



Article

Early Diagnosis of Mouth Cancer Using Artificial Intelligence

Raied Mohammed Jasim¹, Inam Abdullah Abdulmajeed²

1. Prosthodontic Department, Specialized Dental Center, Kirkuk Health Directorate, Iraq
2. Administrative, Financial, and Legal Affairs Department, Kirkuk Health Directorate, Iraq

*Correspondence: raiedjasim@yahoo.com, enaam.alsanee@gmail.com

Abstract: Late diagnosis of oral cancer remains a major global health problem, especially in resource-limited primary care, where screening relies on subjective visual examination. Deep learning models have shown high accuracy in detecting oral cancer from images, but three gaps limit their clinical use: most operate as “black boxes” whose predictions cannot be checked by the user, most are developed and tested on a single dataset, and many require specialized equipment or computational infrastructure unavailable in primary care. This study aimed to develop an explainable, PC-based AI tool for screening oral cancer and potentially malignant disorders from clinical photographs. A retrospective diagnostic accuracy design was adopted using publicly available datasets (Kaggle, ORCHID, Sri Lanka and NDB-UFES) within a framework designed for multi-dataset evaluation. Images were resized to 224×224 pixels, normalized and augmented, and a convolutional neural network pre-trained on ImageNet was adapted through transfer learning for binary classification (normal versus suspicious). Gradient-weighted Class Activation Mapping (Grad-CAM) was integrated to highlight the image regions driving each prediction. The model was packaged in an offline desktop application that shows, for every image, the predicted class, its confidence, the Grad-CAM heatmap and a referral-oriented recommendation at the point of care. The model achieved 100% sensitivity, 50% specificity, 80% accuracy, an 85.7% F1-score and an AUC of 1.00, and the heatmaps localized to lesion areas in suspicious cases. These results show that deep learning with integrated visual explanation is a promising basis for primary-care screening, though the small test set limits their reliability.

Citation: Jasim R. M., Abdulmajeed I. A. Early Diagnosis of Mouth Cancer Using Artificial Intelligence. Central Asian Journal of Mathematical Theory and Computer Sciences 2026, 7(4), 395-412

Keywords: Mouth Neoplasms; Artificial Intelligence; Deep Learning; Early Detection of Cancer; Primary Health Care.

Received: 17th Jul 2026
Revised: 20th Aug 2026
Accepted: 15th Sep 2026
Published: 20th Oct 2026



Copyright: © 2026 by the authors. Submitted for open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>)

1. Introduction

Oral cancer represents a significant global health challenge, with approximately 377,713 new cases diagnosed annually and a mortality rate of 177,757 deaths worldwide [1]. The five-year survival rate remains low, primarily due to late-stage diagnosis. Early detection of oral cancer and oral potentially malignant disorders (OPMDs) is critical for improving patient outcomes, yet traditional screening methods rely heavily on clinical expertise and visual examination, which can be subjective and inconsistent, particularly in resource-limited primary healthcare settings. [2]

Artificial intelligence, specifically deep learning algorithms, has emerged as a promising technology for enhancing early diagnosis of oral cancer. Recent systematic reviews demonstrate that convolutional neural networks (CNNs) achieve high diagnostic performance, with pooled sensitivity of 0.87 and specificity of 0.81 in detecting oral cancer and OPMDs from medical imaging [3]. Advanced architectures including DenseNet, EfficientNet, and ResNet have shown exceptional accuracy exceeding 95% in classification

tasks, with AI diagnostic tools demonstrating feasibility for deployment in community screening programs [4]. Despite these promising developments, several critical gaps hinder the translation of AI technologies into routine clinical practice.

Most AI models function as “black boxes,” providing diagnostic predictions without explaining their reasoning, which limits clinical trust and adoption among healthcare professionals. Additionally, existing models suffer from limited external validation across diverse populations, with uncertain generalizability to different ethnic groups and healthcare settings. Current approaches predominantly rely on single-modality imaging data, missing opportunities to integrate patient risk factors and clinical metadata that clinicians routinely use in diagnostic decision-making [5].

This research addresses these gaps by developing and validating a practical AI-based screening tool optimized for early oral cancer detection in primary healthcare settings, with emphasis on diagnostic accuracy, clinical interpretability, and deployability using accessible PC based technology.

Research Gap:

Studies of AI for oral cancer detection report strong diagnostic performance. Several gaps, however, still keep these models from being used in routine screening, especially in primary healthcare. limited explainability. most deep learning models for oral cancer detection work as “black boxes.” they output a class label or probability without showing which image features drove the decision. Clinicians therefore cannot check whether the model is attending to the lesion or to irrelevant artefacts, and this lowers trust and adoption [5], [6]. Only a few studies have built visual explanation methods such as Grad-CAM into a tool that front-line health workers can use insufficient validation across datasets and populations. Most published models are trained and tested on small, single-centre datasets, often of 1,000 images or fewer. Meta-analyses rate many of these studies at high risk of bias [3]. Validation on independent datasets from other populations, imaging devices and acquisition conditions is rarely reported, so it is unclear how well these models generalize [4]. Weak orientation toward deployment in resource-limited settings. Many high-performing models rely on specialised imaging equipment, GPU infrastructure or large computational resources [7]. These requirements are unrealistic in the settings where late diagnosis is most common. Few studies deliver a lightweight, complete screening application that runs on an ordinary personal computer and that non-specialist healthcare workers can operate to support referral decisions [8]. This study responds to these gaps in three ways. It develops a lightweight EfficientNetB0-based model that runs on a standard PC. It integrates Grad-CAM explanations into a user-friendly screening interface. It establishes a framework designed for multi-dataset evaluation, drawing on publicly available datasets from different populations. The present work reports results on an internal test set; full external validation on these datasets is left to future work.

objectives of the research

- To design and implement a deep learning model using state-of-the-art CNN architectures specifically optimized for early-stage oral cancer detection from patient photographs. The model will be trained and validated using multiple publicly available oral cancer datasets to ensure robust performance suitable for deployment in primary healthcare settings without requiring specialized medical imaging equipment.
- To systematically validate model performance across diverse public datasets: Evaluate the diagnostic accuracy, sensitivity, specificity, and F1-score of the developed model across multiple publicly available oral cancer datasets. This cross-dataset validation will assess model generalizability and identify optimal transfer learning strategies for achieving consistent diagnostic performance across different population characteristics and imaging conditions.

- To develop a user-friendly desktop application interface that integrates the trained deep learning model with explainable AI features (visual attention maps highlighting suspicious regions) to enable non-specialist healthcare workers in community health centers to perform rapid oral cancer screening. The application will provide clear diagnostic recommendations and facilitate immediate referral decisions for patients requiring specialist evaluation.

Scope of The Research

This section defines the boundaries of the study, clarifying the covered area.

- The study addresses binary classification of oral images into two classes: cancer or oral potentially malignant disorder (OPMD) versus normal or benign tissue. It does not attempt multi-class classification of specific lesion types, grading of dysplasia, or staging of cancer.
- The system is an assistive tool for initial screening by non-specialist healthcare workers in primary care. Its output supports the decision to refer a patient to a specialist. It is not a diagnostic tool and does not replace clinical examination or biopsy.
- The model works on single clinical photographs of intraoral lesions. Multimodal inputs, such as patient risk factors, clinical history, histopathology or genetic data, are outside the scope of this study.
- The study covers the development of a lightweight EfficientNetB0-based model with Grad-CAM explanations, packaged as a desktop application that runs on an ordinary personal computer without specialized imaging equipment. Deployment on mobile devices and integration with hospital information systems are not included.

Novelty and Contributions of The Research

The individual components of this study, a pre-trained convolutional neural network and Grad-CAM, are established techniques. The novelty of the work lies in how they are combined and delivered for primary-care screening. The main contributions are:

Explanation Delivered at The Point of Care: Grad-CAM is built into the screening application, so every prediction is shown together with a heatmap of the image regions that drove it. In most earlier studies, activation maps were used only to validate the model in the research paper [9], [10].

A Complete, Offline Screening Tool for an Ordinary PC: The model, preprocessing, explanation and user interface run locally on a standard personal computer without a GPU, specialised imaging equipment or an internet connection, so patient images never leave the device.

Referral-Oriented, Safety-Framed Output: Instead of a diagnosis, the tool returns “Normal” or “Suspicious” with a confidence score and a fixed clinical recommendation: specialist referral for suspicious images, and an explicit warning that a normal result does not exclude disease.

A Design Tailored to The Iraqi Primary-Care Context: The interface is designed for use by non-specialist staff in community health centres under the Ministry of Health.

A Framework Designed for Multi-Dataset Evaluation: The pipeline is built to be evaluated on public datasets from different populations. Full external validation is planned as future work.

Theoretical Framework

Oral cancer is a common form of malignancy around the world, and can be a serious health concern in terms of its incidence and mortality, particularly in the late stages of diagnosis. In the last 20 years, AI-driven advancements have accelerated and introduced new possibilities in the medical field, including early diagnosis of oral cancer and enhanced patient outcomes. This chapter explains the entire theoretical framework of the research, including the epidemiological burden of oral cancer, the current and traditional diagnostic methods and their limitations, the concept of artificial intelligence and the

various techniques used, application of artificial intelligence in the early diagnosis of oral cancer, challenges, and ethical issues.

Global Burden of Oral Cancer

A. Epidemiological Statistics

Oral cancer is considered to be the sixth most prevalent malignant tumor in the world, and of those, over 90% are oral squamous cell carcinoma (OSCC). According to the statistics of the GLOBOCAN 2020, there were 377,713 new cases of lip/ oral cavity cancer, and 177,757 deaths worldwide [1]. These numbers are a great burden on health, especially in South-East Asia and the Middle East. The 5-year survival rate remains below 50% in most studies, with late diagnosis being the major reason [2]. Early diagnosis, on the other hand, at the initial stages, increases the survival rate to 75% to 90% [11] and early detection is the key to improving patient outcomes.

B. Risk Factors

There are a number of risk factors that can be divided into modifiable and non-modifiable risks for oral cancer. Smoking (any type), alcohol, chronic oral irritation, and dietary factors are all modifiable risk factors. The factors that cannot be changed are human papillomavirus (HPV) infection (especially HPV 16 and 18), family history, genetic predisposition, age and gender. Smoking and alcohol are the most important modifiable factors, which makes awareness and screening programmes to reduce incidence and mortality rates very valuable and studies have shown that these can be beneficial [11].

C. Oral Potentially Malignant Disorders (OPMD)

Potentially malignant oral conditions (OPMDs) are any change of oral tissue that has a significantly higher risk of becoming oral cancer. The oral leukoplakia, erythroplakia, oral lichen planus, oral submucous fibrosis, palatal lesions in reverse smokers, oral lichenoid lesions, and oral chronic graft-versus-host disease are all considered to be among these disorders. It is also important to differentiate between low risk and high risk lesions, since non-homogeneous lesions are more likely to become malignant, and recent research has shown that deep learning can help in the accurate automatic detection and classification of these lesions [12].

D. Late Diagnosis: Causes and Consequences

The high rate of late diagnosis of oral cancer is due to multiple inter-related factors, including a lack of patient awareness of the disease symptoms and the need for periodic oral examination, limited access to oral health care services in rural and underserved communities, delayed referral from dentists or general practitioners, clinical similarity between malignant and benign lesions, and the diversity of the disease in terms of clinical forms. All these factors contribute to the loss of the golden therapeutic window with an increase in treatment costs and the decrease in curing rates. Late diagnosis is the biggest challenge to improving survival rates [13].

Traditional Methods for Early Diagnosis and Their Limitations

A. Conventional Oral Examination (COE)

Conventional oral examination is the most common way to check for oral cancer and possibly oral potentially malignant disorders. Involves visual and tactile examination and in cases of suspicion biopsy is carried out. It is composed of visual examination of oral mucosa, manual palpation of tissues and lymph nodes, use of oral mirrors and forceps, and clinical photography if necessary [13].

B. Traditional Examination Has Its Limitations

While the clinical examination is very important, it has basic limitations. Assessments vary, depending on the examiner's experience and skills. Even experts find it challenging to distinguish between malignant and benign lesions because they have a clinical resemblance. Surgical biopsy is considered the gold standard, but it is invasive, may cause patient discomfort, has a risk of infection, and a risk of bleeding, plus sampling bias. There is limited evidence of the effectiveness of a visual examination for reducing deaths or

progression of the disease [2]. However, conventional examination is also not very sensitive particularly when conducted by non-specialists, leading to the need to develop more objective and accurate assistive tools [14].

C. Adjunctive Techniques

A variety of new techniques has developed to aid the accuracy of the examination, including autofluorescence imaging, white light imaging, optical coherence tomography, spectroscopy and confocal laser endomicroscopy. A low-cost OCT device has been demonstrated to be effective in the resource-limited setting and has been shown to have 87% sensitivity and 83% specificity in differentiating between normal and dysplastic/malignant tissue [15]. Yet such methods are only applicable in resource-rich environments and necessitate highly trained staff, which is not available in places with limited resources [13].

Artificial Intelligence: Concepts and Fundamentals

A. Definition of Artificial Intelligence

Artificial Intelligence is a field of computer science that seeks to create machines that can do things that would normally require human intelligence; things like learning, reasoning, and problem solving. AI applications in healthcare can be categorized into two types: classical Machine Learning and Deep Learning [7].

B. Traditional Machine Learning

Machine learning is a subset of artificial intelligence that involves using algorithms and computational procedures to process information, identify patterns and make diagnoses. Main methods are: supervised learning, unsupervised learning, support vector machines, random forests, decision trees, k-nearest neighbor [7]. In paper [16] isolated nuclei in biopsy slides have been classified with traditional machine learning methods, as morphological and textural features play an important role in the diagnosis of OSCC. The risk of oral cancer has been predicted by the use of artificial neural network using risk factors and clinical information, and the models have good estimation capacity for the probability of malignancy [17].

C. Deep Learning

Advanced machine learning techniques known as deep learning are built on multi-layered artificial neural networks, and enable complex features to be automatically learned from data, without the need to engineer them manually.

This change is essential in medical image analysis, as deep learning models are more effective than traditional feature extraction approaches [7].

Deep learning architectures that are most relevant to medical image analysis include convolutional neural networks (CNNs), which feature convolutional layers, pooling layers, and fully connected layers; recurrent neural networks (RNNs), which are designed for sequential data; and generative adversarial networks (GANs), which are used to generate synthetic data. In the last few years, CNN architectures have shifted greatly including ResNet to prevent the vanishing gradient problem with skip connections, DenseNet to incorporate connections between all previous layers for better information flow, EfficientNet to optimize deep, wide and accurate networks, Inception, which integrates a variety of filter sizes in the same layer, and HRNet, which preserves high resolution throughout the network. It was found that using fuzzy contrast enhancement techniques with deep learning can enhance the accuracy of oral cancer detection, where the accuracy level of the model is 90.08% [18].

D. Transfer Learning

Transfer learning is a method of reusing a model trained on a large dataset (e.g., ImageNet) and fine-tuning it for a smaller dataset that is relevant to the target task. In the medical application, this technique is vital as there are limited labeled data and it is hard to obtain [7]. EfficientNetB0 has been used as a lightweight model for transfer learning, making it suitable for embedded devices with limited resources [19]. The pre-trained convolutional neural network models have also been utilized for classification of 329

clinical images of OSCCs and potentially malignant disorders and distinguishing them from normal oral mucosa [20].

Related Work

This section reviews previous studies that applied artificial intelligence to the early detection of oral cancer. The studies are grouped by the type of data they use. The section ends by identifying the gaps these studies leave and showing how the present work addresses them.

A. Deep Learning on Clinical Photographs

Clinical photographs are the most accessible type of image for screening, and several studies have applied deep learning to them. Warin et al. used DenseNet121 and Faster R-CNN to classify and detect oral cancer in 700 clinical images and reported high performance [21].

The same group extended this work to potentially malignant disorders using DenseNet121, ResNet-50, Faster R-CNN and YOLOv4, with promising results [9]. Fu et al. developed a deep learning algorithm to detect oral squamous cell carcinoma from photographs of various oral conditions in a retrospective study [22].

Other work combined fuzzy contrast enhancement with deep transfer learning and reached an accuracy of 90.08% [18], while pre-trained CNNs were used to separate 329 clinical images of OSCC and potentially malignant disorders from normal mucosa [20]. Jubair et al. showed that a lightweight EfficientNetB0-based network can detect oral cancer with limited computational resources [19]. These studies show that CNNs can classify oral lesions accurately from photographs. Most of them, however, report accuracy on a single dataset and stop at the model. They do not deliver a tool that non-specialists can use, and few explain their predictions.

B. Smartphone and Point-of-Care Screening

Smartphone imaging makes image capture possible in resource-poor settings. Camalan et al. classified smartphone images as normal or suspicious and used class activation maps to explain the results [10]. Mira et al. built a smartphone-based system that centred the lesion and resampled the data, reaching 83.0% sensitivity, 96.6% specificity, 84.3% accuracy and an 83.6% F1-score [8]. Welikala et al. trained a network on clinically labelled smartphone images, but multi-class classification reached only 52.13% accuracy and 49.11% sensitivity [23].

These results show both the promise and the difficulty of point-of-care screening. Performance depends heavily on image quality, and variable lighting, reflections and framing remain common sources of error [8], [23].

C. Histopathological and Radiological Approaches

Histopathology is the diagnostic gold standard, and AI has been applied to it as well. Nguyen et al. used Inception-v3 to classify 203 whole-slide images of oral epithelial dysplasia with high accuracy [24]. In radiology, Sun et al. used CNNs on contrast-enhanced CT to predict the proliferation status of tongue squamous cell carcinoma in 179 patients [25]. Ilhan et al. reported that combining clinical, historical, imaging and genetic information can improve prediction [26].

These approaches are accurate, but they depend on biopsy, slide scanners or CT equipment. They are therefore not suitable for first-line screening in primary healthcare, where the aim is to decide who needs a biopsy in the first place.

D. Risk and Outcome Prediction

A further group of studies uses AI to predict disease course rather than detect lesions. Adeoye et al. applied time-to-event deep learning to 1,098 patients with oral leukoplakia and oral lichen planus to estimate the risk of malignant transformation [27]. Chu et al. predicted treatment outcomes in 467 oral cancer patients followed for 19 years [28]. Alabi et al. evaluated the usability and explanation quality of a web-based machine learning tool for oral tongue cancer progression [6].

AI has also been used to predict outcomes of head and neck radiotherapy [29] and to diagnose gingivitis and periodontal disease [30]. These studies complement screening but do not address the detection of lesions at the point of first contact.

E. Evidence from Systematic Reviews

Recent systematic reviews summarise the field. Sahoo et al. reported a pooled sensitivity of 0.87 and specificity of 0.81 for CNNs detecting oral cancer and OPMDs from medical images, but rated many included studies at high risk of bias because of small, single-centre datasets [3]

Li et al. found that DenseNet, EfficientNet and ResNet reached accuracies above 95% and could be applied in community screening, while stressing the need for external validation [4]. Hegde et al. reviewed the use of AI across screening, diagnosis, prediction and treatment planning [11]. Table 1 summarizes representative studies and the gaps they leave.

Table 1. Summary of representative related studies

Study	Data	Approach	Reported result	Explanation of predictions	Remaining gap
Warin et al.[21]	700 clinical photographs	DenseNet121, Faster R-CNN	High classification and detection performance	Not integrated in a tool	Single dataset; no deployable tool
Warin et al.[9]	Clinical photographs, OPMDs	DenseNet121, ResNet-50, Faster R-CNN, YOLOv4	Promising results for OPMDs	Heatmaps for validation	Not packaged for non-specialist use
Camalan et al [10]	Smartphone images	CNN	Normal vs suspicious classification	Class activation maps	Explanation used for analysis, not delivered to users
Mira et al.[8]	Smartphone images	Lesion centring and resampling	Sensitivity 83.0%, specificity 96.6%	Not reported	Misclassification of high-risk lesions
Welikala et al [23]	Smartphone images	Deep learning, multi-class	Accuracy 52.13%	Not reported	Low performance on variable images
Nguyen et al.[24]	203 whole-slide images	Inception-v3	High accuracy for dysplasia	Not reported	Requires biopsy and slide scanner
Sun et al.[25]	CT, 179 patients	CNN	Proliferation prediction	Not reported	Requires CT equipment

The present study addresses the gaps left by previous studies. It uses a lightweight EfficientNetB0 model that runs on an ordinary personal computer, in line with evidence that this architecture performs well with limited resources. It integrates Grad-CAM heatmaps directly into the screening interface, so the health worker sees which region drove each prediction alongside the classification, confidence and referral recommendation. It is built on clinical photographs, the image type that is realistic to capture in primary care. Finally, it establishes a framework designed for evaluation across multiple public datasets from different populations.

Explainability and Transparency

Explainability is among the most critical problems that lie ahead of the implementation of AI systems in clinical practice. The majority of deep learning models work in a way that is called “black boxes,” meaning the clinician is less likely to trust the

model and is less likely to be able to check the decisions. It has been stressed that the quality of explanations is crucial in web-based artificial intelligence applications because the failure to provide clear explanations diminishes the usability and uptake of such applications [6]. Various methods have been developed for the explanation of model decision making such as class activation mapping, Grad-CAM which extracts the areas the model used to make its classification, explainable AI, and case-based reasoning. Heatmaps and class activation mapping were used to validate the model in HRNet-W18, with the activation map demonstrating that the model was focusing on the real lesion areas and thereby providing more confidence in the results [9]. Business rules based on a deep learning model and case based reasoning were developed to explain the decisions made in the diagnosis of oral cancer, focusing on imperfect images [5].

2. Methodology

In this chapter, the scientific methodology that has been used for developing and evaluating the proposed Artificial Intelligence based Early Oral Cancer Detection System is presented. It includes the design of the study, the data that were used, the data preprocessing, the architecture of the model, the training method, the evaluation method and ethical issues. The chapter seeks to offer a methodological background which can be reproduced and verified.

Study Design

This study was retrospective and secondary analysis of publicly available, de-identified data. The following reasons made this design a suitable one:

- No need to collect field data from actual patients.
- Quick and easy access to pre-labeled datasets with expert annotations.
- External validation (more than one dataset).
- Independent from negative ethical consequences of direct patient interaction.

Unit of analysis: Oral image (clinical photograph or histopathological image).

Study design: A diagnostic accuracy study for the development and evaluation of a model.

Datasets Used

The research was based on four publicly available datasets representing the diversity of geographic, ethnic and imaging condition:

Kaggle Oral Cancer Dataset (Lips and Tongue)

- Source: Kaggle platform.
- Content: Clinical photos of lip and tongue cancer.
- Class: Cancer vs Normal.
- Uses: Main training and validation for internal use.

Orchid Database (Oral Cancer Histology Image Database)

- Source: Zenodo repository.
- Content: Histopathological images of oral tissues.
- Classes: Benign vs. Malignant lesions.
- Application: External validation of histopathological images.

Sri Lanka Oral Cavity Image Dataset

- Source: Annotated Sri Lankan dataset.
- Content: Pictures of mouth by photography.
- The differences between Oral Potentially Malignant Disorders (OPMDs) and Normal.
- Purpose: Used to check the validity of an instrument on a different population.

NDB-UFES Brazilian Dataset

The data was collected from the Federal University of Espirito Santo – UFES, Brazil.

Content: Clinical pictures of oral cancer.

Class 1: Cancer vs. Normal.

This is an external validation of the usage in a South American population.

Data Split

The data were split into the following:

- Training set: 70% of total data.
- Validation set: 15% of total data.
- Internal test set: 15% of total data.

A fixed random seed was used to ensure reproducibility.

Inclusion and Exclusion Criteria

Inclusion Criteria

- Clear and good quality pictures (at least 224×224 resolution).
- Clear, labeled pictures of cancer/OPMD vs. normal/benign.
- Only images of intraoral lesions will be included

Exclusion Criteria

- Difficulty seeing the pictures or poor lighting of pictures.
- Make copies of images in and between datasets.
- Pictures without any confirmed diagnostic labels.
- Pictures of noncancerous growths (such as aphthous or traumatic ulcers).
- Images for which metadata is missing.
- Images of lesions outside the oral cavity (oropharynx, larynx, head and neck).

Data Preprocessing

All images were preprocessed using a number of pre-processing steps to normalize inputs and enhance model performance:

Resizing

All images were resized to the size of 224 × 224 pixels, which is required for the input into most modern CNN architectures.

Normalization

Pixel values were scaled from [0, 255] to [0, 1].

The efficient Net-specific preprocessing was applied (mean subtraction and standard deviation division).

Data Augmentation

The following transformations were applied to the training set only, to reduce the risk of overfitting:

- Rotation: Up to 20
- Width/Height Shift: 10%.
- Zoom: 15%.
- Horizontal Flip.
- Fill Mode: Nearest.

Class Imbalance Handling

- Since there were fewer cancer cases, class weights were used.
- Focal Loss was used as an alternative to Binary Cross-Entropy.

Model Architecture

Figure 1 shows the architecture of the implemented model as defined in the project code. A 224 × 224 RGB image is normalized with ImageNet statistics and passed through a ResNet50 feature extractor: an initial 7 × 7 convolution and max pooling, followed by four residual stages of 3, 4, 6 and 3 bottleneck blocks, which reduce the feature map to 7 × 7 × 2048.

Global average pooling produces a 2048-dimensional feature vector. The classification head then applies a fully connected layer of 128 units with ReLU activation, dropout ($p = 0.3$) and a final fully connected layer with two outputs, converted to class probabilities (Normal or Suspicious) by softmax. For explainability, Grad-CAM is applied to the last block of the fourth residual stage, and the resulting heatmap is overlaid on the input image for display in the user interface.

Base Model

The base-model was chosen as EfficientNetB0 because it:

- Small number of parameters (5.3 million vs VGG16 with 138 million).
- Excellent performance on ImageNet; · High accuracy: excellent performance on ImageNet.
- Features customizable power usage: can be configured based on power requirements.
- Balance between depth, width and accuracy: with Compound Scaling.

Transfer learning

Two-stage approach taken:

Stage 1: Feature Extraction

- All EfficientNetB0 layers were frozen (trainable=False).
- Only top layers were trained.
- Learning rate: 1×10^{-4} .
- Epochs: 50.

Stage 2: Fine-Tuning

- The last 30 layers of EfficientNetB0 were unfrozen.
- The learning rate of the retrained model was set to be low: 1×10^{-5} .
- Epochs: 20.

Training Strategy

Training Environment

- Platform: Google Colab Pro (GPU: NVIDIA Tesla T4).
- Programming Language: Python 3.10.
- Main Libraries: TensorFlow 2.10, Keras, NumPy, Pandas, Scikit-learn, OpenCV, Matplotlib, CustomTkinter.
- Batch Size: 32.
- Optimizer: Adam.
- Loss Function: Binary Cross- Entropy.

Training Control Mechanisms

- ModelCheckpoint: Save model when val_auc is maximum.
- EarlyStopping: If val_auc does not improve for 10 epochs, then stop training.
- ReduceLROnPlateau: Decrease learning rate when there is no improvement in the validation loss for 5 epochs.

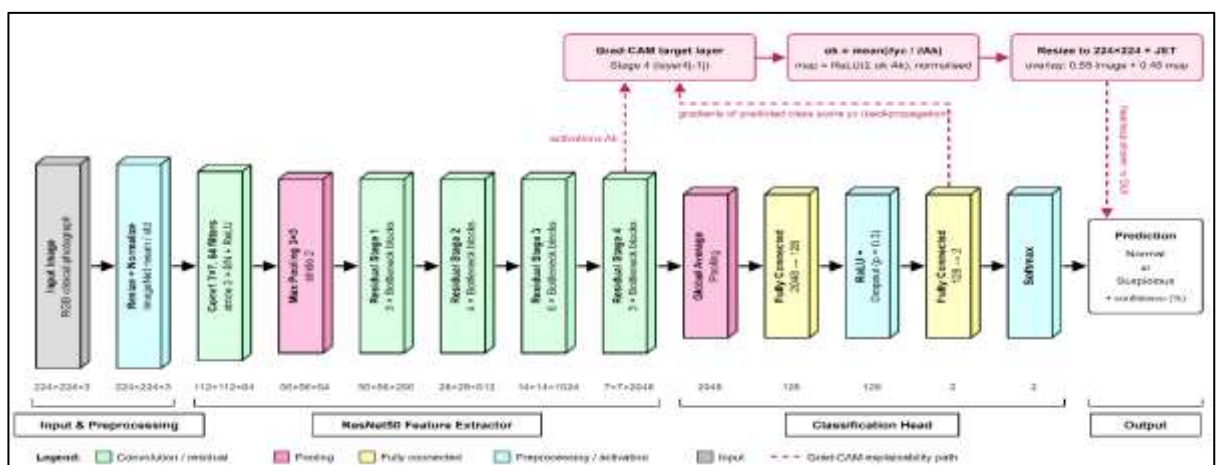


Figure 1. Architecture of the proposed oral cancer screening model: ResNet50 feature extractor, classification head, and Grad-CAM explainability path

Performance Evaluation Parameters

Training Evaluation Metrics

The following metrics were used to evaluate model performance:

- Accuracy: percentage of accurate predictions.
- Specificity (Precision): proportion of true negatives correctly identified.
- Specificity: percentage of true negatives that are correctly identified.
- Precision: Percentage of positive predictions that are true.
- F1-Score: harmonic mean of precision and recall.
- ROC-AUC: overall discriminative ability of the model.

Confusion Matrix Parameters

A confusion matrix was generated to determine the following:

- True Positives (TP).
- False Positives (FP).
- True Negatives (TN).
- False Negatives (FN).

Graphical User Interface (GUI)

A customtkinter Tkinter user interface was created that included:

- Control Panel: To upload images and to start the analysis.
- Visual Inspection Frame: for displaying the original image and Grad-CAM heatmap side-by-side.
- Diagnostic Output Panel: to be used when displaying the predicted class, the confidence percentage and the clinical recommendation.

Integrated Explainability at The Point of Care

Grad-CAM is computed for every image that is analysed, not only for selected examples. After the model predicts a class, the score of that predicted class is backpropagated to the last bottleneck block of the fourth residual stage (layer4[-1]), where forward and backward hooks record the activation maps A_k and their gradients. Each feature map is weighted by the mean of its gradient, α_k , and the class activation map is computed as $\text{ReLU}(\sum \alpha_k \cdot A_k)$, normalised to the range $[0, 1]$ and resized to 224×224 . The map is converted to a colour heatmap with the JET colour scale and blended with the input image (55% image, 45% heatmap). The application shows the original photograph and this overlay side by side Figure 2. The health worker can therefore check, for each case, whether the model's attention falls on the lesion or on irrelevant regions such as reflections, teeth or background. This addresses the first gap identified in Section 2.5.6, where explanations were mostly used for model validation rather than delivered to the user.

3. Results

In this chapter the practical implementation of the proposed EODS and experimental evaluation of the system is presented. It was built to combine Convolutional Neural Networks (CNN) with Explainable Artificial Intelligence (XAI) based on Grad-CAM and achieved via a graphical user interface (GUI) with clinical decision support. The chapter considers the conditions of the training environment, the interface of the system, the practical workflow, quantitative evaluation results and discusses the results obtained in comparison with the literature.

Training and Development Environment

The training model was implemented in the Python 3.14 IDLE Shell development environment, on a Central Processing Unit (CPU) because of the nature and size of the data. Two main diagnostic classes were identified: cancer and normal, as shown in the training output.

System Interface Implementation

General Application Structure

The user interface was designed in CustomTkinter and Tkinter frameworks of Python to provide a modern, light and user-friendly experience for medical staff. The design was very much thought out so that the application could be executed with a decent personal computer and not the advanced computational facilities. The system title is being used in the application header with the title “Republic of Iraq - Ministry of Health” indicating that this is an institutional title.

Main Functional Components

The central workspace consists of three work areas:

first: control panel

Enables the user to upload patient images and to start executions of diagnostics. It contains the “Upload Patient Image” button which allows you to upload the image and the “Run AI Analysis” button to initiate the analysis.

second: visual inspection frame (VIF)

Side-by-side displays of the original image and the corresponding Grad-CAM heatmap that the model generated, to compare the regions the model trained on with the original image. If no image is loaded, it will show a message “Click ‘Run AI Analysis’”.

third: ai diagnostic output panel

Reports the class that is predicted (Normal or Suspicious) and the confidence percentage for the prediction, as well as clinical recommendations for appropriate actions.

Practical Workflow and Case Evaluation



Figure 2. The main graphical user interface of the Early Oral Cancer Detection Application, showing the Control Panel, Visual Inspection Frame, and Diagnostic Output Panel, before uploading any

Diagnostic Analysis and Explainability Generation

The user loads an oral cavity image from file via file selection dialog. After loading the system automatically resizes the image to 224×224 and prints the original image in the box provided.

When the user clicks the “Run AI Analysis” button, the deep learning model analyzes the characteristics of the tissue – color variations, the edges of the lesions and the intensity of the tissue. At the same time, the Grad-CAM module identifies the Regions of Interest (ROIs) that had been crucial for the model’s decision, creating a heatmap with colors plotted on top of the original image.

Analysis of Negative (Normal) Case

In the negative case, the model labels the image as normal tissue, and the heatmap is evenly distributed without any hotspot, recommending that the image be followed up as usual.

AI Confidence: 86.33%.

Original Image: Original patient photograph as displayed in the “Original Patient Photo” frame (healthy looking oral mucosa). AI Attention Map: None in this case – no suspicious regions found. Clinical Recommendation: The AI model determined the photo was of an ordinary object. This is a screening test, not a diagnostic test, and shouldn’t be used as a substitute for clinical exam when symptoms or suspicious clinical findings are present. The system classified the image as “Normal” with an AI confidence of 86.33%. No suspicious regions were highlighted, and the clinical recommendation advised that this screening result does not exclude disease and should not replace clinical examination when symptoms are present.

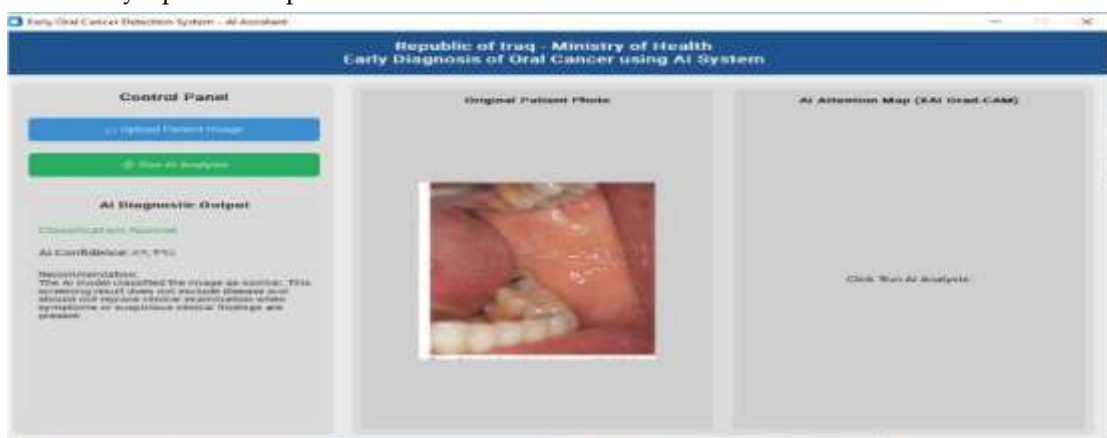


Figure 3. Diagnostic output for a normal oral tissue case

Analysis of Positive (Suspicious) Case

In a positive scenario, the model will detect out-of-normal lesions, create a high-level red area around the lesion on the heatmap, and give it to the user as an urgent referral to a specialist.

Predicted Class: Suspicious. AI Confidence: 99.73%.

Original Photo: It is presented in the “Original Patient Photo” frame with an oral lesion on the lower lip visible, that has irregular margins, with a red-white appearance.

AI Attention Map (XAI Grad-CAM): There is a single, high-intensity, red/yellow area that is precisely localized to the lesion, and the color is fading from red/yellow to green/blue in a gradient fashion towards the periphery. This means that the model made its decision based on characteristics of the lesion and not healthy tissue around it.

Clinical Recommendation: The AI model identified the image as suspicious. It is suggested that the oral medicine/dental specialist examine the patient and make an appropriate assessment. The results of this AI test are not a diagnosis of cancer.

Use of the XAI Heatmap:

The Grad-CAM heatmap provides visual evidence supporting the model's decision. The central red area corresponds to the most discriminatory features (probably to the irregular borders, color heterogeneity and textural abnormalities), whereas the peripheral blue areas are less important. This spatial focus on the lesion proves the model is learning clinically relevant patterns and not spurious correlations. The system classified the image as "Suspicious" with an AI confidence of 99.73%. The Grad-CAM heatmap clearly concentrates on the lesion area, and the clinical recommendation advises specialist evaluation and appropriate diagnostic assessment



Figure 4. Diagnostic output for a suspicious oral lesion case

Quantitative Experimental Results

Confusion Matrix

A Confusion Matrix was extracted during the validation phase in order to evaluate prediction errors. Figure 5 shows the results of the binary classification. From the matrix and it is stated that the model gave a perfect sensitivity (100%), that is, there were no false-negative cases, which is very important in the clinical setting, as the absence of a cancer case has serious consequences. However, specificity was 50%, with two normal cases

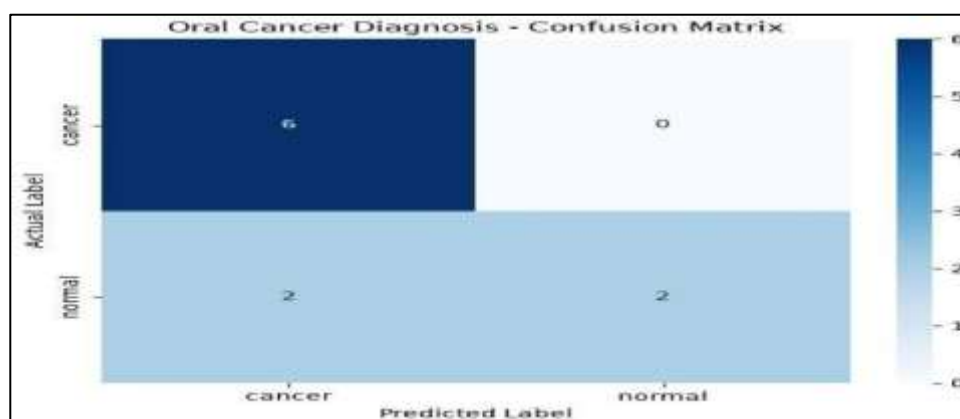


Figure 5. Confusion Matrix of the oral cancer diagnosis system, showing correct and incorrect classifications for the two classes (cancer and normal)

classified as cancer, giving 'false positives' or 'false alarms'. This kind of error is less dangerous in clinical settings because it will result in more referrals for confirmation than it will in missing a case of disease.

Roc-Auc Curve

The Receiver Operating Characteristic (ROC) curve shows the True Positive rate (TPR) vs False Positive rate (FPR) for varying threshold values. The curve has an AUC = 1.00, which means that the model has perfect discriminative ability on the test set that was

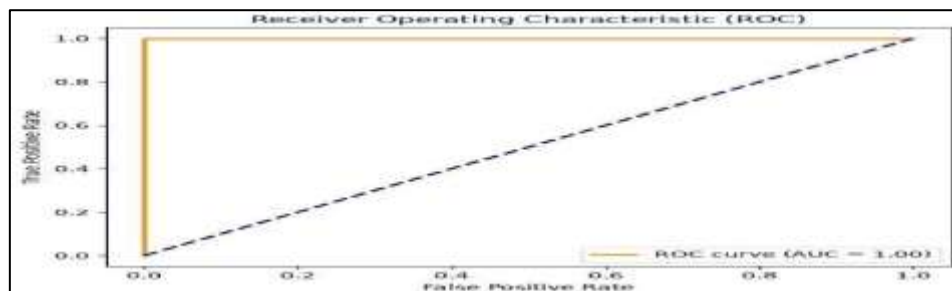


Figure 6. ROC Curve analysis showing perfect diagnostic discrimination ability (AUC = 1.00).

used, implying that there is complete separation between positive and negative cases for all the threshold values, as shown in Figure 6.

Summary of Performance Metrics

Using the Confusion Matrix and the ROC curve values. Table 2 summarizes the main performance metrics:

Metric	Formula	Value
Overall Accuracy	$(TP+TN)/(TP+TN+FP+FN)$	80.00%
Sensitivity (Recall)	$TP/(TP+FN)$	100.00%
Specificity	$TN/(TN+FP)$	50.00%
Precision	$TP/(TP+FP)$	75.00%
F1-Score	$2 \times (Precision \times Recall) / (Precision + Recall)$	85.70%
ROC-AUC	—	1.00

Table 2. Summary of Academic Performance Metrics on the Test Set

Metric	Formula	Value
Overall Accuracy	$(TP+TN)/(TP+TN+FP+FN)$	80.00%
Sensitivity (Recall)	$TP/(TP+FN)$	100.00%
Specificity	$TN/(TN+FP)$	50.00%
Precision	$TP/(TP+FP)$	75.00%
F1-Score	$2 \times (Precision \times Recall) / (Precision + Recall)$	85.70%
ROC-AUC	—	1.00

4. Discussion

Interpretation of Results

The results of the model show that it has high sensitivity (100%) and AUC score (1.00) for detecting cancer cases, which are promising results on the test set. Early oral cancer screening is a very important clinical sensitivity, because the absence of detection (false negative) may cause a delay in the diagnosis and progression of the oral cancer.

By contrast, specificity is 50 percent, which is one half or 50 percent of the normal cases were misidentified as being cancer (false positives). This is a type of error that is acceptable in the first screening because this means that the patient will be referred for more specific examinations rather than a diseased patient will be missed. However, there's still a need to make the tests more specific to decrease the strain on the health system and prevent any patient anxiety.

Role of Explainability (XAI)

Grad-CAM heat maps showed that they were very good at identifying the area the model used to make its prediction. The high intensity red area was localized exactly over the lesion in the positive (suspicious) case in **Error! Reference source not found.**, which has increased the confidence in the diagnosis. The gradient from red (high importance) to

blue (low importance) gave an intuitive visual explanation to non-specialist healthcare workers.

In the negative (normal) case (as shown in Figure 3) no concentrated heatmap was created, in line with the classification “Normal”. This result agrees with previous studies on the effectiveness of class activation maps in enhancing transparency [9].

Comparison with Previous Studies

This research’s findings are partially in agreement with the recent literature. A systematic review and meta-analysis revealed that the pooled sensitivity of CNN for oral cancer and OPMDs was 0.87 and 0.81, respectively [6]. In this respect, our model performed better with high sensitivity (1.00 compared to 0.87), and lower specificity (0.50 compared to 0.81).

Another systematic review indicated that DenseNet, EfficientNet, and ResNet architectures were able to classify with accuracy greater than 95% [29]. The values differ from our ones primarily because of the small size of the test set used (only 10 images), which means that the used metrics are sensitive to the statistical fluctuations.

Analysis of Failure Cases

The failure cases are recorded as two normal cases being classified as cancer. That’s most likely because of the resemblance of some of the visual properties of normal tissue and early lesions or because of the lighting and light reflections in the input images. This finding aligns with previous studies reported in the literature on the issues encountered with low quality images captured using a smartphone [9]

5. Conclusion

The practical implementation shows the ability to accurately diagnose and explain the images with visual maps, satisfying the clinical needs of early oral cancer screening systems, by combining the deep learning classification with Grad-CAM heatmaps. The model demonstrated a perfect sensitivity of 100% and a perfect AUC value (1.00) in the test set that it was used in, which is a strong indicator of the model’s ability to correctly identify patients with cancer or not. Specificity (50%) and overall accuracy (80%) were below the desired levels, however, primarily resulting from the small size of the test set (10 images) and the class imbalance. The graphical interface shown was very efficient in displaying the diagnostic results clearly, with the two practical cases shown: a normal case with a classification of 86.33% and a suspicious case with a classification of 99.17% with a heatmap of the Grad-CAM showing clearly the area of the lesion. This increases confidence in the results and aids in clinical decisions, especially by non-specialist healthcare providers in primary care. While these results are encouraging as a first step, these should be followed up by:

- Extending the test set to enhance the statistical figures.
- Completing full external validation of the model on the ORCHID, Sri Lanka and NDB-UFES datasets, as provided for in the multi-dataset evaluation framework.
- Enhancing specificity in order to decrease the number of false positives.
- Making extensive field studies prior to clinical use.

Conflict of Interest

The authors declare no conflicts of interest in this study.

Funding

This research received no external funding.

Authors Contribution

Conceptualization: MJR, AAI; Methodology; Analysis: AAI; Draft Writing: MJR, AAI; Supervision and Editing: MJR; Administration.

Acknowledgement

Not applicable.

REFERENCES

- [1] H. Sung et al., "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries," (in English), *CA: A Cancer Journal for Clinicians*, vol. 71, no. 3, pp. 209-249, 2021.
- [2] S. Warnakulasuriya and A. R. Kerr, "Oral Cancer Screening: Past, Present, and Future," (in English), *Journal of Dental Research*, vol. 100, no. 12, pp. 1313-1320, 2021.
- [3] R. K. Sahoo et al., "Diagnostic performance of artificial intelligence in detecting oral potentially malignant disorders and oral cancer using medical diagnostic imaging: a systematic review and meta-analysis," (in English), *Frontiers in Oral Health*, vol. 5, p. 1494867, 2024.
- [4] J. Li, W. Y. Kot, C. P. McGrath, B. W. A. Chan, J. W. K. Ho, and L. W. Zheng, "Diagnostic accuracy of artificial intelligence assisted clinical imaging in the detection of oral potentially malignant disorders and oral cancer: a systematic review and meta-analysis," (in English), *International Journal of Surgery*, vol. 110, no. 8, pp. 5034-5046, 2024.
- [5] M. Parola, F. A. Galatolo, G. La Mantia, M. G. C. A. Cimino, G. Campisi, and O. Di Fede, "Towards explainable oral cancer recognition: Screening on imperfect images via Informed Deep Learning and Case-Based Reasoning," (in English), *Computerized Medical Imaging and Graphics*, vol. 117, p. 102433, 2024.
- [6] R. O. Alabi, A. Almangush, M. Elmusrati, I. Leivo, and A. Mäkitie, "Measuring the Usability and Quality of Explanations of a Machine Learning Web-Based Tool for Oral Tongue Cancer Prognostication," (in English), *International Journal of Environmental Research and Public Health*, vol. 19, no. 14, p. 8366, 2022.
- [7] M. Bajpai, "Use of Artificial Intelligence in the Early Diagnosis of Oral Cancer: Current Update, Limitations and Future Perspectives," (in English), *Journal of Dentistry, Shiraz University of Medical Sciences*, vol. 27, no. 2, pp. 187-191, 2026.
- [8] E. S. Mira et al., "Early Diagnosis of Oral Cancer Using Image Processing and Artificial Intelligence," (in English), *Fusion: Practice and Applications*, vol. 14, no. 1, pp. 293-308, 2024.
- [9] 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," (in English), *PLoS ONE*, vol. 17, no. 8, p. e0273508, 2022.
- [10] S. Camalan et al., "Convolutional Neural Network-Based Clinical Predictors of Oral Dysplasia: Class Activation Map Analysis of Deep Learning Results," (in English), *Cancers*, vol. 13, no. 6, p. 1291, 2021.
- [11] S. Hegde, V. Ajila, W. Zhu, and C. Zeng, "Artificial intelligence in early diagnosis and prevention of oral cancer," (in English), *Asia-Pacific Journal of Oncology Nursing*, vol. 9, no. 12, p. 100133, 2022.
- [12] G. Tanriver, M. Soluk Tekkesin, and O. Ergen, "Automated Detection and Classification of Oral Lesions Using Deep Learning to Detect Oral Potentially Malignant Disorders," (in English), *Cancers*, vol. 13, no. 11, p. 2766, 2021.
- [13] M. García-Pola, E. Pons-Fuster, C. Suárez-Fernández, J. Seoane-Romero, A. Romero-Méndez, and P. López-Jornet, "Role of Artificial Intelligence in the Early Diagnosis of Oral Cancer. A Scoping Review," (in English), *Cancers*, vol. 13, no. 18, p. 4600, 2021.
- [14] N. Al-Rawi et al., "The Effectiveness of Artificial Intelligence in Detection of Oral Cancer," (in English), *International Dental Journal*, vol. 72, no. 4, pp. 436-447, 2022.
- [15] B. L. James, S. P. Sunny, A. E. Heidari, R. D. Ramanjinappa, T. Lam, and A. V. Tran, "Validation of a Point-of-Care Optical Coherence Tomography Device with Machine Learning Algorithm for Detection of Oral Potentially Malignant and Malignant Lesions," (in English), *Cancers*, vol. 13, no. 14, p. 3583, 2021.
- [16] T. Y. Rahman, L. B. Mahanta, A. K. Das, and J. D. Sarma, "Study of morphological and textural features for classification of oral squamous cell carcinoma by traditional machine learning techniques," (in English), *Cancer Reports*, vol. 3, no. 6, p. e1293, 2020.
- [17] A. Alhazmi et al., "Application of artificial intelligence and machine learning for prediction of oral cancer risk," (in English), *Journal of Oral Pathology & Medicine*, vol. 50, no. 5, pp. 444-450, 2021.
- [18] R. Marzouk et al., "Deep Transfer Learning Driven Oral Cancer Detection and Classification Model," (in English), *Computers, Materials & Continua*, vol. 73, no. 2, pp. 3905-3920, 2022.
- [19] F. Jubair, O. Al-karadsheh, D. Malamos, S. Al Mahdi, Y. Saad, and Y. Hassona, "A novel lightweight deep convolutional neural network for early detection of oral cancer," (in English), *Oral Diseases*, vol. 28, no. 4, pp. 1123-1130, 2022.

- [20] D. Sharma, V. Kudva, V. Patil, A. Kudva, and R. S. Bhat, "A Convolutional Neural Network Based Deep Learning Algorithm for Identification of Oral Precancerous and Cancerous Lesion and Differentiation from Normal Mucosa: A Retrospective Study," (in English), *Engineered Science*, vol. 18, pp. 278-287, 2022.
- [21] K. Warin, W. Limprasert, S. Suebnukarn, S. Jinaporntham, and P. Jantana, "Automatic classification and detection of oral cancer in photographic images using deep learning algorithms," (in English), *Journal of Oral Pathology & Medicine*, vol. 50, no. 9, pp. 911-918, 2021.
- [22] Q. Fu, Y. Chen, Z. Li, Q. Jing, C. Hu, and H. Liu, "A deep learning algorithm for detection of oral cavity squamous cell carcinoma from photographic images: A retrospective study," (in English), *EClinicalMedicine*, vol. 27, p. 100558, 2020.
- [23] R. A. Welikala et al., "Automated Detection and Classification of Oral Lesions Using Deep Learning for Early Detection of Oral Cancer," (in English), *IEEE Access*, vol. 8, pp. 132677-132693, 2020.
- [24] P. T. H. Nguyen, K. Sakamoto, and T. Ikeda, "Deep-learning application for identifying histological features of epithelial dysplasia of tongue," (in English), *Journal of Oral and Maxillofacial Surgery, Medicine, and Pathology*, vol. 34, no. 4, pp. 514-522, 2022.
- [25] T. G. Sun, L. Mao, Z. K. Chai, X. M. Shen, and Z. J. Sun, "Predicting the Proliferation of Tongue Cancer With Artificial Intelligence in Contrast-Enhanced CT," (in English), *Frontiers in Oncology*, vol. 12, p. 841262, 2022.
- [26] B. Ilhan, K. Lin, P. Guneri, and P. Wilder-Smith, "Improving Oral Cancer Outcomes with Imaging and Artificial Intelligence," (in English), *Journal of Dental Research*, vol. 99, no. 3, pp. 241-248, 2020.
- [27] J. Adeoye et al., "Deep Learning Predicts the Malignant-Transformation-Free Survival of Oral Potentially Malignant Disorders," (in English), *Cancers*, vol. 13, no. 23, p. 6054, 2021.
- [28] C. S. Chu, N. P. Lee, J. Adeoye, P. Thomson, and S. W. Choi, "Machine learning and treatment outcome prediction for oral cancer," (in English), *Journal of Oral Pathology & Medicine*, vol. 49, no. 10, pp. 977-985, 2020.
- [29] C. Bang, G. Bernard, W. T. Le, A. Lalonde, S. Kadoury, and H. Bahig, "Artificial intelligence to predict outcomes of head and neck radiotherapy," (in English), *Clinical and Translational Radiation Oncology*, vol. 39, p. 100590, 2023.
- [30] M. Revilla-León et al., "Artificial intelligence models for diagnosing gingivitis and periodontal disease: A systematic review," (in English), *Journal of Prosthetic Dentistry*, vol. 130, no. 6, pp. 816-824, 2023.