Early Diagnosis of Oral Cancer Using Image Processing and Artificial Intelligence

Article *in* Fusion Practice and Applications · January 2024

DOI: 10.54216/FPA.140122

7 authors, including:

Eman Shawky Mira
Mansoura University
3 PUBLICATIONS **60** CITATIONS
SEE PROFILE

El-Sayed M. El-kenawy
Delta Higher Institute of Engineering and Technology (DHIET)
234 PUBLICATIONS **9,506** CITATIONS
SEE PROFILE

Mohamed Saber
Delta University for Science and Technology
47 PUBLICATIONS **1,531** CITATIONS
SEE PROFILE

Early Diagnosis of Oral Cancer Using Image Processing and Artificial Intelligence

Eman Shawky Mira[*1], Ahmed M. Saaduddin Sapri [2], Rowaa F. Aljehanı[3], Bayan S. Jambı[4], Taseer Bashir[5], El-Sayed M. El-Kenawy[6], Mohamed Saber[7]

[1] Diagnosis & Oral Radiology Department, Faculty of Dentistry , Mansoura University, Mansoura 35516, Egypt
[2] Oral and Maxillofacial Surgery Department, Faculty of Dentistry, Mansoura University, Mansoura 35516, Egypt
[3,4] Dentistry Program, Batterjee Medical College, Jeddah, Saudi Arabia
[5] Department of Oral medicine and radiology, Batterjee Dental College, Jeddah, Saudi Arabia
[6] Department of Communications and Electronics, Delta Higher Institute of Engineering and Technology, Mansoura 35111, Egypt
[7] Electronics and Communications Engineering Department, Faculty of Engineering, Delta University for Science and Technology, Gamasa City 11152, Egypt
Emails: emanshawky@mans.edu.eg; ahmedsaaduddin@mans.edu.eg; Ro2a_aljehani@hotmail.com; bayanja11@gmail.com; taseermds2013@gmail.com; skenawy@ieee.org; Mohamed.Saber@deltauniv.edu.eg

Abstract

There has yet to be a comprehensive investigation on enhancing the diagnostic accuracy of oral disease using handheld smartphone photographic photos. To overcome the difficulties associated with the automatic detection of oral illnesses, we describe an approach based on smartphone image diagnosis powered by a deep learning algorithm. The centered rule method of image capture was offered as a quick and easy way to get high-quality pictures of the mouth. A resampling method was proposed to mitigate the influence of image variability from handheld smartphone cameras, and a medium-sized oral dataset with five types of disorders was developed based on this approach. Finally, we introduce a recently developed deep-learning network to assess oral cancer diagnosis. On 455 test images, the proposed technique showed an impressive 83.0% sensitivity, 96.6% specificity, 84.3% accuracy, and 83.6% F1. The proposed "center positioning" method was about 8% higher than a simulated "random positioning" method; the resampling process had an additional 6% performance improvement. The performance of a deep learning algorithm for detecting oral cancer can be enhanced by capturing oral photos centered on the lesion. Primary oral cancer diagnosis using smartphone-based images with deep learning offers promising potential.

Keywords: Deep learning; Smartphone-based imaging; Image collection; Oral cancer diagnosis; Oral potentially malignant disorders.

1. Introduction and Related Work

Oral cancer has a high incidence and fatality rate, making it a leading cancer killer. There were an expected 377,000 new instances of lip and oral cavity cancer in 2020, with almost 177,000 fatalities globally, as reported by the International Agency for Research on Cancer [1]. Death rates from oral cancer have remained high over the previous few decades despite progress in oncology therapy. Most people diagnosed with oral cancer will not receive adequate care on time. Notably, they were leading to low survival rates in the countryside. Patients diagnosed with oral cancer have a 5-year survival rate of roughly 50%, albeit this varies by race and location [2].

Doi: https://doi.org/10.54216/FPA.140122
Received: July 24, 2023 Revised: October 27, 2023 Accepted: December 19, 2023

Reports of prevalence range from as high as 65% in wealthy countries to as low as 15% in some rural areas, with variation seen even within the mouth. According to the Indian Cancer Statistics Report 2020, almost 66.6% of head and neck cancer cases were discovered at an advanced stage [3]. Patients' chances of survival and prognosis are drastically reduced if their oral cancer is detected at a later stage [4,5]; hence, bettering early diagnosis can significantly improve survival rates [6,7].

Since OSCC is the most common kind of oral cancer [8,9] the two acronyms are sometimes used interchangeably; histopathologically speaking, OSCC begins as an epithelial dysplasia and progresses from precancerous lesions known as oral potentially malignant disorders (OPMDs) [10,11]. Even the most prevalent OPMDs, including oral lichen planus, leukoplakia, and erythroplakia [11], do not necessarily lead to the development of malignancies [10,12]. Because of their heterogeneity, complexity, and shared symptoms, OPMDs are notoriously challenging to diagnose and categorize. According to studies, nonhomogeneous lesions have a higher chance of malignant transformation than homogeneous lesions [10,13,14]. This means that a change in an OPMD's appearance is more likely to be seen as a negative development. Therefore, differentiating the potentially malignant features of OPMD is of greater relevance than arriving at a definite diagnosis. For these reasons, it is crucial to use a reliable clinical prediction approach for early diagnosis of OSCC, especially for staging malignancy at the OPMD stage.

The most common method currently used to treat oral cancer and its precursor condition is the conventional oral examination (COE), which includes a visual and tactile assessment (and, if necessary, a tissue biopsy). However, general dentists may be unable to distinguish between oral cancer and benign conditions such as aphthous ulcers because of the clinical heterogeneity and subtlety of oral cancer symptoms. Second, biopsy is not always the best screening technique due to its invasive nature and sampling bias, which can result in underdiagnosis or misdiagnosis [11]. In addition, the oral cancer burden disproportionately affects low-resource communities even though specialists can recognize the majority of characteristics that differentiate benign and cancerous lesions. As a result, developing a low-cost screening technique to supplement existing practices is gaining traction [12].

In recent years, deep learning algorithms have shown superior to feature-based approaches in medical picture analysis. In numerous disease recognition scenarios, deep learning algorithms have been shown to outperform human specialists, as demonstrated by several studies [12,13,14]. Automatic analysis of pathology, confocal laser endomicroscopy (CLE), and fluorescence pictures all showed promise for use with deep learning in the detection of oral cancer. To classify oral cancer stages in histo-logical pictures, Kumar et al. [15] presented a two-step process, including a segmentation network and a random forest tree classifier. The deep convolutional neural network performed better than the feature-based classification approaches when Aubreville et al. [14] evaluated it for OSCC diagnosis using CLE images. To detect oral disease, Song et al. [15] created mobile-connected devices to gather fluorescence oral images. These techniques, however, rely upon high-priced equipment or a specialized screening platform that only some have access to. In other words, getting a diagnosis for an illness still necessitates a trip to a doctor's office. As image and sensing technology in cameras continue to advance rapidly, more and more smartphone manufacturers are including high-resolution, low-noise, and speedy camera modules.

White light inspection techniques [16] that may be performed with smartphones are practical approaches for collecting oral images. However, the patient sample is small, with only 54 cases, in Camalan et al. [16] CNN-based network for classifying white light images as usual or worrisome. To construct a robust system, Welikala et al. [17] gathered more clinically labeled data for network training and evaluation, increasing the accuracy of their multiclass classification task to 52.13 % while maintaining a specificity of 49.11 %. However, we discovered that the clinical images captured by handheld smartphone cameras may exhibit considerable variability, causing the identification algorithm to perform poorly in diagnostic situations. From a computer vision standpoint, the fields of view and focus distance, for instance, determine the shape and size of imaging lesions. Sadly, not much study has been done on this issue. To overcome these obstacles, we plan to investigate high-performing CNN models for oral illness identification and methods for obtaining white-light images on smartphones. Here is a quick rundown of our most significant contributions.

First, we suggest a straightforward technique for taking photos that reliably preserves the lesion's position and focal distance across photographs. The method enables direct focus on discriminative parts for illness recognition without resorting to area proposal methods [18,19] or depending on bounding box annotation [19] by oral specialists. To address the image unpredictability brought about by the handheld nature of smartphone cameras, we suggest a resampling technique that can also correct the class imbalance issue. Third, we classify oral diseases using one of the most recently suggested convolutional neural networks,

Doi: https://doi.org/10.54216/FPA.140122
Received: July 24, 2023 Revised: October 27, 2023 Accepted: December 19, 2023

outperforming standard classification methods. Our data analysis demonstrates that using a smartphone's typical imaging pattern is valuable for making a prompt diagnosis.

2. Materials and Methods

The research was conducted after the Medical Research Ethics Committee of the Faculty of Dentistry at Mansoura University permitted this work with IRB number (M0106023OM). All patients were collected from the outpatient dental clinics of the Faculty of Dentistry-Mansoura University in the period that extended between (Say, October 2022 and October 2023). A sample size of 232 patients was recruited for this study who had been clinically diagnosed with aphthous ulcers, oral potentially malignant disorders (OPMDs), or oral Cancer. Only healthy patients (ASA-I or ASA-II) older than 18 years old and accepted to sign the informed consent form were included in this study. Children, pregnant women, and adults who are medically or immunocompromised patients and those who refused to sign the consent form were excluded from this study.

A. Image Capturing Method

Imaging was done in outpatient hospital settings using the mobile phone's built-in camera application. From the mobile phone built-in camera's application, the settings panel was selected to activate the photo-grid option that was adjusted into a thin grid divided into nine squares of equal size, which will appear over the captured image. The camera grid was used to ensure the positioning of the lesion so that it was in the center of the picture frame to confirm the proportionality between the different parts of the lesion (Figure 1) as well as to enable a more accurate extraction for the discriminative features and to remove the extra backdrops as the CNN's performance in image recognition is enhanced when the main object is placed in this way. Accordingly, the area of the covered lesion (on the screen) is smaller than the center square of the phone's camera grid. Then, the diagnostic images were transferred into the computer running the diagnostic software via a wired or wireless connection.

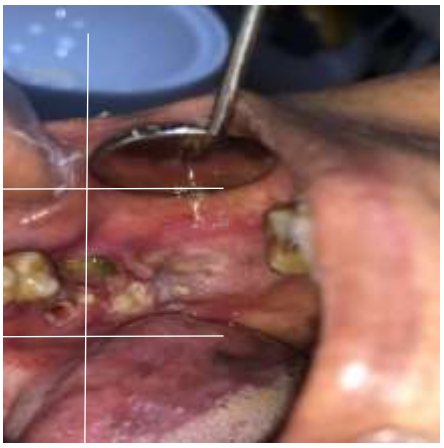


Figure 1: The image has an aspect ratio of 4:3, it displays a lesion in the affected area.

B. Creation of the Oral Dataset

The demographic and clinical data of 232 adult healthy subjects classified as either (ASA-I) or (ASA-II) and clinically diagnosed with aphthous ulcers, OPMD, or oral Cancer were collected and introduced into the computer software. The rule of thumb is to take one picture for each lesion; if a single patient has numerous forms of oral lesions, then multiple pictures will be taken. Otherwise, we merely take one photo of the patient's mouth (Table 1 shows the breakdown of patient counts, image counts, and image counts per patient).

Multiple pictures of a healthy person's mouth can be taken from various anatomical locations without worrying about capturing anything abnormal. These locations include the tongue (dorsal or ventral), palate, labial mucosa (upper or lower), buccal mucosa (left or right), and floor of the mouth. Multifocal lesions are also more common in people with low- or high-risk OPMD than those with aphthous ulcers or malignancy. Accordingly, to the initial clinical impression, the mucosal photographic presentations were assigned under five categories namely normal mucosa, aphthous ulcer, low-risk OPMD, high-risk OPMD, and oral Cancer (Figure. 2)

- Normal mucosa: it looks smooth, pink, and shiny, without any white or red spots (Figure 3. a).

Doi: https://doi.org/10.54216/FPA.140122
Received: July 24, 2023 Revised: October 27, 2023 Accepted: December 19, 2023

- Aphthous ulcer lesion: It looks like a shallow, round, or oval ulcer with a yellow pseudo membrane and an erythematous halo around it (Figure 3. b).
- Low-risk (homogeneous) OPMD: presentation is characterized by a uniform flat white patchy lesion with a thin, smooth, or fissured surface (Figure 3. c).
- High-risk (heterogeneous) OPMD: presentation is characterized by notably distinctive nonuniform surface texture caused by red components secondary to atrophy or erosion [19] (Figure 3.d).
- Oral Cancer: presents with either a nodular form lesion or a deep ulcerative lesion with a rough surface (Figure 3. e). Both forms are embodied by hazy borders [19].

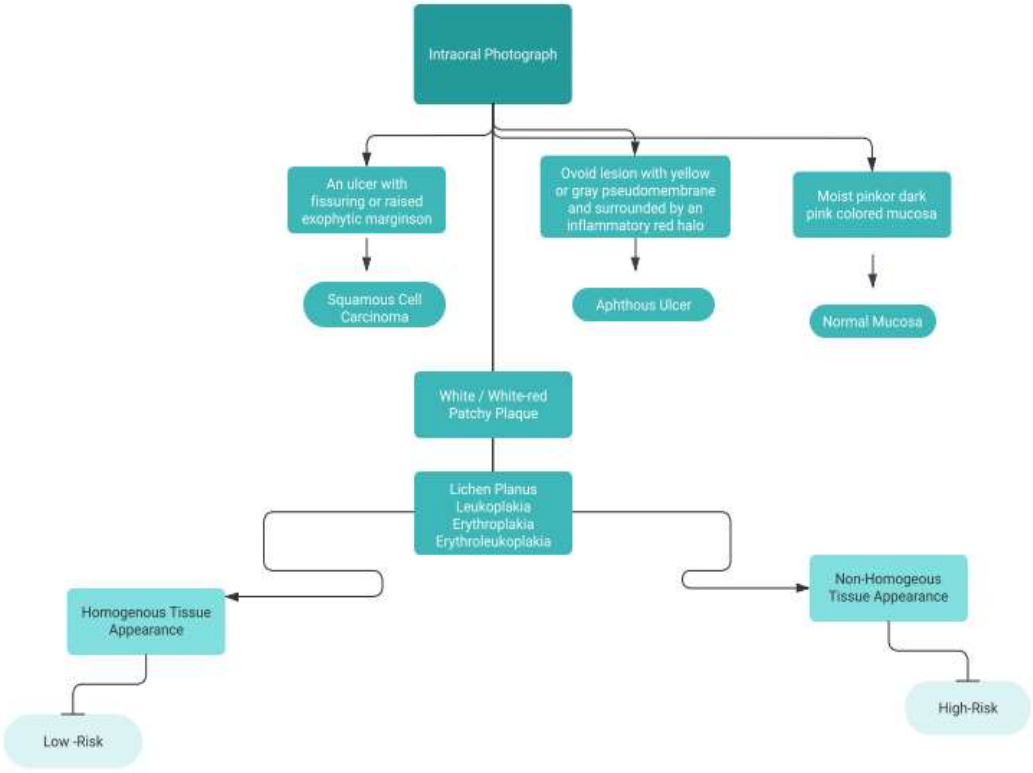
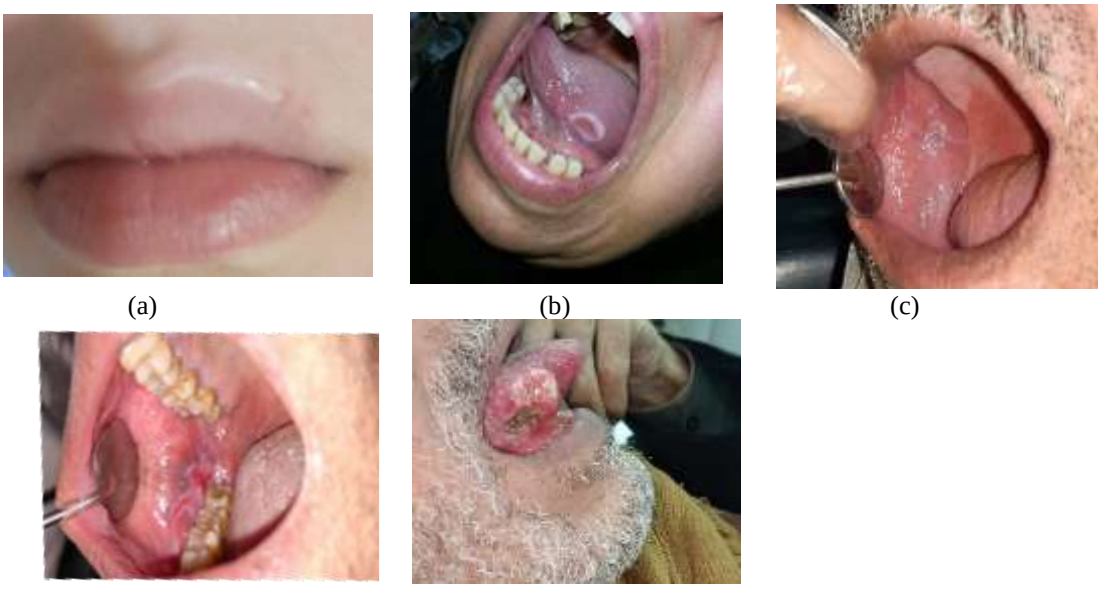


Figure 2: Summary of the five diagnostic groups established on a physician's first impression. The depicted lesions were categorized into five groups based on the following assumption regarding their clinical findings

The same surgeon collected biopsies, "author number XX," who has more than ten years' experience in oral and maxillofacial surgery to ensure a proper collection of samples. Histopathologic examination had been performed on all patients with an initial oral Cancer diagnosis. Histopathology-documented examples serve as the benchmark for comparison. Aphthous ulcers and low-risk OPMD are two more examples of minor symptomatic lesions that might not need a histopathological report. Three oral medicine specialists dedicated to the early detection and treatment of mouth cancer and other potentially malignant illnesses will personally annotate photographs to guarantee a high level of accurate annotation of collected data. Initially, two senior specialists annotate each photograph based on the presentation of symptoms and clinical records or histopathologic reports. Cases that are initially assessed as having numerous oral disease conditions or as being disputed are removed from the dataset, and a second expert in mucosal diseases repeats this process (Figure. 4). 688 and 760 photos of oral lesions and healthy mucosa successively made up the final oral data set.

Doi: https://doi.org/10.54216/FPA.140122
Received: July 24, 2023 Revised: October 27, 2023 Accepted: December 19, 2023



(a) (b) (c)

(d) (e)

Figure 3: Clinical characteristics of representative photos from each of the five classes. (a) Mucosal tissue from a healthy person's upper lip and lower labial mucosa. (b) Ventrolateral tongue surface aphthous ulceration. (c) The buccal mucosa on the right side of the mouth has a uniform white patch. (d) A large white lesion with a red component affects the entire right buccal mucosa. (e) A cancerous growth along the tongue's side

Table 1: Statistics of patient in our oral disease dataset

Statistics	Normal	Aphthous Ulcer	Low risk OPMD	High risk OPMD	Oral cancer
Number of Images	760	251	231	141	65
Number of patients	232	218	165	107	65
Image / patient (mean)	3.28	1.15	1.4	1.32	1

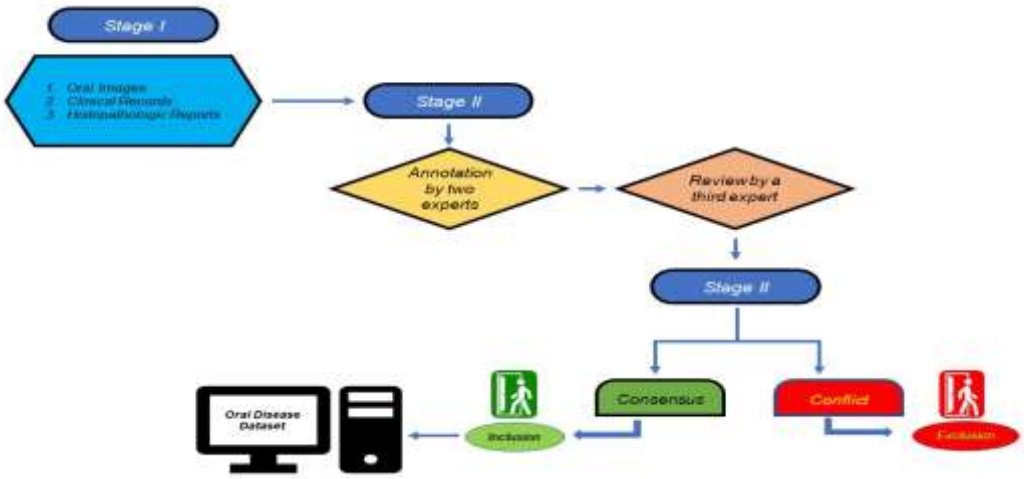


Figure 4: Flow of data annotations from multiple experts

Doi: https://doi.org/10.54216/FPA.140122
Received: July 24, 2023 Revised: October 27, 2023 Accepted: December 19, 2023

The pictures were taken from a diverse group of people, aged 18 to 82, representing a wide range of intraoral anatomical locations (Table. 2 contemplated the photographic clinical presentations against the various anatomical regions).

The dataset is split into two groups, one based on patients and the other on the percentage of patients in each group. Since we are particularly interested in developing methods for the early detection of oral cancer, since high-risk OPMDs and cancers are more severe diseases than the others in the dataset, we intentionally increased the split ratio of these two groups in the test set. The 455 patients in the test set include:

- 30 cases of Cancer (46%).
- 52 cases of high-risk OPMD (37%).
- 69 cases of low-risk OPMD (30%).
- 228 cases of normal (30%).

Table 2: Overview of image numbers for different anatomic sites in our dataset

Site	Normal	Aphthous Ulcer	Low risk OPMD	High risk OPMD	Oral cancer
Lip & labial mucosa	75	114	4	7	7
Buccal mucosa	314	35	147	93	13
Gingiva & alveolus	36	17	22	9	11
Tongue	136	56	41	28	23
Palate	119	13	16	3	11
Floor of the mouth	80	16	1	1	0
Total	760	251	231	141	65

C. Diagnostic Aids to Clinical Cases

The clinical diagnosis of the aphthous ulcer cases was supported by follow-up of each patient for less than one month. All subjects showed clinical improvement in both the symptoms and the lesion size without even needing medications. The clinical diagnosis of OPMD cases (figure 5,6,7,8,9,10) was supported by biopsies examined under electron microscopic study [20]. The biopsy site was selected from patients with leukoplakia and lichen planus (Table. 3).

Every possible effort had been carried out to protect the collected specimens and to ensure a proper histologic examination, such as collection under nerve block local anesthesia to avoid the distortion artifacts in the sample that can be caused by direct injection of the anesthetic solution into the biopsy site, delicate surgical maneuverers for tissue manipulation as well as collecting a representative tissue sample with at least 5 mm in diameter and 3 mm in depth to overcome the possible shrinkage caused by the fixation solution (1% glutaraldehyde and 4% formaldehyde, pH = 7.4) [21].

Histopathological assessment for hematoxylin and eosin (H&E)-stained sections examined under the transmitted light-microscope as well as examination under the transmission Electron Microscopic. Pathologic evaluation of the presence and degree of epithelial dysplasia (mild, moderate, severe, or carcinoma in situ) was used to assess the malignant risk of oral premalignant lesions. The World Health Organization (WHO) has established dysplasia criteria, including the epithelium's architectural and cytological changes. Architectural characteristics include:

- Irregular epithelial stratification.
- Loss of polarity of basal cells.
- Drop-shaped rete ridge.
- Increased number of mitotic figures.
- Abnormally superficial mitoses.
- Keratin pearls within rete pegs.

Doi: https://doi.org/10.54216/FPA.140122
Received: July 24, 2023 Revised: October 27, 2023 Accepted: December 19, 2023

Cellular characteristics include anisonucleosis (nuclear pleomorphism), anisocytosis (cellular pleomorphism), increased nuclear-cytoplasmic ratio, dyskeratosis, and atypical mitotic figures [21].

Grading dysplasia depends on the extent of dysplastic changes in the epithelial layers. In the case of mild dysplasia, cytological and architectural differences are confined to the lower third of the thickness of the epithelium. In moderate dysplasia, changes are seen up to two-thirds of the thickness of the epithelium. In severe dysplasia, these dysplastic changes occupy more than two-thirds of the thickness but less than the entire thickness of the epithelium. The dysplastic carcinoma cells in situ (CIS) occupy the whole thickness of the epithelium (bottom to top changes), but the basement membrane is still intact. Invasive squamous cell carcinoma (SCC) involves dysplastic cells invading the underlying connective tissue stroma through the basement membrane [21].

Transmission electron microscopic features were considered to be characteristic of dysplastic changes in the epithelium where malignant transformation is likely to occur, as shown in Table (4) [20,21]. These features are discontinuous basal lamina., ruptured hemidesmosomes, presence of pathologic cytoplasmic processes, increased ton fibrils, widened and disrupted intercellular junctions, presence of intracellular vacuolization, nuclear alterations, nucleolar alterations, degenerated mitochondria, increased ribosomes.

Table.3: Clinical type of oral leukoplakia

Type of Leukoplakia	N = 372	%
Homogenous leukoplakia	260	70.0%
Verrucous leukoplakia	75	20.0%
Erythroleukoplakia (speckled)	37	10.0%

Figure.5: Clinical photograph of homogenous leukoplakia at the floor of the mouth

Figure.6: Clinical photograph of erythroleukoplakia (speckled) at the commissure

Doi: https://doi.org/10.54216/FPA.140122
Received: July 24, 2023 Revised: October 27, 2023 Accepted: December 19, 2023

Figure.7: Clinical photograph of erosive oral lichen planus of the buccal mucosa

Table 4: Degree of dysplasia in studied cases

Degree of Dysplasia	Group I (oral leukoplakia)		Group II (Erosive oral lichen planus)		Monte Carlo test	Total n = 20(%)	
	N	%	N	%		N	%
Mild dysplasia	7	70%	6	60%	MC p = 0.87	13	65%
Moderate dysplasia	2	20%	3	30%		5	25%
Severe dysplasia	1	10%	1	10%		2	10%
Carcinoma in situ	0	0%	0	0%		0	0%
Invasive squamous cell carcinoma	0	0%	0	0%		0	0%

(a) (b)

Figure.8: clinical photos showing Oral Squamous cell carcinoma cases (a,b) of different parts of the mouth

3. Data Processing

Our CNN-based algorithm takes a single oral picture as input and generates a probability distribution for each disease. The most reliable group is chosen for the final projection. We created a unique diagnostic software package in Python, utilizing the publicly available PyQT and PyTorch modules, to facilitate research for the end user. As can be seen in Fig. 14, there are three stages to our CNN-based system.

Doi: https://doi.org/10.54216/FPA.140122
Received: July 24, 2023 Revised: October 27, 2023 Accepted: December 19, 2023

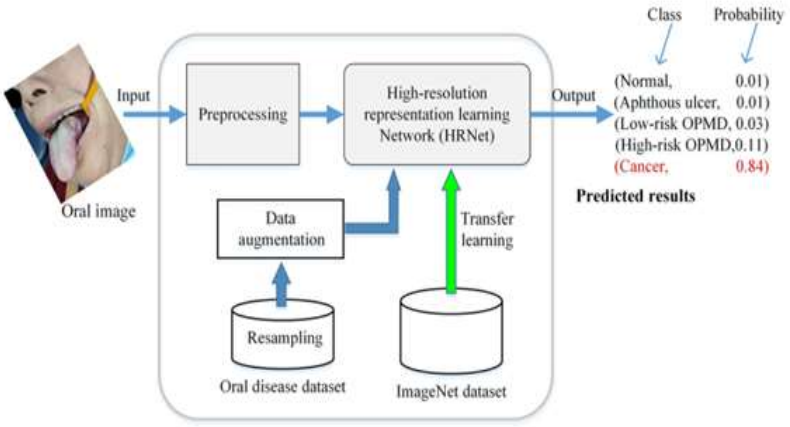


Figure 9: Concise explanation of the suggested oral illness image classification approach.

The initial step is to take a picture of the lesion, following the procedure outlined in the previous section. With some assistance, the patient can even perform this with a smartphone camera. The photographs could have clutter; therefore, it's essential to filter out the unimportant details before moving on after that comes preprocessing. To get the ROI picture, we crop just the central part out. Both human specialists and deep learning algorithms require looking at surrounding picture regions for more information when identifying an ROI around a lesion.

We did this by making the ROI bigger than the central grid area. Data normalization is also used in preprocessing to adjust the image's pixel brightness and contrast. To be more precise, we first transform each pixel in the input image from the range [0, 255] to the range [0, 1.0]. Then, one takes the input and subtracts the channel mean before dividing by the channel standard deviation. For the next step, we lower the processing cost by resizing the ROI photos to 512 512 pixels. CNN-based picture categorization is the final process. The images are analyzed using the recently introduced high-resolution representation learning network (HRNet) pre-trained on the ImageNet [22,23]. Then, all the weights are fine-tuned on our oral dataset (transfer learning). The proposed resampling method is used to increase the size of the training set during the training phase. The following part will go into further detail.

A. Resampling

We suggest the resampling technique to reduce the impact of inherent variation in the photos captured using our approach. Only the training data is subjected to the resampling procedure. The oral images may be acquired at various angles although the above strategy roughly centers each lesion. To fix this, we rotate the photos. As shown in Fig15(a), each oral image is rotated in two different directions, once by a step size of deg and once by a step size of deg.

Doi: https://doi.org/10.54216/FPA.140122
Received: July 24, 2023 Revised: October 27, 2023 Accepted: December 19, 2023

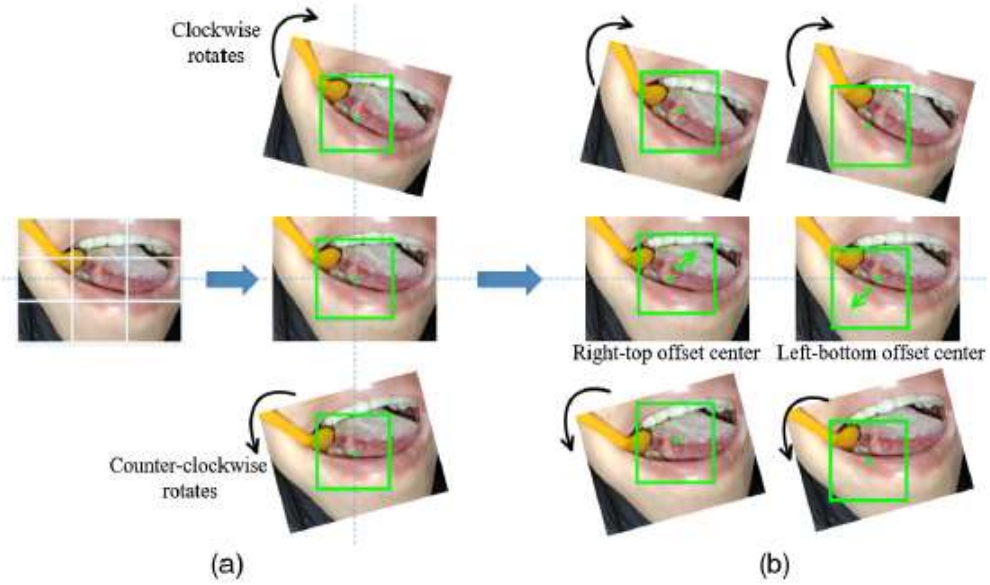


Figure 10: The captured image is expanded by an optional offset center of rotation. The green squares mark the cropped areas that will be used to create patches for training. Rotate the image around(a) its center and (b) its offset center

We then reduce the scale of the photos so that the lesion and enough distinguishing features are visible. In the experiments, we determined that. 1596, 525, 486, 267, and 105 samples are taken from each of the five classes, with each patch measuring H by W and centered on the images' respective centers. The captured lesion center may also be off in practice, so we have to be careful while using it.

As illustrated in Fig. 15(b), we can approximate the nucleus by iteratively multiplying each lesion centroid to produce points. Then, we employ the strategy mentioned earlier once more to extract the nearby regions locally, and therefore, we need to boost the amount of images we're sampling even more. Since we have collected nearly 5 times as many ulcer samples as cancer samples, for example, training deep networks with this data can produce a positive bias in the large rate class due to the imbalance in sample sizes in datasets. There are a variety of approaches to addressing the class imbalance, such as incorporating GANs [24,25] to generate data scarcity samples and employing cost-modifying methods [25] to even out learning.

Here, we deal with biased centralization and uneven data by adjusting the value of. It's worth noting that we skipped the offset resampling process in favor of just rotating the regular samples. We determined to be 3, 3, 5, and 15 for the ulcer, low-risk OPMD, high-risk OPMD, and cancer samples, respectively. There are 1575 cancer cases, 1335 OPMD cases with high risk, 1458 OPMD cases with low risk, and 1596 typical cases for the training set.

In contrast to the standard practice of augmenting training data with additional examples, the proposed resampling approach differs in two key respects. To start, our solution avoids the drawbacks of zero-padding or image size shifts that plague the more common rotational augment approach. Second, by adjusting the resampling rate and the offset angle, we can fine-tune the increased sample size for each group.

B. Network Architecture and Applications

Let $x_i \in R^{H\times W\times 3}$ be an input image containing a lesion to be classified. The predicted class c_i for input image x_i with corresponding output $F(x_i)$ is given as

$$c_i = argmax\ P_j\left(F(x_i)\right) \quad (1)$$

Where $P\left(F(x_i)\right)$ is a softmax function define by

Doi: https://doi.org/10.54216/FPA.140122
Received: July 24, 2023 Revised: October 27, 2023 Accepted: December 19, 2023

$$P\big(F(x_i)\big) = \frac{e^{\big(F(x_i)\big)}}{\sum_k e^{\big(F(x_i)\big)}} \quad (2)$$

In recent years, high-resolution representation learning has been proposed [26] due to its promising results in a wide range of visual applications. HRNet employs high-resolution representation learning to preserve semantic depth and spatial understanding. The authors' reported improvements in ImageNet categorization are better than those of previously published networks like ResNet50 [27], VGG16 [28], and DenseNet169 [28]. Since the HRNet-W18 represents the widths of the high-resolution subnetworks in the final three stages, it is the CNN of choice for illness diagnosis. Images of the mouth are fed into the HRNet-W18, and the network then calculates the likelihood that a disease is present. We change the HRNet-W18's representation head using the web in network strategy [28] to simplify the network's architecture and lower the computational cost. In particular, we reduce the convolution layer's channel output from 2048 to 5. Then, as shown in Fig. 16, we employ a globally average-pooling layer that directly outputs log probability. This contrasts with the conventional design, which utilizes an average-pooling layer and a wholly connected layer as the last layer.

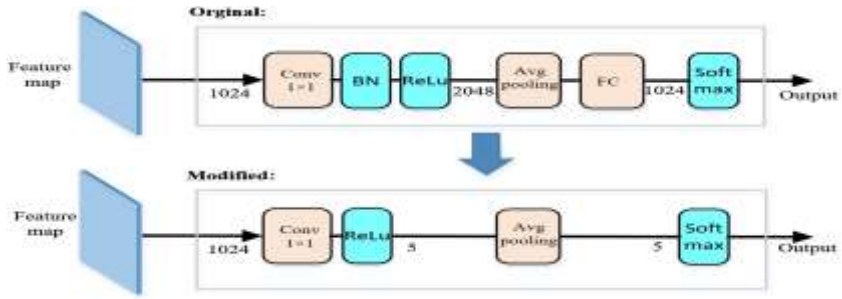


Figure 11: Illustration of modified representation head in HRNet-W18.
The number in each layer is the number of channels.
BN is the abbreviation of "batch normalization," FN is the abbreviation for "fully connected."

4. Results

One Nvidia GeForce GTX3090 GPU was installed in a desktop PC with an i5-8400 CPU and 24GB of RAM. The HRNet-W18 was developed in Python, constructed using PyTorch [28], and tested on Ubuntu18.04. HRNet was developed using a supervised learning technique to reduce the discrepancy between the probabilities it generates and the actual labels assigned to classes. For the supervised learning, a standard cross-entropy loss function was utilized. We employed pre-initialization using the ImageNet dataset to speed up the training process and boost performance.

We did not preserve any of the network's layers by freezing them, and we retrained the entire system. At epochs 15, 30, and 45, the learning rate was lowered by 30% from its initial value of 0.001. We used SGD [29] with a Nesterov momentum of 0.9 and a weight decay 0.0001. Using a mini-batch size 24, HRNet-W18 was trained on the training set using a five-fold cross-validation approach and ended after a preset number of epochs (100). For our last model, we only considered the one that outperformed the rest on the hold-out validation set.

In spite of the fact that the resampling strategy was applied to the collected cases to generate more training data, the training samples still needed to be bigger, leaving the deep learning algorithm vulnerable to overfitting and poorly generalizable to new examples. CNNs with a small training set can benefit from data augmentation, which has been shown to boost their performance [29]. For the model training, we also used standard techniques for enhancing the data, such as left-right flipping, random resizing cropping with a ratio of 0.7 to 1.0, and arbitrary brightness and contrast. Data augment approaches were not used in our testing because they did not improve our ability to anticipate. Only one pass was completed downfield. Since individual images were processed to produce a single score, this method has the potential to be more effective than ensemble methods.

A . The Effectiveness of the Image-Capturing Method and Resampling Method

Our image-gathering strategy was tested by contrasting the results of two identical networks: one trained on images with centered lesions (our acquired oral dataset) and the other trained on images without

Doi: https://doi.org/10.54216/FPA.140122
Received: July 24, 2023 Revised: October 27, 2023 Accepted: December 19, 2023

prerequired lesion centering (random positioning). These latter photos are from a different dataset, which we derived from the first set using a simulated image-capturing technique that avoided the need to center lesions. To achieve a more diverse group of lesion locations in the photos, we randomly cropped each original image to form a smaller image, then resized the cropped images to a size of 512 512 for network training and testing. The first two rows of Table 4 detail the outcomes of training and testing the HRNet-W18 on the two datasets. About 8% more improvement in F1 score was seen for the model trained on the centered rule dataset compared to the model trained on the random positioning dataset. As a result of our picture-capturing method's ability to precisely pinpoint the ROI (the lesion region) and the center cropping method's ability to direct CNN attention solely to the discriminative areas rather than the entire cavity image, which can contain distracting extraneous elements, we can achieve such impressive results. This indicates the efficiency of the suggested image-capturing technology in detecting oral diseases.

First, the random over-sampling method [30] was employed to generate additional small class samples to maintain the training class balance; this allowed us to assess the performance of the resampling approach qualitatively. However, we increased the size of the training set by employing the proposed resampling technique. The HRNet-W18 was retrained using these two datasets, and the results of our tests can be seen in the third and fourth rows of Table 3. Due to its susceptibility to over-fitting of the minority classes [30], we found that the over-sampling strategy reduces the model's performance. Still, our resampling method improves it by roughly 6% in F1 score. The outcomes prove the usefulness of the resampling approach in enhancing the network's efficiency in diagnosing oral diseases.

Table 4: Comparison of classification performance with different methods

Method	SE_{macro}	SP_{macro}	Pr_{macro}	F1
Random positioning	0.717	0.938	0.753	0.730
Center positioning	0.786	0.954	0.790	0.787
Center positioning + over sampling	0.776	0.954	0.791	0.778
Venter positioning + resampling	0.830	0.966	0.843	0.836

B. Comparison with Other Networks

We used our data to analyze and compare different neural network architectures, including VGG16, ResNet50, DenseNet169, and HRNet-W18. Table 5 displays the collected data. Comparatively, HRNet-W18 has more parameters than DenseNet169 but fewer than VGG16 or ResNet50.

Table 5: Comparison of classification performance with different methods on oral disease dataset

Method	Params	SE_{macro}	SP_{macro}	Pr_{macro}	F1
VGG16	568M	0.728	0.950	0.767	0.745
ResNet50	24M	0.770	0.954	0.788	0.771
DenseNet169	12M	0.810	0.965	0.835	0.817
HRNet-W18	17M	0.830	0.966	0.843	0.836

All network designs performed well on this dataset despite the scarcity of labeled training material. We found that HRNet- W18 has an 83% sensitivity, 96% specificity, 84.3 % accuracy, and 83.6 % F1 score. Except for specificity, where it performed similarly to the DenseNet169, it outperformed all other models. This contrast demonstrates the superiority of the updated HRNet-W18 network's high-resolution representations learned for classifying oral diseases. Applying the HRNet-W18 model to the validation set resulted in the confusion matrix in Figure 12. The percentage of instances correctly categorized and cases incorrectly categorized across the five categories were reported. Table 6 shows the results for each group.

Doi: https://doi.org/10.54216/FPA.140122
Received: July 24, 2023 Revised: October 27, 2023 Accepted: December 19, 2023

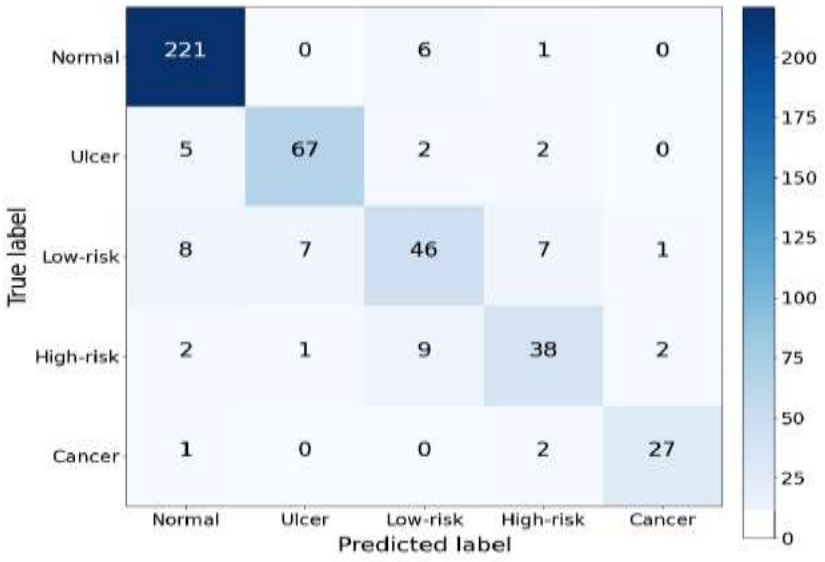


Figure 12: Confusion matrix for multiclass performance for the proposed method. The numbers of correct classifications are shown in the diagonal. The numbers of wrong categories due to overdiagnosis and underdiagnosis are found in the top-right and bottom-left triangles, respectively

Table 6: Diagnostic multiclass performance of the proposed method.
The corresponding AUCs were computed in each case using one versus another strategy.
The confidence intervals (CIs) of AUC at 95% was calculated by the DeLong method

Class	AUC (95% CI)	F1
Normal	0.949 (0.929:0.970)	0.951
Ulcer	0.930 (0.907: 0.953)	0.887
Low-risk	0.811 (0.766 :0.857)	0.697
High-risk	0.851 (0.808: 0.893)	0.745
Cancer	0.946 (0.923: 0.970)	0.9

The results showed that the F1 for the standard class was 95.0%. In contrast, the F1 for the other courses was much lower, demonstrating that classes with fewer samples are typically more challenging to categorize. Additionally, we found that the suggested method's error rates are significantly more excellent for low-risk and high-risk lesions than for the other groups. Because the symptoms of these two disorders are so similar and because even expert dentists may have trouble telling them apart, these findings are not surprising [30].

C. Visualization

We implemented class activation mapping in our HRNet-W18 for prediction visualization to enhance our understanding of the CNN diagnosis. First, we omitted HRNet-W18's final layer (average-pooling layer), and then we linearly interpolated 16 times to restore the output class activation maps to their original size. Images with "important" regions highlighted indicated that the model's predictions focused there. Most lesion sites were appropriately located, as shown in Figure 13. This suggests that identifying item sizes is crucial to making accurate diagnoses

Doi: https://doi.org/10.54216/FPA.140122
Received: July 24, 2023 Revised: October 27, 2023 Accepted: December 19, 2023

.

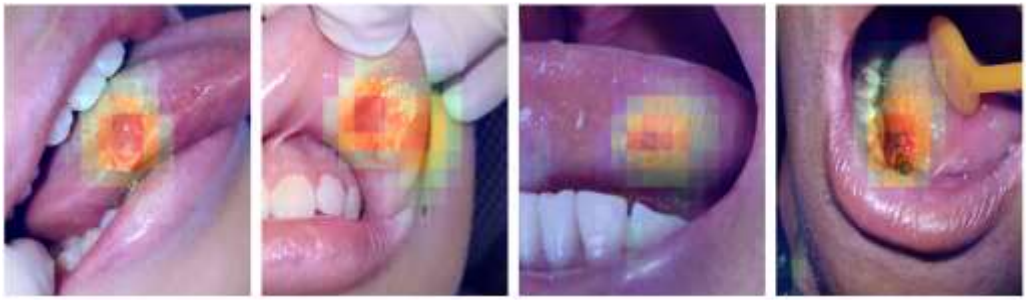


Figure 13: Examples of oral images with visualization.
The heatmaps highlight the class-discriminative regions used for oral classification

5. Discussion

The machine is an essential part of artificial intelligence (AI). AI is not a myth; it is dentistry's future as it is considered a valuable instrument that will help dentists lessen their diagnostic workload. AI has been developed dramatically across virtually every sector to carry out tasks that typically need human intelligence that is based on the cumulative acquired knowledge and the experiences of humans. Traditional dentistry is gradually shifting to digital dentistry, which is flourishing throughout the different fields of dentistry. [Verma P. Artificial intelligence: Third eye in dentistry. Journal of Oral Medicine, Oral Surgery, Oral Pathology and Oral Radiology. However, artificial intelligence cannot be a safely proofed diagnostic tool despite this study's fantastic progression and promising results. Some of the outcomes need to be meticulously investigated to avoid the catastrophes associated with missing the appropriate treatment window and even fatalities secondary to erroneous misdiagnosis from inaccurate recognition.

Our study depicted several examples of failure (Figure 14). The high-risk example of OPMD seen in Figure 14(a) is misdiagnosed as an ulcer. Misdiagnosis, missing the appropriate treatment window, or even death could result from inaccurate recognition; hence, this type of missed instance should be avoided even though the model's misclassification rate is low (1.9%). An ulcer lesion that was incorrectly identified as high-risk OPMD is shown in Figure 14 (b).

Due to the identical lesion size, dentists had no trouble misdiagnosing the considerable area of ulcer misclassified by the algorithms. Another case of OPMD with a high potential for malignancy [Figure 14(c)] Yet it is misdiagnosed as cancer. Indeed, in some circumstances, separating high-risk OPMD from cancer is inherently challenging. However, this misclassified lesion was likely malignant or precancerous; this situation is not as dire as the first. In light of the data above, biopsy remains an essential confirmatory investigative tool for making a final diagnosis. The image of reticular oral lichen planus [shown in Figure 14(d)], which was affected by reflecting light, has features that mimic the findings of the normal tissue that affected the leading algorithm to label it as normal mucosa incorrectly. Accordingly, we recommend calibrating the photographic acquisition parameters to avoid false negatives.

(a) (b)

Doi: https://doi.org/10.54216/FPA.140122
Received: July 24, 2023 Revised: October 27, 2023 Accepted: December 19, 2023

(c) (d)

Figure 14: Representative examples of failure cases: (a) high-risk OPMD misclassified as an ulcer lesion; (b) ulcer lesion misclassified as high-risk OPMD; (c) high-risk OPMD misclassified as being cancerous; and (d) low-risk OPMD misclassified as usual.

There are some caveats to our study. (1) Four separate smartphone cameras were used to acquire most of the data used in the studies. However, despite the promising results, further validation of the suggested approach on other smartphone cameras is needed. (2) Many forms of oral disease are not considered in our study because there are only five categories for evaluation. Since the network may be trained using photos of the oral cavity from many individuals or phones, the model's generalization performance can be guaranteed by teaching the network to recognize the signs of disease variations. To thoroughly define the performance features of our AI diagnosis method, additional validation on other oral diseases kinds (such as oral thrush) and multiple smartphone cameras is still necessary. Thus, our approach may prove helpful in settings where dentists have less experience or for self-prediagnosis by patients, but it is likely to fail to catch on with specialists.

We recommend the researchers carry out a study involving a more significant number of photographs with more standardizations for the acquisition and involving more intraoral lesions. We anticipate a considerable improvement in performance with increased available data, as deep learning algorithms thrive on massive datasets. Meanwhile, we'll be concentrating on enhancing the robustness, efficiency, and accuracy of image classification by integrating the appearance characteristics of lesions with the potent learning ability of convolutional neural networks.

6. Conclusion

We describe a simple yet successful image-collection methodology and resampling method that takes advantage of the benefits of deep learning algorithms to tackle the problem of automatically recognizing oral disease in white-light photographs captured with a smartphone. Recent HRNet, pre-trained on the ImageNet, is tested on our collected images across five illness categories. The outcomes show that our techniques can significantly enhance the prediction performance of smartphone photography images for early cancer diagnosis.

Funding: "This research received no external funding"

Conflicts of Interest: "The authors declare no conflict of interest."

References

[1] H. Sung et al., Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countrie, CA Cancer J. Clin., 71(3), 209-249, 2021.
[2] A. Jemal et al., Annual report to the nation on the status of cancer, featuring survival, JNCI-J. Natl. Cancer Inst. 109(9), 1975-2014, 2017.
[3] P. Mathur et al., Cancer statistics, 2020: report from national cancer registry programme, India, JCO Global Oncol. 6(6), 1063–1075, 2020.
[4] W. Sungwalee et al., Comparing survival of oral cancer patients before and after launching of the universal coverage scheme in Thailand, Asian Pac. J. Cancer Prev., 17(7), 3541-3544, 2016.

Doi: https://doi.org/10.54216/FPA.140122
Received: July 24, 2023 Revised: October 27, 2023 Accepted: December 19, 2023

[5] T. S. Dantas et al., Influence of educational level, stage, and histological type on survival of oral cancer in a Brazilian population: a retrospective study of 10 years observation, Medicine (Baltimore), 95(3), e2314, 2016.
[6] D. Chakraborty, N. Chandrasekaran, and A. Mukherjee, Advances in oral cancer detection, Adv. Clin. Chem. 91, 181-200, 2019.
[7] D. K. Zanoni et al., Survival outcomes after treatment of cancer of the oral cavity (1985-2015), Oral Oncol. 90, 115-121, 2019.
[8] S. Choi and J. N. Myers, Molecular pathogenesis of oral squamous cell carcinoma: implications for therapy, J. Dent. Res. 87(1), 14-32, 2008.
[9] B. Ayaz et al., A clinico-pathological study of oral cancers. Biomedica, 27, 29-32, 2011.
[10] P. M. Speight, S. A. Khurram, and O. Kujan, Oral potentially malignant disorders: risk of progression to malignancy, Oral Surg. Oral Med. Oral Pathol. Oral Radiol., 125(6), 612-627 2018.
[11] G. Yardimci et al., Precancerous lesions of oral mucosa, World J. Clin. Cases, 2(12), 866-872 2014.
[12] B. W. Neville and T. A. Day, Oral cancer and precancerous lesions, CA-Cancer J. Clin., 52(4), 195-215, 2002.
[13] F. Dost et al., A retrospective analysis of clinical features of oral malignant and potentially malignant disorders with and without oral epithelial dysplasia, Oral Surg. Oral Med. Oral Pathol. Oral Radiol., 116 (6), 725-733, 2013.
[14] M. Aubreville et al., Automatic classification of cancerous tissue in laser endo microscopy images of the oral cavity using deep learning, Sci. Rep., 7(1), 11979, 2017.
[15] B. Song et al., Automatic classification of dual-modalilty, smartphone-based oral dysplasia and malignancy images using deep learning, Biomed. Opt. Express 9(11), 5318–5329, 2018.
[16] S. Camalan et al., Convolutional neural network-based clinical predictors of oral dysplasia: class activation map analysis of deep learning results, Cancers 13(6), 1291, 2021..
[17] R. A. Welikala et al., Fine-tuning deep learning architectures for early detection of oral cancer. in Mathematical and Computational Oncology, 25-31, 2020.
[18] Poh CF, Ng S, Berean KW, Williams PM, Rosin MP, and Zhang L , Biopsy and histopathologic diagnosis of oral premalignant and malignant lesions. Journal of the Canadian Dental Association, 74(3), 283-288, 2008.
[19] Branes L, Eveson J, and Reichart P, Tumors of the oral cavity and oropharynx. Pathol Genet, 67, 177-179, 2005.
[20] Lumerman H, Freedman P, and Kerpel S, Oral epithelial dysplasia and the development of invasive squamous cell carcinoma. Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology and Endodontics, 1995; 79(3): 321-329.
[21] Hayat MA. Principles and techniques of electron microscopy, Biological applications. 1981: Edward Arnold. Buajeeb W, Poomsawat S, Punyasingh J, and Sanguansin S, Expression of p16 in oral cancer and premalignant lesions. Journal of oral pathology & medicine, 38(1), 104-108, 2009.
[22] D. Jia et al., ImageNet: a large-scale hierarchical image database, in Proc. Conf. Comput. Vis. Pattern Recognit., 248-255, 2009.
[23] I. J. Goodfellow et al., Generative adversarial networks, in Adv. Neural Inf. Proc. Sys., (NIPS), 2672-2680, 2014.
[24] T. Y. Lin et al., Focal loss for dense object detection, in Proc. Conf. Comput. Vis. Pattern Recognit., 2980–2988, 2017.
[25] C. Elkan, The foundations of cost-sensitive learning, in Proc. Joint Conf. Artif. Intell., 973-978 2001.
[26] K. Sun et al., High-resolution representations for labeling pixels and regions, arXiv:1904.04514 2019.
[27] J. Wang et al., Deep high-resolution representation learning for visual recognition, in IEEE Trans. Pattern Anal. Mach. Intell., 1–1, 2020.
[28] K. He et al., Deep residual learning for image recognition," in Proc. Conf. Comput. Vis. Pattern Recognit., 770-778, 2016.
[29] K. Simonyan and A. Zisserman, Very deep convolutional networks for large-scale image recognition, in Int. Conf. Learn. Repr., (ICLR), 2014.
[30] G. Huang et al., Densely connected convolutional networks, in Proc. Conf. Comput. Vis. Pattern Recognit., 4700–4708, 2017.

Doi: https://doi.org/10.54216/FPA.140122
Received: July 24, 2023 Revised: October 27, 2023 Accepted: December 19, 2023