

Dual Deep Learning and Feature-Based Models for Classification of Laryngeal Squamous Cell Carcinoma Using Narrow Band Imaging



J Sharmila Joseph , Abhay Vidyarthi 

School of Electrical and Electronics, VIT Bhopal University, Bhopal 466114, India

Corresponding Author Email: abhay.vidyarthi@vitbhopal.ac.in

Copyright: ©2024 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/ts.410119>

ABSTRACT

Received: 22 May 2023

Revised: 25 July 2023

Accepted: 8 September 2023

Available online: 29 February 2024

Keywords:

laryngeal cancer, narrow band imaging, contrast enhancement, SqueezeNet, statistical features

Laryngeal Squamous Cell Carcinoma (LSCC) is a prevalent form of laryngeal cancer that originates from the mucosal surface of the larynx. The visual analysis of laryngeal tissue vascular patterns poses a significant challenge, as it heavily relies on the expertise and experience of medical practitioners. This paper proposes a dual approach for the early diagnosis of LSCC by employing a lightweight Deep Convolutional Neural Network (CNN) and statistical features. It further delves into feature visualization and interpretation of the proposed classification models. Methods: The initial step involves enhancing image quality through Contrast Limited Adaptive Histogram Equalization (CLAHE). In the first approach, we employ a modified SqueezeNet for classifying laryngeal tissues. In the second approach, we extract a combination of first-order statistical features – Percentile-25, Percentile-50, Percentile-75, Mean, and Standard Deviation of each RGB channel – and second-order statistical features such as Contrast, Energy, Homogeneity, and Correlation from the Gray-Level Co-Occurrence Matrix (GLCM). These features are then classified using the Extreme Gradient Boosting (XGBoost) classification model. Results: The proposed models are trained and validated using an augmented publicly available dataset, prepared for both binary and multiclass classifications. The results indicate that the proposed models demonstrate exceptional accuracy and efficiency in classifying types of laryngeal cancer.

1. INTRODUCTION

Laryngeal cancer accounts for 30% of all head and neck cancers, typically developing in the larynx, an organ critical for speaking, breathing, and swallowing. The primary form of laryngeal cancer is squamous cell carcinoma, which originates in the epithelial cells of laryngeal tissues. Persistent hoarseness of voice that remains unabated after a few weeks is the most common symptom. If not addressed early, the cancerous cells can multiply, invading nearby tissues through blood vessels, ultimately leading to fatal outcomes. Early detection of laryngeal cancer is challenging due to the identical macroscopic appearance of the mucosal tissue vessels' microarchitecture, despite potential pathological variations. Therefore, an accurate assessment of the pathological nature of laryngeal tissues is crucial in determining appropriate treatment and prognostic outcomes.

Recent meta-analyses have confirmed the effectiveness of narrow-band imaging (NBI) for diagnosing laryngeal cancer, demonstrating notable diagnostic accuracy and clinical applicability [1, 2]. NBI, a technique that utilizes short wavelengths (415 and 540 nm) of light, allows for the observation of the mucosal vascular pattern. Research [3] has demonstrated that NBI endoscopy can facilitate early and accurate detection of head and neck malignancies by examining the intrapapillary capillary loop (IPCL), characterized by elongated hypertrophic and dot-like vessels.

NBI enhances the visibility of the mucosal surface's microvascular architecture through endoscopic imaging, aiding in the identification of early recurring cancer lesions [4]. A study [5] successfully classified vocal cord leukoplakia, characterized by the thickening and whitening of the epithelial layer, from NBI imaging. The changes in IPCL indicate the initial stage of cancerous tissue development, while changes in leukoplakia suggest precancerous tissues [6]. However, there is limited research on laryngeal tissue classification from NBI images for early cancer diagnosis.

Initial computational methods for diagnosing early-stage squamous cell carcinoma were proposed, based on blood vessel segmentation and statistical characteristics such as tortuosity, thickness, and density [7]. Laryngeal disorders have been detected based on Histogram-Oriented Gradient (HOG) descriptors [8]. Recent studies have utilized Deep Convolutional Neural Networks (CNN) for laryngeal cancer detection [9-11]. A hybrid feature-based early squamous cell carcinoma detection model was proposed, leveraging Local Binary Pattern (LBP) and fine-tuned ResNet V2 pretrained CNN for feature extraction and multiclass One Against All SVM for classification [6]. An ensemble model based on the You-Only-Look-Once (YOLO) CNN was introduced to detect laryngeal squamous cell carcinoma (LSCC) in both white light (WL) and NBI laryngoscopies [12]. This model combined two effective models, YOLOv5s and YOLOv5m, with the Test Time Augmentation (TTA) technique to enhance detection

rates [13]. A novel Deep-Learning-Based Mask R-CNN Model was presented, which identified Laryngeal Cancer from CT images [14]. The Xception model was used to classify three classes: normal vocal folds, abnormal, and no finding from laryngoscopy images [15]. An early glottic cancer detection model was proposed, employing ensemble learning of Convolutional Neural Network classifiers based on voice and laryngeal imaging [16]. Faster R-CNN was utilized to differentiate between malignant and benign vocal lesions [17]. A Depth Domain Adaptive Network (DDANet) was proposed, which merges gradient CAM and guided attention for enhancing the effectiveness of classification and interpretability of tumors by incorporating the pathologist's experience at high magnification into the depth model [18]. A segmentation model was proposed for detecting laryngeal diseases [19].

Despite these advancements, there is a dearth of studies focusing on early-stage laryngeal cancer detection. To our knowledge, this study is the first to aim at developing a cost-effective model with high predictive capabilities for early-stage laryngeal cancer diagnosis. The main objectives of this study are:

- To develop deep learning and feature-based models for early classification of LSCC using Narrow Band Imaging.
 - To propose and validate two cost-effective yet efficient models for the task at hand. The first model leverages a modified version of the smallest CNN, SqueezeNet, while the second model utilizes a combination of statistical features and the XGBoost classifier.
 - To perform feature engineering visualizations to interpret the predictions of our proposed models.
 - To compare the performance of the models for binary and multi-class classification of laryngeal images obtained using NBI.
- Other contributions of this work include:
- Enhancement of image contrast to better understand the tissue patterns of laryngeal images.
 - Implementation of data augmentation techniques to

improve the performance of deep learning models.

- Execution of multiple experiments to test the robustness of the proposed methods for binary and multi-class classification.

This research paper is structured as follows: Section 2 outlines the methodologies of the proposed work. Section 3 presents the results, along with discussions. Finally, Section 4 provides the conclusion.

2. METHODS

The objective of this study is to provide two efficient and computationally simple models for laryngeal cancer detection. The method involves three important steps such as Preprocessing, Feature extraction and fusion and classification. The proposed classification framework is shown in Figure 1.

2.1 Dataset

Laryngeal dataset [20] is collected from <https://zenodo.org/record/1003200>. It consists of 1320 patches with different types of laryngeal tissue, including Hypertrophic tissue, Healthy tissue, IPCL-like tissue, and Leukoplakia tissue. These NBI image patches were produced from 33 NBI videos of 33 Squamous cell cancer patients. By carefully choosing 10 images from each video, a total of 330 images were acquired. From every image, four patches of size 100×100 pixels were cropped and therefore, a balanced distribution of 1320 patches were acquired. Few sample images from the dataset are depicted in Figure 2. Offline data augmentation was adopted to increase the dataset size. The settings for data augmentation are given in Table 1. For the evaluation of proposed method 30% of data used for testing and 70% of data used for training using CV partition method that defines a random partition on dataset. The Table 2 shows the data distribution after augmentation.

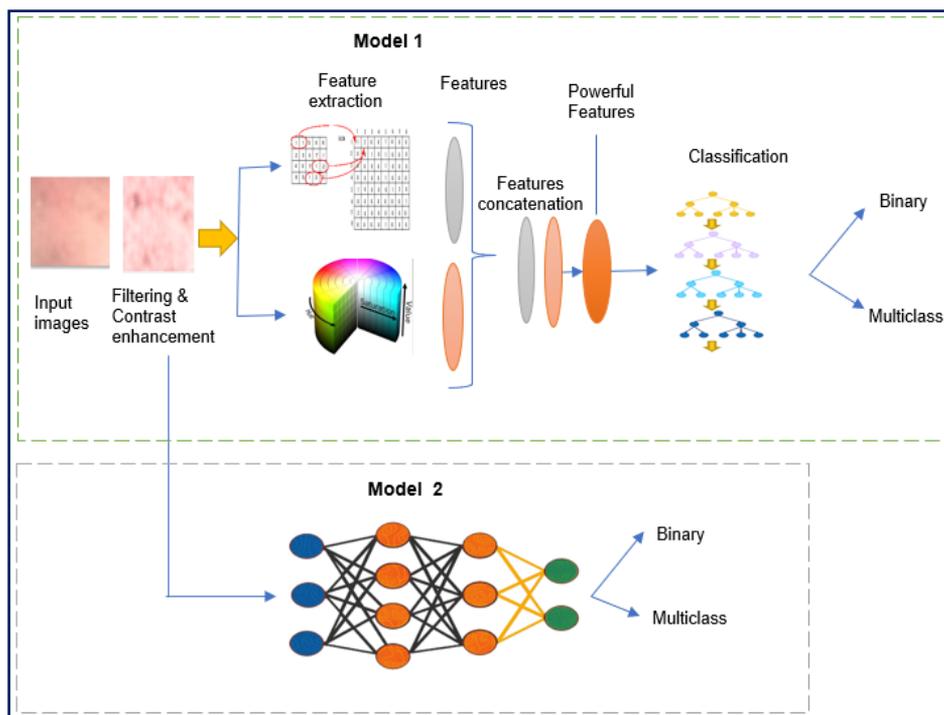


Figure 1. Proposed methodology

Table 1. Augmented settings

Data Augmentation Settings	
Rotation	[-45 45]
XShear	[0 25]
YShear	[0 25]
YReflection	true
XReflection	true

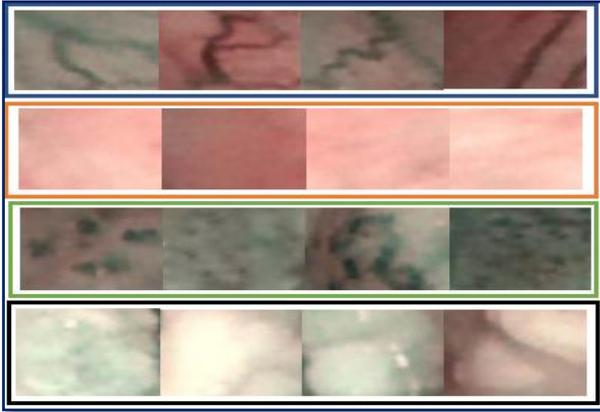


Figure 2. Each class is distinguished by a unique color outline: Blue: hypertrophic vessels tissue, Orange: normal tissue, Green: tissue with IPCL vessels, Black: leukoplakia tissue

2.2 Preprocessing

Before implementing an effective system, the laryngeal images are needs to be preprocessed to improve the sharpness of edges and local contrast effectively. We adopted CLAHE to improve image quality and then smoothen the resultant images using median filter.

2.2.1 Contrast enhancement technique

CLAHE is a significant method for enhancing digital images, particularly in medical imaging [21, 22] and provided better outcomes compared to Contrast Stretching (CS) Histogram Equalization (HE) [23, 24]. In two classes such as IPCL and HBV, the blood vessels pattern has more discrimination power, hence to understand blood vessels patterns in laryngeal tissue, the RGB images is concerted to L*a*b* color space. The L channel contains Lightness information hence the L channel of images is enhanced using CLAHE. The chosen clip limit for enhancing details of L channel is 0.006 and *Num Tiles* is 8x8, where the clip limit is the contrast factor that adjust the enhancement limit and the latter one is the number of rectangular regions that an images has to be partitioned for interpolation. The enhanced image still contains noise and is reduced using anisotropic filtering which significantly improves the image detail, reduces noise, and gets rid of false borders created by the CLAHE algorithm. An improved final image is generated by converting the L*a*b* color space image to RGB color space.

Table 2. Data distribution before and after augmentation

Before Augmentation		After Data Augmentation			
		Binary Task		Multiclass	
Healthy	330	Normal	5000	Healthy	5000
Hypertrophic	330			Hypertrophic	5000
Intra-papillary capillary	330	Abnormal	5000	Intra-papillary capillary	5000
Leukoplakia	330			Leukoplakia	5000
Total	1320	Total	1000	Total	20,000

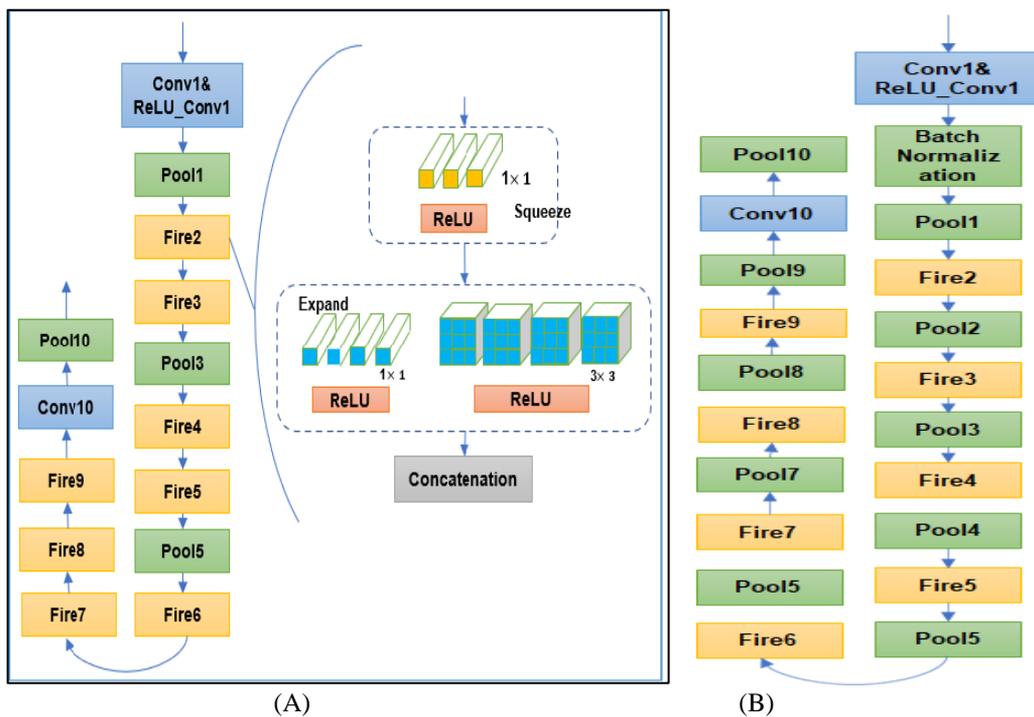


Figure 3. A. Traditional SqueezeNet architecture, B. Modified SqueezeNet

2.3 SqueezeNet

The proposed method used pretrained SqueezeNet model, which is a smaller CNN with fewer parameters that might fit in computer memory and be communicated more readily across a computer network. The architecture of classical SqueezeNet is represented in Figure 3A. The first layer is the convolutional layer, conv1 which is followed by eight fire modules. In every fire module, there is squeeze convolutional layer with a single 1×1 filter. It passes to an expand layer, which is made up of a combination of 3×3 and 1×1 convolutional filters. Configuration of the fire module is shown in Figure 3A.

The reason behind the fire module configuration is to reduce the model size while retaining the prediction accuracy. The number of filters in each fire module steadily increases from the starting till end of the network. Max pooling is done with stride 2 after 4, 8 fire modules and convolutional 10 layer. Max pooling layers would not increase the model's size since they lack any trainable weights. Moreover, it tends to lessen the model's overfitting [25]. The convolutional layer at last results in large activation maps which improves classification accuracy.

2.3.1 Modified SqueezeNet

The modified SqueezeNet architecture is shown in Figure 3B. In modified SqueezeNet, after every fire module, max pooling layer is added to avoid overfitting. In the last conv10 layer the number of filters was set as 2 for binary task and 4 for multiclass task. Batch Normalization [26] is a technique that normalizes the inputs of the layers by re-centering and re-scaling, thereby it makes training of artificial neural networks more quickly and stable. This architecture improves the accuracy compare to the traditional architecture. Further the last classification layers in classical SqueezeNet are removed and fine-tuned according to our objective. The training options are given in Table 3.

Table 3. Settings for training modified SqueezeNet model

Solver	SGDM
Initial Learning Rate	0.01
Validation Frequency	20iterations
Max Epochs	15
Mini Batch Size	15
L2Regularization	0.0001

2.4 Handcrafted features

Handcrafted features refer to the features or characteristics extracted using the traditional machine learning approaches. There are several popular handcrafted feature extraction methods but, in our work, we utilized the features which are most important to discriminate cancer tissue patterns. The following is an overview of the extracted feature descriptors.

2.4.1 Color features

Color features are widely used visual features. The significant benefits of color features are the ability to convey visual content in images, the relative strength with which images may be distinguished from one another, independent of picture dimensions and orientations and reasonably resistant to background complexity [27]. We extracted the histogram statistical color information from each channel of RGB color space. The features like mean (μ), standard deviation (σ), 25% percentile, 50% percentile, 75% percentile of each channel are calculated. The equations and descriptions of all extracted features are given in Table 4.

Table 4. First order statistical features

Feature	Description	Formula
Mean	Measure of average color in image	$\mu = \frac{1}{m} \sum_{j=1}^m X_j$
Standard Deviation	Provides a measurement of the distribution of grey level intensities in images	$\sigma = \sqrt{\frac{1}{m} \left(\sum_{j=1}^m (X_j - \mu)^2 \right)}$
75 th Percentile	Give 75 th percentile of image	$P_{75} = \frac{75}{100} (n + 1)$
50 th Percentile	Give 50 th percentile of image	$P_{50} = \frac{50}{100} (n + 1)$
25 th Percentile	Give 25 th percentile of image	$P_{25} = \frac{25}{100} (n + 1)$

where, m denotes the number of pixels in an image and X_j denotes the pixel j value of R, G and B channels. The total number of gray levels is denoted by n . Thus, we calculated total fifteen statistical color characteristics from the images.

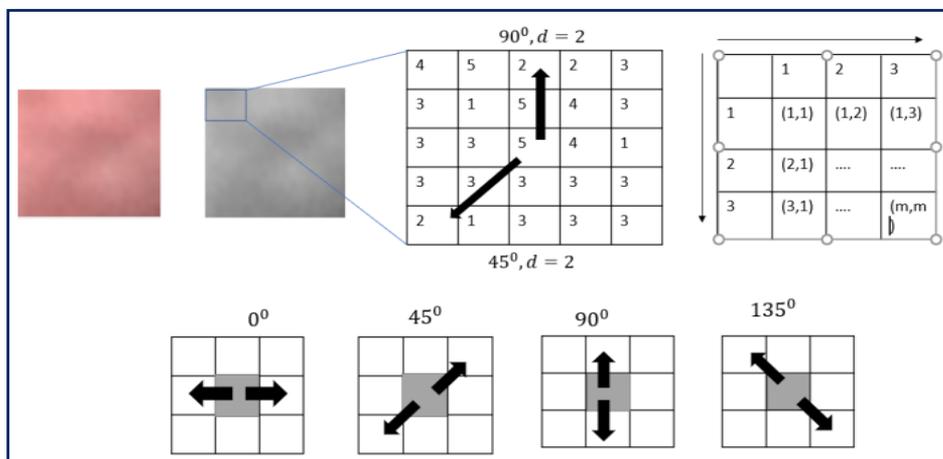


Figure 4. Illustrations of GLCM

Table 5. Extracted GLCM features

Feature	Description	Formula
Contrast	Calculate the intensity contrast of each pixel and its surrounding pixels	$\sum_s \sum_t (s - t)^2 C(s - t)$
	Determines the uniformity in neighborhood pixels intensity	$\sum_s \sum_t C(s, t)^2$
Homogeneity	Evaluates the closeness between the distributed GLCM elements to diagonal element	$\sum_s \sum_t \frac{C(s, t)}{1 + (s - t)^2}$
	Gives a measurement of how correlated each pixel is with each of its neighbors throughout the whole image	$\sum_{s, t} \frac{(s - \mu_s)(t - \mu_t)C(s, t)}{\sigma_s \sigma_t}$

2.4.2 GLCM features

The first-order statistics-generated features give details about the image's distribution of grey levels. The relative placements of the various grey levels within the image aren't made clear by them [28]. These attributes cannot determine if all low grey levels are grouped together or whether they are changed with high-value grey levels. Gray-level co-occurrence matrix (GLCM) is a statistical technique that generates a symmetric matrix of relative distance $C_{\theta, r}(s, t)$ and determines how frequently pairs of pixels (s, t) with certain values and in a particular spatial relationship occur in an image. The co-occurrence matrix depends on two factors: the relative distance (d) between the pixels and the orientation of those pixels in relation to one another. The orientation is quantized into four directions, with 0 representing horizontal, 45 representing diagonal, 90 representing vertical, and 135 representing anti-diagonal.

The extracted second order statistical features are given in Table 5. Figure 4 illustrates the different orientations of GLCM.

2.4.3 Classification

Boosting is an ensemble approach that aims to create a powerful learner by combining several weak classifiers. Initially a model is created using the training set of data. Following that, a second model is created in an effort to fix the errors in the previous one. Models are added in this manner until either the whole training data set is successfully predicted or the maximum number of models is added. One common approach for boosting is gradient boosting. Each predictor in gradient boosting corrects the mistake of its predecessor.

XGBoost [29] is a popular machine learning algorithm that makes predictions by combining gradient boosting with a group of decision trees. Strong generalization capability, great expandability, and quick computation speed are benefits of the XGBoost algorithm [30]. It also reduces overfitting [31]. In our study we used number of estimators as 100 and max depth 4.

3. RESULTS AND DISCUSSIONS

Several experiments were carried out in order to evaluate the performance and efficacy of our proposed model for

laryngeal cancer image classification. We performed our experiments in MATLAB 2020a and Python in HPC system. Our proposed model was evaluated using following evaluation metrics.

$$Accuracy = \frac{True\ Positive(TP) + True\ Negative(TN)}{True\ Positive(TP) + True\ Negative(TN) + False\ Positive(FP) + False\ Negative(FN)}$$

$$Precision = \frac{True\ Positive(TP)}{True\ Positive(TP) + False\ Positive(FP)}$$

$$Recall = \frac{True\ Positive(TP)}{True\ Positive(TP) + False\ Negative(TN)}$$

$$F1score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

TP is the result occurs when the model accurately identifies the positive class.

TN is the result occurs when the model accurately identifies the negative class.

FP is the result occurs when the model wrongly identifies the positive class.

FN is the result occurs when the model wrongly identifies the negative class.

In this study two deep learning-based and handcrafted based approaches have been proposed for early-stage laryngeal cancer detection using NBI images with less training time and high prediction capability. In this section the obtained results were discussed.

The aim of this study is to build a simple but efficient model. SqueezeNet is the smallest network that has fewer parameters, thereby it lessens computation time and memory requirements. Due to this reason SqueezeNet network was selected in this work. In approach 1, SqueezeNet architecture is modified to predict the type of images accurately. Though CNN can easily distinguish the structural features, but not much effective with the statistical textures. Therefore, in approach 2, first and second order statistical features are combined with to increase the prediction capability. Experiments, are validated using augmented publicly available dataset that has 10,000 images for binary classification and 20,000 images for multiclass classification. The proposed models have been tested for Multiclass and Binary classification.

Before feature extraction, preprocessing step has been carried to enhance the quality of the images. For that, RGB images are converted to CIE lab color space and the luminance channel is enhanced to investigate the texture and structural information of blood vessels in tissues.

3.1 Approach one for binary classification

In this experiment binary classification using modified SqueezeNet has performed. The distribution of the dataset is shown in the Table 1. For this experiment, the images were partitioned into two classes as normal tissues and abnormal tissues. A greater understanding of the model's operation may be gained by visualizing the inner operation of how CNN learns to recognize various elements contained in images. In Figure 5 some sample images and their important features learned by SqueezeNet for making classification decision.

Figure 6 illustrates the feature from different layers of SqueezeNet. It shows in earlier layer like fire 4 layer the network learned the horizontal and vertical lines, when it goes deeper specific features are learned.

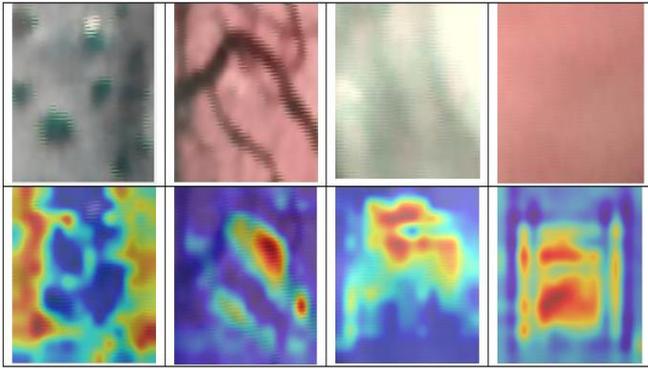


Figure 5. Visualization of CNN activations using grad CAM interpretability technique

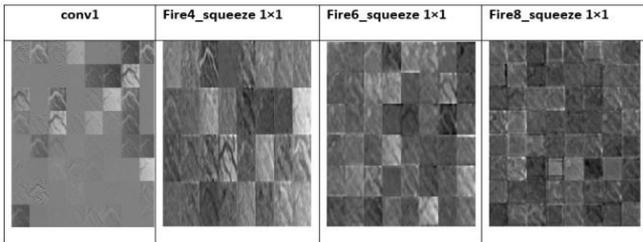


Figure 6. Visualization of features from different squeeze net layers

Figure 7 depicts the accuracy and loss curves of training and validation data using modified squeeze net for binary classification. The blue and orange curves indicate the accuracy and loss curves of training data. The dotted curves indicate the accuracy and loss curves of validation data. The results obtained using modified Squeeze net for binary classification is given in Table 6. The obtained validation accuracy is 99.9% and testing accuracy is 99% in 15 epochs. The hyperparameter learning rate is chosen as 0.01. The prediction capability of model 1 for binary classification is shown using confusion matrix in Figure 8 and ROC curve in Figure 9.

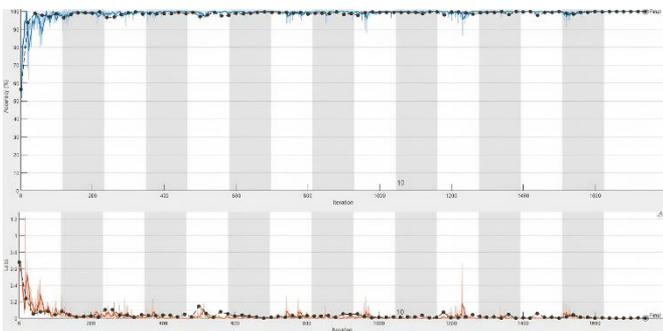


Figure 7. Training, validation accuracy and loss curves of modified SqueezeNet for binary classification

Though, DL has grown in popularity and that computing power has been advancing the computation time is the critical consideration. The validation accuracy of modified Squeeze Net is approximately 3% higher than accuracy of traditional SqueezeNet. In case of computation time, the training time of modified SqueezeNet is slightly higher (approximately equal to 25 sec) than traditional SqueezeNet.

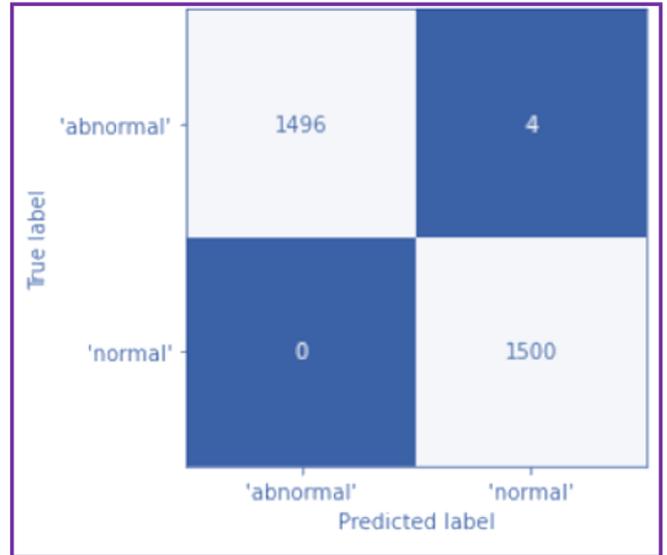


Figure 8. Confusion matrix of approach 1 for binary classification

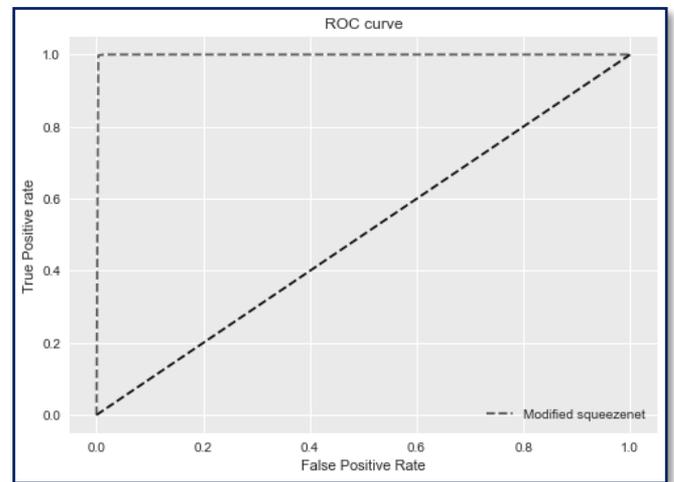


Figure 9. ROC curve of approach 1 for binary classification

Table 6. Comparison of traditional Squeeze net and Modified Squeeze net for binary classification

CNN	Training Accuracy	Validation Accuracy	Training Time
Modified SqueezeNet	99.9	99.7	16min 1sec
SqueezeNet	97.05	96.6	15 min 35 sec

3.2 Approach one for multiclass classification

Multiclass classification of four different classes of laryngeal tissue is performed in this experiment. The images are equally distributed in each class. Data distribution is given in section 2. The hyperparameter settings is same as in binary classification task. The modified SqueezeNet has given its supreme results in multiclass task by giving validation accuracy of 99.1% and testing accuracy of 98.9%. The training and validation accuracy with classical and modified SqueezeNet for multiclass classification is shown in Table 7. In case of multiclass classification task, the validation accuracy of modified SqueezeNet is approximately 3.5% higher than the accuracy of traditional SqueezeNet. But in

terms of computational time, the training time of modified SqueezeNet is 3 sec more than the traditional SqueezeNet.

The accuracy and loss curves for training and validation data is illustrated in Figure 10.

Figure 11 and Figure 12 illustrate the prediction performance of model 1 in multiclass classification through confusion matrix and ROC curve respectively. By comparing the overall results of normal and traditional SqueezeNet for binary and multiclass classification, the performance of modified SqueezeNet is better than the traditional one.

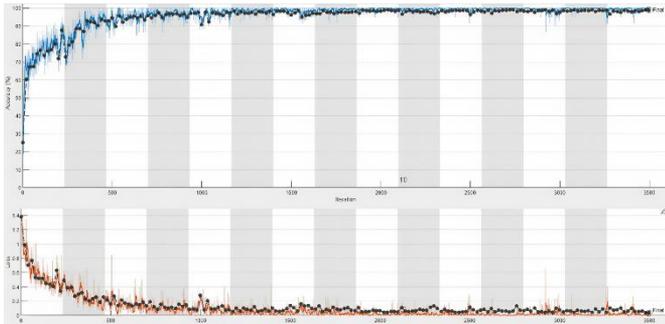


Figure 10. Accuracy and loss curves of modified SqueezeNet for Multiclass classification

True Label	Healthy	1500	0	0	0
	Hypertrophic	7	1471	5	17
	Leukoplakia	2	2	1495	1
	IPCL	4	5	0	1491
		Healthy	Hypertrophic	Leukoplakia	IPCL
		Predicted Label			

Figure 11. Confusion matrix of approach 1 for multiclass classification

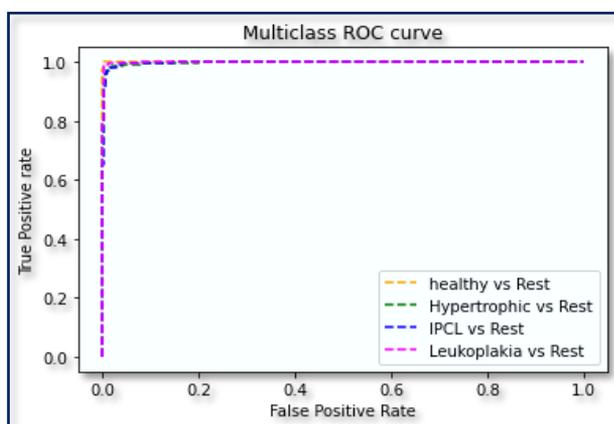


Figure 12. ROC curve of approach 1 for multiclass classification

Table 7. Comparison of traditional SqueezeNet and Modified SqueezeNet for multiclass classification

	Training Accuracy	Validation Accuracy	Training Time
Modified SqueezeNet	99.8	99.1	43min 38 sec
SqueezeNet	96.3	95.2	40min 33sec

3.3 Approach two for binary classification

The model 2 is based on Hybridized statistical features with XGBoost classifier. The images are partitioned into 70% for training and 30% for testing. The statistical features were extracted and fused to form optimal feature subset. Eventually classified with XGBoost classifier. Fifteen first order statistical features and sixteen second order statistical color features are extracted and integrated. The resultant 31 features are trained with XGBoost classifier and acquired the highest classification accuracy of 100% for binary classification. It is observed that performance of XGBoost classifier is very powerful compared to other ML classifiers. The individual features performance and hybrid feature performances with XGBoost has been calculated. All feature combinations have provided satisfactory results. The performance of hybrid statistical feature is equal to the hybrid deep and handcrafted features. Table 8 summarized the results for binary classification with different features and its confusion matrices of approach 2 using different features with XGBoost classifier are depicted in Figure 13(a)-Figure 13(g). 13(a) with only first order stats 13(b) with second order stats 13(c) fusion of first order and second order stats 13(d) with SqueezeNet features 13(e) with SqueezeNet and first order stats 13(f) with SqueezeNet and second order stats 13(g) with fusion of first order, second order and SqueezeNet features.

Table 8. Performance metrics of various features with XGBoost for binary classification

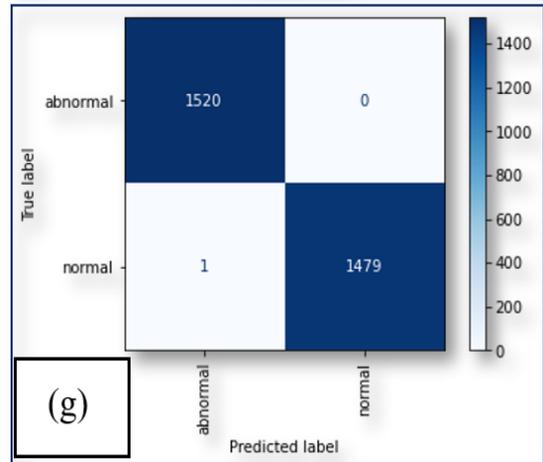
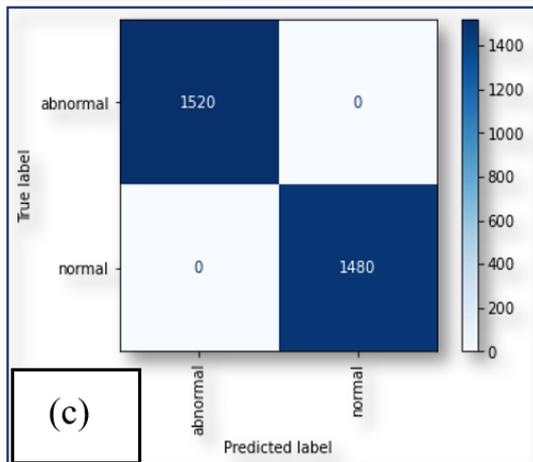
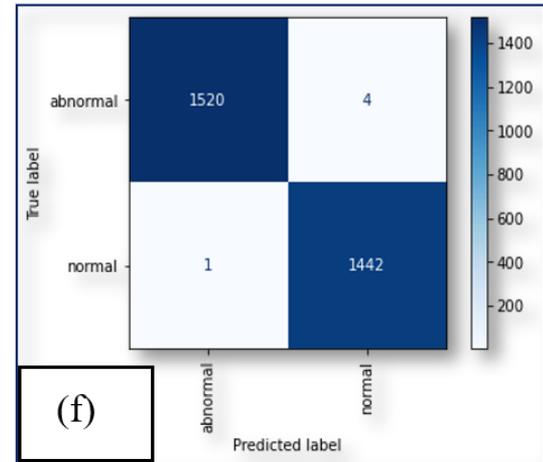
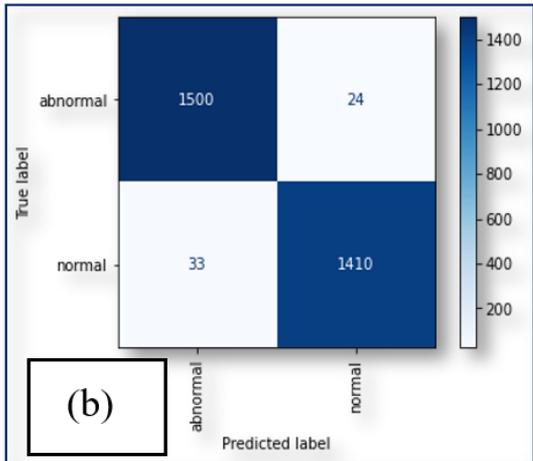
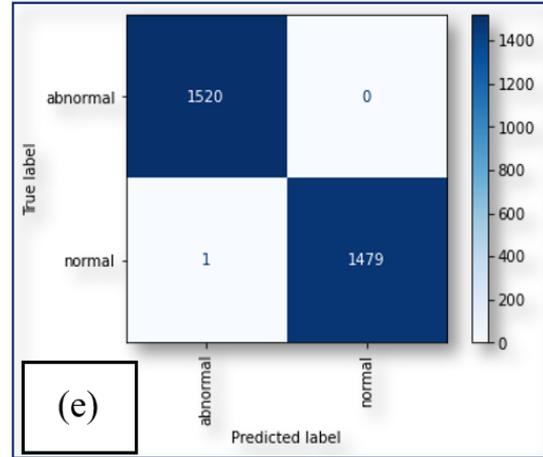
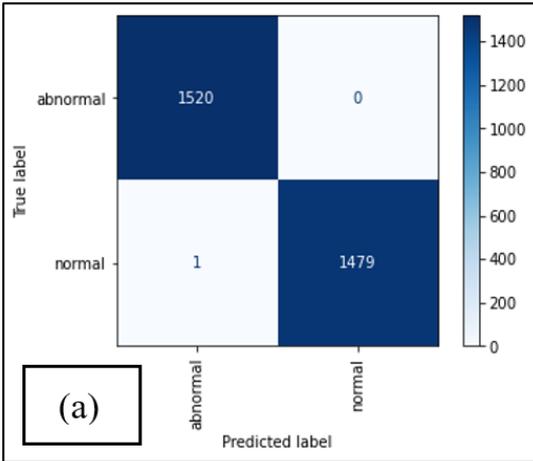
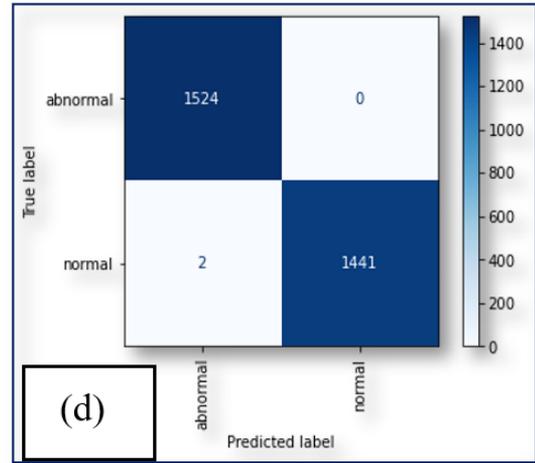
Methods	Accuracy	Precision	Recall	F1-Score
First order stats	99.9	100	100	100
Second order stats	98	98	98	98
First +Second order stats	100	100	100	100
SqueezeNet	99.9	100	100	100
Squeeze+ First order stats	99.9	100	100	100
Squeeze+ second order stats	99.8	100	100	100
Squeeze +first +second order stats	99.9	100	100	100

3.4 Approach two for multiclass classification

The multiclass classification of four different types of laryngeal tissues using approach 2 is performed in this experiment. The confusion matrices for multiclass classification task using various feature combinations with XGBoost classifier are displayed in Figure 13(h)-13(n). Figure 13(h) With only first order stats 13(i) with second order stats 13(j) fusion of first order and second order stats 13(k) with SqueezeNet features 13(l) with SqueezeNet and first order stats 13(m) with SqueezeNet and second order stats 13(n) with

fusion of first order, second order and SqueezeNet features. The result shows the performance hybrid stats is slightly better than hybrid deep stats by giving classification accuracy of 99.55%. Table 9 summarized the results for multiclass classification with different features and XGBoost classifier.

From the results it is observed that the discrimination ability of stat features and SqueezeNet features in combination with XGBoost classifier is high. By comparing the confusion matrices, the performance of first order and second order statistics with XGBoost is superior to other methods. To interpret the handcrafted statistical features utilized in our model, Shapely Additive Explanations (SHAP) feature importance approach was carried out. This is an approach that helps to predict the contribution of every feature for making the predictions.



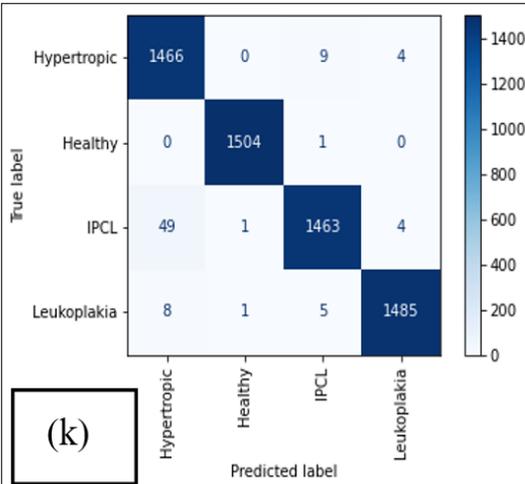
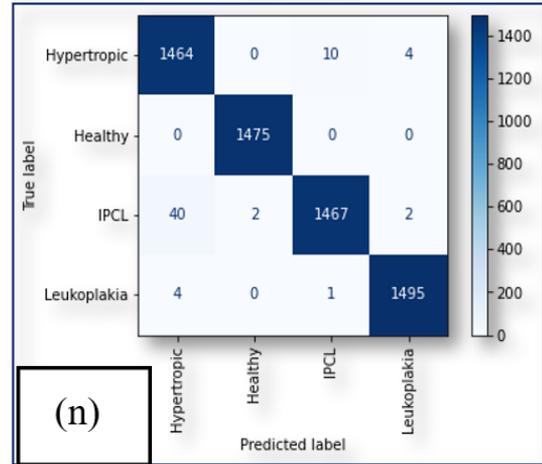
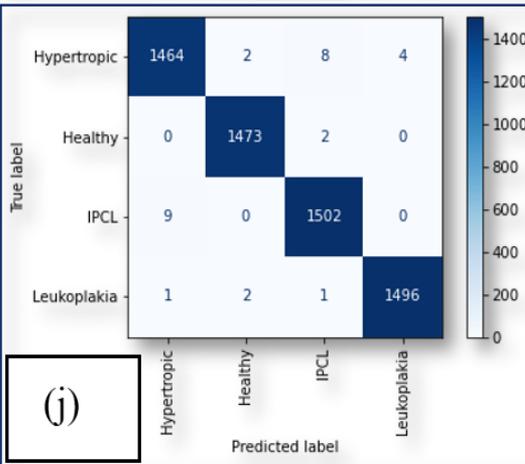
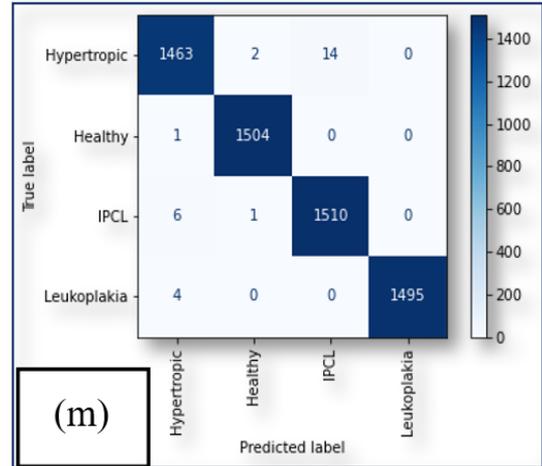
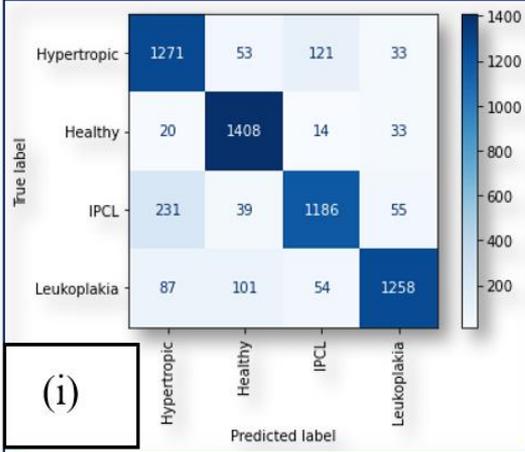
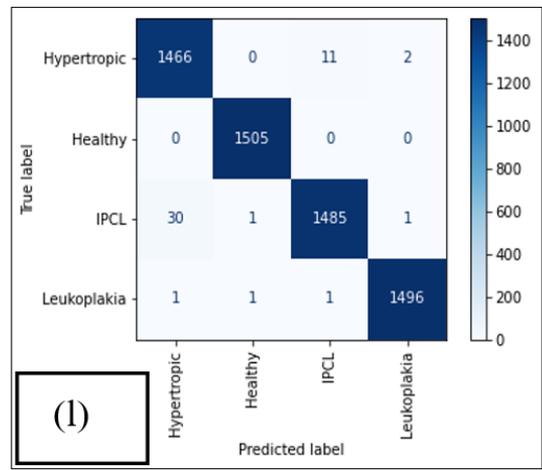
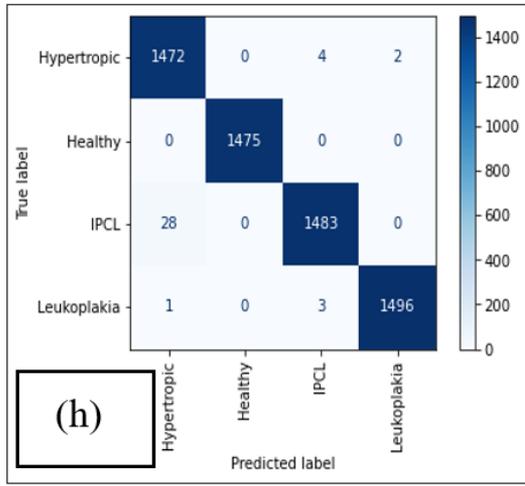
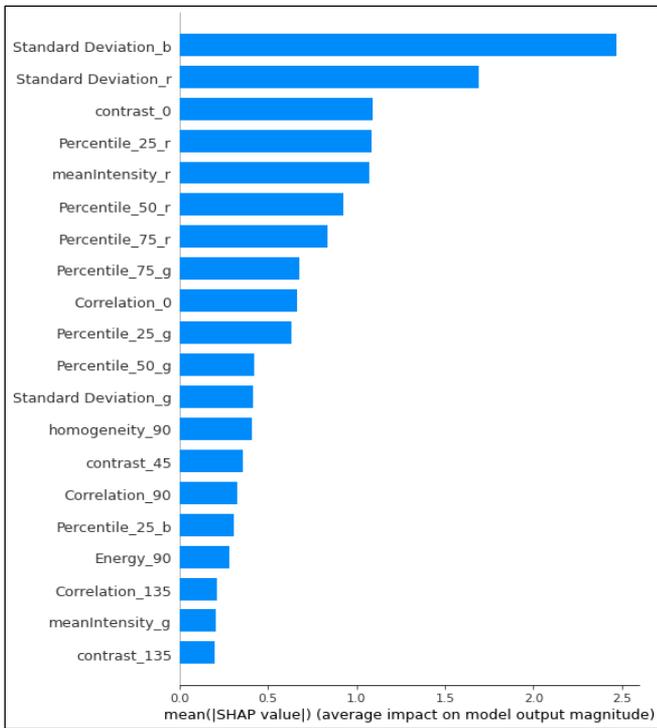


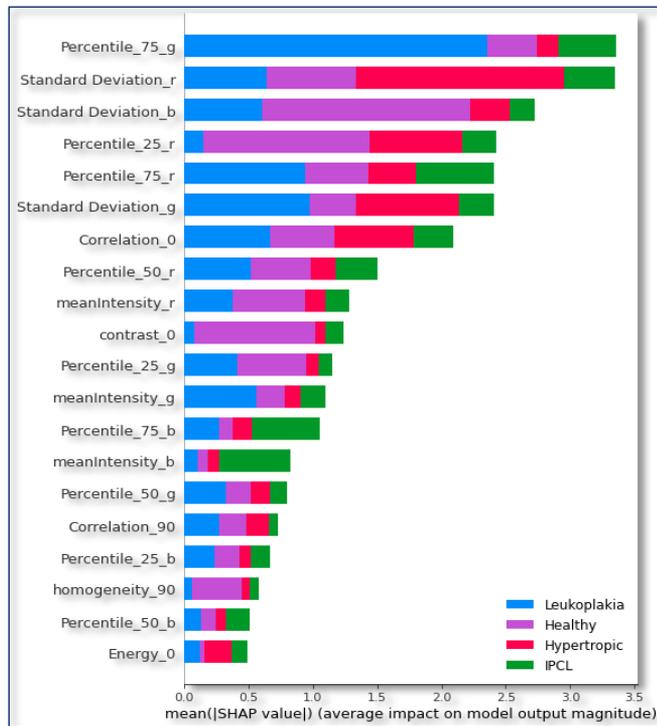
Figure 13. Confusion matrices of XGBoost classifier with various feature combinations

Table 9. Performance metrics of various features with XGBoost for multiclass classification

Methods	Accuracy	Precision	Recall	Recall
First order stats	99.3	99	99	99
Second order stats	85	86	86	86
First+ Second order stats	99.55	100	100	100
SqueezeNet	98.63	99	99	99
SqueezeNet + first order stats	99.2	99	99	99
SqueezeNet + second order stats	98.9	99	99	99
SqueezeNet + first + second order stats	99.5	100	100	100



(a)



(b)

Figure 14. SHAP feature importance bar chart for multiclass (a) and Binary (b) classification

The average of the absolute Shapley values is used to determine the importance of SHAP characteristics. According to the concept of SHAP feature significance, the most significant features are those with high absolute Shapley values. Figure 14 shows the most essential features for the trained XGBoost model that were chosen and ranked in accordance with their significance using the SHAP approach. Figure 14a illustrates the SHAP feature importance bar plot for binary classification and Figure 14b shows the summary bar plot for multiclass classification. In case of multiclass

classification, 75th percentile of green channel feature has more importance in predicting the classes and energy at 0° orientation has less importance for multiclass predictions whereas, in binary classification case, the standard deviation of blue channel is the highest significant feature and contrast feature 135° orientation has less significance.

4. CONCLUSION

This study presented a deep transfer learning and statistical feature-based models for early-stage laryngeal cancer detection with less computational resources and high prediction accuracy. In this study two computer-based models have been developed for the Laryngeal cancer image classification. The model one introduced a simplified version of SqueezeNet, which is the smallest pretrained CNN model that uses fewer parameters but still maintaining better accuracy with reduced training time compared to other pretrained models and acquired maximum classification accuracy of 99.7% for binary classification and 99.1% for multiclass classification. The model two utilized the hybrid statistical features in combination with XGBoost classifier and attained maximum classification accuracy of 100% and 99.55% for binary and multiclass classification respectively.

Through several experiments it has been observed that both models exhibit equal level of effectiveness. and their results show no statistically significant difference in their performance metrics. This indicates that they possess an equivalent level of effectiveness and are equally proficient in achieving the desired outcomes. As, SqueezeNet has limited layers, it might possess less expressive capability in comparison to deeper and more intricate CNN models. As a consequence, this could present challenges when handling complex and diverse datasets that require higher-level feature representations. In future, segmentation based on CNN will be incorporated to precise segmentation of the blood vessels in the tissues.

REFERENCES

- [1] Sun, C., Zhang, Y., Han, X., Du, X. (2018). Diagnostic performance of narrow band imaging for nasopharyngeal cancer: A systematic review and meta-analysis. *Otolaryngology–Head and Neck Surgery*, 159(1): 17-24. <https://doi.org/10.1177/0194599818758302>
- [2] Kraft, M., Fostiropoulos, K., Gürtler, N., Arnoux, A., Davaris, N., Arens, C. (2016). Value of narrow band imaging in the early diagnosis of laryngeal cancer. *Head & Neck*, 38(1): 15-20. <https://doi.org/10.1002/hed.23838>
- [3] Cosway, B., Drinnan, M., Paleri, V. (2016). Narrow band imaging for the diagnosis of head and neck squamous cell carcinoma: A systematic review. *Head & Neck*, 38(S1): E2358-E2367. <https://doi.org/10.1002/hed.24300>
- [4] Deva, F.A.L. (2023). Narrow band imaging technology: Role in the detection of recurrent laryngeal and hypopharyngeal cancers post-radiotherapy. *Indian Journal of Otolaryngology and Head & Neck Surgery*, 75: 753-759. <https://doi.org/10.1007/s12070-022-03457-8>
- [5] Ni, X.G., Zhu, J.Q., Zhang, Q.Q., Zhang, B.G., Wang, G.Q. (2019). Diagnosis of vocal cord leukoplakia: The role of a novel narrow band imaging endoscopic

- classification. *The Laryngoscope*, 129(2): 429-434. <https://doi.org/10.1002/lary.27346>
- [6] Araújo, T., Santos, C.P., De Momi, E., Moccia, S. (2019). Learned and handcrafted features for early-stage laryngeal SCC diagnosis. *Medical & Biological Engineering & Computing*, 57: 2683-2692. <https://doi.org/10.1007/s11517-019-02051-5>
- [7] Barbalata, C., Mattos, L.S. (2014). Laryngeal tumor detection and classification in endoscopic video. *IEEE Journal of Biomedical and Health Informatics*, 20(1): 322-332. <https://doi.org/10.1109/JBHI.2014.2374975>
- [8] Turkmen, H.I., Karşlıgil, M.E., Kocak, I. (2015). Classification of laryngeal disorders based on shape and vascular defects of vocal folds. *Computers in Biology and Medicine*, 62: 76-85. <https://doi.org/10.1016/j.compbiomed.2015.02.001>
- [9] Hu, R., Zhong, Q., Xu, Z.G., Huang, L.Y., Cheng, Y., Wang, Y.R., He, Y.D. (2021). Application of deep convolutional neural networks in the diagnosis of laryngeal squamous cell carcinoma based on narrow band imaging endoscopy. *Chinese Journal of Otorhinolaryngology Head and Neck Surgery*, 56(5): 454-458. <https://doi.org/10.3760/cma.j.cn115330-20200927-00773>
- [10] He, Y., Cheng, Y., Huang, Z., Xu, W., Hu, R., Cheng, L., He, S., Yue, C., Qin, G., Wang, Y., Zhong, Q. (2021). A deep convolutional neural network-based method for laryngeal squamous cell carcinoma diagnosis. *Annals of Translational Medicine*, 9(24): 1797. <https://doi.org/10.21037/atm-21-6458>
- [11] Esmacili, N., Sharaf, E., Gomes Ataide, E.J., Illanes, A., Boese, A., Davaris, N., Arens, C., Navab, N., Friebe, M. (2021). Deep convolution neural network for laryngeal cancer classification on contact endoscopy-narrow band imaging. *Sensors*, 21(23): 8157. <https://doi.org/10.3390/s21238157>
- [12] Azam, M.A., Sampieri, C., Ioppi, A., Africano, S., Vallin, A., Mocellin, D., Fragale, M., Guastini, L., Moccia, S., Piazza, C., Mattos, L.S., Peretti, G. (2022). Deep learning applied to white light and narrow band imaging video laryngoscopy: Toward real-time laryngeal cancer detection. *The Laryngoscope*, 132(9): 1798-1806. <https://doi.org/10.1002/lary.29960>
- [13] Zhou, X., Tang, C., Huang, P., Mercaldo, F., Santone, A., Shao, Y. (2021). LPCANet: Classification of laryngeal cancer histopathological images using a CNN with position attention and channel attention mechanisms. *Interdisciplinary Sciences: Computational Life Sciences*, 13(4): 666-682. <https://doi.org/10.1007/s12539-021-00452-5>
- [14] Sahoo, P.K., Mishra, S., Panigrahi, R., Bhoi, A.K., Barsocchi, P. (2022). An improvised deep-learning-based mask R-CNN model for laryngeal cancer detection using CT images. *Sensors*, 22(22): 8834. <https://doi.org/10.3390/s22228834>
- [15] Tran, B.A., Dao, T.T.P., Dung, H.D.Q., et al. (2023). Support of deep learning to classify vocal fold images in flexible laryngoscopy. *American Journal of Otolaryngology*, 44(3): 103800. <https://doi.org/10.1016/j.amjoto.2023.103800>
- [16] Kwon, I., Wang, S.G., Shin, S.C., Cheon, Y.I., Lee, B.J., Lee, J.C., Lim, D.W., Jo, C., Cho, Y., Shin, B.J. (2022). Diagnosis of early glottic cancer using laryngeal image and voice based on ensemble learning of convolutional neural network classifiers. *Journal of Voice*. <https://doi.org/10.1016/j.jvoice.2022.07.007>
- [17] Yan, P., Li, S., Zhou, Z., Liu, Q., Wu, J., Ren, Q., Chen, Q., Chen, Z., Chen, Z., Chen, S., Scholp, A., Jiang, J.J., Kang, Ge, P. (2023). Automated detection of glottic laryngeal carcinoma in laryngoscopic images from a multicentre database using a convolutional neural network. *Clinical Otolaryngology*, 48(3): 436-441. <https://doi.org/10.1111/coa.14029>
- [18] Huang, P., Zhou, X., He, P., Feng, P., Tian, S., Sun, Y., Mercaldo, F., Santone, A., Qin, J., Xiao, H. (2023). Interpretable laryngeal tumor grading of histopathological images via depth domain adaptive network with integration gradient CAM and priori experience-guided attention. *Computers in Biology and Medicine*, 154: 106447. <https://doi.org/10.1016/j.compbiomed.2022.106447>
- [19] Chen, I.M., Yeh, P.Y., Hsieh, Y.C., Chang, T.C., Shih, S., Shen, W.F., Chin, C.L. (2023). 3D VOSNet: Segmentation of endoscopic images of the larynx with subsequent generation of indicators. *Heliyon*, 9(3). <https://doi.org/10.1016/j.heliyon.2023.e14242>
- [20] Moccia, S., De Momi, E., Guarnaschelli, M., Savazzi, M., Laborai, A., Guastini, L., Peretti, G., Mattos, L.S. (2017). Confident texture-based laryngeal tissue classification for early stage diagnosis support. *Journal of Medical Imaging*, 4(3): 034502-034502. <https://doi.org/10.1117/1.JMI.4.3.034502>
- [21] Sahu, S., Singh, A.K., Ghreera, S.P., Elhoseny, M. (2019). An approach for de-noising and contrast enhancement of retinal fundus image using CLAHE. *Optics & Laser Technology*, 110: 87-98. <https://doi.org/10.1016/j.optlastec.2018.06.061>
- [22] Li, L., Si, Y., Jia, Z. (2018). Medical image enhancement based on CLAHE and unsharp masking in NSCT domain. *Journal of Medical Imaging and Health Informatics*, 8(3): 431-438. <https://doi.org/10.1166/jmih.2018.2328>
- [23] Zhou, M., Jin, K., Wang, S., Ye, J., Qian, D. (2017). Color retinal image enhancement based on luminosity and contrast adjustment. *IEEE Transactions on Biomedical Engineering*, 65(3), 521-527. <https://doi.org/10.1109/TBME.2017.2700627>
- [24] Alwazzan, M.J., Ismael, M.A., Ahmed, A.N. (2021). A hybrid algorithm to enhance colour retinal fundus images using a Wiener filter and CLAHE. *Journal of Digital Imaging*, 34(3): 750-759. <https://doi.org/10.1007/s10278-021-00447-0>
- [25] Nirthika, R., Manivannan, S., Ramanan, A., Wang, R. (2022). Pooling in convolutional neural networks for medical image analysis: A survey and an empirical study. *Neural Computing and Applications*, 34(7): 5321-5347. <https://doi.org/10.1007/s00521-022-06953-8>
- [26] Ioffe, S., Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International Conference on Machine Learning*, pp. 448-456.
- [27] Alamdar, F., Keyvanpour, M. (2011). A new color feature extraction method based on QuadHistogram. *Procedia Environmental Sciences*, 10: 777-783. <https://doi.org/10.1016/j.proenv.2011.09.126>
- [28] Haryanto, T., Pratama, A., Murni, A., Suhartanto, H., Pidanic, J., Arymurthy, A.M. (2020). Multipatch-GLCM for texture feature extraction on classification of the colon histopathology images using deep neural network

- with GPU acceleration. *Journal of Computer Science*, 16(3): 280-294. <https://doi.org/10.3844/JCSP.2020.280.294>
- [29] Chen, T., Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785-794. <https://doi.org/10.1145/2939672.2939785>
- [30] Li, S., Zhang, X. (2020). Research on orthopedic auxiliary classification and prediction model based on XGBoost algorithm. *Neural Computing and Applications*, 32: 1971-1979. <https://doi.org/10.1007/s00521-019-04378-4>
- [31] Sarker, I.H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science*, 2(3): 160. <https://doi.org/10.1007/s42979-021-00592-x>