesions using an optimized classifier. The proposed pipeline involves six key steps as shown in Figure 1. It comprises i) data collection, ii) pre-processing, iii) GT, iv) feature extraction and feature scaling, v) feature validation, and vi) binary classification using GT-SVM. The entire framework was developed using several Python packages such as OpenCV, NumPy, Pandas, and scikit-learn to ensure consistent methodology and reproducible results.

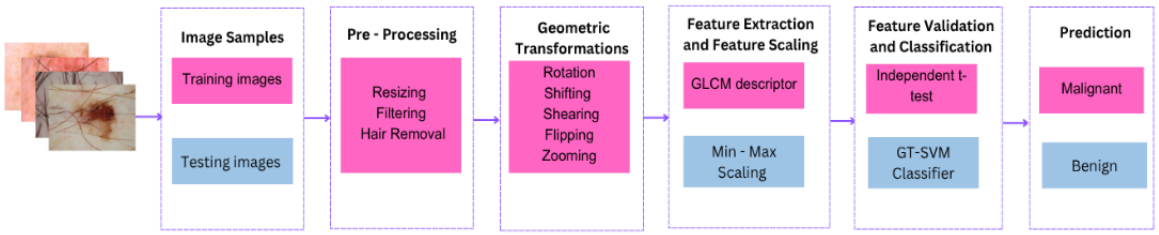


Figure 1. Flow diagram of the proposed GT-SVM framework for skin cancer classification

2.1. Data acquisition

This work selected two benchmark dermoscopic image datasets: ISIC-2016 [20] and PH^2 [21]. The ISIC-2016 dataset is gathered from the ISIC archive database. 1,279 dermoscopic images of this dataset comprise 1,031 non-melanoma (benign) and 248 melanoma (malignant) lesions. The PH^2 dataset is developed by the Dermatology Department, Portugal Hospital, and comprises 200 dermoscopy images of resolution 768×560. 160 benign samples of this dataset contain 80 common nevi and 80 atypical nevi, and 40 images are confirmed melanoma cases. This dataset includes segmentation masks, annotations, and dermoscopic features verified by experts and is perfectly suitable for segmentation as well as classification. The statistical data of both datasets are presented in Table 1.

Table 1. ISIC-2016 and PH^2 dataset statistics

Dataset	Total images	Benign samples	Malignant samples
ISIC-2016	1,279	1,031	248
PH2	200	160	40

2.2. Pre-processing

The pre-processing stage is the first important step required to remove image defects and sustain the morphological features of lesions. It was implemented in three phases, consisting of image resizing, followed by noise removal and unwanted hair artifacts elimination. All these were carefully implemented to improve the quality of the images while preserving the lesion's morphological features [22]. Figure 2 presents the sequence of preprocessing operations applied to benign and malignant samples.

The first step in pre-processing was a resizing operation in which images of varying resolutions were modified to maintain consistency in the datasets. All image samples were standardized to an input size of 224×224 pixels. This created a uniform spatial size and minimized the processing time required for feature analysis and classification. This step was performed using the OpenCV library with the cv2.resize() function.

After resizing, filtering was performed to reduce pixel-level noise without modifying meaningful texture information. Instead of a median filter, a Gaussian smoothing filter with a 3×3 kernel was preferred as it better preserves lesion edges and texture variations, which are essential for feature extraction. The Gaussian parameters were automatically tuned on the basis of kernel size. This preserved clinically meaningful texture information and effectively reduced high-frequency noise. The cv2.GaussianBlur() function available in the OpenCV library was used to implement the Gaussian filtering operation.

Hair artifacts removal was the next important part in preprocessing for clear visualization of lesion borders, which were hidden by unwanted hairlines in dermoscopic images. The hair artifacts affect the accurate extraction and classification of features. The hair artifact removal process was performed in multiple morphological stages, which involved grayscale conversion, black-hat morphology, adaptive thresholding, and image inpainting. First, the dermoscopic images were converted into grayscale. This was done by selecting the parameter cv2.COLOR_RGB2GRAY, which is available in the cv2.cvtColor() function from the OpenCV library. This operation helped in removing unnecessary color information and reduced the computational cost. The Black-hat morphological operation was the next step to be implemented via 9×9 elliptical kernels. The parameters for this operation were selected with the help of cv2.getStructuringElement() and cv2.morphologyEx() functions. This step separated the background and hair lines without affecting the lesion details. Adaptive thresholding was then applied to detect thin hair lines very precisely. The operation was implemented using the cv2.adaptiveThreshold() function. Finally, Telea's inpainting algorithm was applied to restore the areas covered by the mask by estimating them using neighboring pixel values. This reconstructed the natural appearance of lesions. The cv2.inpaint() function

with the cv2.INPAINT_TELEA method was used to execute this step. The Telea inpainting technique was employed as it is computationally efficient and preserves the original appearance of the lesions.

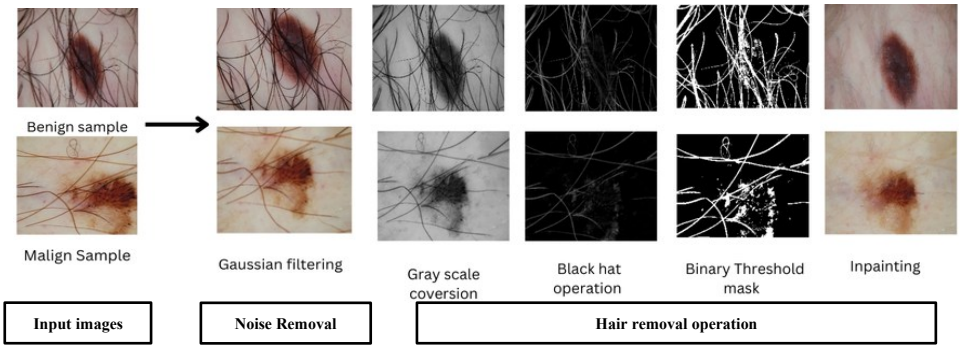


Figure 2. Sequence of pre-processing operations applied to benign and malignant samples

2.3. Geometric transformations

The PH^2 and ISIC-2016 datasets are both significantly imbalanced, as benign classes are higher than malignant ones. There is approximately a 4:1 benign-to-malignant ratio in both datasets. There are 160 benign (negative cases), and 40 malignant (positive cases) samples in the PH^2 dataset, and ISIC-2016 comprises of 1,031 benign and 248 malignant samples. As observed in repeated experiments, classification models always favored the majority (benign) cases due to this imbalance. This resulted in reduced sensitivity while detecting malignant lesions. To solve this issue, GT were applied to enhance malignant minority images without modifying the count of benign images. These transformations improved dataset diversity and ensured that lesion boundaries, pigment networks, and structural patterns are clinically valid for balanced and reliable skin lesion detection [23]. The GT applied were rotations randomly varying between 0° and 25°, vertical and horizontal translation up to 20% of image size, random shearing limited to 20%, horizontal flipping, and zooming up to 20%. These transformations were implemented in TensorFlow/Keras with the support of image data generator modules. As shown in Figure 3, each malignant image generated four new augmented versions. Figure 3(a) shows a sample image from ISIC-2016, and Figure 3(b) presents its four new augmented versions resulting from GT. These transformations effectively expanded the training datasets without altering the morphological details of skin lesions. Consequently, the minority class samples quadrupled from 248 to 992 for ISIC-2016 and from 40 to 160 samples in the PH^2 dataset after the application of GT. Table 2 presents the statistics of both datasets after the application of GT.

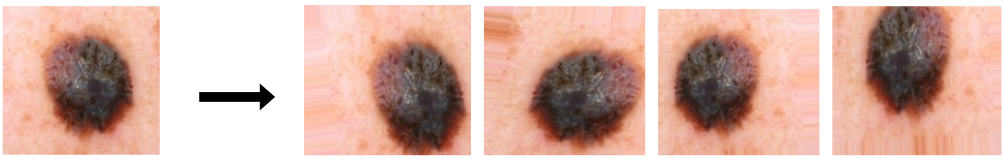


Figure 3. Illustration of GT operations of (a) ISIC-2016 sample image and (b) four outputs after GT

Table 2. Statistics of ISIC-2016 and PH^2 datasets after application of GT

Dataset	Malignant images (Before transformation)	Malignant images (After transformation)	Benign images	Total
ISIC-2016	248	992	1,031	2,023
PH^2	40	160	160	320

2.4. Feature extraction and feature scaling

After GT, feature extraction was performed by extracting texture and statistical features from the GLCM method proposed by Haralick [24]. GLCM approach is a second-order statistical texture analysis

technique that measures the relationship between neighboring pixel pairs in the spatial domain. This approach captures more detailed texture information than first-order histogram methods. It evaluates how frequently specific intensity pairs occur at a defined distance (s) and orientation (α) thereby representing structural and morphological characteristics of lesions. The step-by-step procedure for constructing the GLCM is detailed in Algorithm 1.

Algorithm 1. GLCM algorithm
1: Initialize gray levels r, spatial distance s and angle α.
2: Initialize $r * r$ GLCM matrix P.
3: for each pixel (x,y) in input image I:
compute neighbor (x', y') based on s and angle α
let pixel $I(x, y)$ intensity = i and pixel $I(x', y')$ intensity = j
$i, j \in \{0,1, \dots r-1\}$
increment Pij by 1
4: Normalize P represents probability matrix P
5: Extract features from normalized GLCM P

In this study, statistical features, including maximum probability, mean, variance, standard deviation, skewness, and kurtosis, together with Haralick texture features such as correlation, contrast, energy, homogeneity, entropy, dissimilarity, and angular second moment, were derived from the normalized GLCM P. The GLCM matrices were computed at four orientations (α = 0°, 45°, 90°, and 135°) with $s = 1$ pixel, and the average of the resulting feature values was used to obtain a rotation-invariant texture representation. This method guaranteed that the extracted features remain stable with respect to lesion orientation, thereby improving the model's classification performance. Table 3 summarizes the 13 selected texture and statistical descriptors extracted for classification.

Table 3. Brief description of 13 extracted statistical and Haralick features

Features	Definition	Calculation method
Maximum probability (Pmax)	Measure of the peak response	$Max\ probability\ (Pmax) = \max_{i,j} P_{ij}$
Mean (μ)	Information about the average gray level relationship within the image	$Mean\ (\mu_i) = \sum_{i,j=0}^{r-1} iP_{ij}$
Variance (σ^2)	Amount of deviation of the gray level from the average value	$Varaince\ (\sigma^2) = \sum_{i,j=0}^{r-1} (i-\mu)^2 P_{ij}$
Standard deviation (σ)	Clear measure of variability in pixel intensity values	$Standard\ devaiation\ (\sigma) = \sqrt{Variance}$
Skewness (γ)	Quantifies the pixel intensity variation relative to the mean	$Skewness\ (\gamma) = \sum_{i,j=0}^{r-1} \frac{(i-\mu)^3\ P_{ij}}{\sigma^3}$
Kurtosis (k)	Evaluate the peakiness of a pixel intensity distribution compared to a standard normal distribution	$Kurtosis\ (k) = \sum_{i,j=0}^{r-1} \frac{(i-\mu)^4\ P_{ij}}{\sigma^4}$ -3
Correlation (ρ)	Quantifies the degree of correlation among a pixel and its neighboring pixels	$Correlation\ (\rho) = \frac{\sum_{i,j=0}^{r-1} (i-\mu_i)(j-u_j)P_{ij}}{\sigma_i \sigma_j}$
Contrast (C)	Measures the local differences or variations in pixel intensities	$Contrast\ (C) = \sum_{i,j=0}^{r-1} (i-\mu)^2\ P_{ij}$
Energy or uniformity (E)	Measures how evenly or homogeneously the intensities of the pixels in an image are distributed	$Energy\ (E) = \sum_{i,j=0}^{r-1} {P_{ij}}^2$
Homogeneity (H)	Indicates how closely the arrangement of elements in the GLCM aligns with its diagonal	$Homogeneity\ (H) = \sum_{i,j=0}^{r-1} \frac{P_{ij}}{1+(i-j)^2}$
Entropy (E)	Evaluates the randomness or complexity in pixel intensities	$Entropy\ (E) = -\sum_{i,j=0}^{r-1} P_{ij}\ log(P_{ij})$
Dissimilarity (D)	Measures the average absolute intensity difference between adjacent pixels	$Dissimilarity\ (D) = \sum_{i,j=0}^{r-1} \lvert i-j \rvert P_{ij}$
Angular second moment (ASM)	Higher ASM demonstrates uniform pixel pair distribution	$ASM = (Energy)^2$

The extracted 13 statistical and texture feature sets were saved as a comma-separated value (.csv) file for further analysis. The feature values were normalized to a standard scale using the min-max scaling approach. This ensured all features contribute uniformly in classification. The min-max scaler function was available in the Python scikit-learn package. After normalization, all features were converted to values between [0, 1]. The min-max scaling function for a given feature f in the dataset is expressed by (1) and (2).

$$f_{std} = \frac{f - f_{min}}{f_{max} - f_{min}} \quad (1)$$

$$f_{scaled} = f_{std} * (max - min) + min \quad (2)$$

Where f_{scaled} is the final normalized feature value varying in range between [min, max].

2.5. Feature validation

An independent *t*-test [25] was performed to examine the prediction ability of the extracted features. It measured the mean differences between benign and malignant categories. This statistical validation method ensured that only important and unique features impacted the classification. There is very limited attention on this statistical aspect in earlier studies.

The hypotheses were defined as follows:

– Null hypothesis (H_0): the mean feature values were equal in both diagnostic classes.
– Alternative hypothesis (H_a): there existed a statistically meaningful difference between the mean feature values of classes.

The *t*-statistic is calculated using the (3).

$$t = \frac{\overline{X_1} - \overline{X_2}}{SD\ pooled . \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (3)$$

The pooled standard deviation $SD\ pooled$ is computed using the (4).

$$SD\ pooled = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \quad (4)$$

Where $\overline{X_1}$ represent the sample mean of the benign feature class and $\overline{X_2}$ represents the malignant class sample mean, n_1 and n_2 denote their respective sample sizes and s_1^2 and s_2^2 are the corresponding sample variances.

The independent *t*-test was carried out using the *ttest_ind()* function from the SciPy library. Results showed statistically meaningful differences ($p < 0.05$) for 12 among the 13 features, except for entropy. Although entropy did not meet the significance threshold ($p > 0.05$), it was selected because it effectively captured relevant texture pattern variations in dermoscopic images. Among all features, correlation ($t = 16.45$) and homogeneity ($t = 15.77$) demonstrated the highest *t*-scores, which proved they have a strong ability to differentiate the classes. Overall, the *t*-test results validated that the extracted features are statistically significant and suitable for the proposed GT-SVM framework.

2.6. Binary classification using geometric transformation–augmented support vector machine

To execute binary classification using the proposed GT-SVM framework, the validated feature set was separated into a train and test set at a split ratio of 80:20. The train subset was applied to optimize the classifier's parameters, while the test data was mainly employed for assessing the model's performance. The GT-SVM algorithm classified each sample as a benign or malignant category depending on the statistically extracted significant features. GT-SVM is developed from the conventional SVM model, which is proven effective for both classification and regression operations. SVM focuses on determining the most suitable hyperplane that maximizes the separation gap between distinct classes [26].

Given a train set $\{(x_i, y_i)\}_{i=1}^{n}$, where $x_i \in \mathbb{R}^m$ represents an *m*-dimensional feature matrix and $y_i \in \{-1,1\}$ are two-class labels, the optimization problem is formulated as in (5).

$$\min_{w,b,\xi} \frac{1}{2} \| w \|^2 + C \sum_{i=1}^{n} \xi_i \text{ limited by } y_i(w.x_i + b) \geq 1 - \xi_i,\ \xi_i \geq 0, i = 1, \ldots n \quad (5)$$

Where w indicates the weight parameters, b is the offset factor, ξ_i are slack variables for handling misclassified instances, and C is a penalty coefficient regulating the balance between maximizing margin and minimizing classification error. To handle nonlinear separations, radial basis function (RBF) kernel was

applied in the GT-SVM framework. This technique implicitly projects the input features into a large feature space using the (6).

$$K\ (x_i, x_j) = exp(-\gamma \parallel\ x_i - x_j \parallel^2) \quad (6)$$

Where γ is a hyperparameter that determines the kernel's sensitivity to pairwise distances between feature vectors x_i .and x_j.

Grid search approach with five-fold cross-validation (cv =5) was used to optimize hyperparameters by implementing scikit-learn's GridSearchCV operation. The best configuration was obtained with C = 1,000, γ = 1, and the RBF kernel. This yielded superior classification accuracy by effectively balancing margin width and error minimization. The final GT-SVM framework was trained on morphology preserving and rotation-invariant texture features and demonstrated reliable discrimination between benign and malignant lesions.

3. EXPERIMENTAL RESULTS AND DISCUSSIONS

All computational experiments were carried out on a Windows 10 operating system equipped with a 2.5 GHz processor and 8 GB of RAM. This ensured stable and efficient performance throughout the experiments. This configuration enabled the smooth execution of ML workflows and helped with quicker training and prediction of the proposed framework. The Python 3.11 in the Spyder environment and Google Colab was used for performing experiments both locally and in the cloud. The standard Python packages like NumPy, Pandas, OpenCV, and scikit-learn were applied to perform the simulation work.

The performance parameters measured were accuracy, precision, TPR, specificity, and F1-score. Their brief description is provided in Table 4. The true positives (TP) mean correctly classified malignant (positive) samples, while false positives (FP) denote incorrectly identified benign (negative) samples. Conversely, true negatives (TN) represent accurately recognized benign samples, and false negatives (FN) indicate incorrect classification of malignant samples.

Table 4. Summary of performance parameters

Metric	Calculation	Clinical meaning
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$	Measures overall prediction performance
Precision	$\frac{TP}{TP+FP}$	Capability to reduce false positives
TPR	$\frac{TP}{TP+FN}$	Measure the detection rate of the positive class
Specificity	$\frac{TN}{TN+FP}$	Measures the correct benign cases.
F1-score	$\frac{2*PRE*Recall}{PRE+Recall}$	Balances model's precision and recall values

3.1. Key findings and interpretation

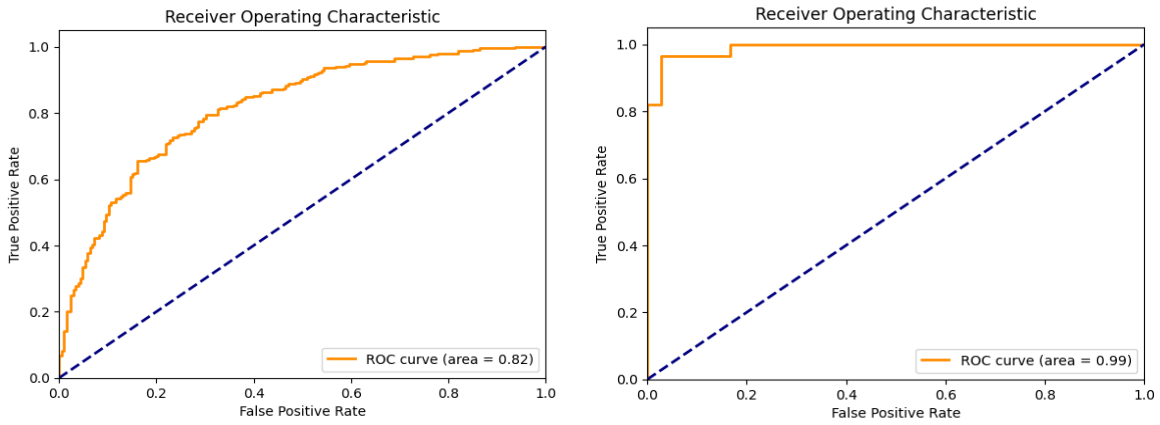
The classification performance of the proposed GT-SVM framework on the ISIC-2016 and PH^2 datasets was compared with the existing ML algorithms, mainly logistic regression (LR), random forest (RF), KNN, MLP, and extreme gradient boosting (XGBoost), and presented in Table 5. The simulation results highlight that the GT-SVM framework trained on 13 extracted features showed superior classification performance compared to existing baseline methods, with an accuracy score of 80% on ISIC-2016 and 95.3% on the PH^2 dataset. The TPR of 81.1% on ISIC-2016 and 93% on PH^2 indicates that the framework is correctly identifying positive cases. The specificity score of 78.1% on ISIC-2016 and 97.2% on PH^2 datasets confirms precise detection of benign cases. The precision values of 82% and 96.3% reveal that the model has fewer false-positive predictions, while the F1-score of 81.3% and 95% indicate an effective balance between precision and recall parameters.

Notably, the GT-SVM framework successfully attained high sensitivity without compromising specificity. This enables accurate identification of positive cases with fewer false predictions. The TPR values across both datasets were consistently high, particularly the 93% achieved on PH^2. This confirmed that morphology preserving GT effectively alleviated the class imbalance issue, which resulted in high TPR. Overall, the framework achieved its key objective of enhancing malignant lesion detection and is interpretable and computationally efficient.

Table 5. Comparison of GT-SVM with conventional models for classification

Dataset	Models	Accuracy (%)	Precision (%)	TPR (%)	Specificity (%)	F1-score (%)
ISIC-2016	LR	76	78	77.9	73	77.7
	RF	75.1	77.8	76.3	73.7	77
	KNN	74.5	76	77.9	70.3	76.9
	MLP	76	77.6	77.9	72	77.7
	XGBoost	77.1	79.5	78.3	75.7	78.8
	Proposed GT-SVM	80	82	81.1	78.1	81.3
PH^2	LR	91	89.2	89.2	92	89.2
	RF	91	84.3	92	86.1	90
	KNN	91	87	89	70.3	90
	MLP	92.1	90	93	92	91.2
	XGBoost	91	84.3	92.2	86.1	90
	Proposed GT-SVM	95.3	96.3	93	97.2	95

The receiver operating characteristic (ROC) curves are plotted in Figure 4 at different threshold values to further assess the model's classification performance. Figure 4(a) displays the ROC curve across the ISIC-2016 dataset, and Figure 4(b) displays the ROC curve corresponding to the PH^2 dataset. It attained a higher area under the curve (AUC) score of 0.99 on the PH^2 dataset, that prove that the model has a strong capability of identifying malignant cases correctly. In comparison, the AUC of 0.82 on the ISIC-2016 dataset was slightly lower because image samples in this dataset are highly diverse, which challenges the classification performance of the model. Overall, these results confirmed that the GT-SVM framework can effectively distinguish all classes on the PH^2 dataset while maintaining high sensitivity and specificity.



(a) (b)

Figure 4. ROC plot of the proposed framework across (a) ISIC-2016 dataset and (b) PH^2 dataset

3.2. Comparative evaluation with existing methods

The proposed GT-SVM classification framework was evaluated against existing studies on the ISIC-2016 and PH^2 datasets. Table 6 presents a TPR comparison across the ISIC-2016 dataset. Simulation results demonstrated that the GT-SVM framework achieved a TPR of 81.1% on ISIC-2016 and outperformed most existing benchmark models. The ensemble model proposed by Alizadeh and Mahloojifar [11] did not effectively minimize class imbalance, which resulted in low TPR. Although the models proposed in [12], [17] incorporated GT and deep classification architectures, their focus was more on improving specificity rather than balanced detection. In contrast, the ARDT-DenseNet model proposed by Wu *et al*. [18] achieved a slightly higher TPR of 81.6%, but the deep and complex design of the model made it less interpretable and computationally inefficient. Unlike these models, the proposed GT-SVM framework is lightweight, interpretable, and computationally efficient that is maintaining high sensitivity while effectively addressing class imbalance.

Table 7 depicts a comparison of GT-SVM and other baseline studies on the PH^2 dataset. The ABCD method implemented in [7], [8] did not incorporate any ML algorithms and was only based on conventional image processing. Thus, their models were less capable of working reliably in physical clinical environments. Haider *et al*. [13] achieved a TPR of 89.2% by applying active contour segmentation, LBP descriptors, and

an SVM classifier. However, they worked on imbalanced data, which made the model clinically unreliable. An optimized framework proposed by Balaha and Hassan [19] combined DenseNet-201 architecture and the SSO algorithm. It attained 94.75% accuracy and 92.74% TPR. However, the framework is not computationally efficient because it includes multiple deep learning and optimization layers. In comparison, the proposed GT-SVM model achieved 95.3% accuracy, 93% TPR, and 97.2% specificity without the integration of any complex layer. Therefore, its performance is superior with low computational cost. The model is capable of detecting malignant lesions reliably by effectively minimizing false positives. This validates it as a robust and interpretable approach for automated detection of skin cancer.

Table 6. Comparative TPR assessment of the proposed framework against baselines on ISIC-2016

Model	TPR (%)
CNN + Haralick features [11]	52
Multi-CNN and SVM [12]	60.1
GP-CNN-DTEL [17]	32
ARDT-DenseNet [18]	81.6
Proposed GT-SVM	81.1

Table 7. Comparison of the GT-SVM framework against baselines on the PH^2 dataset

Model	Accuracy (%)	TPR (%)	Specificity (%)
ABCD [7]	90	85	92.22
ABCD [8]	84.00	60.50	89.50
LBP and SVM [13]	89.2	89.2	84
DenseNet-201 and SSO [19]	94.75	92.74	96.20
Proposed GT-SVM	95.3	93	97.2

3.3. Practical implications and future scope

The GT-SVM framework exhibited a robust, interpretable, and computationally efficient diagnostic approach for the binary classification of dermoscopic images for the detection of skin cancer. The findings revealed that model results are highly superior compared with several baseline approaches. It is lightweight and provides high TPR, which enhances its ability for deployment in physical healthcare practice. However, the current framework is suitable only for binary classification and cannot be validated across diverse clinical datasets. To further enhance its generalization and the clinical applicability of this framework, future developments shall explore the model for the classification of multiple lesion categories, incorporating additional hybrid feature descriptors such as color, shape, and deep learning features, and using diverse clinical datasets will improve its generalization and clinical applicability.

4. CONCLUSION

The GT-SVM framework presented in this study is a robust and interpretable ML approach for the binary classification of dermoscopic images for skin cancer detection. The rotation-invariant texture descriptors derived from the GLCM approach and morphology preserving GT have been implemented for the GT-SVM framework. The framework improved the TPR by addressing class imbalance issues in an effective way. The framework achieved a TPR of 81.1% on ISIC-2016 and 93% on PH^2 with accuracy values of 80% and 95.3%, respectively. Compared to complex deep learning models available for the classification of dermoscopic images, the GT-SVM framework is a lightweight, interpretable, and computationally efficient framework with high sensitivity and without compromising specificity. This framework is able to detect malignant cancers in an efficient way; therefore, it is suitable for the timely assessment of skin cancer in physical clinical environments. To further support the clinical applications of this framework, future developments shall explore multi-class lesion classification, validation across diverse clinical datasets, and hybrid feature fusion.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Shikha Malik	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓			
Vaibhav V. Dixit			✓		✓	✓		✓	✓	✓	✓	✓	✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The author declares that there are no known conflicts of interest associated with this publication. There are no financial or personal relationships that could inappropriately influence or bias the content of this work.

INFORMED CONSENT

Not applicable. This study did not involve human participants, human data, or any personally identifiable information. All data used were either publicly available, fully anonymized, or derived from non-human sources, and therefore no informed consent was required from individuals.

ETHICAL APPROVAL

Not applicable. This research did not involve human subjects, human biological materials, or experimental procedures on animals. The work was conducted solely on computational models, publicly available datasets, or non-sensitive data that did not require intervention with living organisms. Therefore, ethical approval from an institutional review board or animal ethics committee was not necessary for this study.

DATA AVAILABILITY

The data that support the findings of this study are openly available in the International Skin Imaging Collaboration (ISIC) archive at http://doi.org/10.1016/j.media.2021.102305, reference number [20] and in PH^2 database at http://doi.org/10.1016/j.bspc.2022.104186, reference number [21].

REFERENCES

[1] H. Sung *et al.*, "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 71, no. 3, pp. 209–249, May 2021, doi: 10.3322/caac.21660.
[2] T. K. Nikolouzakis *et al.*, "Current and future trends in molecular biomarkers for diagnostic, prognostic, and predictive purposes in non-melanoma skin cancer," *Journal of Clinical Medicine*, vol. 9, no. 9, pp. 1–21, Sep. 2020, doi: 10.3390/jcm9092868.
[3] M. Arnold *et al.*, "Global burden of cutaneous melanoma in 2020 and projections to 2040," *JAMA Dermatology*, vol. 158, no. 5, pp. 495–503, May 2022, doi: 10.1001/jamadermatol.2022.0160.
[4] M. A. M. Almeida and I. A. X. Santos, "Classification models for skin tumor detection using texture analysis in medical images," *Journal of Imaging*, vol. 6, no. 6, Jun. 2020, doi: 10.3390/JIMAGING6060051.
[5] M. Mirbabaie, S. Stieglitz, and N. R. J. Frick, "Artificial intelligence in disease diagnostics: a critical review and classification on the current state of research guiding future direction," *Health and Technology*, vol. 11, no. 4, pp. 693–731, Jul. 2021, doi: 10.1007/s12553-021-00555-5.
[6] T. M. Alam *et al.*, "An efficient deep learning-based skin cancer classifier for an imbalanced dataset," *Diagnostics*, vol. 12, no. 9, Aug. 2022, doi: 10.3390/diagnostics12092115.
[7] I. Zaqout, "Diagnosis of skin lesions based on dermoscopic images using image processing techniques," in *Pattern Recognition - Selected Methods and Applications*, London, UK: IntechOpen, 2019, doi: 10.5772/intechopen.88065.
[8] E. M. Senan and M. E. Jadhav, "Analysis of dermoscopy images by using ABCD rule for early detection of skin cancer," *Global Transitions Proceedings*, vol. 2, no. 1, pp. 1–7, Jun. 2021, doi: 10.1016/j.gltp.2021.01.001.
[9] F. Warsi, R. Khanam, S. Kamya, and C. P. S.-Araujo, "An efficient 3D color-texture feature and neural network technique for melanoma detection," *Informatics in Medicine Unlocked*, vol. 17, 2019, doi: 10.1016/j.imu.2019.100176.
[10] P. P. Tumpa and M. A. Kabir, "An artificial neural network based detection and classification of melanoma skin cancer using hybrid texture features," *Sensors International*, vol. 2, 2021, doi: 10.1016/j.sintl.2021.100128.
[11] S. M. Alizadeh and A. Mahloojifar, "Automatic skin cancer detection in dermoscopy images by combining convolutional neural networks and texture features," *International Journal of Imaging Systems and Technology*, vol. 31, no. 2, pp. 695–707, Jun. 2021, doi: 10.1002/ima.22490.
[12] Z. Yu *et al.*, "Convolutional descriptors aggregation via cross-net for skin lesion recognition," *Applied Soft Computing Journal*, vol. 92, Jul. 2020, doi: 10.1016/j.asoc.2020.106281.

[13] S. T. Haider *et al.*, "Illustration and identification of skin lesions in dermoscopic dataset using pattern recognition techniques," in *Advances in Intelligent Systems and Computing*, vol. 1304, Cham, Swiss: Springer, 2021, pp. 976–983, doi: 10.1007/978-3-030-63784-2_118.
[14] A. Pavate and R. Bansode, "Design and analysis of adversarial samples in safety–critical environment: disease prediction system," in *Lecture Notes in Computational Vision and Biomechanics*, vol. 37, Singapore: Springer, 2023, pp. 349–361, doi: 10.1007/978-981-19-0151-5_29.
[15] C. Ding, Y. Li, Z. Lu, S. Wang, and S. Guo, "A resource-efficient feature extraction framework for image processing in IoT devices," *IEEE Transactions on Mobile Computing*, vol. 23, no. 1, pp. 42–55, Jan. 2024, doi: 10.1109/TMC.2022.3218402.
[16] Q. A. Memon, N. Musthafa, M. Masud, and G. Al Ameri, "Enhanced skin cancer classification via Xception model," *International Journal of Advances in Applied Sciences*, vol. 14, no. 1, pp. 69–76, Mar. 2025, doi: 10.11591/ijaas.v14.i1.pp69-76.
[17] P. Tang, Q. Liang, X. Yan, S. Xiang, and D. Zhang, "GP-CNN-DTEL: global-part CNN model with data-transformed ensemble learning for skin lesion classification," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 10, pp. 2870–2882, Oct. 2020, doi: 10.1109/JBHI.2020.2977013.
[18] J. Wu, W. Hu, Y. Wen, W. Tu, and X. Liu, "Skin lesion classification using densely connected convolutional networks with attention residual learning," *Sensors*, vol. 20, no. 24, pp. 1–15, Dec. 2020, doi: 10.3390/s20247080.
[19] H. M. Balaha and A. E. S. Hassan, "Skin cancer diagnosis based on deep transfer learning and sparrow search algorithm," *Neural Computing and Applications*, vol. 35, no. 1, pp. 815–853, Jan. 2023, doi: 10.1007/s00521-022-07762-9.
[20] B. Cassidy, C. Kendrick, A. Brodzicki, J. J.-Korjakowska, and M. H. Yap, "Analysis of the ISIC image datasets: usage, benchmarks and recommendations," *Medical Image Analysis*, vol. 75, Jan. 2022, doi: 10.1016/j.media.2021.102305.
[21] H. K. Gajera, D. R. Nayak, and M. A. Zaveri, "A comprehensive analysis of dermoscopy images for melanoma detection via deep CNN features," *Biomedical Signal Processing and Control*, vol. 79, Jan. 2023, doi: 10.1016/j.bspc.2022.104186.
[22] S. Kumar and A. Kumar, "Lesion preprocessing techniques in automatic melanoma detection system—a comparative study," in *Lecture Notes on Data Engineering and Communications Technologies*, vol. 75, Singapore: Springer, 2022, pp. 357–368, doi: 10.1007/978-981-16-3728-5_27.
[23] A. Mumuni and F. Mumuni, "Data augmentation: a comprehensive survey of modern approaches," *Array*, vol. 16, Dec. 2022, doi: 10.1016/j.array.2022.100258.
[24] A. Magdy, H. Hussein, R. F. A.-Kader, and K. A. El Salam, "Performance enhancement of skin cancer classification using computer vision," *IEEE Access*, vol. 11, pp. 72120–72133, 2023, doi: 10.1109/ACCESS.2023.3294974.
[25] K. Okoye and S. Hosseini, "T-test statistics in R: independent samples, paired sample, and one sample t-tests," in *R Programming*, Singapore: Springer Nature Singapore, 2024, pp. 159–186, doi: 10.1007/978-981-97-3385-9_8.
[26] M. A. Chandra and S. S. Bedi, "Survey on SVM and their application in image classification," *International Journal of Information Technology*, vol. 13, no. 5, pp. 1–11, Oct. 2021, doi: 10.1007/s41870-017-0080-1.

BIOGRAPHIES OF AUTHORS

Shikha Malik received her M.Tech. degree in Electronics and Communication Engineering from Kurukshetra University in 2010 and is currently pursuing a Ph.D. at G.H. Raisoni College of Engineering and Management, Pune, India. She is presently working as an assistant professor at Thakur College of Engineering, Mumbai, India. With 15 years of teaching experience, her research interests include biomedical image processing and ML. She can be contacted at email: shikha.malik.phdetc@ghrcem.raisoni.net.

Vaibhav V. Dixit received his Ph.D. in Electronics and Telecommunication Engineering from Savitribai Phule Pune University (SPPU), Pune, India, in 2015. He currently serves as the principal of Rasiklal M. Dhariwal Sinhgad School of Engineering and as director of the Rasiklal M. Dhariwal Sinhgad Technical Institutes Campus, Warje, Pune, India. His research interests include pattern recognition, neural networks, and biomedical applications. He brings with him over 24 years of academic experience and 5 years of industry exposure. He can be contacted at email: vvdixit73@gmail.com.