



A Pyramid-Based Feature Fusion Framework for Keratoconus Detection Using LightGBM

Prabhu Teja Geddada^{*}, Rajesh Kumar Pullagura^{*}

Department of Electronics and Communication Engineering, Andhra University College of Engineering (A), Andhra University, Visakhapatnam 530003, India

Corresponding Author Email: prabhuteja.rs@andhrauniversity.edu.in

Copyright: ©2025 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/isi.300815>

ABSTRACT

Received: 9 July 2025

Revised: 19 August 2025

Accepted: 24 August 2025

Available online: 31 August 2025

Keywords:

keratoconus, multi-scale image pyramids, HOG, LightGBM, multi-class classification

Keratoconus is a degenerative eye disorder that affects the cornea. It is a progressive eye disorder that leads to irregular astigmatism and decrease in visual acuity as the cornea deforms and protrudes outward assuming a cone shape. The early diagnosis can be challenging as the disease can be asymptomatic. This study proposes a machine learning pipeline for the classification of the keratoconus using multi-scale feature extraction from Pentacam derived corneal topographic maps. A labeled dataset comprising 2961 images, categorized into Keratoconus, Normal and Suspect classes, is used in this study. Multi scale image representations are generated using Gaussian and Laplacian pyramids, alongside a patch-based pyramid. Gradient-based features are extracted from multi-scaled images using Histogram of Oriented Gradients (HOG) and L1-regularised Logistic Regression is used for feature selection. An optimized Light Gradient Boosting Machine (LightGBM) classifier is employed for classification. Experimental results show that Gaussian pyramid based multi-scale HOG features consistently outperformed Laplacian and patch-based approaches with an overall accuracy of 85.62%, F1-score of 0.85 and AUC score of 0.95, confirming the effectiveness of multi-scale analysis in corneal disease classification.

1. INTRODUCTION

Keratoconus is a progressive, bilateral disorder [1] of the cornea characterized by thinning and bulging of the corneal tissue, resulting in conical shape of the cornea [2]. This change in the corneal shape leads to irregular astigmatism and vision impairment [3]. The disease usually manifests during adolescence and progresses until the fourth decade of life [4]. Unlike the early stages of the disease, moderate to advanced stages are easily diagnosed due to the presence of classic signs observable via slit-lamp and clinical examination [5]. The early detection of the disease is crucial as progression can be halted using medical intervention [6]. However, the detection of the disease in its incipient stage is often challenging as the early signs are often subtle and not easily detectable through conventional methods [7].

In recent years, Machine Learning (ML) approaches have shown promise as an automated diagnostic tool in healthcare [8]. These techniques offer objective, data-driven classification by identifying subtle patterns and abnormalities, potentially enhancing diagnostic accuracy and aiding healthcare professionals when integrated into automated screening systems [9].

In this paper, multi-scale representation of corneal maps using image pyramids is proposed. Histogram of Oriented Gradients (HOG) is then applied to extract features across these scales of corneal maps in $L^*a^*b^*$ colour space. Gradient based features like HOG measure the rate and direction of change in an image [10]. Since, Keratoconus is defined by a

change in corneal shape, HOG descriptors are well suited to detect these changes in cornea. At the fine scale, HOG captures localized structural changes such as steepening and surface irregularities, which are often indicative of early-stage Keratoconus. At the coarse level, it highlights broader deformation patterns, including generalized thinning and changes in corneal curvature that are characteristics of the advanced disease [11]. L1-penalized Logistic Regression is used for feature selection and hyperparameter tuned LGB classifier is used to evaluate the classification performance of the model.

The main aim of this study is to develop an automated multiscale image analysis framework for keratoconus detection from corneal topographic maps using multiscale image representation and feature extraction via image pyramids and Histogram of Oriented Gradient (HOG) Descriptors and LightGBM for classification.

The structure of this paper is outlined as follows. section 2 reviews previous studies on the detection of Keratoconus, focusing on applications of machine learning techniques as well as image analysis approaches based on pyramid structures and HOG in medical imaging. Section 3 describes the proposed methodology in detail, covering aspects such as dataset collection and pre-processing, imbalance handling, multi scale representation of images using Gaussian, Laplacian and Patch-based pyramids, feature extraction using HOG Descriptors in $L^*a^*b^*$ color space, classification strategy using LightGBM classification framework and statistical analysis. Section 4 discusses experimental setup,

performance evaluation methods and results. A thorough analysis and interpretation of the findings are provided in section 5. Finally, section 6 summarizes the main conclusions of this study and offers suggestions for future research directions.

2. LITERATURE SURVEY

In a study done by Anwar and Özbilge [12] YOLOv8 is implemented to capture the ROI from the 1000 eye images, and the extracted region is used to train pre-trained CNN models. In their study, Xception and InceptionResNetV2 outperformed with accuracies of 93.80% and 94.23% when combined with ROI, and 91.43% and 91.45% without ROI in classifying Normal, mild and advanced keratoconus stages. Hashim and Mazinani [13] in their study have utilized histogram equalization and PCA based feature extraction to train ML and CNN models for keratoconus detection based on 400 images and achieved varied accuracies ranging from 70.9% to 99% for a binary classification of the disease. Awwad et al. [14] proposed a study for keratoconus detection which included 349 eyes and six parameters from corneal thickness progression maps. They have trained over 7 ML algorithms and have achieved varied accuracies of 87%, 91% and 100% for four, three and two-class classification respectively. Li et al. [15] employed 11 predictors including age, gender and ocular biometrics derived from 1523 eyes to train ML regression models to predict first applanation (SP-A1) and corvis biomechanical index for Chinese populations (cCBI) for keratoconus detection. They have achieved AUCs of 0.939 and 0.881 for predicted SP-A1 and cCBI, respectively between keratoconus and normal classes. In a study done by Lu et al. [16] multiple imaging modalities such as air puff tonometry, Scheimpflug tomography and spectral domain optical coherence tomography (SD-OCT) are analyzed with Random Forest and neural networks. They have included 599 eyes to train models and have reported the highest AUC of 0.902 when Random Forest algorithm was used to select features from SD-OCT and air puff tonometry. Al-Timemy et al. [17] proposed a hybrid deep learning model for keratoconus detection by training 7 efficient-netb0 models and extracted 7000 deep features to train SVM classifier and achieved an accuracy of 81.6% for classification of normal, suspect and keratoconus classes. Ahmed et al. [18] compared different deep learning models on an augmented corneal map dataset of 16016 samples and achieved an accuracy of 98% with MobileNet-v2. However, it should be noted that all the train validation and test sets have been augmented with data augmentation techniques which could have contributed to inflated classification accuracy. In a study done by Gandhi et al. [19] 3962 corneal topographic maps based on 10-class Amsler Krumeich classification with augmentation are used to train deep learning models, they have reported a performance accuracy of 77.43% with VGG19.

Jawad et al. [20] in their work suggested a retinal image enhancement by segmenting the retinal image regions from RGB to $L^*a^*b^*$ space and used adaptive histogram equalization on L channel for image enhancement. In a comparative study [21] of L1-regularised logistic regression methods, the performance of IRLs and Glmnet is evaluated across high dimensional scenarios with varying sample sizes and predictor correlations. The results indicated that both methods effectively selected the informative features in low to

moderately correlated settings. Ozyurt et al. [22] employed an exemplar pyramid-based feature extraction for covid-19 detection using deep learning algorithms.

Most of the research done on keratoconus detection based on corneal imaging or raw data is based on region-specific or institution-specific data which limits the generalizability of the model [23].

3. METHODOLOGY

3.1 Data collection and pre-processing

In this work, a total of 2961 corneal images obtained from 542 eyes using the Oculus Pentacam imaging system is utilized. This dataset, which had been previously compiled and made publicly accessible through earlier publications [17], includes three categories: Keratoconus, Normal and Suspect. Image labeling was performed by three experienced corneal specialists following established guidelines as reported in prior studies [17].

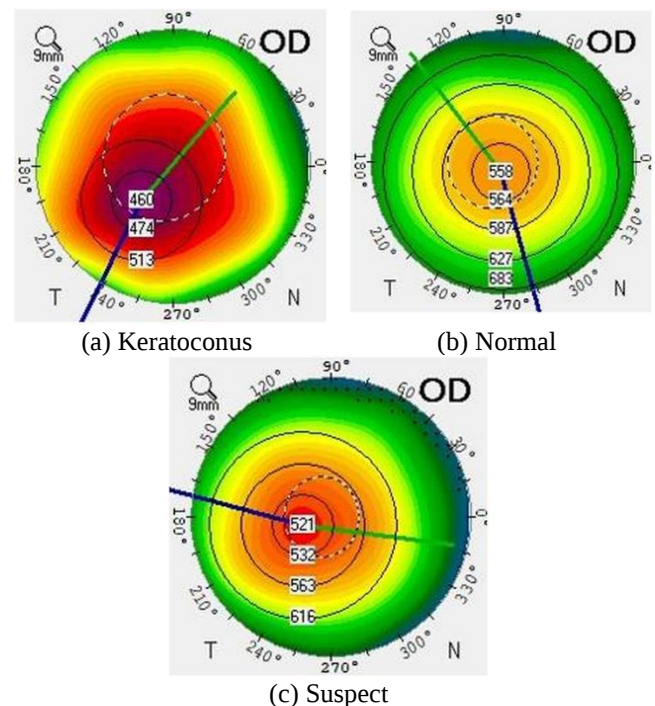


Figure 1. Corneal topography images

Specifically, the dataset comprises of 1050 images of each Keratoconus and Normal categories, while the Suspect group contains 861 images. For model development and evaluation, the entire dataset is partitioned into stratified split of 70% for training and 15% each for validation and testing. This resulted in 2072 images in the train set, 444 images in the validation set and 445 images in the test set. The size of each corneal map is $224 \times 224 \times 3$. Figure 1. shows the corneal topographic maps of keratoconic, normal and suspect eyes derived from Pentacam device.

To address the issue of class imbalance within the dataset, class weights are computed using a balanced weighting strategy. This technique assigns proportionally higher weights to classes with fewer samples, thereby promoting fairer learning during the model training. The weight for each class w_j is determined using the formula:

$$w_j = \frac{N_{\text{train}}}{C \times N_j} \quad (1)$$

where,

w_j represents the weight assigned to class j ,

N_{train} is the total number of images in the train set,

C is the total number of categories, and

N_j is the number of training images belonging to class j .

Besides class weighting, Synthetic Minority Over-Sampling Technique (SMOTE) is also employed to handle data imbalances. SMOTE [24] is used to equalize the number of samples in the suspect class to that of normal and keratoconus classes in the training set. Validation and test sets are left untouched. An extra of 133 synthetic samples from suspect class are generated to balance the dataset.

To enhance the discriminative quality of the corneal maps, all images are converted from RGB to L*a*b* color space prior to feature extraction. This conversion is inspired by prior work [20].

3.2 Multi-scale representation using Image pyramids

To capture multi-scale features from corneal topography, three distinct pyramid-based image representations were constructed. Gaussian pyramids, Laplacian pyramids [25] and patch-based pyramid [22]. Each image was first converted to the L*a*b* colour space to enhance the corneal maps. The Gaussian pyramid is constructed by iteratively applying Gaussian smoothing followed by a downsampling with factor of 2 at each level. A total of 3 Gaussian levels is generated beyond the original image. A Laplacian pyramid is derived by subtracting the upsampled lower-level Gaussian image from its corresponding higher-level image, emphasizing localized edge and texture details at each size. Three Laplacian levels corresponding to the three Gaussian levels are computed. Additionally, a patch-based image pyramid is constructed by directly resizing image with downsampling factor of 2 per level, without applying any smoothing operation. At each resolution, the downsampled image is partitioned into non-overlapping 28×28 patches.

3.3 Histogram of Oriented Gradients

The Histogram of Oriented Gradients (HOG) is a feature descriptor in computer vision and image analysis tasks [26]. HOG features are independently extracted from each pyramid type to capture texture and edge information. For both the Gaussian and the Laplacian pyramid, HOG features are computed from each image level independently across all scales. In the case of the patch-based pyramid, each downsampled image is subdivided into non-overlapping patches of size 28×28 pixels, and HOG features are extracted individually from each patch. A stride of 28 pixels is used to ensure complete spatial coverage without overlapping patches.

Table 1 presents the HOG descriptor parameters used for feature extraction at each pyramid level.

The parameters in Table 1 in are chosen in a way to strike a balance between capturing discriminative features and maintaining computational efficiency. 7 orientation bins were chosen to capture major edge directions while keeping computational cost to a minimum. 4x4 pixels per cell is chosen to capture finer details when compared to the default of 8×8 pixels per cell. A 2×2 cells per block was found to be the most effective in capturing local variations and small enough to not

average out the important detail. L2-norm with hysteresis is chosen as it prevents few high contrast gradients from dominating the entire feature vector [10].

Table 1. HOG descriptor parameters

Parameter	Description	Value
Orientations	Number of gradient orientation bins per cell	7
Pixels Per Cell	Size of each cell over which the gradient histogram is computed	4×4
Cells per Block	Number of cells in each block for local normalization	2×2
Block Normalization	Normalization applied to improve invariance	L2-Hys

3.4 Feature selection

To reduce the feature dimensionality and retain only the most discriminative features, a coordinate descent based L1-regularized Logistic Regression is employed for feature selection with solver set to 'liblinear', max iterations are set to 10000 and the parameter 'C' is set to 1.0. L1 penalty encourages sparsity by driving the coefficients of less informative features to zero effectively selecting a subset of relevant features for classification [21]. The model is trained on scaled HOG feature vectors extracted from each pyramid representation, and the selected features are subsequently used as input to the final LightGBM classifier. This approach improved computational efficiency and reduced the risk of overfitting in high dimensional feature spaces.

3.5 LightGBM classifier

LightGBM [27] is an efficient gradient boosting framework designed for machine learning applications. Unlike conventional gradient boosting models that use a level-wise approach, LightGBM adopts a leaf-wise tree growth strategy. In this approach, the leaf with the highest potential gain is selected and split at each iteration, enabling the model to focus on the most informative regions of the feature space. The selected features by the L1-Logistic Regression are used as input to the LightGBM classifier. To identify the optimal hyperparameters for the LightGBM classifier, a random search cross-validation (RandomizedSearchCV) strategy is employed. This approach samples a predefined number of random combinations from specified hyperparameter distributions, offering a more computationally efficient alternative. In this study, 30 randomly selected hyperparameter configurations are evaluated for each feature extraction pipeline.

Table 2 presents description of the hyperparameters employed in the model training process along with their respective search ranges utilized during optimization.

To maintain data integrity and prevent overlap between training and validation subsets, a PredefinedSplit (ps) is utilized during cross-validation. The model selects the best performing hyperparameter set after training on the train set based on the validation accuracy, evaluated on a reserved validation set containing 444 samples. Once the best configuration of the hyperparameter set is found, the model's classification performance is evaluated finally on the test set containing 445 samples. The workflow for proposed machine learning framework for keratoconus classification is depicted in Figure 2.

Table 2. Hyperparameters and search ranges for model optimization

Hyperparameter	Description	Search Range
colsample_bytree	Subsample ratio of columns when constructing each tree	0.5 to 1.0
learning_rate	Step size shrinkage used in updates	0.01 to 0.2
max_depth	Maximum depth of each tree	3 to 10
n_estimators	Number of boosting iterations	50 to 500
num_leaves	Number of leaves in full trees	20 to 150
subsample	Subsample ratio of training instances	0.5 to 1.0

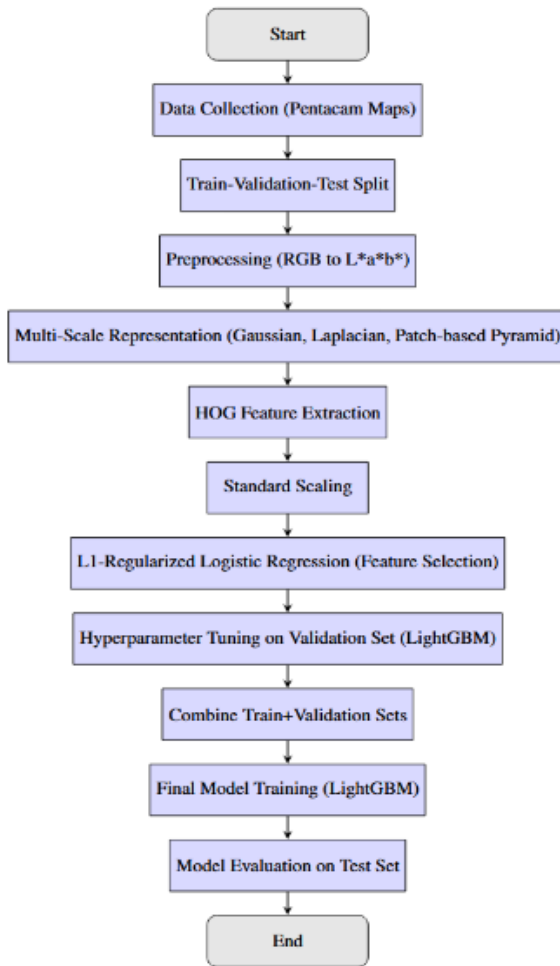
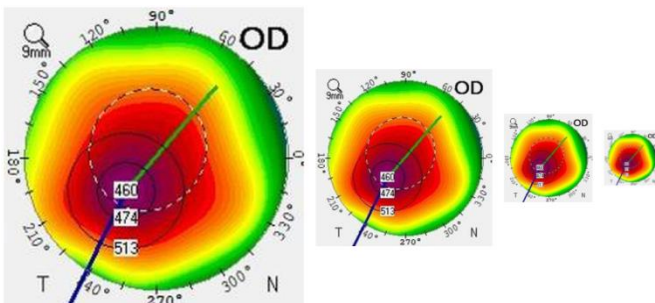
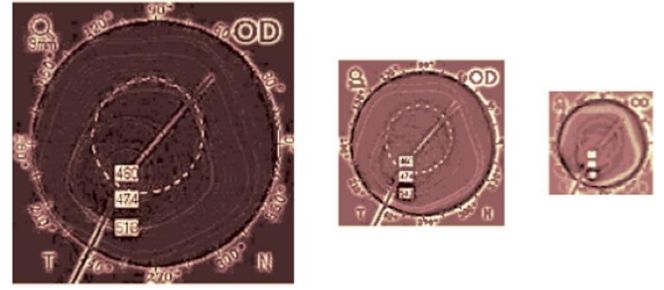


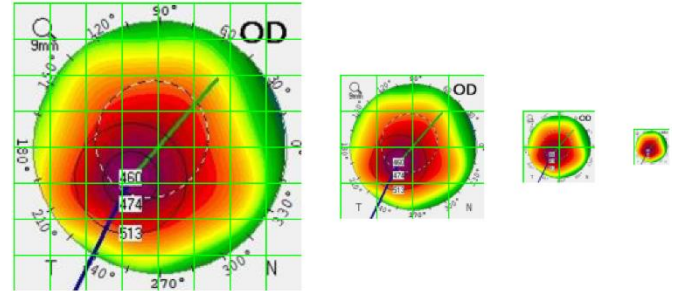
Figure 2. Workflow of the proposed machine learning framework for keratoconus detection



(a) Gaussian pyramid constructed using 3 downsampling steps resulting in 4 levels with resolutions: 224×224, 112×112, 56×56 and 28×28



(b) Laplacian pyramid constructed using 3 downsampling steps resulting in 3 levels with resolutions: 224×224, 112×112, and 56×56



(c) Patch-based pyramid constructed using 3 downsampling steps without smoothing resulting in 4 levels with resolutions: 224×224, 112×112, 56×56 and 28×28

Figure 3. The Gaussian, Laplacian and Patch-based pyramids

Furthermore, bootstrapping with $n = 1000$ resamples is employed to account for variability. For each resample, models are trained and evaluated. The results of each iteration are recorded and stored in .csv files. A one-way Analysis of Variance (ANOVA) is performed across the three pyramid-HOG pipelines to test whether the mean performance metrics differed significantly. Tukey's Honestly Significant Difference (HSD) post-hoc test is used for pairwise comparison between the three pyramid-HOG pipelines that are statistically significant. Pairwise t-tests are conducted on bootstrap derived distributions to compare class weighting and SMOTE imbalance handling. These statistical methods are used in machine learning research to compare model performance [28]. Figure 3 shows the corneal maps represented at different scales using Gaussian, Laplacian and Patch-based pyramid.

4. PERFORMANCE EVALUATION AND RESULTS

To evaluate the performance of the proposed automated machine learning framework for classification of Keratoconus, the metrics like accuracy, sensitivity, specificity, f1-score and AUC score along with confusion matrix plots and AUROC are considered.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (3)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (4)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (5)$$

$$F1 - score = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6)$$

Table 3 presents the optimal hyperparameter configurations determined for the LightGBM classifier associated with the multi-scale image representation and feature extraction pipelines.

Table 3. Best LightGBM hyperparameters obtained for each pyramid-HOG pipeline (Randomized Search)

Hyperparameter	Gaussian Pyramid - HOG	Laplacian Pyramid - HOG	Patch-based Pyramid - HOG
colsample_bytree	0.5103	0.8737	0.8926
learning_rate	0.2040	0.1179	0.0499
max_depth	6	4	9
n_estimators	463	266	293
num_leaves	57	143	83
subsample	0.5004	0.6380	0.7334

Table 4. Class-wise performance metrics for each pyramid type with HOG features

Method	Class	Precision	Recall (Sensitivity)	Specificity	F1 Score	AUC (OVR)
Gaussian Pyramid-HOG	Keratoconus	0.9536	0.9114	0.9756	0.9320	0.9907
	Normal	0.8776	0.8165	0.9373	0.8459	0.9617
	Suspect	0.7347	0.8372	0.8766	0.7826	0.9450
Laplacian Pyramid-HOG	Keratoconus	0.8931	0.8987	0.9408	0.8959	0.9830
	Normal	0.8947	0.7532	0.9512	0.8179	0.9512
	Suspect	0.6863	0.8140	0.8481	0.7447	0.9296
Patch-based Pyramid-HOG	Keratoconus	0.9363	0.9304	0.9652	0.9333	0.9858
	Normal	0.8487	0.8165	0.9199	0.8323	0.9563
	Suspect	0.7426	0.7829	0.8892	0.7623	0.9348

Table 4 Summarizes the performance metrics for each class, Keratoconus, Normal and Suspect across the three pyramid-HOG pipelines. This comparative evaluation highlights the strengths and limitations of each pyramid-HOG technique in classification of Keratoconus.

Table 5. Overall and averaged performance metrics for each pyramid type with HOG features LightGBM hyperparameters obtained for each pyramid-HOG pipeline

Metric	Gaussian Pyramid - HOG	Laplacian Pyramid - HOG	Patch-based Pyramid - HOG
Weighted Avg Precision	0.8632	0.8337	0.8491
Weighted Avg Recall	0.8562	0.8225	0.8472
Weighted Avg F1 Score	0.8581	0.8244	0.8479
Macro Avg Precision	0.8553	0.8247	0.8425
Macro Avg Recall	0.8550	0.8220	0.8433
Macro Avg F1 Score	0.8535	0.8195	0.8426
Overall Specificity	0.9298	0.9134	0.9248
Weighted Avg AUC	0.9671	0.9562	0.9606
Macro Avg AUC	0.9658	0.9546	0.9590

Table 5 provides an overall comparative assessment of classification performance of all three pipelines. It reports weighted and macro-averaged values for precision, Sensitivity (Recall), F1-score, AUC score and overall specificity.

Table 6 presents the statistical comparison of all three pyramid-HOG pipelines based on one-way ANOVA and

Tukey's Honestly Significant Difference (HSD) post-hoc test.

Table 6. Statistical analysis of each pyramid-HOG pipeline

Metric	F-statistic	Methods Compared	Mean difference
Accuracy	959.5	Gaussian vs. Laplacian	-0.0336
		Gaussian vs. Patch-based	-0.0090
		Laplacian vs. Patch-based	0.0246
F1-Macro	941.79	Gaussian vs. Laplacian	-0.0340
		Gaussian vs. Patch-based	-0.0110
		Laplacian vs. Patch-based	0.0230
Precision-Macro	789.09	Gaussian vs. Laplacian	-0.0305
		Gaussian vs. Patch-based	-0.0128
		Laplacian vs. Patch-based	0.0177
Recall-Macro	874.52	Gaussian vs. Laplacian	-0.0330
		Gaussian vs. Patch-based	-0.0118
		Laplacian vs. Patch-based	0.0212
AUC-Macro	739.04	Gaussian vs. Laplacian	-0.0112
		Gaussian vs. Patch-based	-0.0069
		Laplacian vs. Patch-based	0.0043

Note: All p-values for the one-way ANOVA and adjusted p-values for Tukey's HSD post-hoc test are < 0.001 , indicating a statistically significant difference for all comparisons.

Table 7. Comparison of class weighting and SMOTE performance

Method	Metric	Mean score (Class Weighting)	Mean score (SMOTE)	Mean difference
Gaussian	Accuracy	0.8564	0.8631	0.0067
	F1-Macro	0.8533	0.8601	0.0068
	Precision-Macro	0.8555	0.8616	0.0061
	Recall-Macro	0.8553	0.8620	0.0067
	AUC-Macro	0.9659	0.9638	-0.0021
	Accuracy	0.8228	0.8230	0.0002
Laplacian	F1-Macro	0.8193	0.8168	-0.0025
	Precision-Macro	0.8250	0.8186	-0.0064
	Recall-Macro	0.8223	0.8183	-0.0040
	AUC-Macro	0.9547	0.9515	-0.0032
	Accuracy	0.8474	0.8266	-0.0208
	F1-Macro	0.8423	0.8214	-0.0209
Patch-based	Precision-Macro	0.8427	0.8219	-0.0208
	Recall-Macro	0.8435	0.8225	-0.0210
	AUC-Macro	0.9590	0.9537	-0.0053

Note: All p -values for the paired t-test are < 0.001 , indicating a statistically significant difference between SMOTE and Class Weighting technique except for the accuracy metric in the Laplacian method where $p > 0.05$.

Table 8. Summary of related studies and their classification performance

Study	Classes	Dataset	Technique Used	Accuracy (%)
Ahmed et al. [18]	3-class	16,016 augmented images	MobileNet v2	98.00
Al-Sharify et al. [29]	5-class	491 subjects/8 parameters	Decision Tree/Nearest Neighbour Analysis	65.7/62.6
Gandhi et al. [19]	10-class	3962 maps with augmentation	VGG19	77.43
Al-Timemy et al. [17]	3-class	542 eyes/7 maps	EfficientNet-b0+SVM	81.60
This Study	3-class	542 eyes/7maps	Image pyramid+HOG+LightGBM	85.62

Table 7 provides statistical analysis of all three pyramid-HOG pipelines using class weighting and SMOTE techniques.

Table 8 gives a comparison of recent studies on keratoconus detection based on the number of classes, dataset characteristics, techniques employed and achieved accuracy.

Table 9 gives the classification accuracies achieved on the validation and independent test sets across the three multi-scale feature extraction pipelines.

Table 9. Validation and test accuracies (%) with pyramid levels for each multi-scale feature extraction pipeline

Pyramid Type	Downsampling Steps	Gaussian Levels (incl. Original)	Laplacian Levels	Validation Accuracy (%)	Test Accuracy (%)
Gaussian Pyramid	3	4	N/A	82.21	85.62
Laplacian Pyramid	3	4	3	80.63	82.25
Patch-based Pyramid	3	N/A	N/A	80.63	84.72

5. DISCUSSION

In this study, an automated machine learning framework for classification of Keratoconus using image pyramids, HOG and LightGBM classifier is proposed. A stratified split of 70%-15%-15% of train validation and test split is employed on a dataset of 2961 pentacam derived corneal maps containing keratoconus, normal and suspect classes. The dataset is imbalanced with the “suspect” class having relatively fewer instances. To address this, class weighting based on inverse class frequency is applied during training, ensuring that minority class errors are penalized more heavily during optimization. RGB to L*a*b* conversion is used as the preprocessing step.

Three distinct multi-scale feature extraction approaches are investigated: Gaussian pyramid-HOG Laplacian Pyramid and a Patch-based Pyramid, with 3 levels each and downsampling factor of 2 per each level. Each technique aims to enhance feature representation by decomposing spatial information at multiple levels for detecting keratoconus. These decompositions are followed by HOG feature extraction,

which captured the relevant discriminative features in the image. These extracted features are scaled using ‘StandardScaler’ and LightGBM classifier is trained with hyperparameter optimization via randomized search. From Table 3. It is evident that Gaussian Pyramid and HOG required a relatively higher learning rate and fewer leaves suggesting its multi-scale smoothed features resulted in a simple decision boundary. In contrast, the patch-based method required a ‘max_depth’ of 9 with a lower learning rate. All simulations are performed on Pop!_OS 22.04 (Ubuntu based Linux distribution) with 32GB of RAM and Ryzen 5600H Hexacore CPU. Model evaluation metrics in Table 4. revealed that all three pyramid strategies achieved strong classification performance, particularly for the keratoconus class, where precision and recall scores are consistently high. For instance, the patch-based pyramid model achieved a recall of 0.9304 for keratoconus class. However, suspect class detection remained challenging across all pipelines with precision ranging from 0.6863 to 0.7426. This underperformance is likely due to the intermediate morphological characteristics of suspect cases, which share traits with both normal and keratoconus corneas.

Nevertheless, the patch-based pyramid exhibited slightly improved performance for suspect class classification, likely due to its ability to focus on localized regions where early ectatic changes are or likely to appear.

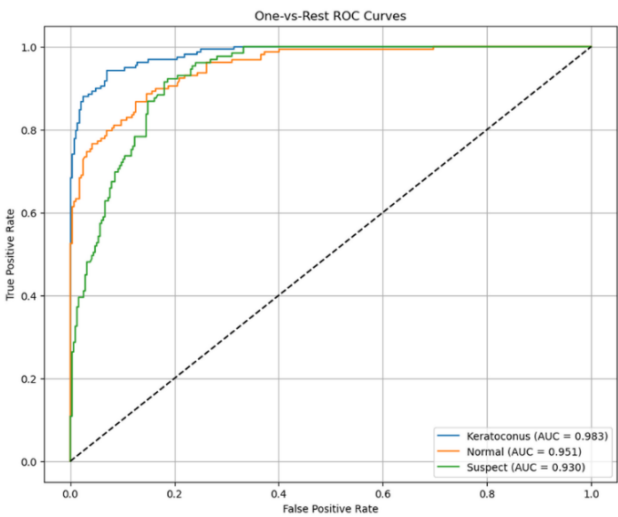
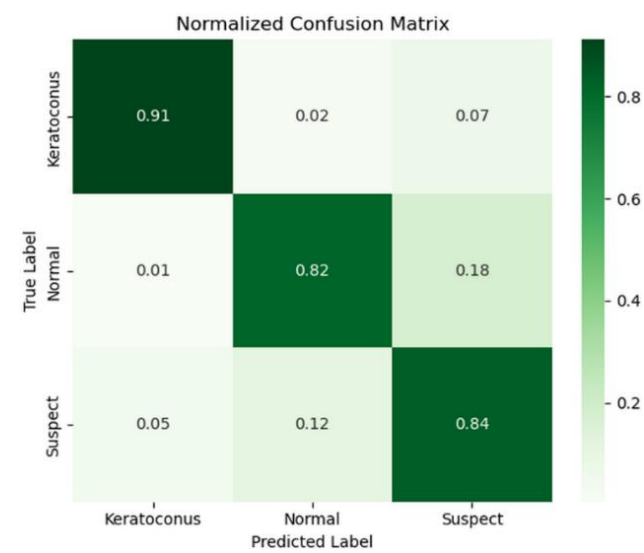


Figure 5. The confusion matrix plot and AUROC for Laplacian pyramid-based HOG feature extraction pipeline

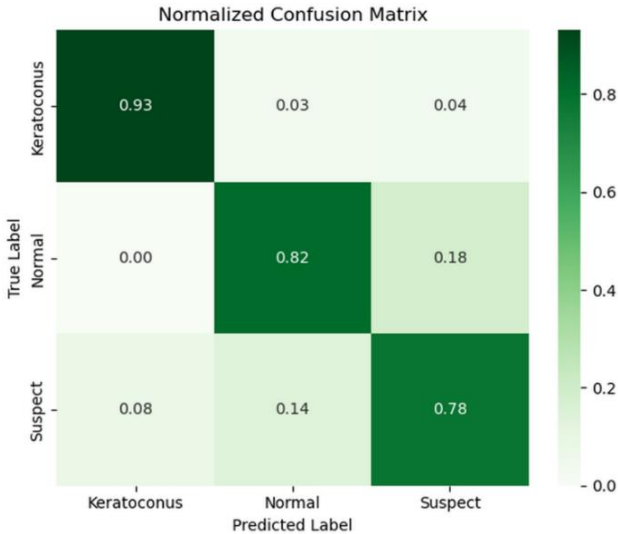
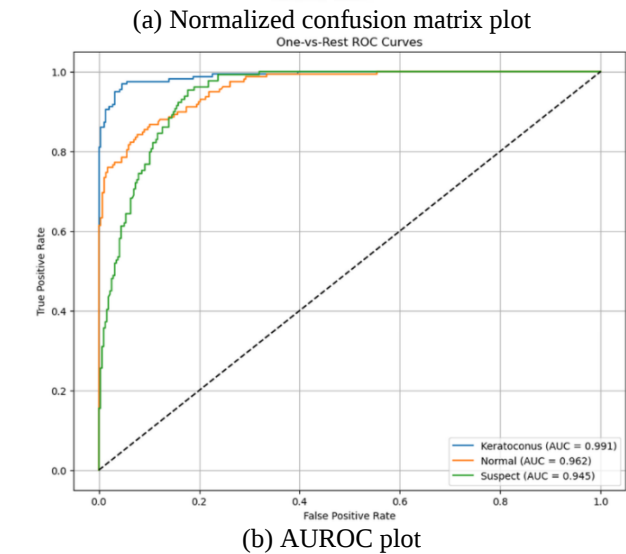


Figure 4. The confusion matrix plot and AUROC plot for Gaussian pyramid-based HOG feature extraction pipeline

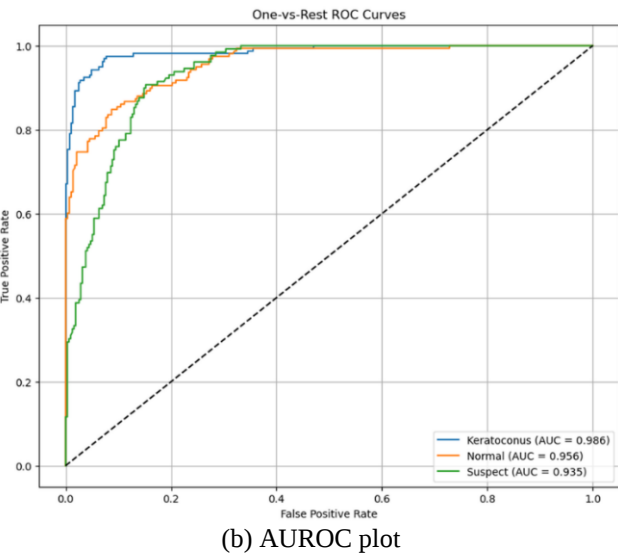
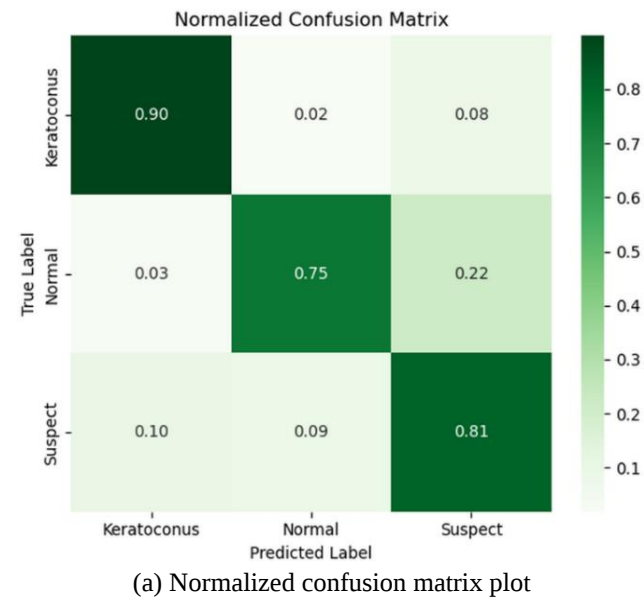


Figure 6. The confusion matrix plot and AUROC for patch-based pyramid and HOG feature extraction pipeline

Table 5 provides the macro and weighted average metrics

for each model. The Gaussian Pyramid and HOG pipeline delivered the best overall performance with a weighted average F1 score of 0.8581 and macro average AUC of 0.9568. Although, patch-based method trailed slightly in macro and weighted averages, it maintained a balanced performance across all classes, suggesting improving robustness in handling class imbalance. These results emphasize the effectiveness of image pyramids in enhancing class

separability in corneal maps and validate the suitability of HOG descriptors in keratoconus classification. Figures 4, 5 and 6 show the confusion matrix plot and AUROC plot for all the three pipelines.

Figure 7 provides feature maps of corneal images across different scales and color channels. These feature maps confirm that model’s decisions are based on gradient-based information extracted by HOG.

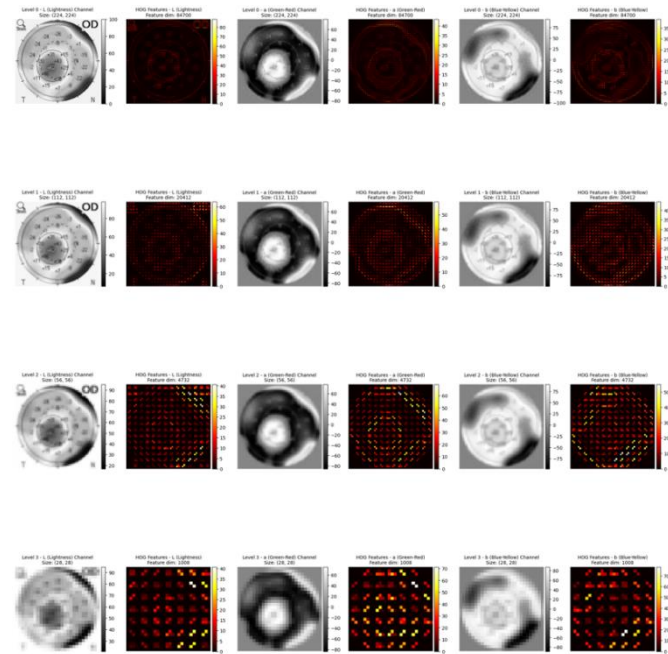


Figure 7. The feature maps of the corneal image at multi scales

Table 6 provides statistical evidence for the performance differences observed in Tables 4 and 5. The F-statistic shown for each metric is very high (all are over 739). The high F-statistic shows that the probability of average performance of the three pyramid-HOG pipelines being the same is very low. The high F-statistic coupled with p -value < 0.001 indicates that the differences are significant. Tukey’s HSD post-hoc test confirms that the difference between each pair of pipelines (Gaussian-HOG vs. Laplacian-HOG, Gaussian-HOG vs. Laplacian-HOG and Laplacian vs. Patch-based-HOG) is also statistically significant with adjusted p -values < 0.001 . The positive mean difference indicates patch-based-HOG outperformed Laplacian-HOG in Laplacian vs. Patch-based HOG. The negative mean difference in Gaussian-HOG vs. Laplacian-HOG and Gaussian-HOG vs. Patch-based-HOG indicates that Gaussian-HOG has outperformed both the other methods.

Table 7 presents the results of a pairwise t-test comparing the performance of Class-Weighting and SMOTE within each of the three pyramid-HOG pipelines. This analysis statistically validates the effectiveness of each imbalance handling technique. The negative mean difference between Class Weighting and SMOTE indicates that Class Weighting has performed significantly better than SMOTE. The positive mean difference indicates that SMOTE has outperformed Class Weighting technique. For the Gaussian pyramid-HOG SMOTE slightly outperforms Class Weighting across most metrics with a positive mean difference with the exception of AUC-Macro. The p -values are less than 0.001, showing that the small differences are significant. In the Laplacian pyramid-HOG the difference between the two methods is very little.

The p -value for accuracy metric is > 0.05 indicates that mean difference in accuracy is not significant. For all other metrics, the p -values are < 0.001 . For the patch-based pyramid-HOG Class Weighting significantly outperforms the SMOTE in all metrics with mean differences being negative and substantial and corresponding p -values < 0.001 . This indicates that the Class Weighting is the superior technique for handling class imbalance when using patch-based pyramid-HOG pipeline.

Compared to existing studies in Table 8. The proposed approach achieved a classification accuracy of 85.62%, outperforming the classification models applied to corneal imaging, such as the study by Al-Timemy et al. [17] and the study by Gandhi [19]. While Ahmed et al. [18] reported a higher accuracy of 98% using MobileNet-v2, their model was trained on over 16,000 augmented images, whereas the current study is done on a non-augmented dataset. This suggests that hand crafted features, when combined with optimized models, could match deep learning models in data-limited scenarios.

As shown in Table 9, all models maintained high generalization capability, with test accuracy above 82%. The Gaussian pyramid-HOG has achieved a test accuracy of 85.62%. To assess the reliability of this result, bootstrap resampling method is employed with 1000 samples, which yielded a mean accuracy of 85.64% (95% CI: 82.25%-88.76%) and a standard deviation of 0.0170. The relatively lower performance of Laplacian pyramid in both validation and test sets may stem from its emphasis on high-frequency details which, while important, may introduce noise.

Despite promising results, the proposed pipeline has limitations. First, the lower sensitivity in suspect class across all models indicates the inherent difficulty of detecting suspect

classes. This challenge is consistent with clinical findings, where subclinical keratoconus often overlaps morphologically with normal corneas. Second, while the use of HOG feature enables interpretability, they may lack the expressiveness of learned deep features. For early diagnosis of the disease, the detection of suspect class from normal is crucial. Instead of relying solely on HOG features, attention networks like wavelet guided attention nets, multi-scale attention layers within CNN or a hybrid of CNN models and handcrafted features can be employed to detect the subtle variations in the corneal topography map of a suspect class. In addition, explainable AI, Grad cam visualization might help the model focus on the regions of the corneal maps that are highly discriminative in classifying suspect versus normal cases.

6. CONCLUSION

In this study, a comparative evaluation of three multi-scale representation techniques, Gaussian Pyramid, Laplacian Pyramid and Patch-based Pyramid, is conducted for the task of Keratoconus classification using Pentacam derived corneal topographic maps. Each multi scale representation is coupled with Histogram of Oriented Gradients feature extraction and classification is performed using optimized LightGBM classifier.

The experimental results show that Gaussian Pyramid-HOG consistently achieved better performance with an accuracy of 85.62%, F1 score of 0.85 and AUC score of 0.96. The patch-based pyramid showed competitive results, while Laplacian pyramid exhibited comparatively lower classification performance. The bootstrap CIs, ANOVA and Tukey's post-hoc tests proved that choice of pyramid-HOG pipeline has a statistically significant impact on performance.

This study also performed pairwise t-tests on two imbalance handling techniques. The results showed that, while SMOTE is more effective with Gaussian-HOG pipeline, Class Weighting performed significantly better with patch-based pyramid-HOG pipeline.

Overall, the findings confirm the importance of multi-scale texture analysis in medical image classification tasks and highlight the effectiveness of combining HOG descriptors with image pyramids for robust and efficient feature extraction in Keratoconus classification. Future work will explore the integration of LSTM, wavelet guided attention networks and meta-heuristic algorithms to enhance classification accuracy and generalizability in clinical applications.

REFERENCES

- [1] Gomes, J.A.P., Rapuano, C.J., Belin, M.W., Ambrósio, R.J. (2015). Global consensus on keratoconus diagnosis. *Cornea*, 34(12): e38-e39. <https://doi.org/10.1097/ICO.0000000000000623>
- [2] Garcia-Ferrer, F.J., Akpek, E.K., Amescua, G., Mah, F.S., Dunn, S.P. (2019). Corneal ectasia preferred practice pattern®. *Ophthalmology*, 126(1): 170-215. <https://doi.org/10.1016/j.ophtha.2018.10.021>
- [3] Bui, A.D., Truong, A., Pasricha, N.D., Indaram, M. (2023). Keratoconus diagnosis and treatment: Recent advances and future directions. *Clinical Ophthalmology*, 17: 2705-2718. <https://doi.org/10.2147/OPTH.S392665>
- [4] Ng, J.M., Lin, K.K., Lee, J.S., Chen, W.M., Hou, C.H., See, L.C. (2024). Incidence and prevalence of keratoconus in Taiwan during 2000–2018 and their association with the use of corneal topography and tomography. *Eye*, 38(4): 745-751. <https://doi.org/10.1038/s41433-023-02767-7>
- [5] de Sanctis, U., Loiacono, C., Richiardi, L., Turco, D., Mutani, B., Grignolo, F.M. (2008). Sensitivity and specificity of posterior corneal elevation measured by Pentacam in discriminating keratoconus/subclinical keratoconus. *Ophthalmology*, 115(9): 1534-1539. <https://doi.org/10.1016/j.ophtha.2008.02.020>
- [6] Said, A.G.A., Piñero, D.P., Shneor, E. (2023). Revisiting the oil droplet sign in keratoconus: Utility for early keratoconus diagnosis and screening. *Ophthalmic and Physiological Optics*, 43(1): 83-92. <https://doi.org/10.1111/opo.13066>
- [7] Gordon-Shaag, A., Millodot, M., Ifrah, R., Shneor, E. (2012). Aberrations and topography in normal, keratoconus-suspect, and keratoconic eyes. *Optometry and Vision Science*, 89(4): 411-418. <https://doi.org/10.1097/OPX.0b013e318249d727>
- [8] Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., et al. (2019). A guide to deep learning in healthcare. *Nature Medicine*, 25: 24-29. <https://doi.org/10.1038/s41591-018-0316-z>
- [9] Korot, E., Guan, Z., Ferraz, D., Wagner, S.K., et al. (2021). Code-free deep learning for multi-modality medical image classification. *Nature Machine Intelligence*, 3: 288-298. <https://doi.org/10.1038/s42256-021-00305-2>
- [10] Dalal, N., Triggs, B. (2005). Histograms of oriented gradients for human detection. In 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), San Diego, CA, USA, pp. 886-893. <https://doi.org/10.1109/CVPR.2005.177>
- [11] Balaji, K., Gobalakrishnan, N. (2025). An optimized multi-scale dilated attention layer for keratoconus disease classification. *International Ophthalmology*, 45: 318. <https://doi.org/10.1007/s10792-025-03688-y>
- [12] Anwar, M.S., Özbilge, E. (2025). Detection of keratoconus through YOLOv8 region of interest preprocessing and pre-trained convolutional neural networks using 2D images. *Cyprus Journal of Medical Sciences*, 10(1): 50-54. <https://doi.org/10.4274/cjms.2025.2024-137>
- [13] Hashim, A.A., Mazinani, M. (2025). Detection of keratoconus disease depending on corneal topography using deep learning. *Kufa Journal of Engineering*, 16(1): 463-478. <https://doi.org/10.30572/2018/KJE/160125>
- [14] Awwad, S.T., Hammoud, B., Assaf, J.F., Asrouiet, L., et al. (2025). Thickness speed progression index: Machine learning approach for keratoconus detection. *American Journal of Ophthalmology*, 271: 188-201. <https://doi.org/10.1016/j.ajo.2024.11.011>
- [15] Li, L.H., Xiang, Y.F., Chen, X., Lin, D.R., et al. (2025). Machine learning model for predicting corneal stiffness and identifying keratoconus based on ocular structures. *Intelligent Medicine*, 5(1): 66-72. <https://doi.org/10.1016/j.imed.2024.09.006>
- [16] Lu, N.J., Koppen, C., Hafezi, F., Ní Dhubhghaill, S., Aslanides, I.M., Wang, Q.M., Cui, L.L., Rozema, J.J. (2023). Combinations of Scheimpflug tomography, ocular coherence tomography and air-puff tonometry improve the detection of keratoconus. *Contact Lens and*

- Anterior Eye, 46(3): 101840.
<https://doi.org/10.1016/j.clae.2023.101840>
- [17] Al-Timemy, A.H., Mosa, Z.M., Alyasseri, Z., Lavric, A., Lui, M.M., Hazarbassanov, R.M., Yousefi, S. (2021). A hybrid deep learning construct for detecting keratoconus from corneal maps. *Translational Vision Science & Technology*, 10(14): 16.
<https://doi.org/10.1167/tvst.10.14.16>
- [18] Ahmed, N., Rahman, M.M., Ishrak, M.F., Joy, M.I.K., Sabuj, M.S.H., Rahman, M.S. (2024). Comparative performance analysis of transformer-based pre-trained models for detecting keratoconus disease. *arXiv preprint arXiv:2408.09005*.
<https://doi.org/10.48550/arXiv.2408.09005>
- [19] Gandhi, S.R., Satani, J., Jain, D. (2022). Classification of keratoconus using corneal topography pattern with transfer learning approach. In *ICT with Intelligent Applications. Smart Innovation, Systems and Technologies*. https://doi.org/10.1007/978-981-19-3571-8_18
- [20] Jawad, E.M., Daway, H.G., Mohamad, H.J. (2022). Retinal image enhancement by using adapted histogram equalization based on segmentation and lab color space. *International Journal of Intelligent Engineering and Systems*, 15(3): 614-622.
<https://doi.org/10.22266/ijies2022.0630.52>
- [21] El Guide, M., Jbilou, K., Koukouvinos, C., Lappa, A. (2020). Comparative study of L1 regularized logistic regression methods for variable selection. *Communications in Statistics - Simulation and Computation*, 51(9): 4957-4972.
<https://doi.org/10.1080/03610918.2020.1752379>
- [22] Ozyurt, F., Tuncer, T., Subasi, A. (2021). An automated COVID-19 detection based on fused dynamic exemplar pyramid feature extraction and hybrid feature selection using deep learning. *Computers in Biology and Medicine*, 132: 104356.
<https://doi.org/10.1016/j.compbiomed.2021.104356>
- [23] Muhsin, Z.J., Qahwaji, R., Ghafir, I., AlShawabkeh, M., Al Bdour, M., AlRyalat, S., Al-Tae, M. (2025). Advances in machine learning for keratoconus diagnosis. *International Ophthalmology*, 45: 128.
<https://doi.org/10.1007/s10792-025-03496-4>
- [24] Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16(1): 321-357.
- [25] Ravisankar, P., Sree Sharmila, T., Rajendran, V. (2018). Acoustic image enhancement using Gaussian and laplacian pyramid – A multiresolution based technique. *Multimedia Tools and Applications*, 77: 5547-5561.
<https://doi.org/10.1007/s11042-017-4466-7>
- [26] Sharma, A.K., Nandal, A., Dhaka, A., Polat, K., Alwadie, R., Alenezi, F., Alhudhaif, A. (2023). HOG transformation based feature extraction framework in modified Resnet50 model for brain tumor detection. *Biomedical Signal Processing and Control*, 84: 104737.
<https://doi.org/10.1016/j.bspc.2023.104737>
- [27] Ke, G.L., Meng, Q., Finley, T., Wang, T.F., Chen, W., Ma, W.D., Ye, Q.W., Liu, T.Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, California, USA, pp. 3149-3157.
- [28] Vo, T., Ibrahim, A.K., Zhuang, H.Q. (2025). A multimodal multi-stage deep learning model for the diagnosis of Alzheimer's Disease using EEG measurements. *Neurology international*, 17(6): 91.
<https://doi.org/10.3390/neurolint17060091>
- [29] Al-Sharify, N.T., Yussof, S., Ghaeb, N.H., Al-Sharify, Z.T., Naser, H.Y., Ahmed, S.M., See, O.H., Weng, L.Y. (2024). Advances in corneal diagnostics using machine learning. *Bioengineering*, 11(12): 1198.
<https://doi.org/10.3390/bioengineering11121198>