IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 6 February 2024, accepted 19 February 2024, date of publication 26 February 2024, date of current version 4 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3370157

RESEARCH ARTICLE

Classification of Oral Cancer Into Pre-Cancerous Stages From White Light Images Using LightGBM Algorithm

BIBEK GOSWAMI[1], M. K. BHUYAN[1], (Senior Member, IEEE), SULTAN ALFARHOOD[2], AND MEJDL SAFRAN[2]

[1]Department of Electronics and Electrical Engineering, Indian Institute of Technology Guwahati, Guwahati 781039, India
[2]Department of Computer Science, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia

Corresponding authors: Bibek Goswami (b.goswami@iitg.ac.in) and Sultan Alfarhood (sultanf@ksu.edu.sa)

This work was supported by the Researchers Supporting Project, King Saud University, Riyadh, Saudi Arabia, under Grant RSPD2024R890.

ABSTRACT Cancer is one of the foremost reasons for death worldwide, with nearly 10 million deaths noted in 2020. Globally, oral cancer ranks sixth when compared to other cancers. It is lethal because most cases are noticed at advanced stages, which can be prevented if screened for or treated early in the pre-cancerous stages, successively leading to a significant decrease in the mortality rate. In this work, a method is proposed that can effectively differentiate between benign and malignant oral cavity lesions and also classify their pre-cancerous stages. The method involves exploring five distinct color spaces and extracting color and texture features, which are then classified using a machine learning technique called Light Gradient Boosting Machine (LightGBM). The overall performance is promising, outperforming the state-of-art methods for the task of oral cancer classification, with a testing accuracy of 99.25%, precision of 99.18%, recall of 99.31%, f1-score of 99.24% and specificity of 99.31% for the binary classification, and testing accuracy of 98.88%, precision of 98.86%, recall of 97.92%, f1-score of 98.38% and specificity of 99.03% for multi-class classification. The proposed method used hand-crafted features and a machine-learning classifier, which uses limited resources and is less time-consuming.

INDEX TERMS Binary and multi-class classification, color spaces, early detection, feature importance, oral cancer, white light images.

I. INTRODUCTION

Any abnormal growth of cells that invades and spreads into other parts of the body could be termed cancer [1]. Oral cancer (OC) is a variety of cancer that develops in the oral cavity that includes the gingiva(gums), buccal mucosa, floor of the mouth or under the tongue, lips, anterior part of the tongue, hard palate, retromolar trigone and may spread to the oropharynx as well, the next part of oral cavity consisting of the walls of the throat, back part of the tongue, soft palate and tonsils [2]. The oral cavity and oropharynx are mostly linked with epithelial tissues, epithelial cells undergo pre-malignant pathological changes and form some typical pre-malignant lesions such as leukoplakia, erythro-leukoplakia, erythroplakia that develop into Squamous Cell Carcinoma (SCC) [3] which constitutes roughly 90% of all oral cancers. Leukoplakia is a white, flat and raised patch that cannot be characterized by any pathological disease or disorder with no apparent underlying cause but the risk factors include tobacco and excessive alcohol abuse. It is usually painless and could be quickly scrapped away [4]. Cellular-level leukoplakia shows a thick keratin layer that looks white when it's flat and shows some dysplasia which is a pre-cancerous condition distinguishable generally at the microscopic level. These dysplastic cells gradually grow into malignant cells. As cells become increasingly dysplastic, the lesions generally develop red spots known as erythro-leukoplakia. Cells in the red area have suffered severe damage

The associate editor coordinating the review of this manuscript and approving it for publication was Carmelo Militello.

to their DNA and become premature and, therefore, cannot produce keratin. As a result, it gets more delicate and blood vessels could be perceptible through the mucosa known as erythroplakia, which is severe dysplasia or early cancer [5].

Globally OC has the sixth rank when compared to other cancers. India on the other hand has the most significant number of OC cases, about one-third of the entire load globally [6]. Analogized to other regions of the world, the crisis of OC is more in India because 70% of the cases in India are reported late in the advanced stages (American Joint Committee on Cancer, Stage III-IV). Due to the delay in detection, the chances of getting curated are negligible, leading to a very high mortality rate [7]. Early diagnosis and proper treatment could improve the survival rate to 90%. Therefore, there is a serious need for OC diagnosis techniques that are easy to use, fast, accurate and non-invasive [8], [9]. For the detection of OC, initially the primary traces of the abnormal tissues of the oral cavity are adequately observed by a dentist. Upon surmise of some potential malignancy, it is further referred to an oral or maxillographic surgeon to conduct further tests [10]. However, in the conventional techniques used, the diagnosis is subjective to the knowledge and expertise of the clinician. Human diagnosis may be prone to errors, and with no definitive parameter to set the threshold, it becomes more difficult and subjective to classify the oral cavity condition. Hence, an automatic system would greatly help the community with the early identification of OC and lessen false-positive rates. Therefore, we propose a method that uses color and texture features from different color spaces to classify oral cancer into its pre-cancerous stages.

Texture features are foremost in segmentation, classification and synthesis of images. However, a fixed definition of texture is yet to be considered. The texture is usually termed as repetitive visual patterns appearing in the image. Numerous techniques are presented over time for texture feature extraction and the research arena is yet vulnerable to many analyses [11], [12], [13], [14], [15], [16], [17], [18], [19], [20]. Newly, seven segments were suggested to categorize the texture feature extraction procedure [11]: statistical procedures (for instance, the co-occurrence matrices), structural procedures, transform-based procedures (Fourier transform-based procedures, and a few others), model-based procedures (random field models), graph-based procedures (local graph structures), learning-based procedures, and entropy-based procedures. Apart from texture, color is also an essential aspect of human perception of images and image processing [21], [22], [23], [24], [25], [26], [27]. Unlike intensity which is summarized as scalar in grayscale images, color can be seen as a vectorial feature that is fixed to each pixel of a colored image [21]. In disparity with grayscale images that could be processed in simple elementary ways, colored images could be processed in diverse ways. Thus, it becomes evident to determine the need to study color and texture features individually or both should be combined using some transformation and studied together [21], [27], [28]. Limited research has been conducted on colored texture analysis, most of the studies carried out were conducted by assimilating the application of grayscale texture analysis methods [29], [30], [31], [32]. The application areas of Machine Learning (ML) and Deep Learning (DL) have expanded vast in the past few years due to the technological advances that created opportunities for data digitization which became helpful in digitizing the data of patients through electronic case narratives and image files especially in pathological and radiological areas. A contemporary principle can be seen in the increasing benefit of radiomics, a computational standard that assists in manifesting the detection and radically influences imaging data transformation to catch the discriminatory characteristics not noticeable to the naked eye. Such unique imaging markers may be of diagnostically, prognostically, and restoratively beneficial [33], [34], [35], [36], [37], [38].

ML and DL approaches have achieved notable success in classifying oral cancer lesions into their stages. Some previous works that used white light images for detecting oral cancer are discussed. Thomas et al. [39] extracted texture discrimination features such as Gray Level Co-occurrence Matrix (GLCM) and Gray Level Run-length Matrix (GLRM) and adopted a Back Propagation based Artificial Neural Network (BPANN) for classifying the tissues of the oral cavity into six categories such as Carcinoma Retromolar area as Group 1, Carcinoma angle of mouth – ulcero-proliferative as group 2, Verrucous carcinoma buccal mucosa as group 3, Carcinoma Buccal mucosa – Ulcero-proliferative as group 4, Carcinoma Tongue – lateral border of the tongue as group 5 and Carcinoma Palate – Ulcero-proliferative growth as group 6. The performance was considerable and increased by combining both GLCM and GLRM features. However, the dataset contained only 16 images, which is very diminutive for a multi-class classification problem, and other performance matrices other than accuracy were also not considered. Song et al. [40] captured white light and autofluorescence images using a low-cost, dual-modality, intra-oral screening device and used them to classify oral dysplasia and malignancy among patients. Different DL techniques based on Convolutional Neural Network (CNN) such as VGG-CNN-M, VGG-CNN-S, and VGG-16 were tested and it was concluded that the fusion of modalities of imaging produced better results than the individual ones. Nevertheless, the performance potentially degraded because of the variation in the tissue structure of the oral cavity. Welikala et al. [41] adopted a novel strategy in which bounding box annotation from multiple clinicians was considered. Further, a Deep Neural Network was used to build an automated system that extracted critical patterns for classification. Two DL computer vision based techniques, ResNet-100 and Faster R-CNN were used to automatically detect and classify the oral lesion into six sub-groups. However, the accuracy was not considerable and some data annotations were not appropriately done.

Fu et al. [42] used a Cascaded CNN-based classifier to detect Oral Squamous Cell Carcinoma (OSCC) that acquired high performance in various scenarios, including early detection and real-time smartphone application. However, the classifier could not generalize other types and areas of oral diseases in the oral cavity. Song et al. [43] developed an easy to use android smartphone application that implemented mobile-based dual-modality image classification for differentiating normal, potentially malignant and malignant lesions. An end-to-end dual-modality deep convolutional neural network was implemented for the automatic classification of dysplasia and malignancy images. However, the accuracy and time consumption was inversely inter-dependent, and a trade-off was needed between the two to get a considerable outcome. Lin et al. [44] developed a smartphone-based classifier to classify white light images into five categories of OC. The recent HRNet, trained on ImageNet, was validated on the collected image dataset with five data categories and the performance was acceptable. However, image acquisition is complex and requires domain knowledge. Nanditha et al. [45] considered a dataset with a considerable size of well-annotated oral lesions and used machine and deep learning models to build an automated diagnosis system for classifying the lesion into benign or malignant. Classifiers such as Naive Bayes, K-Nearest Neighbour (KNN), Support Vector Machine (SVM), Artificial Neural Network (ANN) and CNN were explored, and the results were compared. Also, a new CNN classification model was proposed that served as the new state-of-art for the classification task. Prachi et al. [46] developed a method for classifying normal tissues, leukoplakia and erythroplakia where RGB and YCbCr color spaces were considered and corresponding the box plots of red and Y color components, a threshold is estimated to divide the lesions into categories. Later, texture features were derived using the Gray Level Co-occurrence Matrix (GLCM) to make the final decisions. The efficiency was notable but could not generalize all the types of lesions due to the presence of both white and red regions. Also, the proposed model could not perform efficiently for both binary and multi-class classification. Kevin et al. [47] introduced a two-stage segmentation and classification approach for OC detection. Random oversampling of the data, followed by data augmentation was carried out to increase the number of samples. Finally, a CNN-based network Guided Attention Based Inference Network (GAIN) was considered for the classification and segmentation of OC images. GAIN network works on three streams for achieving the goal, classification stream for classification of the samples, attention mining stream for generating the attention maps and bounding box stream for improving accuracy through pixel-to-pixel supervision. The performance achieved was acceptable and GAIN could effectively show the affected regions. However, performance could be further improved and data duplication was carried out to solve the issue of class imbalance, followed by data augmentation on the duplicated data, which may have led to over-fitting. Also, there was no external validation considered for validating the proposed architecture. Song et al. [48] evaluated the performance of remote and onsite diagnosis of OC. The evaluation was done for specialists evaluating patients on-site and remotely compared to AI-based diagnosis systems. A CNN-based architecture, MobileNet, pre-trained on ImageNet was used for on-device diagnosis on a smartphone device, which required no external resources. Again, another module was developed for cloud-based diagnosis considering the Bayesian Deep Learning (BDL) network trained using VGG19. The adaptive histogram equalization technique was utilized to improve the brightness and contrast of the input images in the pre-processing step. However, the performance was not acceptable compared to the diagnosis of the specialists and dual-modality images were not combined to produce better outcomes. Warin et al. [49] evaluated numerous DL algorithms for OC diagnosis considering white light images of normal, OPMD and OSCC lesions. For the classification of OC lesions, DenseNet-169, ResNet-101, SqueezeNet and Swim-M Transformer were considered. For multi-class detection of the lesions, faster R-CNN, YOLOv5, RetinaNet and CenterNet2 were considered. Pre-trained model using transfer learning was considered except for SqueezeNet and Swim-S Transformer, which were trained from scratch. Also, attention maps for the affected regions were generated. The performance stood adequate for OPMD and OSCC classification from normal tissues but no external validation or cross-validation was considered for validating the model, and classification and detection were also considered independently. Al Duhayyim et al. [50] introduced a diagnosis system for the detection of oral cancer as benign or malignant from white light images, where a fusion-based feature extractor was considered and passed into an extreme learning machine (ELM) for categorization. The performance was notable but efficacy could be further improved and the issue of overfitting also needs to be clarified. Flugge et al. [51] developed a DL architecture based on Swim-Transformer for detecting OSCC. Also, gradient-weighted class activation (Grad-CAM) was applied to generate visual explanations, highlighting the important regions for classifying OSCC. The performance was satisfactory but neither validation was done on external datasets nor cross-validation was considered. Begum et al. [52] implemented a DL-CNN based technique to focus on the cancerous regions of the lip, tongue and cheeks and classify them to be cancerous or not. For this, a region proposal network (RPN) was proposed to extract the region of consideration which were then classified using a DenseNet201-based classifier. The performance was considerable but only accuracy as a metric was considered which cannot be completely acceptable when dealing with the diagnosis of cancer. Pahadiya et al. [53] proposed an algorithm for the segmentation and classification of tongue cancer using CNN. Initially, image enhancement techniques such as histogram equalization, contrast stretching and sharpening were considered for image enhancement,

after which the Gabor filter was considered to remove the unwanted noise present. Consequently, hybrid algorithms such as Firefly K-Means Watershed (FKW), and Hybrid Particle Swarm Optimization and Firefly K-Means Watershed (HPSOFKW) were implemented for segmenting smaller affected tissues and a comparison between the algorithms was highlighted. Finally, a two-layer CNN model was developed for classifying the segmented regions. Validation accuracy was considerable, but other performance metrics essential for biomedical image classification were not considered. Furthermore, there was no testing conducted on data that had not been previously seen, nor was there validation performed using an external dataset or cross-validation. Talwar et al. [54] aimed to evaluate DL algorithms for early diagnosis of OC. CNN architectures such as VGG19, Inception-ResNet-v2, MobileNet-v2, DenseNet-121, DenseNet-169, DenseNet-201 and transformer architectures such as Vanilla Vision Transformer (VVT), Data-Efficient Transformers, Swim Transformers were evaluated in this work. A 5-fold cross-validation was considered and heatmaps for the affected regions were generated for better visualization. The performance was considerable for DenseNet-201 in the case of CNN-based architecture and Swim Transformer for transformer-based architectures. However, the performance decreased when tested on images captured by non-trained front-line health workers (FLHWs) which shows the lack of generability in the model for the task. Also, due to the complexity and high resource requirements, these networks could not be considered for real-time applications.

It is observed from the literature survey that there is an essential need for a non-invasive technique for the diagnosis of OC in real-time with acceptable performance and negligible false positives. It is also observed that separate color spaces are significantly less explored. Also, colored texture features are significantly less explored while considering the white light images in OC classification. Therefore, in this paper, we aim to develop a technique for the classification of OC into its pre-cancerous stages using white light images to reduce the bias among physicians and classify the lesion in real-time with nominal resource utilization.

II. DATASET

In this work, two datasets were considered, a publicly available dataset and another dataset created in-house. The public dataset used [55] consisted of 630 oral images, a few collected from internet sources, while others collected from collaborating with numerous hospitals and consulting oral specialists. From those images, 1200 lesion areas were pulled and individual images for lesion areas were obtained. Having the 1200 lesion images obtained, 600 were malignant images and 600 were normal images of the oral cavity. The in-house dataset was created by using leukoplakia, erythro-leukoplakia and erythroplakia as keywords in Google and Mozilla search engines to collect images of three pre-cancerous stages: leukoplakia, erythro-leukoplakia and erythroplakia. The dataset created consists of 200 images having 100 images of leukoplakia and 50 images each of erythro-leukoplakia and erythroplakia.

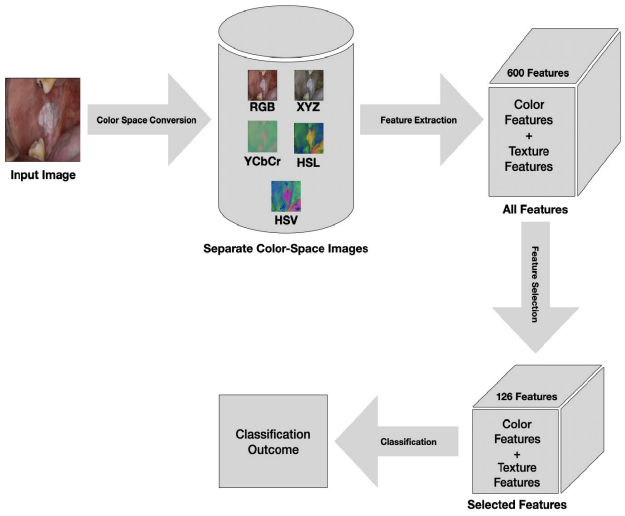

FIGURE 1. Flow diagram of the proposed methodology.

III. METHODOLOGY

The proposed method in this article comprises of first pre-processing the input image by converting the image into separate color spaces. Secondly, color and texture features were pulled from each channel of every color space, and feature selection was carried out to extract significant features. Finally, a classifier was trained and tested on the selected features. The flow diagram of the model is shown in Figure 1.

A. DATA AUGMENTATION

As the number of images is significantly less and unevenly distributed among the classes in the multi-class classification dataset, data augmentation was necessary to bring balance into the dataset and increase the number of images for training and testing. Augmentation was carried out on erythro-leukoplakia and erythroplakia images by performing horizontal flipping and ninety degree clockwise rotation operations to tackle the data imbalance in the dataset. After the augmentation step, the dataset consisted of 300 images having 100 images of each class.

B. PRE-PROCESSING

In the pre-processing step, all images were resized into 264 $\times$ 264 pixels and transformed from Red Green Blue (RGB) color space to four other color spaces: Hue Saturation Value (HSV) color space, Hue Saturation Lightness (HLS) color space, YCrCb color space, XYZ color space [56], [57]. Different color spaces were explored to better understand the changes occurring in different pre-cancerous stages in the oral cavity. Finally, label encoding was performed on the dataset where every class was assigned a number, making the classification task simple.

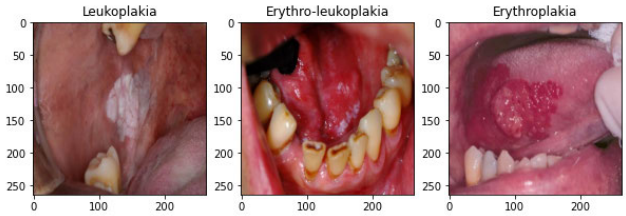

FIGURE 2. Pre-cancerous stages in RGB color space.

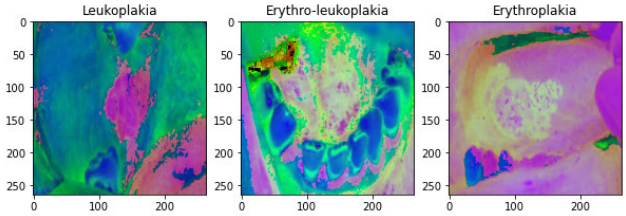

FIGURE 3. Pre-cancerous stages in HSV color space.

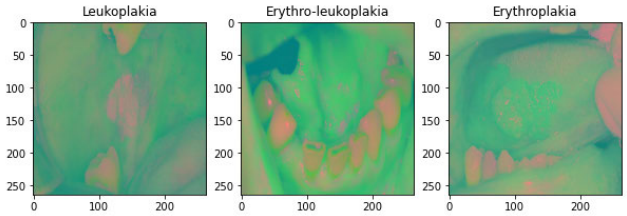

FIGURE 4. Pre-cancerous stages in YCbCr color space.

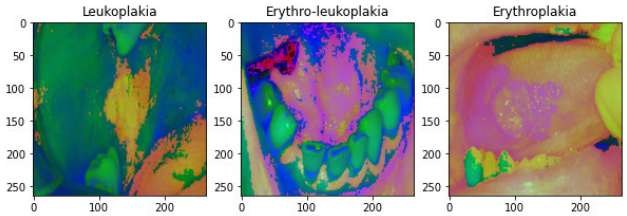

FIGURE 5. Pre-cancerous stages in HLS color space.

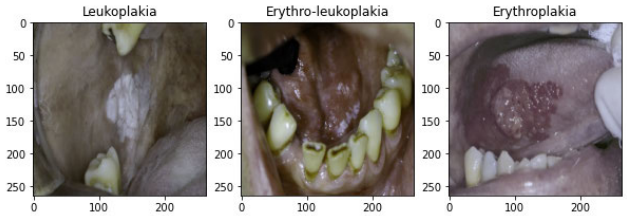

FIGURE 6. Pre-cancerous stages in XYZ color space.

1) RGB COLOR SPACE

The RGB color model is a preservative color model in which the red, green and blue primary colors of light are counted together in diverse ways to propagate an expansive array of colors. However, RGB is a device-dependent color model where RGB levels may differ for particular devices or even in the exact device over some time. Thus, an RGB value does not represent the exact color across instruments without some color control. The images from second dataset in RGB color space is shown in Figure 2.

2) HSV COLOR SPACE

This color model does not use primary colors straight. It utilizes color in the manner humans perceive them. Hue is the color element. Saturation represents the percentage of the color, this value typically varies between 0-1. While saturation represents the gray color, the value denotes the intensity of the color chosen. The images from second dataset in HSV color space is shown in Figure 3.

3) YCbCr COLOR SPACE

YCbCr is a system of color spaces utilized as a portion of the color image channel in video and digital photography approaches. Y is luminance, Cb and Cr are the blue-difference and red-difference chroma features, suggesting that light intensity is non-linearly ingrained and established on gamma-corrected RGB primaries. YCbCr is a suitable estimation to color processing and intuitive uniformity, where the primary colors compared approximately to red, green and blue are processed into intuitive significant details. The images from the second dataset in YCbCr color space is shown in Figure 4.

4) HLS COLOR SPACE

This color model does not use primary colors directly. It uses color in the way humans perceive them. The HSL color model is represented how different paints are mixed to create a shade in the real world, with the lightness proportions corresponding to the irregular amount of black or white paint in the combination. The distinction between HSV and HLS is that the color holding the utmost lightness in HSL is pure white, but the color holding the utmost lightness in HSV is identical to a shining white light on a colored entity. The images from the second dataset in HLS color space is shown in Figure 5.

5) XYZ COLOR SPACE

The XYZ color space was constructed by the International Commission on Illumination (CIE) in 1931. The CIE XYZ color space surrounds all color phenomena visible to a person with moderate eyesight. That is the very reason CIE XYZ is a device-independent portrayal of color. It is a typical concern through which numerous additional color spaces are determined. Y is the luminance, Z is quasi-equal to blue (of CIE RGB), and X is a combination of the three CIE RGB curves selected to be non-negative. The images from the second dataset in XYZ color space is shown in Figure 6.

C. FEATURE EXTRACTION

From the five different color spaces, color and texture features were extracted to better understand and classify the lesions into appropriate pre-cancerous stages.

1) COLOR FEATURES

The statistical color features such as the median, peak position, energy, entropy and skewness were extracted from the histogram of all three channels for the five different color spaces.

- **Median** - The median is a statistical feature that tells us about the central value of a signal. It indicates the general brightness of the image.
- **Peak Position** - The position at which the peak is obtained suggests the change in the histogram distribution across different classes.
- **Energy** - The energy measure reveals how the gray level is distributed along each channel.
- **Entropy** - Entropy measures how many bits are required to encode the data. It gives an intuition about the details present in the image, the more the value of entropy indicates a more detailed picture.
- **Skewness** - Skewness measures the symmetry about the mean of the gray level distribution.

2) TEXTURE FEATURES

The texture could be termed as a repeating design of regional variations in image intensity illustrated by the spatial dissemination of intensity levels in an encompassing. Statistical texture features consider texture as a perceptible estimation of the arrangement of intensities in a region. In this work, statistical texture descriptors such as the Gray Level Co-occurrence Matrix (GLCM) and the Gray Level Run Length matrix (GLRLM) are considered to better understand the lesions in the image.

- **GLCM**
  A GLCM holds information about the position of the pixels having comparable gray-level values. It is a two-dimensional array in which rows and columns represent potential resemblance values. A GLCM of size N x N represents the second-order joint probability function of an image area restrained by the mask characterized as $G(x, y|d, \theta)$. The $(x, y)^{th}$ element of this matrix depicts the number of times the combination of levels x and y occur in the image that is separated by a distance of d pixels along angle θ. Quantitative texture measures used in this work to measures the properties of texture calculated from GLCM are as follows:
  - **Energy**- Energy measures the uniformity present in the image in the range [0,1]. It is usually high when all the elements of GLCM are equal. The equation for energy is shown in equation-1.

$$E = \sum_x \sum_y G(x, y|d, \theta)^2 \tag{1}$$

  - **Correlation**- Correlation measures the amount of linearity present in the image. A high correlation indicates that the image has a considerable amount of linear structures. The equation for correlation is shown in equation-2.

$$Corr = \sum_x \sum_y \frac{(x - \mu_x)(y - \mu_y)G(x, y|d, \theta)}{\sigma_x \sigma_y} \tag{2}$$

    where μ_x, μ_y, σ_x and σ_y denotes the means and standard deviations along the x and y labels respectively
  - **Dissimilarity**- Dissimilarity estimates the distance between pairs of entities (pixels) in the area of interest. The equation for dissimilarity is shown in equation-3.

$$Dissimilarity = \sum_x \sum_y |x - y| G(x, y|d, \theta) \tag{3}$$

  - **Homogeneity**- Local homogeneity estimates the likeness of the distribution of entities in the GLCM to the GLCM diagonal. The equation for homogeneity is shown in equation-4.

$$Homog. = \sum_x \sum_y \frac{G(x, y|d, \theta)}{1 + |x - y|^2} \tag{4}$$

  - **Contrast**- Contrast estimates the intensity difference between a pixel and its neighbor over the entire image. The contrast is usually zero for a constant image. The equation for contrast is shown in equation-5.

$$Contrast = \sum_x \sum_y |x - y|^2 G(x, y|d, \theta) \tag{5}$$

  In this work, the above-mentioned statistical features were extracted from GLCM for each channel for a distance of 1,3 and 5 pixels along different angles of 0°, 45°.
- **GLRLM**
  GLRLM gives us information about the number of consecutive pixels with the same gray level value. In other words, it quantifies the gray level run. In a gray level run length matrix $Gr(x, y|\theta)$, the $(x, y)^{th}$ element represents the number of runs with gray level x and length y that appear in the area along angle θ. Let N_g be the number of discrete intensity values in the image, N_r be the number of discrete run lengths in the image, N_p be the number of voxels in the image, $N_r(\theta)$ be the number of runs in the image along angle θ, which is equal to $\sum_{x=1}^{N_g} \sum_{y=1}^{N_r} Gr(x, y|\theta)$ and $1 \leq N_r(\theta) \leq N_p$. Quantitative texture measures used in this work to measure the properties of texture calculated from GLRLM are as follows:
  - **Short Run Emphasis (SRE)**- SRE measures the distribution of the short run lengths. A large value of SRE indicated more fine structural textures. The equation for SRE is shown in equation-6.

$$SRE = \frac{\sum_{x=1}^{N_g} \sum_{y=1}^{N_r} \frac{Gr(x,y|\theta)}{y^2}}{N_r(\theta)} \tag{6}$$

- **Long Run Emphasis (LRE)**- LRE measures the distribution of the long run lengths. A large value of LRE indicates a more coarse structural texture. The equation for LRE is shown in equation-7.

$$LRE = \frac{\sum_{x=1}^{N_g}(\sum_{y=1}^{N_r} Gr(x,y|\theta))^2}{N_r(\theta)} \tag{7}$$

- **Gray Level Uniformity (GLU)**- GLU estimates the likeness of gray level intensity values. Lower GLU indicates more difference between the intensities. The equation for GLU is shown in equation-8.

$$GLU = \frac{\sum_{x=1}^{N_g}(\sum_{y=1}^{N_r}(1 - Gr(x,y|\theta)))^2}{N_r(\theta)} \tag{8}$$

- **Run Length Uniformity (RLU)**- RLU estimates the likeness of run lengths. A smaller value indicates less homogeneity among run lengths. The equation for RLU is shown in equation-9.

$$RLU = \frac{\sum_{y=1}^{N_r}(\sum_{x=1}^{N_g}(1 - Gr(x,y|\theta)))^2}{N_r(\theta)} \tag{9}$$

- **Run Percentage (RPC)**- RPC estimates the coarseness of the texture by calculating the ratio of the number of runs to the number of voxels in the ROI. A higher value indicates that a more significant part of ROI consists of short runs or have for fine textures. The equation for RPC is shown in equation-10.

$$RPC = \frac{N_r(\theta)}{N_p}; \quad \frac{1}{N_p} \leq RPC \leq 1 \tag{10}$$

In this work, the above-mentioned statistical features were extracted from GLRLM for each channel considering the average of the different angles such as 0°, 45°, 90° and 135° and creating a feature map.

D. FEATURE SELECTION

During this step, one of the initial tasks was to eliminate any duplicate features that were present. After that, feature selection was carried out estimating the feature importance, determined by considering the information gain of each feature calculated using a decision tree based approach from which significantly contributing features were selected by setting a threshold. The information gain could be considered as the ability of the feature to contribute towards the separation of the classes. Feature selection based on the information gain eases the identification and retention of the most instructive features while discarding less instructive ones, directing to more methodical and decipherable models. The formula of the feature importance score I_f for an individual feature f is shown in equation-11.

$$I_f = \sum_{i=1}^{N} \frac{Gain_i}{Total\ Gain} \tag{11}$$

Here, N denotes the total number of trees in the ensemble, $Gain_i$ denotes the information gain at the i^{th} node for the feature f and Total Gain denotes the sum of the information gains across all nodes and features. The ratio assures that the importance scores are standardized, delivering a relative measure of the contribution of each feature. After identifying these features, classification was performed using those selected parameters.

E. CLASSIFIER

The classifier considered in this work is Light Gradient Boosting Machine (LightGBM), a gradient boosting framework based on decision trees to increase capability and decrease memory usage. Two distinguished approaches, Gradient-based One Side Sampling (GOSS) and Exclusive Feature Bundling (EFB) form the characteristics of the LightGBM algorithm [58]. These components meet the requirements of the histogram-based algorithm that is mainly used in all Gradient Boosting Decision Tree (GBDT) frameworks. They work together to construct the model function effectively, giving it an edge over other GBDT frameworks.

1) GRADIENT-BASED ONE SIDE SAMPLING FOR LightBGM ALGORITHM

In the analysis of information gain, different data samples play varying roles. When the models have significant gradients, they tend to impart significantly to the information gain. GOSS algorithm identifies those occurrences with substantial gradients, usually exceeding a premeditated threshold or among the top percentiles. However, it unmethodically excludes instances with trivial gradients to ensure the accuracy of information gain estimation. This approach results in more precise gain estimation compared to uniformly random sampling, predominantly when the result of information gain has a wide radius.

2) EXCLUSIVE FEATURE BUNDLING FOR LightBGM ALGORITHM

Sparse high-dimensional data offer opportunities for developing lossless techniques leading to the deduction in the number of features without compromising accuracy. In an infrequent feature space, innumerable features are mutually exclusive, i.e., they never have nonzero outcomes simultaneously. Such features can be grouped together safely into an Exclusive Feature Bundle, thereby reducing the intricacy of the histogram structure from Order of data × feature to Order of data × bundle, where bundle is much smaller than feature. This approach enhances the training framework's speed while maintaining accuracy.

3) ARCHITECTURE

Instead of growing decision trees level-by-level like other boosting algorithms, LightGBM breaks the tree leaf-wise. It single outs the leaf with the most significant delta loss

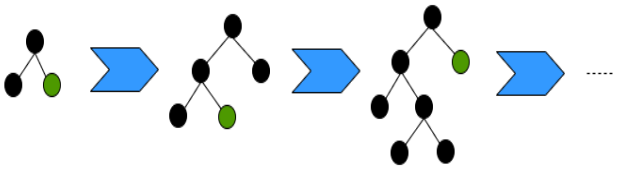

FIGURE 7. Leaf-wise tree growth.

to expand, ensuring less optimal loss than the level-wise algorithm. Additionally, it's worth mentioning that the leaf-wise growth approach may increase the algorithm's complexity and lead to overfitting in datasets with substantially less data. The diagrammatic representation of Leaf-Wise Tree Growth is shown in Figure 7.

IV. IMPLEMENTATION

For classification purposes, both datasets were divided such that 90% of the data formed the training set and the remaining 10% formed the testing set. The 90% of the data considered for the training set was utilized to perform 10-Fold Cross Validation where iteratively 90% of the data was considered for training and the rest 10% was considered for validation. The model was evaluated at every iteration for 10 folds and finally, the average performance of all the iterations was considered.

For the first dataset used in [45], the LightGBM classifier was trained using the training set for 5000 iterations using a learning rate of 0.01, boosting type was set as gradient boosting with the number of leaves set to 10 and the maximum depth also set to 10. Binary-logloss was considered the evaluation metric for validating the model using the validation set.

For the second dataset created in this work, the LightGBM classifier was trained using the training set for 100 iterations using a learning rate of 0.05, boosting type was set as gradient boosting with the number of leaves set to 10 and maximum depth also set to 10. Here, multi-logloss was considered the evaluation metric for validating the model using the validation set. The equations of logloss for a particular sample and the total logloss are shown in equation-12 and equation-13 respectively.

$$LogLoss_i = -[y_i \ln(p_i) + (1 - y_i) \ln(1 - p_i)] \tag{12}$$

Here, i denotes the sample number, ln denotes the natural logarithm, y_i denotes the actual outcome and p_i denotes the predicted value.

$$LogLoss = \frac{1}{N} \sum_{i=1}^{N} Logloss_i \tag{13}$$

Here, N denotes the total number of samples.

V. PERFORMANCE METRICS

The following metrics were calculated for quantifying the performance: Accuracy, Precision, Recall, F1-score and Specificity. These criterion counts agreement between the actual label and the predicted levels of false-positive, false-negative, true-negative and true-positive enumerations. In our proposition, true-positives (TP) correspond to images accurately labeled as belonging to the category under analysis, in other words, a person having oral cancer is predicted to have oral cancer. False positives (FP) are images incorrectly identified as belonging to the category under investigation, in other words, an individual not having oral cancer is predicted to have it. True-negatives (TN) images are correctly classified as not belonging to the relevant category, in other words, an individual not having oral cancer is predicted as not having it. In comparison, false negatives (FN) are pixels that are incorrectly classified, in other words, an individual having oral cancer is predicted as not having it.

The parameters computed are as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{14}$$

$$Precision = \frac{TP}{TP + FP} \tag{15}$$

$$Recall = \frac{TP}{TP + FN} \tag{16}$$

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{17}$$

$$Specificity = \frac{TN}{TN + FP} \tag{18}$$

Accuracy is a significantly used performance metric that is merely the ratio of accurately classified observations to all observations, as shown in equation-14. Here, it indicates how well the algorithm categorizes oral cancer into its pre-cancerous stages. However, when it comes to identifying diseases like oral cancer, accuracy cannot be considered the only performance measure as accuracy does not give any intuition about the correctly and wrongly classified stages. To counter this issue, precision, recall, f1-score and specificity were computed. Precision provides the intuition about the proportion of individuals diagnosed with oral cancer have it, as shown in equation-15. In contrast, recall shows that the proportion of individuals with oral cancer was diagnosed the same by the model, as shown in equation-16. Specificity gives the intuition about the ratio of individuals that did not have oral cancer were predicted by the model as non-cancerous, as shown in equation-17. F1-Score is the harmonic mean of precision and recall which is significant when dealing with unbalanced data, as shown in equation-18.

VI. RESULTS AND DISCUSSION

A. CLASSIFIER EVALUATION

After the feature selection step, both datasets were divided such that 90% of the data formed the training set and the remaining 10% formed the testing set. To find the optimal classifier, both datasets were evaluated using several machine learning classifiers such as the Naive Bayes Classifier (NB)

using the Gaussian kernel [59], a Support Vector Machine Classifier (SVM) using the Radial Basis Function kernel [60], a Random Forest Classifier (RF) [61] with a maximum depth of 10 layers, a 2 layer Multilayer Perceptron (MLP) [62] with ReLU activation function and Adam optimizer with learning rate of 0.001, and a LightGBM Classifier [58] with the maximum depth of 10 layers. Due to the sparse amount of data, deep learning techniques could not be evaluated.

For the binary classification, the aforementioned classifiers were trained and tested on the color, texture and the combination of both as shown in Table 1, Table 2 and Table 3 respectively. Similarly, for the multi-class classification, the aforementioned classifiers were trained and tested on the color, texture and the combination of both as shown in Table 4, Table 5 and Table 6 respectively.

TABLE 1. Classifier evaluation for color features in binary classification.

Classifier	Accuracy	Precision	Recall	F1-score	Specificity
NB	58.59%	57.65%	56.21%	56.92%	56.21%
SVM	65.19%	64.69%	64.52%	64.60%	64.52%
RF	71.80%	72.98%	72.83%	72.90%	72.83%
MLP	56.38%	61.74%	59.14%	60.41%	59.14%
LightGBM	**89.95%**	**90.56%**	**90.77%**	**90.66%**	**90.77%**

TABLE 2. Classifier evaluation for texture features in binary classification.

Classifier	Accuracy	Precision	Recall	F1-score	Specificity
NB	64.76%	66.38%	65.99%	66.18%	65.99%
SVM	65.19%	65.33%	65.51%	65.42%	65.51%
RF	98.68%	98.56%	98.81%	98.68%	98.81%
MLP	65.19%	69.28%	67.27%	68.26%	67.27%
LightGBM	**99.21%**	**99.13%**	**99.28%**	**99.02%**	**99.27%**

TABLE 3. Classifier evaluation for the combination of color and texture features in binary classification.

Classifier	Accuracy	Precision	Recall	F1-score	Specificity
NB	68.28%	68.17%	68.38%	68.27%	68.38%
SVM	66.96%	66.98%	67.19%	66.57%	67.19%
RF	96.92%	97.02%	96.78%	96.89%	97.02%
MLP	75.33%	75.65%	75.91%	75.78%	75.91%
LightGBM	**99.25%**	**99.18%**	**99.31%**	**99.24%**	**99.31%**

TABLE 4. Classifier evaluation for color features in multi-class classification.

Classifier	Accuracy	Precision	Recall	F1-score	Specificity
NB	62.22%	38.95%	41.43%	40.15%	70.97%
SVM	66.67%	62.78%	46.19%	53.22%	72.92%
RF	96.92%	97.02%	96.78%	96.89%	95.83%
MLP	80%	71.25%	70.32%	70.78%	84.72%
LightGBM	**98.66%**	**98.06%**	**89.97%**	**90.01%**	**99.05%**

From the above analysis, it was evident that LightGBM Classifier performs the best for both the datasets and hence, it was considered the optimal classifier for this work.

B. FEATURE ANALYSIS

The feature analysis was carried out in-depth with the aim of selecting the optimal number of features that could bring balance to resource utilization and performance. This was done to ensure that the system operates efficiently without compromising on performance. Additionally, individual color and texture features were analyzed for binary and multi-class classification in order to gain insight into the significant features and color spaces that should be considered to detect changes in the oral cavity caused by malignancy. Initially, there were 75 color features and 525 texture features, combining to a total of 600 features. Based on the information gain of the features, features were reduced to find the optimal number of features in all the cases and finally, cross validation was carried out.

TABLE 5. Classifier evaluation for texture features in multi-class classification.

Classifier	Accuracy	Precision	Recall	F1-score	Specificity
NB	71.11%	55.55%	51.27%	53.32%	76.94%
SVM	60%	28.11%	35.24%	31.27%	67.77%
RF	97.78%	97.78%	96.67%	97.22%	97.92%
MLP	70.33%	60.93%	60.32%	60.62%	80.69%
LightGBM	**98.67%**	**98.28%**	**98.19%**	**98.23%**	**98.89%**

TABLE 6. Classifier evaluation for the combination color and texture features in multi-class classification.

Classifier	Accuracy	Precision	Recall	F1-score	Specificity
NB	66.66%	47.52%	45.92%	46.71%	74.24%
SVM	68.88%	60.82%	60%	60.41%	77.86%
RF	97.77%	95.24%	96.29%	95.76%	98.61%
MLP	86.66%	79.87%	79.99%	79.93%	89.40%
LightGBM	**98.88%**	**98.86%**	**97.92%**	**98.38%**	**99.03%**

For the binary classification of oral cancer as benign or malignant, while considering only the color features, it was observed that 46 features significantly had the highest information gain and after applying 10-fold cross validation achieved an accuracy of 89.95%, precision of 90.56%, recall of 90.77%, f1-score of 90.66% and specificity of 90.77% as shown in Table 7 achieving a reduction in the number of features by 38.66%. Similarly, while considering texture features it was observed that 62 features significantly had the highest information gain and after applying 10-fold cross validation achieved an accuracy of 99.21%, precision of 99.13%, recall of 99.28%, f1-score of 99.02% and specificity of 99.27% as shown in Table 8 achieving a reduction in the number of features by 88.19%. Now, combining both color and texture features as shown in Table 9, it was observed that 61 features significantly had the highest information gain and after applying 10-fold cross validation achieved an accuracy of 99.25%, precision of 99.18%, recall of 99.31%, f1-score of 99.24% and specificity of 99.31% achieving a reduction in the number of features by 88.19%. By virtue of this experiment, it was observed that texture features contribute maximum towards predicting benign or malignant tissues in the oral cavity. Also, it has been observed that YCbCr color space is vital for understanding the differences and similarities of the same.

For the multi-class classification of oral cancer as benign and malignant, while considering only the color features,

TABLE 7. 10-fold cross validation on test set for color features in binary classification.

Iterations	Accuracy	Precision	Recall	F1-score	Specificity
Iteration-1	91.63%	91.91%	92.36%	92.13%	92.36%
Iteration-2	91.63%	91.91%	92.36%	92.13%	92.36%
Iteration-3	92.07%	92.15%	92.66%	92.40%	92.66%
Iteration-4	89.43%	89.27%	89.69%	89.48%	89.69%
Iteration-5	92.07%	92.15%	92.66%	92.40%	92.66%
Iteration-6	89.87%	90.15%	90.58%	90.36%	90.58%
Iteration-7	89.87%	90.15%	90.58%	90.36%	90.58%
Iteration-8	88.54%	89.76%	89.68%	89.72%	89.68%
Iteration-9	86.34%	87.69%	88.26%	87.97%	87.69%
Iteration-10	88.10%	89.45%	89.28%	89.36%	89.28%
Average	**89.95%**	**90.56%**	**90.77%**	**90.66%**	**90.77%**

TABLE 8. 10-fold cross validation on test set for texture features in binary classification.

Iterations	Accuracy	Precision	Recall	F1-score	Specificity
Iteration-1	99.56%	99.51%	99.60%	99.55%	99.60%
Iteration-2	99.56%	99.51%	99.60%	99.55%	99.60%
Iteration-3	99.12%	99.11%	99.11%	99.11%	99.11%
Iteration-4	99.56%	99.51%	99.60%	99.55%	99.60%
Iteration-5	99.12%	99.02%	99.21%	99.09%	99.20%
Iteration-6	99.12%	99.02%	99.21%	99.09%	99.20%
Iteration-7	98.68%	98.56%	98.81%	98.68%	98.81%
Iteration-8	98.68%	98.56%	98.81%	98.68%	98.81%
Iteration-9	99.56%	99.51%	99.60%	99.55%	99.60%
Iteration-10	99.12%	99.03%	99.21%	99.12%	99.21%
Average	**99.21%**	**99.13%**	**99.28%**	**99.02%**	**99.27%**

TABLE 9. 10-fold cross validation on test set for the combination of color and texture features in binary classification.

Iterations	Accuracy	Precision	Recall	F1-score	Specificity
Iteration-1	99.56%	99.51%	99.60%	99.55%	99.60%
Iteration-2	99.56%	99.51%	99.60%	99.55%	99.60%
Iteration-3	99.11%	99.11%	99.11%	99.11%	99.11%
Iteration-4	99.56%	99.51%	99.60%	99.55%	99.60%
Iteration-5	99.56%	99.51%	99.60%	99.55%	99.60%
Iteration-6	99.11%	99.03%	99.20%	99.11%	99.20%
Iteration-7	98.68%	98.56%	98.81%	98.68%	98.81%
Iteration-8	99.12%	99.03%	99.20%	99.11%	99.20%
Iteration-9	99.12%	99.03%	99.20%	99.11%	99.20%
Iteration-10	99.12%	99.03%	99.20%	99.11%	99.20%
Average	**99.25%**	**99.18%**	**99.31%**	**99.24%**	**99.31%**

it was observed that 45 features significantly had the highest information gain and after applying 10-fold cross validation achieved an accuracy of 98.66%, precision of 98.06%, recall of 89.97%, f1-score of 90.01% and specificity of 99.05% as shown in Table 10 achieving a reduction in the number of features by 40%. Similarly, while considering texture features it was observed that 65 features significantly had the highest information gain and after applying 10-fold cross validation achieved an accuracy of 98.67%, precision of 98.28%, recall of 98.19%, f1-score of 98.23% and specificity of 98.89% as shown in Table 11 achieving a reduction in the number of features by 87.61%. Now, combining both color and texture features as shown in Table 12, it was observed that 126 features significantly had the highest information gain and after applying 10-fold cross validation achieved an accuracy of 98.88%, precision of 98.86%, recall of 97.92%, f1-score of 98.38% and specificity of 99.03% achieving a reduction in the number of features by 79%. By virtue of this experiment, it was observed that both color and texture features contribute significantly towards predicting the pre-cancerous stages of OC. Also, it has been observed that HSV and HSL color spaces are vital for understanding the differences and similarities of the oral cancer tissues.

TABLE 10. 10-fold cross validation on test set for color features in multi-class classification.

Iterations	Accuracy	Precision	Recall	F1-score	Specificity
Iteration-1	100%	100%	100%	100%	100%
Iteration-2	100%	100%	100%	100%	100%
Iteration-3	100%	100%	100%	100%	100%
Iteration-4	100%	100%	100%	100%	100%
Iteration-5	100%	100%	100%	100%	100%
Iteration-6	100%	100%	100%	100%	100%
Iteration-7	100%	100%	100%	100%	100%
Iteration-8	97.78%	96.97%	97.62%	97.29%	98.33%
Iteration-9	93.33%	92.31%	92.86%	92.58%	95%
Iteration-10	95.55%	91.67%	95.24%	93.42.11%	97.22%
Average	**98.66%**	**98.06%**	**89.97%**	**90.01%**	**99.05%**

TABLE 11. 10-fold cross validation on test set for texture features in multi-class classification.

Iterations	Accuracy	Precision	Recall	F1-score	Specificity
Iteration-1	100%	100%	100%	100%	100%
Iteration-2	100%	100%	100%	100%	100%
Iteration-3	100%	100%	100%	100%	100%
Iteration-4	97.78%	97.78%	96.67%	97.22%	97.92%
Iteration-5	97.78%	97.78%	96.67%	97.22%	97.92%
Iteration-6	100%	100%	100%	100%	100%
Iteration-7	97.78%	97.78%	96.67%	97.22%	97.92%
Iteration-8	100%	100%	100%	100%	100%
Iteration-9	95.55%	91.67%	95.24%	93.42%	97.22%
Iteration-10	97.78%	97.78%	96.67%	97.22%	97.92%
Average	**98.67%**	**98.28%**	**98.19%**	**98.23%**	**98.89%**

TABLE 12. 10-fold cross validation on test set for the combination of color and texture features in multi-class classification.

Iterations	Accuracy	Precision	Recall	F1-score	Specificity
Iteration-1	100%	100%	100%	100%	100%
Iteration-2	100%	100%	100%	100%	100%
Iteration-3	97.77%	96.66%	97.77%	97.21%	98.41%
Iteration-4	97.77%	98.24%	96.29%	97.25%	98.61%
Iteration-5	100%	100%	100%	100%	100%
Iteration-6	100%	100%	100%	100%	100%
Iteration-7	97.77%	97.92%	96.29%	97.09%	97.77%
Iteration-8	100%	100%	100%	100%	100%
Iteration-9	97.77%	97.92%	94.44%	96.15%	97.77%
Iteration-10	97.77%	97.92%	94.44%	96.15%	97.77%
Average	**98.88%**	**98.86%**	**97.92%**	**98.38%**	**99.03%**

C. OUTCOME AND COMPARISON

It is observed from the previous sections that a combination of color and texture features yielded the best performance for

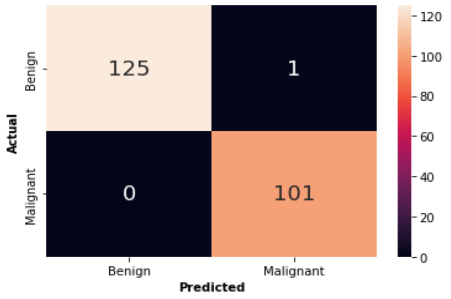

FIGURE 8. Confusion matrix for test data of fifth iteration for binary classification.

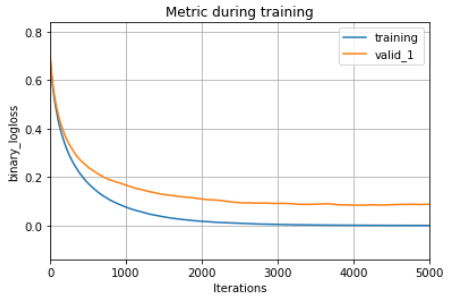

FIGURE 9. Logloss curve of fifth iteration for binary classification.

the proposed algorithm. The proposed algorithm performed well in the first dataset with a testing accuracy of 99.25%, precision of 99.18%, recall of 99.31%, f1-score of 99.24% and specificity of 99.31%. For the combination on color and texture features, the confusion matrix generated on testing data of the fifth iteration in binary classification is shown in Figure 8. The loss curve for 5000 iterations for training and validation data of the fifth iteration are shown in Figure 9. The performance of the proposed method compared to the state-of-the-art algorithms is shown in Table 13. The structure of the decision tree from the fifth iteration could also be visualized in Figure 10.

TABLE 13. Testing performance measures of ML models for binary classification.

Evaluation measure	ANN [45]	CNN [45]	Swim Trans-former [51]	Proposed Algorithm
Accuracy	97.00%	97.51%	98.58%	**99.25%**
Precision	98.18%	97.35%	98.57%	**99.18%**
Recall	98.54%	97.67%	96.57%	**99.31%**
F1-Score	98.18%	97.51%	97.53%	**99.24%**
Specificity	95.55%	97.34%	98.59%	**99.31%**

For the second dataset, the proposed algorithm performed well in the testing set with none of the images overlapping with the training set, with an accuracy of 98.88%, precision of 98.86%, recall of 97.92%, f1-score of 98.38% and specificity of 99.03%. For the combination on color and texture features, the confusion matrix generated on the testing data of the fifth iteration for multi-class classification is shown in Figure 11.

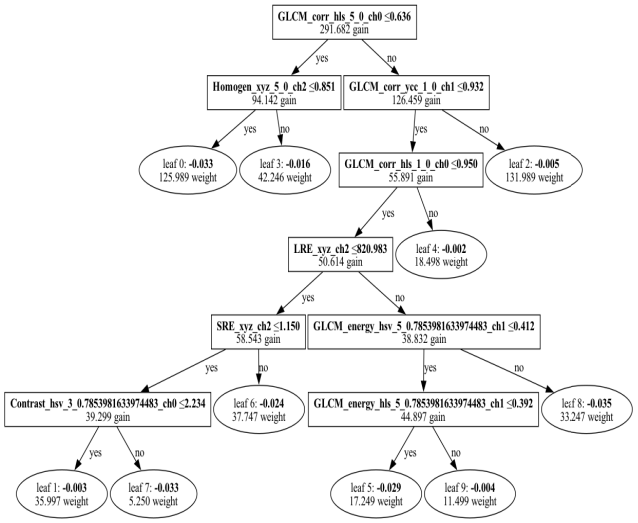

FIGURE 10. Decision tree of fifth iteration for binary classification.

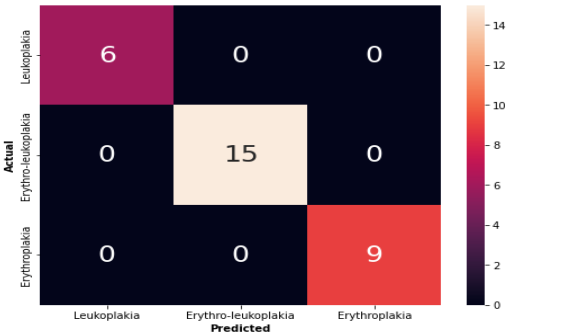

FIGURE 11. Confusion matrix for test data of fifth iteration for multi-class classification.

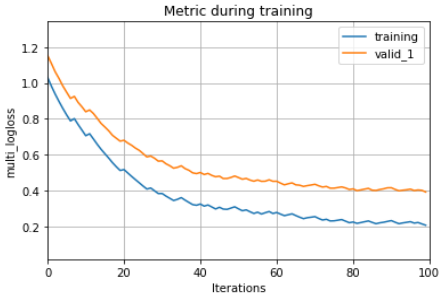

FIGURE 12. Logloss curve of fifth iteration for multi-class classification.

TABLE 14. Testing performance measures of ML models for multi-class classification.

Evaluation measure	BAPNN [39]	Dual Modality CNN [40]	DL. Algo [44]	Proposed Algorithm
Accuracy	97.92%	81.00%	95.07%	**98.88%**
Precision	-	79%	84.3%	**98.86%**
Recall	-	-	83%	**97.92%**
F1-Score	-	-	83.6%	**98.38%**
Specificity	-	82%	96.6%	**99.03%**

The loss curve for 100 iterations for training and validation data of the fifth iteration is shown in Figure 12. The testing

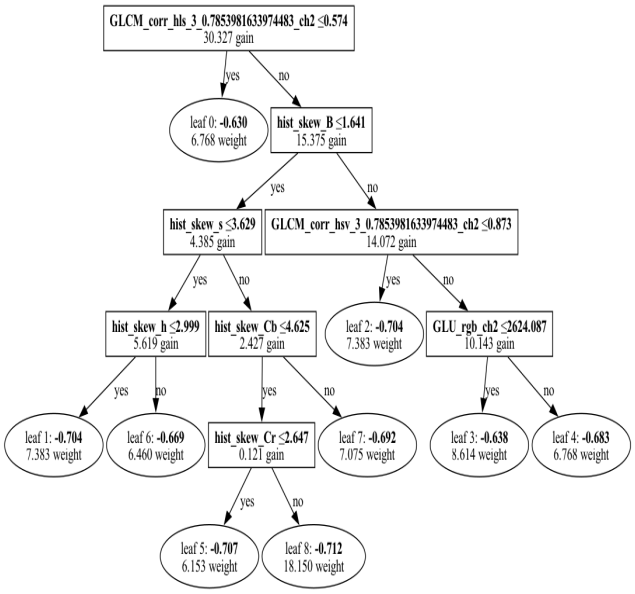

FIGURE 13. Decision tree of fifth iteration for multi-class classification.

performances of the classification models presented in the literature compared to our proposed model are shown in Table 14. The structure of the final decision tree from the fifth iteration could also be visualized in Figure 13.

VII. CONCLUSION

The study presents a generalized method for classifying oral cavity lesions into benign or malignant for binary classification and their pre-cancerous stages for multi-class classification. In this work, a novel technique of exploring different color spaces and extracting features from all the color spaces is employed for classifying the stages of oral cancer. The images are converted into five different color spaces, color and texture features are extracted and classified using Light Gradient Boosting Machine (LightGBM) algorithm. The overall performance is satisfactory, with a testing accuracy of 99.25%, precision of 99.18%, recall of 99.31%, f1-score of 99.24% and specificity of 99.31% for the binary classification of benign or malignant tissues and a testing accuracy of 98.88%, precision of 98.86%, recall of 97.92%, f1-score of 98.38% and specificity of 99.03% for multi-class classification into the pre-cancerous stages of oral cancer. The generalized algorithm put forward has outperformed the state-of-art algorithms for both binary and multi-class classification of oral cancer with sparse resources and data. Also, the method has been implemented using hand-crafted feature extraction and machine learning classifier, making it significantly less time and resource-consuming compared to deep learning methods.

ACKNOWLEDGMENT

The authors extend their appreciation to King Saud University for funding this research through Researchers Supporting Project Number (RSPD2024R890), King Saud University, Riyadh, Saudi Arabia.

REFERENCES

[1] S. Sarkar, G. Horn, K. Moulton, A. Oza, S. Byler, S. Kokolus, and M. Longacre, "Cancer development, progression, and therapy: An epigenetic overview," *Int. J. Mol. Sci.*, vol. 14, no. 10, pp. 21087–21113, Oct. 2013.

[2] P. H. Monter and S. G. Patel, "Cancer of the oral cavity," *Surgical Oncol. Clinics North Amer.*, vol. 24, no. 3, pp. 491–508, Apr. 2015.

[3] S. Warnakulasuriya, N. W. Johnson, and I. Van Der Waal, "Nomenclature and classification of potentially malignant disorders of the oral mucosa," *J. Oral Pathol. Med.*, vol. 36, no. 10, pp. 575–580, Nov. 2007.

[4] B. W. Neville and T. A. Day, "Oral cancer and precancerous lesions," *CA, A Cancer J. Clinicians*, vol. 52, no. 4, pp. 195–215, Jul. 2002.

[5] J. Bagan, G. Sarrion, and Y. Jimenez, "Oral cancer: Clinical features," *Oral Oncol.*, vol. 46, no. 6, pp. 414–417, Jun. 2010.

[6] S. Sudha, A. Veluthattil, S. Kandasamy, and S. Chakkalakkoombil, "Effect of hypofractionated, palliative radiotherapy on quality of life in late-stage oral cavity cancer: A prospective clinical trial," *Indian J. Palliative Care*, vol. 25, no. 3, p. 383, 2019.

[7] U. Mangalath, S. Aslam, A. H. Abdul Khadar, P. Francis, M. K. Mikacha, and J. Kalathingal, "Recent trends in prevention of oral cancer," *J. Int. Soc. Preventive Community Dentistry*, vol. 4, no. 6, p. 131, 2014.

[8] I. van der Waal, R. de Bree, R. Brakenhoff, and J. Coebergh, "Early diagnosis in primary oral cancer: Is it possible?" *Medicina Oral Patología Oral y Cirugia Bucal*, vol. 16, no. 3, pp. e300–e305, 2011.

[9] J. Seoane, B. Takkouche, P. Varela-Centelles, I. Tomás, and J. M. Seoane-Romero, "Impact of delay in diagnosis on survival to head and neck carcinomas: A systematic review with meta-analysis," *Clin. Otolaryngology*, vol. 37, no. 2, pp. 99–106, Apr. 2012.

[10] M. A. Weinberg and D. J. Estefan, "Assessing oral malignancies," *Amer. Family Physician*, vol. 65, no. 7, pp. 84–1379, Apr. 2002.

[11] A. Humeau-Heurtier, "Texture feature extraction methods: A survey," *IEEE Access*, vol. 7, pp. 8975–9000, 2019.

[12] T. Song, J. Feng, S. Wang, and Y. Xie, "Spatially weighted order binary pattern for color texture classification," *Expert Syst. Appl.*, vol. 147, Jun. 2020, Art. no. 113167.

[13] L. Liu, J. Chen, P. Fieguth, G. Zhao, R. Chellappa, and M. Pietikäinen, "From BoW to CNN: Two decades of texture representation for texture classification," *Int. J. Comput. Vis.*, vol. 127, no. 1, pp. 74–109, Jan. 2019.

[14] L. Liu, P. Fieguth, Y. Guo, X. Wang, and M. Pietikäinen, "Local binary features for texture classification: Taxonomy and experimental study," *Pattern Recognit.*, vol. 62, pp. 135–160, Feb. 2017.

[15] T. P. Nguyen, N.-S. Vu, and A. Manzanera, "Statistical binary patterns for rotational invariant texture classification," *Neurocomputing*, vol. 173, pp. 1565–1577, Jan. 2016.

[16] X. Qi, G. Zhao, L. Shen, Q. Li, and M. Pietikäinen, "LOAD: Local orientation adaptive descriptor for texture and material classification," *Neurocomputing*, vol. 184, pp. 28–35, Apr. 2016.

[17] S. Wang, Q. Wu, X. He, J. Yang, and Y. Wang, "Local N-ary pattern and its extension for texture classification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 9, pp. 1495–1506, Sep. 2015.

[18] J. Zhang, J. Liang, C. Zhang, and H. Zhao, "Scale invariant texture representation based on frequency decomposition and gradient orientation," *Pattern Recognit. Lett.*, vol. 51, pp. 57–62, Jan. 2015.

[19] A. R. Backes, A. S. Martinez, and O. M. Bruno, "Texture analysis using graphs generated by deterministic partially self-avoiding walks," *Pattern Recognit.*, vol. 44, no. 8, pp. 1684–1689, Aug. 2011.

[20] M. K. Ghalati, A. Nunes, H. Ferreira, P. Serranho, and R. Bernardes, "Texture analysis and its applications in biomedical imaging: A survey," *IEEE Rev. Biomed. Eng.*, vol. 15, pp. 222–246, 2022.

[21] C. Palm, "Color texture classification by integrative co-occurrence matrices," *Pattern Recognit.*, vol. 37, no. 5, pp. 965–976, May 2004.

[22] A. R. Backes, D. Casanova, and O. M. Bruno, "Color texture analysis based on fractal descriptors," *Pattern Recognit.*, vol. 45, no. 5, pp. 1984–1992, May 2012.

[23] A. Drimbarean and P. F. Whelan, "Experiments in colour texture analysis," *Pattern Recognit. Lett.*, vol. 22, no. 10, pp. 1161–1167, Aug. 2001.

[24] Q. Xu, J. Yang, and S. Ding, "Color texture analysis using the wavelet-based hidden Markov model," *Pattern Recognit. Lett.*, vol. 26, no. 11, pp. 1710–1719, Aug. 2005.

[25] V. Arvis, C. Debain, M. Berducat, and A. Benassi, "Generalization of the co-occurence matrix for colour images: Application to colour texture classification," *Image Anal. Stereology*, vol. 23, no. 1, pp. 63–72, May 2011.

[26] I.-U.-H. Qazi, O. Alata, J.-C. Burie, A. Moussa, and C. Fernandez-Maloigne, "Choice of a pertinent color space for color texture characterization using parametric spectral analysis," *Pattern Recognit.*, vol. 44, no. 1, pp. 16–31, Jan. 2011.
[27] T. Mäenpää and M. Pietikäinen, "Classification with color and texture: Jointly or separately?" *Pattern Recognit.*, vol. 37, no. 8, pp. 1629–1640, Aug. 2004.
[28] F. Bianconi, "Theoretical and experimental comparison of different approaches for color texture classification," *J. Electron. Imag.*, vol. 20, no. 4, Oct. 2011, Art. no. 043006.
[29] L. F. S. Dos Santos, L. A. Neves, G. B. Rozendo, M. G. Ribeiro, M. Z. do Nascimento, and T. A. A. Tosta, "Multidimensional and fuzzy sample entropy (SampEnMF) for quantifying H&E histological images of colorectal cancer," *Comput. Biol. Med.*, vol. 103, pp. 148–160, Dec. 2018.
[30] R. Furlong, M. Hilal, V. O'Brien, and A. Humeau-Heurtier, "Parameter analysis of multiscale two-dimensional fuzzy and dispersion entropy measures using machine learning classification," *Entropy*, vol. 23, no. 10, p. 1303, Oct. 2021.
[31] B. S. Manjunath, J.-R. Ohm, V. V. Vasudevan, and A. Yamada, "Color and texture descriptors," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 6, pp. 703–715, Jun. 2001.
[32] M. Hilal, A. S. F. Gaudêncio, C. Berthin, P. G. Vaz, J. Cardoso, L. Martin, and A. Humeau-Heurtier, "Bidimensional colored fuzzy entropy measure: A cutaneous microcirculation study," in *Proc. 5th Int. Conf. Adv. Biomed. Eng. (ICABME)*, Oct. 2019, pp. 1–4.
[33] J. A. M. Sidey-Gibbons and C. J. Sidey-Gibbons, "Machine learning in medicine: A practical introduction," *BMC Med. Res. Methodology*, vol. 19, no. 1, pp. 1–18, Mar. 2019.
[34] A. Rajkomar, J. Dean, and I. Kohane, "Machine learning in medicine," *New England J. Med.*, vol. 380, no. 14, pp. 1347–1358, 2019.
[35] D. W. Kim, S. Lee, S. Kwon, W. Nam, I.-H. Cha, and H. J. Kim, "Deep learning-based survival prediction of oral cancer patients," *Sci. Rep.*, vol. 9, no. 1, p. 6994, May 2019.
[36] B. Ilhan, K. Lin, P. Guneri, and P. Wilder-Smith, "Improving oral cancer outcomes with imaging and artificial intelligence," *J. Dental Res.*, vol. 99, no. 3, pp. 241–248, Mar. 2020.
[37] H. Shimizu and K. I. Nakayama, "Artificial intelligence in oncology," *Cancer Sci.*, vol. 111, no. 5, pp. 1452–1460, 2020.
[38] R. Cuocolo, M. Caruso, T. Perillo, L. Ugga, and M. Petretta, "Machine learning in oncology: A clinical appraisal," *Cancer Lett.*, vol. 481, pp. 55–62, Jul. 2020.
[39] B. Thomas, V. Kumar, and S. Saini, "Texture analysis based segmentation and classification of oral cancer lesions in color images using ANN," in *Proc. IEEE Int. Conf. Signal Process., Comput. Control (ISPCC)*, Solan, India, Sep. 2013, pp. 1–5.
[40] B. Song, S. Sunny, R. D. Uthoff, S. Patrick, A. Suresh, T. Kolur, G. Keerthi, A. Anbarani, P. Wilder-Smith, M. A. Kuriakose, P. Birur, J. J. Rodriguez, and R. Liang, "Automatic classification of dual-modalilty, smartphone-based oral dysplasia and malignancy images using deep learning," *Biomed. Opt. Exp.*, vol. 9, no. 11, p. 5318, Oct. 2018.
[41] R. A. Welikala, P. Remagnino, J. H. Lim, C. S. Chan, S. Rajendran, T. G. Kallarakkal, R. B. Zain, R. D. Jayasinghe, J. Rimal, A. R. Kerr, R. Amtha, K. Patil, W. M. Tilakaratne, J. Gibson, S. C. Cheong, and S. A. Barman, "Automated detection and classification of oral lesions using deep learning for early detection of oral cancer," *IEEE Access*, vol. 8, pp. 132677–132693, 2020.
[42] Q. Fu et al., "A deep learning algorithm for detection of oral cavity squamous cell carcinoma from photographic images: A retrospective study," *EClinicalMedicine*, vol. 27, Oct. 2020, Art. no. 100558.
[43] B. Song et al., "Mobile-based oral cancer classification for point-of-care screening," *J. Biomed. Opt.*, vol. 26, no. 6, Jun. 2021, Art. no. 065003.
[44] H. Lin, H. Chen, L. Weng, J. Shao, and J. Lin, "Automatic detection of oral cancer in smartphone-based images using deep learning for early diagnosis," *J. Biomed. Opt.*, vol. 26, no. 8, Aug. 2021, Art. no. 086007.
[45] B. R. Nanditha and A. Geetha, "Oral cancer detection using machine learning and deep learning techniques," *Int. J. Current Res. Rev.*, vol. 14, no. 1, pp. 64–70, 2022.
[46] P. Shah, N. Roy, and P. Dhandhukia, "Algorithm mediated early detection of oral cancer from image analysis," *Oral Surgery, Oral Med., Oral Pathol. Oral Radiol.*, vol. 133, no. 1, pp. 70–79, Jan. 2022.
[47] K. C. Figueroa, B. Song, S. Sunny, S. Li, K. Gurushanth, P. Mendonca, and N. Mukhia, "Interpretable deep learning approach for oral cancer classification using guided attention inference network," *J. Biomed. Opt.*, vol. 27, no. 1, Jan. 2022, Art. no. 015001.
[48] B. Song, S. P. Sunny, P. Mendonca, N. Mukhia, S. Li, S. Patrick, T. Imchen, S. T. Leivon, T. Kolur, V. Shetty, V. Bhushan, D. Vaibhavi, S. Rajeev, S. Pednekar, A. D. Banik, R. M. Ramesh, V. Pillai, P. W. Smith, A. Sigamani, and M. A. Kuriakose, "Field validation of deep learning based Point-of-Care device for early detection of oral malignant and potentially malignant disorders," *Sci. Rep.*, vol. 12, no. 1, pp. 1–11, Aug. 2022.
[49] K. Warin, W. Limprasert, S. Suebnukarn, S. Jinaporntham, P. Jantana, and S. Vicharueang, "AI-based analysis of oral lesions using novel deep convolutional neural networks for early detection of oral cancer," *PLoS ONE*, vol. 17, no. 8, Aug. 2022, Art. no. e0273508.
[50] M. Al Duhayyim, A. A. Malibari, S. Dhahbi, M. K. Nour, I. Al-Turaiki, M. Obayya, and A. Mohamed, "Sailfish optimization with deep learning based oral cancer classification model," *Comput. Syst. Sci. Eng.*, vol. 45, no. 1, pp. 753–767, 2023.
[51] T. Flügge, R. Gaudin, A. Sabatakakis, D. Tröltzsch, M. Heiland, N. van Nistelrooij, and S. Vinayahalingam, "Detection of oral squamous cell carcinoma in clinical photographs using a vision transformer," *Sci. Rep.*, vol. 13, no. 1, pp. 1–7, Feb. 2023.
[52] S. H. Begum and P. Vidyullatha, "Automatic detection and classification of oral cancer from photographic images using attention maps and deep learning," *Int. J. Intell. Syst. Appl. Eng.*, vol. 11, no. 11s, pp. 221–229, Jul. 2023.
[53] P. Pahadiya, R. Vijay, K. K. Gupta, S. Saxena, and T. Shahapurkar, "Digital image based segmentation and classification of tongue cancer using CNN," *Wireless Pers. Commun.*, vol. 132, no. 1, pp. 609–627, Aug. 2023.
[54] V. Talwar, P. Singh, N. Mukhia, A. Shetty, P. Birur, K. M. Desai, C. Sunkavalli, K. S. Varma, R. Sethuraman, C. V. Jawahar, and P. K. Vinod, "AI-assisted screening of oral potentially malignant disorders using smartphone-based photographic images," *Cancers*, vol. 15, no. 16, p. 4120, Aug. 2023.
[55] H. S. Chandrashekar, A. G. Kiran, S. Murali, M. S. Dinesh, and B. R. Nanditha, "Oral images dataset," Mendeley Data Elsevier, V2, Tech. Rep., 2021, doi: 10.17632/mhjyrn35p4.2.
[56] I. Noor, H. Mokhtar, K. R. Zaman, and M. Pramod, "Understanding color models: A review," *ARPN J. Sci. Technol.*, vol. 2, no. 3, pp. 265–275, Apr. 2012.
[57] B. Sharma and R. Nayyer, "Use and analysis of color models in image processing," *Int. J. Adv. Scientific Res.*, vol. 1, no. 8, p. 329, Oct. 2015.
[58] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "LightGBM: A highly efficient gradient boosting decision tree," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 3146–3154.
[59] M. V. Anand, B. KiranBala, S. R. Srividhya, K. C., M. Younus, and M. H. Rahman, "Gaussian Naïve Bayes algorithm: A reliable technique involved in the assortment of the segregation in cancer," *Mobile Inf. Syst.*, vol. 2022, pp. 1–7, Jun. 2022.
[60] X. Ding, J. Liu, F. Yang, and J. Cao, "Random radial basis function kernel-based support vector machine," *J. Franklin Inst.*, vol. 358, no. 18, pp. 10121–10140, Dec. 2021.
[61] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5–32, Oct. 2001.
[62] P. Marius-Constantin, B. Valentina, P.-P. Liliana, and M. Nikos, "Multilayer perceptron and neural networks," *WSEAS Trans. Circuits Syst.*, vol. 8, pp. 579–588, Jul. 2009.

BIBEK GOSWAMI received the B.Tech. degree in electronics and communication engineering from the Indian Institute of Information Technology (IIIT) Guwahati, India, in 2020. He is currently pursuing the Ph.D. degree with the Department of Electronics and Electrical Engineering, Indian Institute of Technology (IIT) Guwahati, India. He has a special interest in the early diagnostics of head and neck cancer exploring different imaging modalities with the help of computer vision and artificial intelligence. His research interests include biomedical image/video analysis, machine and deep learning, computer vision, and application of artificial intelligence.

M. K. BHUYAN (Senior Member, IEEE) received the Ph.D. degree in electronics and communication engineering from the Indian Institute of Technology (IIT) Guwahati, India, in 2006. He was with the School of Information Technology and Electrical Engineering, The University of Queensland, St. Lucia, QLD, Australia, where he was involved in postdoctoral research. Subsequently, he was a Researcher with the SAFE Sensor Research Group, NICTA, Brisbane, QLD. He was an Assistant Professor with the Department of Electrical Engineering, IIT Roorkee, India, and the Jorhat Engineering College, Assam, India. He also worked in Indian engineering services. He is currently a Professor with the Department of Electronics and Electrical Engineering, IIT Guwahati, where he is also the Associate Dean of Infrastructure, Planning and Management. In 2014, he was a Visiting Professor with Indiana University and Purdue University, IN, USA. He is also a Visiting Professor with the Department of Computer Science, Chubu University, Japan. He has almost 25 years of industry, teaching, and research experience. His current research interests include image/video processing, computer vision, machine and deep learning, human–computer interactions (HCI), virtual reality and augmented reality, and biomedical signal processing. He was a recipient of the National Award for Best Applied Research/Technological Innovation, which was presented by the Honorable President of India, in 2012; the prestigious Fullbright-Nehru Academic and Professional Excellence Fellowship; and the BOYSCAST Fellowship.

SULTAN ALFARHOOD received the Ph.D. degree in computer science from the University of Arkansas. He is an Assistant Professor with the Department of Computer Science, King Saud University (KSU). Since joining KSU in 2007, he has made several contributions to the field of computer science through his research and publications. Sultan's research spans a variety of domains including machine learning, recommender systems, linked open data, text mining and ML-based IoT systems. His work includes proposing innovative approaches and techniques to enhance the accuracy and effectiveness of these systems. His recent publications have focused on using deep learning and machine learning techniques to address challenges in these domains. Sultan's research continues to make significant contributions to the field of computer science and machine learning. His work has been published in several high-impact journals and conferences.

MEJDL SAFRAN received the bachelor's degree in computer science from King Saud University, in 2007, and the master's and Ph.D. degrees in computer science from Southern Illinois University, Carbondale, in 2013 and 2018, respectively. His doctoral dissertation was on developing efficient learning-based recommendation algorithms for top-N tasks and top-N workers in large-scale crowdsourcing systems. He is a passionate researcher and an educator in the field of artificial intelligence, with a focus on deep learning and its applications in various domains. He is currently an Assistant Professor of computer science with King Saud University, where he has been a Faculty Member since 2008. He has published more than 20 papers in peer-reviewed journals and conference proceedings, such as *ACM Transactions on Information Systems*, *Applied Computing and Informatics*, *Mathematics*, *Sustainability*, *International Journal of Digital Earth*, IEEE Access, *Biomedicines*, *Sensors*, IEEE International Conference on Cluster, IEEE International Conference on Computer and Information Science, International Conference on Database Systems for Advanced Applications, and International Conference on Computational Science and Computational Intelligence. He has been leading grant projects in the fields of AI in medical imaging and AI in smart farming. His current research interests include developing novel deep learning methods for image processing, pattern recognition, natural language processing, predictive analytics, modeling and analyzing user behavior, and online platforms. He has been an AI consultant for several national and international agencies, since 2018.