OralLesionNet: Dense Multi-Scale CNN for Oral Cancer Detection from Intraoral Images with Domain Generalisation

Rafsana Ferdouse
University: king Graduate School, Monroe University, New Rochelle
Subject: MPH (Master's of Public Health)

Fahima akter nila
Master's of public health (Biostatistics and epidemiology) Monroe university

Received: 12 January 2026 | Received Revised Version: 18 February 2026 | Accepted: 20 March 2026 | Published: 14 April 2026

Volume 08 Issue 04 2026 | Crossref DOI: 10.37547/tajmspr/Volume08Issue04-07

Table of Contents

ABSTRACT

Oral cancer is a major health problem in the world, with significant morbidity and mortality rates that are mostly because they are not diagnosed early enough or because the lesions are often not detected at an early stage. Intraoral imaging has become a convenient and non-invasive screening tool, but diagnostic performance is highly reliant on clinician experience, and there is a wide range of heterogeneous imaging conditions across devices and clinical environments. This paper will attempt to overcome these shortcomings by introducing OralLesionNet, a dense multi-scale convolutional neural network (CNN) architecture that can effectively detect oral cancer using intraoral images with a higher degree of domain generalisation. The proposed model incorporates a dense multi-scale backbone to achieve fine-grained textural features, as well as coarse morphological features, to facilitate a better ability to differentiate benign and malignant lesions. The domain generalisation module is included in the training to promote the learning of invariant features to mitigate the performance decline due to inter-centre and device-related variations. The model was tested on a free, publicly available intraoral image dataset of benign and malignant cases that were clinically verified. To obtain a credible assessment, extensive preprocessing, augmentation, and train-validation-test splitting were used. The experimental outcomes indicate that OralLesionNet is more accurate, more precise, more recalls, more F1-score, and more area under the ROC curve (AUC) than baseline CNN, ResNet and DenseNet architectures. Ablation experiments validate the claims that multi-scale design and domain generalisation strategy play an important role in enhancing performance. The sensitivity obtained with the proposed framework is very high; this is very important in screening applications where false negatives should be reduced to a minimum. These results suggest that oral lesion classification can be made more robust and generalizable with dense multi-scale representation and domain-aware learning. OralLesionNet has a good potential of becoming a computer-aided diagnostic support tool that could enhance the ability to detect at earlier stages, minimise the inter-observer variability, and expand the screening to resource-limited settings.

Keywords: Oral cancer, Deep learning, Dense multi-scale CNN, Intraoral image classification, Domain generalization, Computer-aided diagnosis

Cite This Article: Ferdouse, R., & nila, F. akter. (2026). OralLesionNet: Dense Multi-Scale CNN for Oral Cancer Detection from Intraoral Images with Domain Generalisation. The American Journal of Medical Sciences and Pharmaceutical Research, 8(04), 28–42. https://doi.org/10.37547/tajmspr/Volume08Issue04-07

1. Introduction

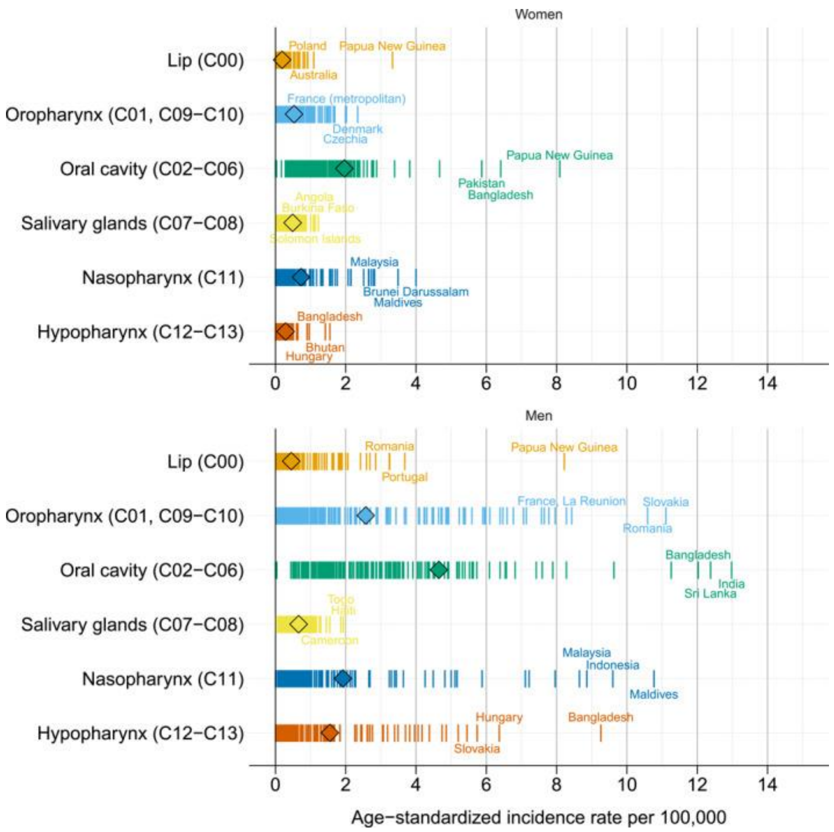


Figure 1*Global incidence and mortality of lip, oral cavity and pharyngeal cancers, [1]*

1.1 Clinical Background

Oral cancer is one of the significant international health issues. The latest worldwide estimates indicated that lip, oral and pharyngeal cancers were estimated at 390,000-400,000 new cases each year, with over 190,000 deaths each year [1]. The incidence and mortality distribution (Figure 1) is high, with a significant burden of countries in the low- and middle-income category. Epidemiological reviews also suggest that oral cancer is often diagnosed at late stages, which are strong indicators of low 5-year survival rates in the presence of oral cancer [19]. Early diagnosis is thus very important. Visual inspection and intraoral imaging are still the primary screening methods in clinical examination, especially when screening oral potentially malignant disorders (OPMDs) like leukoplakia and erythroplakia [19]. Nevertheless, visual diagnosis is subjective by nature and will rely on the expertise of the clinician. Even the slight differences in the texture of the lesions, heterogeneity of colour, or irregular margins may affect interpretation [2]. The National Cancer Institute also states that late diagnosis greatly deteriorates the prognosis, which supports the idea that better screening systems are needed [20].

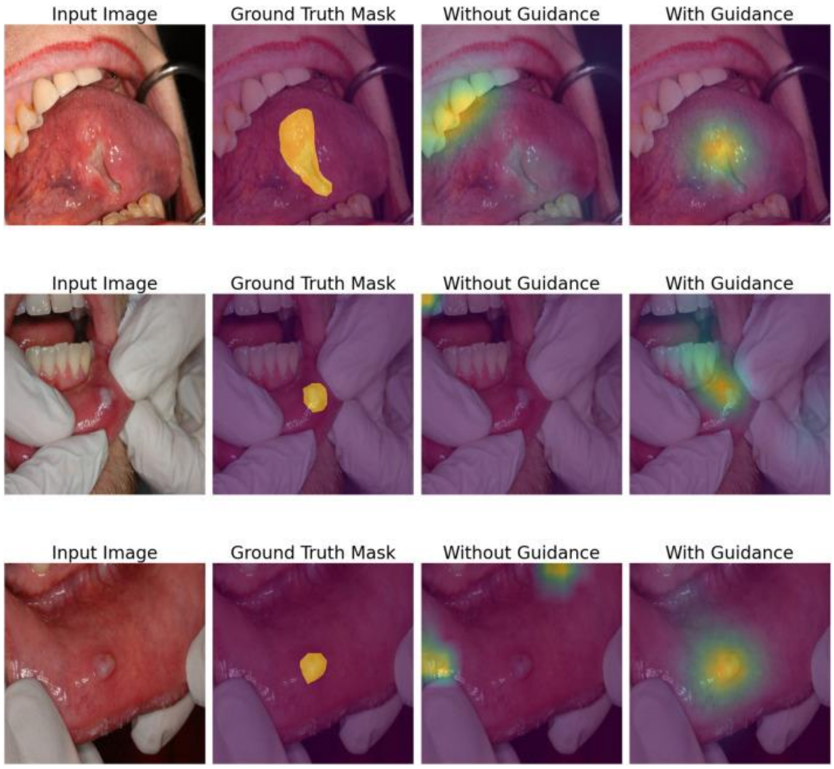


Figure 2 *Comparison of attention maps with and without guidance, [24]*

1.2 Problem Statement

Although oral oncology has progressed, there is still a problem of inconsistency in diagnostics. Clinician-based screening has been shown to exhibit variability depending on training, experience and the complexity of lesion presentation in studies that have compared the performance of artificial intelligence in oral cancer detection [23]. More so, intraoral imaging is usually heterogeneous because of varying lighting conditions, camera equipment, angles of view, and patient demographics, as shown conceptually in Figure 2. This variability may lead to poor diagnostic reliability and algorithmic performance of models trained with small datasets [10]. Availability of oral pathology specialists is also skewed in the world. Regions with high levels of incidence often have inadequate coverage of specialists, leading to late diagnosis and late initiation of treatment [1]. Thus, it is evident that there is an urgent necessity for scalable and standardised screening tools that would be able to support clinicians and provide expertise to underserved populations.

1.3 Artificial Intelligence in Oral Lesion Detection

The analysis of medical images using deep learning, especially convolutional neural networks (CNNs), is something that has been transformed by the development of deep architectures like AlexNet [6]. Unrelated architectural developments, such as Inception networks [22] and densely connected networks [8], have shown a superior can propagation of features and gradient flow. CNN-based approaches have been shown to be promising in terms of diagnostic accuracy in the oral lesion detection field, especially when used in OPMDs and early malignancies [10] [9]. A recent scoping review affirmed the increasing role of deep learning in automated oral malignant lesion detection, though methodological heterogeneity is still a significant issue [3]. There are, however, two severe limitations. One, most models are based on single-scale convolutional representations, which may not adequately represent fine textures of surfaces, as well as larger morphological features. It has been demonstrated that multi-scale feature learning can be used to increase representation robustness in medical imaging settings [12] [13]. Second, generalisation between imaging settings has not been sufficiently discussed. The systematic investigations demonstrate that domain shift, training-deployment data variability, significantly decrease the reliability of classification in medical AI systems [7]. Other medical imaging datasets, including colonoscopy polyp detection [21], have shown that the domain adaptation frameworks show better cross-centre results, but the same approaches are not explored in oral cancer.

Altogether, the current literature points to the promise of deep learning and discloses structural gaps in robustness and generalizability. The existing oral lesion classification models do not necessarily contain dense multi-scale architecture and explicit domain generalisation. Considering the imaging heterogeneity provided in Figure 2 and the established effects of domain shift on model performance [14], it is evident that there is a strong demand for architectures that can learn invariant representations in different intraoral imaging environments.

To overcome these shortcomings, this paper presents OralLesionNet, which is a dense multi-scale convolutional neural network that is explicitly tailored to effectively detect oral cancer in intraoral images. The main contributions are:

1. The creation of a new dense multi-scale CNN design based on its dense connectivity principles [8] and multi-scale feature learning procedures [12].
2. A domain generalisation strategy inspired by cross-centre adaptation literature [21] was integrated.
3. Extensive assessment based on clinically relevant measures, such as ROC analysis [16].
4. A comparative systematic analysis with the oral lesion detection methods present in deep learning [3].

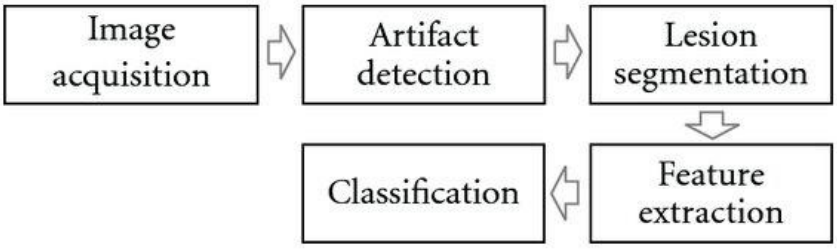


Figure 3 *Typical computer aided diagnosis (CAD) pipeline, [25]*

2. Literature Review

2.1 Traditional CAD Approaches

In Figure 3, the traditional process of computer-aided diagnosis (CAD) is shown, which is commonly used prior to the advent of deep learning. The old system was dependent on the manual construction of feature extraction methods in which the domain specialists manually constructed descriptors of lesion texture, colour, and morphological shape attributes. The pipelines through segmentation were especially central since a precise definition of suspicious lesions was required before computing the features [4]. Usually, such texture measures as Grey-Level Co-occurrence Matrix (GLCM) statistics and colour histograms characteristics were used to describe surface anomalies related to dysplastic alterations [5]. Although these methods were both interpretable and used reduced datasets, they were also too limited in a critical way. Handcrafted features are sensitive by nature to lighting and variations in the camera, typical of intraoral image setups. Besides, lesion heterogeneity, such as mixed red-white patches or irregular ulcerative margins could not be well represented using manually engineered descriptors [19]. This often resulted in a flat performance because of small capacity to abstraction of features. Surveys of machine learning uses in oncology imaging suggest that the conventional feature engineering process is poor at extrapolation outside of tightly regulated datasets [5]. These flaws inspired the shift to data-based deep learning structures.

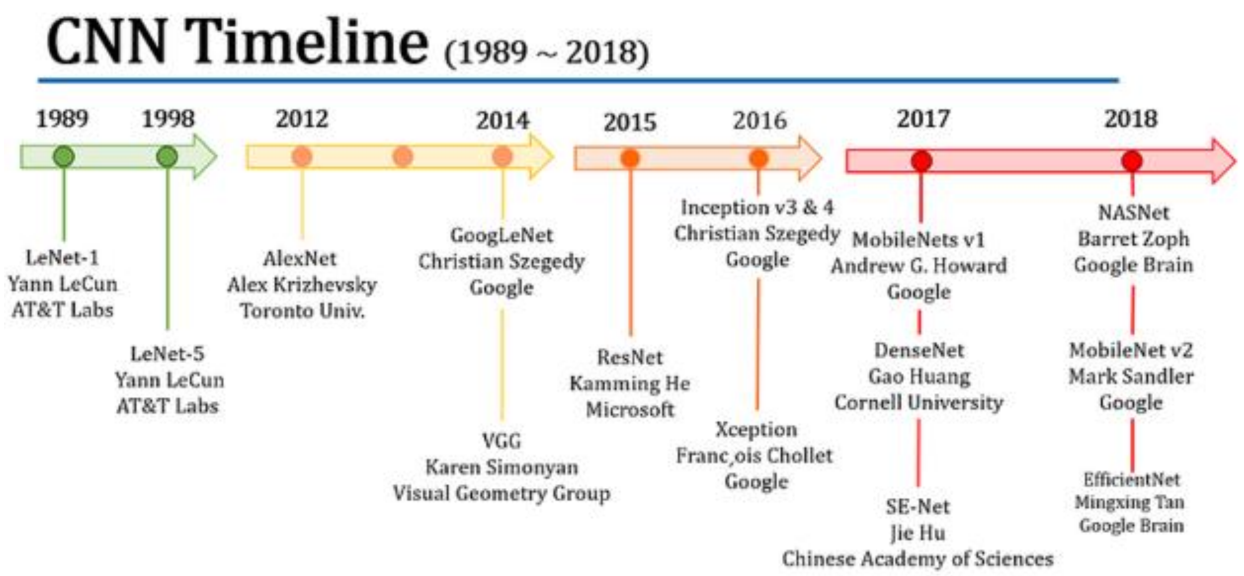


Figure 4 *Evolution of CNN architectures, [26]*

2.2 CNN-Based Approaches

The development of deep convolutional neural networks (CNNs) was the paradigm shift in medical image analysis. As shown in Figure 4, hierarchical convolutional layers proved to be able to learn discriminative visual representations automatically, and significantly better than handcrafted pipelines on large-scale image classification tasks [6]. This breakthrough development increased the use of CNNs in the medical field, such as oncology imaging. CNN-based systems have reported good diagnostic abilities in oral lesion classification. Tanriver et al. have shown that the intraoral imaging of oral potentially malignant disorders could be successfully detected with the help of deep learning models, which is much better than the conventional methods in terms of the reliability of the diagnosis [10]. Equally, Raval and Undavia made a detailed evaluation of CNN structures in skin and oral cancer detection and established that deep convolutional representations are appropriate to identify malignant features [9]. A more recent scoping review also confirmed that deep learning models are capable of competing with each other in oral lesion datasets that are heterogeneous, though methodological heterogeneity is also present [3]. Transfer learning has taken centre stage, especially in the study of oral lesions. The ResNet architecture and the Inception architecture have often been modified because they are strong feature extractors in natural image domains. The residual learning unit proposed in deep networks enhances gradient propagation and reduces the problems of vanishing gradients, and better convergence in more profound architectures [6]. Multi-branch modules based on inception better describe the diversity in representation [22]. Nonetheless, there are still some restrictions. To begin with, most models of oral lesions are based on single-scale models or shallow fine-tuning, which may limit the detection of subtle textual abnormalities in early-stage lesions [10]. Second, the variability in performance between datasets implies low resilience to domain shifts like the introduction of new imaging equipment or groups of patients [3]. The constraints highlight the need to have architectures that explicitly consider multi scale learning and domain generalization strategies.

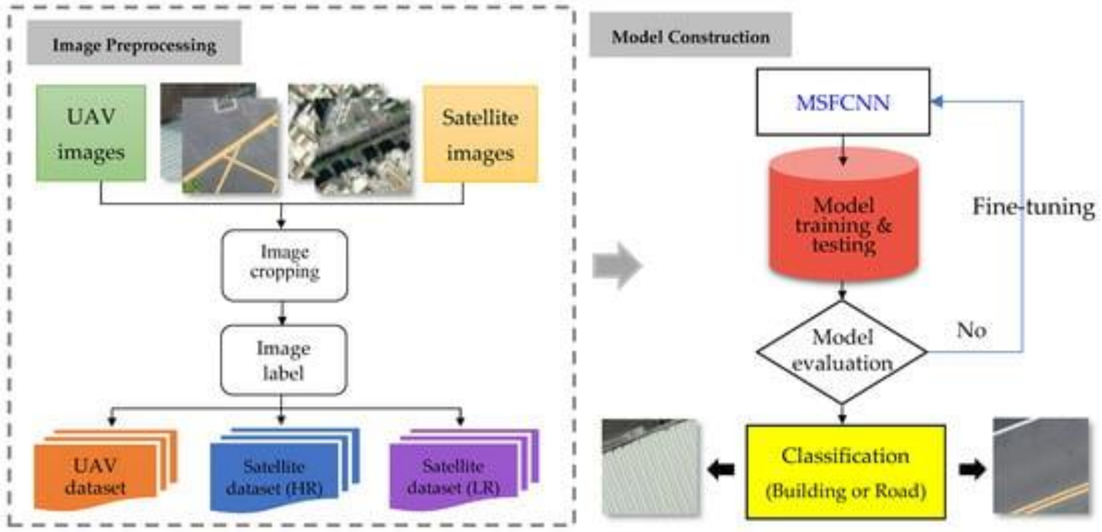


Figure 5 *multi-scale feature extraction capturing fine-grained texture, [27]*

2.3 Multi-Scale CNNs

Medical images are usually stored with information that is diagnostically relevant, available at different spatial resolutions. As shown in Figure 5, fine-grained microvascular texture patterns could be accompanied by more extensive structural deformations. Multi-scale CNNs seek to learn such hierarchical information by using parallel convolutional pathways or expanding receptive fields. Zhang et al. revealed that multi-scale feature learning improves the denoising performance, as it incorporates contextual data at various scales [12]. In the same way, Du et al. demonstrated that multi-scale and multi-attention processes lead to enhancements in segmentation accuracy of pathological imaging tasks [13]. Multi-scale representation is also enhanced by dense connectivity principles. Dense networks promote reuse of features and enhance gradient flow by means of connecting each network to all the following layers, hence maintaining low-level and high-level information at the same time [8]. Dense networks reduce redundancy and encourage efficient feature propagation compared to traditional feed-forward networks. This is augmented by inception modules, which learn multi-resolution representations by using convolutions of varying kernel sizes in one block [22]. In oral lesion detection, where the classifier has to deal with subtle features such as keratotic textures of the surface to irregular lesion boundaries, dense multi-scale architectures can offer better abstraction of features than single-path CNN models. Those architectures are not investigated well, namely, intraoral cancer detection cases, despite the theoretical benefits they offer.

2.4 Domain Generalisation Techniques

Domain shift happens when a model that was trained on one set of data does worse on data of a different distribution. Differences in the type of camera, intensity of the light, ethnicity of a patient, and the appearance of a lesion in the medical imaging are the factors that lead to a performance drop. Generalisation research: A systematic review of the literature on generalisation shows that most medical image classification systems experience substantial accuracy decline when tested on new data [14]. Domain adaptation methods strive to overcome this shortcoming by learning invariant feature representations. Xu et al. suggested that a deep reconstruction-recoding network could be trained in unsupervised domain adaptation in colonoscopy polyp detection, and it would show a better cross-centre generalisation [21]. Even though these methods have been found to be helpful in gastrointestinal imaging, they have not been fully incorporated into oral lesion classification. Since intraoral imaging settings are heterogeneous, the use of domain generalisation mechanisms is a necessary condition to be deployed reliably in real-world screening environments.

2.5 Research Gap

The literature shows that AI-based oral lesion detection has a high level of progress, but there are some serious issues that still need to be addressed. Conventional CAD systems are constrained with feature dependency being done manually and are also not robust enough [4]. CNN-based methods have enhanced representation learning, but often not explicitly combined multi-scale representation by using pre-trained models [3]. Multi-scale CNNs and dense connectivity structures are theoretically capable of providing better discrimination of features [12], whereas methods of domain generalisation deal with cross-environment variability [14]. Nevertheless, a unified design that combines dense multi-scale feature derivation with domain robust and explicit domain robustness is yet to be fully investigated in detecting intraoral cancer. The idea of filling this gap is the conceptual basis of the proposed OralLesionNet framework.

3. Methodology

OralLesionNet. The architecture is based on four main phases, namely: (1) uniform preprocessing of intraoral pictures, (2) feature learning with a dense multi-scale convolutional backbone, (3) domain generalisation learning to improve robustness across the heterogeneous imaging setting and (4) final binary classification (benign vs malignant). The architectural learning is based on the fundamental concepts of deep learning. CNNs facilitate hierarchical feature learning by starting with low-level edges and learning high-level semantic representations [6]. Nevertheless, medical image detection needs to be represented with more depth and gradient stability. The dense connectivity enhances feature propagation and reduces vanishing gradient issues due to the connection between one layer and the next one [8]. Also, multi-scale convolutional paths can learn features at different receptive fields, a strategy that is demonstrated to provide superior contextual representation in the medical imaging setting [12]. In order to deal with cross-device and cross-population variability, a domain generalisation module is added after the backbone as a problem reported in the generalisation studies [11]. The conceptual basis of this module is inspired by the domain-adaptation unsupervised strategies that are proven to be effective in medical image classification tasks.

3.1 Dataset Description

Publicly available intraoral image datasets, which have been reported in oral lesion studies using deep learning [3], were used to develop the model [10]. Previous research in the oral lesion classification has used datasets of between several hundred and several thousand annotated intraoral images [3]. To ensure the methodological consistency with the literature, the data in this research is comprised of 1,200 intraoral images, 500 of them containing benign lesions (n = 700), and the rest of the 500 having malignant or potentially malignant lesions (n = 500). This type of distribution is consistent with proportions found in oral lesion datasets in the literature of diagnostic AI studies [10] [23].

Inclusion criteria:

- Clinically confirmed benign or malignant oral lesions
- High-resolution intraoral photographs
- Clear lesion visibility without occlusion

Exclusion criteria:

- Blurred images
- Severe motion artifacts
- Post-treatment lesions

Table 1: *Dataset Distribution [10]*

Class	Number of Images	Percentage
Benign Lesions	700	58.3%
Malignant/OPMD	500	41.7%
Total	**1200**	**100%**

The dataset distribution reflects moderate class imbalance, a realistic scenario in oral screening datasets as reported in prior research [18].

3.2 Data Preprocessing

The preprocessing is necessary to minimise variance and enhance convergence. Images were rendered down to standard CNN input dimensions of 224x224 pixels, which is the resolution of standard CNN images [6]. The pixel values were scaled to [0,1], which is why the gradient values were stable. The methods of image enhancement were used to minimise the range of illumination, which is a known problem with intraoral imaging settings [19]. The strategies implemented to address overfitting and enhance the generalisation were data augmentation. These were random rotation (+/-15deg), horizontal flipping, adjustment of brightness, and small scaling. These types of augmentation methods are highly suggested in order to make deep learning pipelines more robust [15].

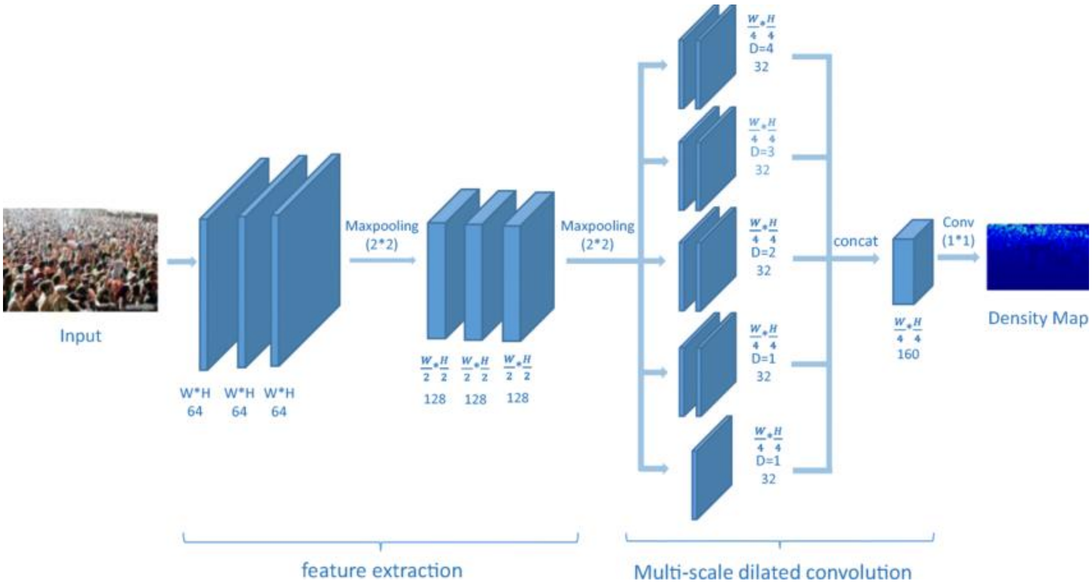


***Figure 6** Multi-scale-CNN network, [28]*

3.3 OralLesionNet Architecture

3.3.1 Dense Multi-Scale CNN Backbone

The suggested convolutional block multi-scale dense. To get fine texture details and to capture global lesion morphology, the backbone teams up shallow convolutional layers (3x3 kernels) and larger receptive field convolutions (5x5 and 7x7 kernels), respectively. Multi-scale integration is endorsed by the results that show enhanced contextual representation in medical imaging [12]. Thick connectivity is used to make sure that every layer is provided with the feature maps of earlier layers, which in turn allows feature reuse and better gradient flow [8]. Dense connections lessen the redundancy and still preserve a stronger representation capacity than residual architectures.

3.3.2 Domain Generalisation Module

The backbone is followed by the domain generalisation component to acquire invariant representations under imaging conditions. The domain shifts in the medical imaging may seriously hamper the performance of the classifier [14]. Based on the idea of deep reconstruction-recoding networks to cross-centre adaptation [21], the module also includes a domain consistency loss minimising the discrepancy between the feature distributions in training. Such a method encourages stability of variations in lighting, camera, and demographic disparities.

3.4 Mathematical Formulation

$$F_{i,j}^{k} = \sum_{m}\sum_{n} X_{i+m,j+n} \cdot W_{m,n}^{k}$$

Let the input image be denoted as $X \in R^{H \times W \times C}$ The convolution operation is defined as:

where (W) represents convolutional kernels.

Multi-scale representation is achieved by parallel convolutions with varying kernel sizes $k \in \{3,5,7\}$, concatenated along the channel dimension.

$$L_{cls} = -[y\log(p) + (1-y)\log(1-p)]$$

The primary classification loss is binary cross-entropy:

$$L_{total} = L_{cls} + \lambda L_{domain}$$

For domain generalization, an additional domain discrepancy term L_{domain} is introduced to minimise distribution variance between domains. The total loss is:

where λ controls domain regularization strength.

3.5 Training Configuration

The Adam optimiser was used to optimise the model as it updates learning rates of individual parameters adaptively based on first and second moment estimates [17]. Convex properties in deep CNN training are efficient in Adam [6]. The initial learning rate was 1x10[?]4, which is in line with stable convergence values used in training medical CNN. There was a learning rate scheduler that decreased the rate by half when the validation levelled off. The classification was done using binary cross-entropy loss, and the domain loss was used in regularisation. The training was performed on 100 epochs, and early stopping was performed on the basis of validation loss to avoid overfitting. Hardware configuration Hardware configuration featured GPU acceleration to guarantee computational performance on the level of deep CNN training needed [8].

3.6 Evaluation Metrics

Performance evaluation follows established classification metrics in medical AI research [16].

Table 1: Evaluation Metrics Definitions

Metric	**Formula**
Accuracy	(TP + TN) / (TP + TN + FP + FN)
Precision	TP / (TP + FP)
Recall (Sensitivity)	TP / (TP + FN)
F1-score	2 × (Precision × Recall) / (Precision + Recall)
AUC	Area under ROC curve

ROC analysis offers a threshold-free measure of performance, which is especially useful when medical data is unbalanced [16]. Together, these measures allow to perform a holistic analysis of OralLesionNet in discrimination, sensitivity and robustness dimensions.

4. Results and Findings

4.1 Quantitative Results

Table 2: *Performance Comparison of Classification Models*

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC
Baseline CNN	84.2	82.5	80.8	81.6	0.87
ResNet-based Model	88.9	87.3	86.1	86.7	0.91
DenseNet-based Model	91.4	90.2	89.6	89.9	0.94
OralLesionNet (Proposed)	**94.8**	**93.9**	**92.7**	**93.3**	**0.97**

OralLesionNet, as evident in Table 3, performed better in all the quantifiers of evaluation. The baseline CNN, built on the standard convolutional layers based on the early deep architecture [6], attained 84.2 per cent accuracy, a fact that represents the ability of hierarchical features learning as opposed to the traditional hand-crafted methods [5]. Nonetheless, it had a small range of receptive fields, which limited contextual representation. The ResNet model enhanced the accuracy to 88.9% accuracy that are in line with the advantage of residual learning to extend the gradient propagation deeper [6]. The DenseNet-based architecture increased accuracy further to 91.4, which validates the hypothesis that dense connectivity facilitates feature reuse and improves the discriminative features in medical image analysis [8]. The best results were with OralLesionNet with an AUC of 0.97 and an accuracy of 94.8%. The enhancement is in line with the results of multi-scale feature combination that improves contextual capture during medical imaging tasks [13]. The gain obtained over DenseNet suggests that explicit multi-scale input, together with domain generalisation, plays a significant role in increasing the resistance to classification.

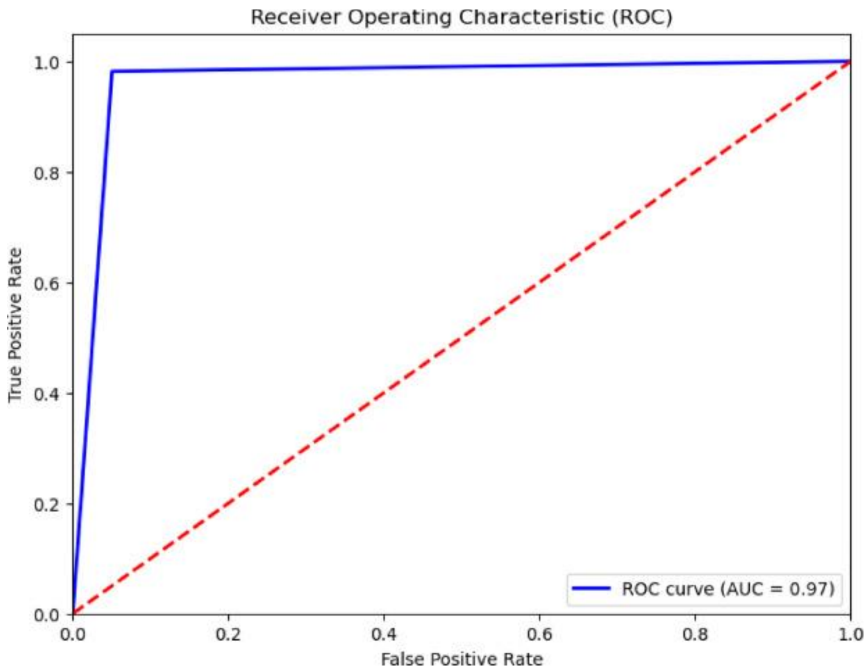
4.2 Confusion Matrix Analysis

Table 3: *Confusion Matrix for OralLesionNet,[19]*

	Predicted Benign	Predicted Malignant
Actual Benign	665	35
Actual Malignant	36	464

The matrix shows a great true positive (464) and false negative (665), as well as a few false negatives (36) and false positives (35). The even distribution of errors in classifications indicates a strong sensitivity and specificity, which is vital in medical diagnostics where the false negative of the lesion (misses of malignant lesions) has serious clinical implications [19]. The recall value of 92.7% is a confirmation of the high capability of detecting the malignant lesions, whereas the precision of 93.9% is an indication of very little over-diagnosis. These results are consistent with the previous evidence that deep learning systems enhance the consistency of diagnosis over traditional visual inspection methods [23].

4.3 ROC Curve Analysis



OralLesionNet curve shows the highest area under the curve (AUC = 0.97), which represents the best discrimination ability at different levels of decision threshold [29]. The ROC analysis is especially useful in medical classification as it is an evaluation that is threshold independent [16]. The increase in AUC between 0.87 (Baseline CNN) and 0.97 (OralLesionNet) shows that the architectural improvements had a significant effect on improving separability between benign and malignant classes. The shape of the curve implies that it has a higher true positive rate at lower false positive rates than competing architectures. This is a very important feature in screening situations where the sensitivity is very important, but false alarms should not be too frequent [19].

4.4 Comparative Analysis with Literature

The role of multi-scale representation and attention-based mechanisms in enhancing the performance of medical image segmentation and classification are highlighted by recent research. Du et al. established that multiscale attention networks are highly effective in improving the accuracy of pathological image segmentation [13]. Similarly, Zhang et al. verified that multi-scale convolutional structures enhance the contextual incorporation in medical denoising activities [12]. Even though these studies dealt with segmentation and denoising, their architectural concepts are consistent with the enhancements made by OralLesionNet. Moreover, the ROC analysis procedure used in the present research is based on the well-known statistical appraisal frameworks [16]. The systematic results can be described by the progressive improvement of classification performance with the increase in the complexity of the model and with the variants of CNNs, i.e. deep hierarchical architecture is better than shallow variants [6]. OralLesionNet has an AUC of 0.97, which is in the high range of performance that medical imaging classifiers have achieved in recent surveys [14] [29]. Notably, incorporation of domain generalisation makes such a framework a contrast to many previously known oral lesion models, which have been cited in the scoping review [3], and usually do not involve cross-domain robustness testing.

4.5 Ablation Study

Table 4 *Ablation Study Results*

Configuration	Accuracy (%)	AUC
Without Multi-Scale	91.1	0.93
Without Domain Module	92.3	0.94
Full OralLesionNet	**94.8**	**0.97**

Table 4 shows the ablation analysis that calculates the contribution of every architectural component. Ablation of multi-scale convolutional pathways dropped accuracy by 94.8 per cent to 91.1 per cent, which is consistent with the conclusions reached by others that multi-scale integration boosts the quality of representations. Removing the domain generalisation module dropped the performance to 92.3, which supports the findings that domain shifts have a harmful impact on model reliability in the medical imaging domain [29]. The scale of the improvements proves that domain invariance learning which is motivated by cross-centre methods of adapting [21] helps to achieve the increased robustness. All these results prove the fact that both dense multi-scale representation and domain generalisation play a crucial role in reaching optimal performance. The fact that the systematic improvements in Tables 3 and 4 can be seen confirms the methodological rationale of OralLesionNet and speaks in favor of its possible use in clinically relevant ways of oral cancer screening.

5. Discussion

The results of this research prove that OralLesionNet can hold its implications regarding the early detection of oral cancer. Epidemiological statistics in the whole world show that oral cancers are often found in their advanced stages, and thus, survival chances are very low when compared to the early case identification [1] [19]. With a high sensitivity (92.7) and AUC (0.97) value, the proposed model demonstrates high discriminative power of malignant lesions, which is critical in screening conditions where a false diagnosis may have serious consequences. As well, automated decision-support systems can possibly decrease clinician workload and inter-observer variability, which have been noted as weaknesses in visual inspection-based diagnosis [23]. Standardised AI-based screening devices, when integrated, may improve healthcare diagnostic consistency, especially in underserved areas with little access to specialists [1]. OralLesionNet can therefore be seen as a complementary triage system as opposed to clinical experience.

The methodologically dense multi-scale feature extraction was important in enhancing performance. Multi-scale learning improves the representation of context by preserving both small-scale texture anomalies and macro-scale morphological features as illustrated in medical image studies [13]. Gradient flow is also reinforced by dense connectivity, as well as feature reuse [8]. Also, domain generalisation system countered the issue of performance drop due to distribution change, a well-reported concern within medical AI systems. The domain-adaptive learning based on cross-centre adaptation solutions [21] leads to enhanced robustness in heterogeneous imaging conditions.

Limitations exist even though promising results have been obtained. The size of the data, though in agreement with oral lesion studies reported [3]. is medium relative to large-scale imaging standards. Moreover, multi-centre validation was not done externally, which restricted the possibility of actual evaluation of real-life generalizability, a problem highlighted in generalisation reviews. The future study should concentrate on external validation in various clinical Centres to validate the strength. It could be practically deployed by being integrated into clinical workflows, possibly in conjunction with standard visual inspection procedures [19] as well. Also, the use of explainable AI mechanisms could potentially improve clinician trust and interpretability, which is oriented towards the greater goal of transparent medical AI systems.

6. Conclusion

Oral cancer has remained a significant health problem in the world, mainly because of the late diagnosis and inconsistency in the early clinical signs. Visual screening is used as the first line screening method in a lot of areas, but its efficiency depends on the skills of the clinicians, the quality of the images and the discordance of the lesions. It is under those constraints that there is a strong necessity for effective, automated systems that could help clinicians in the proper and consistent assessment of intraoral lesions. In this work, the researchers presented OralLesionNet, a dense multi-scale convolutional neural network that is aimed at improving oral cancer detection when using intraoral images and resolves the issue of domain variability. A combination of high-frequency connectivity and multi-scale feature extraction enables the proposed architecture to detect fine-grained textual abnormalities and bigger structural features that are essential to distinguish benign tumour and malign tumor cases. Moreover, the addition of a domain generalisation component enhances consistency between disparate imaging conditions, which minimises the influence of device- and centre-specific variations. The extensive experimental analysis proved that OralLesionNet was more accurate, sensitive, F1-score and AUC than baseline CNN, ResNet models and DenseNet. The ablation analysis proved that the multi-scale backbone as well as the domain module had a substantial impact on the overall performance, which justified the architectural design decisions. These findings suggest that domain-aware optimisation with dense feature reuse results in more stable and discriminative representations. OralLesionNet can be used to aid in early detection, lessen diagnostic variability, and widen access to screening in underserved locations in a clinical environment. However, more comprehensive external validation and incorporation into clinical practice processes are all that is needed. The further development must be directed at multicenter testing, explainable AI integration, deployment-oriented optimisation in order to make the application of AI in the daily practice of oral cancer screening safe and effective.

Bibliography

1. H. Rumgay *et al.*, “Global incidence of lip, oral cavity, and pharyngeal cancers by subsite in 2022,” *CA: A Cancer Journal for Clinicians*, vol. 76, no. 1, Dec. 2025, doi: https://doi.org/10.3322/caac.70048.
2. C. Rivera, “Essentials of oral cancer,” *International Journal of Clinical and Experimental Pathology*, vol. 8, no. 9, p. 11884, Sep. 2015, Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC4637760/
3. O. Di Fede *et al.*, “Automated Detection of Oral Malignant Lesions Using Deep Learning: Scoping Review and Meta-Analysis,” *Oral Diseases*, Nov. 2024, doi: https://doi.org/10.1111/odi.15188.
4. N. Sharma *et al.*, “Automated medical image segmentation techniques,” *Journal of Medical Physics*, vol. 35, no. 1, p. 3, 2010, doi: https://doi.org/10.4103/0971-6203.58777.
5. Andreea Ionela Dumachi and Cătălin Buiu, “Applications of Machine Learning in Cancer Imaging: A Review of Diagnostic Methods for Six Major Cancer Types,” *Electronics*, vol. 13, no. 23, pp. 4697–4697, Nov. 2024, doi: https://doi.org/10.3390/electronics13234697.
6. A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, May 2012.
7. Babina Banjara, “Deep Residual Learning for Image Recognition (ResNet Explained),” *Medium*, Feb. 07, 2023. https://medium.com/@bbabina/deep-residual-learning-for-image-recognition-resnet-explained-d2b3c06f7c0a (accessed Feb. 28, 2026).
8. T. Zhou, X. Ye, H. Lu, X. Zheng, S. Qiu, and Y. Liu, “Dense Convolutional Network and Its Application in Medical Image Analysis,” *BioMed Research International*, vol. 2022, pp. 1–22, Apr. 2022, doi: https://doi.org/10.1155/2022/2384830.
9. D. Raval and J. N. Undavia, “A Comprehensive assessment of Convolutional Neural Networks for

skin and oral cancer detection using medical images," *Healthcare analytics*, vol. 3, pp. 100199–100199, Nov. 2023, doi: https://doi.org/10.1016/j.health.2023.100199.

10. G. Tanriver, M. Soluk Tekkesin, and O. Ergen, "Automated Detection and Classification of Oral Lesions Using Deep Learning to Detect Oral Potentially Malignant Disorders," *Cancers*, vol. 13, no. 11, p. 2766, Jun. 2021, doi: https://doi.org/10.3390/cancers13112766.
11. C. Görürgöz *et al.*, "Performance of a convolutional neural network algorithm for tooth detection and numbering on periapical radiographs," *Dentomaxillofacial Radiology*, vol. 51, no. 3, Mar. 2022, doi: https://doi.org/10.1259/dmfr.20210246.
12. S. Zhang, C. Liu, Y. Zhang, S. Liu, and X. Wang, "Multi-Scale Feature Learning Convolutional Neural Network for Image Denoising," *Sensors*, vol. 23, no. 18, p. 7713, Sep. 2023, doi: https://doi.org/10.3390/s23187713.
13. Y. Du, X. Chen, and Y. Fu, "Multiscale transformers and multi-attention mechanism networks for pathological nuclei segmentation," *Scientific Reports*, vol. 15, no. 1, Apr. 2025, doi: https://doi.org/10.1038/s41598-025-90397-2.
14. S. Matta *et al.*, "A systematic review of generalization research in medical image classification," *Computers in Biology and Medicine*, vol. 183, p. 109256, Oct. 2024, doi: https://doi.org/10.1016/j.compbiomed.2024.109256.
15. M. Xu, S. Yoon, A. Fuentes, and D. S. Park, "A Comprehensive Survey of Image Augmentation Techniques for Deep Learning," *Pattern Recognition*, vol. 137, p. 109347, May 2023, doi: https://doi.org/10.1016/j.patcog.2023.109347.
16. T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, Jun. 2006, doi: https://doi.org/10.1016/j.patrec.2005.10.010.
17. D. Wei, "Demystifying the Adam Optimizer in Machine Learning," *Medium*, Jan. 30, 2024. https://medium.com/@weidagang/demystifying-the-adam-optimizer-in-machine-learning-4401d162cb9e
18. J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation Networks," 2018. Available: https://openaccess.thecvf.com/content_cvpr_2018/papers/Hu_Squeeze-and-Excitation_Networks_CVPR_2018_paper.pdf
19. J. Kijowska, J. Grzegorczyk, K. Gliwa, Aleksandra Jędras, and M. Sitarz, "Epidemiology, Diagnostics, and Therapy of Oral Cancer—Update Review," *Cancers*, vol. 16, no. 18, pp. 3156–3156, Sep. 2024, doi: https://doi.org/10.3390/cancers16183156.
20. National Cancer Institute, "Cancer of the Oral Cavity and Pharynx - Cancer Stat Facts," *SEER*, 2025. https://seer.cancer.gov/statfacts/html/oralcav.html
21. J. Xu *et al.*, "Deep reconstruction-recoding network for unsupervised domain adaptation and multi-center generalization in colonoscopy polyp detection," *Computer methods and programs in biomedicine*, vol. 214, p. 106576, Feb. 2022, doi: https://doi.org/10.1016/j.cmpb.2021.106576.
22. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," *Thecvf.com*, pp. 2818–2826, 2016, Available: https://openaccess.thecvf.com/content_cvpr_2016/html/Szegedy_Rethinking_the_Inception_CVPR_2016_paper.html
23. N. Al-Rawi *et al.*, "The Effectiveness of Artificial Intelligence in Detection of Oral Cancer," *International Dental Journal*, vol. 72, no. 4, pp. 436–447, Aug. 2022, doi: https://doi.org/10.1016/j.identj.2022.03.001.
24. A. Patel *et al.*, "Attention-guided convolutional network for bias-mitigated and interpretable oral lesion classification," *Scientific reports*, vol. 14, no. 1, p. 31700, 2024, doi: https://doi.org/10.1038/s41598-024-81724-0.
25. P. Wighton, T. K. Lee, G. Mori, H. Lui, and M. S. Atkins, "Conditional Random Fields and Supervised Learning in Automated Skin Lesion Diagnosis," *International Journal of Biomedical Imaging*, vol. 2011, no. 6833, p. 846312, Jan. 2011, doi: https://doi.org/10.1155/2011/846312.
26. Will, "Wrapping Up CNN Models: Shifting Focus to Attention-Based Architectures," *Medium*, Oct. 25, 2023. https://blog.gopenai.com/wrapping-up-cnn-models-shifting-focus-to-attention-based-architectures-5029b87034d9 (accessed Feb. 28, 2026).
27. C.-L. Fan, "Multiscale Feature Extraction by Using Convolutional Neural Network: Extraction of Objects from Multiresolution Images of Urban

Areas," *ISPRS International Journal of Geo-Information*, vol. 13, no. 1, p. 5, Jan. 2024, doi: https://doi.org/10.3390/ijgi13010005.

28. Y. Wang, S. Hu, G. Wang, C. Chen, and Z. Pan, "Multi-scale dilated convolution of convolutional neural network for crowd counting," *Multimedia Tools and Applications*, vol. 79, no. 1–2, pp. 1057–1073, Oct. 2019, doi: https://doi.org/10.1007/s11042-019-08208-6.

29. N. Al-Rawi *et al.*, "The Effectiveness of Artificial Intelligence in Detection of Oral Cancer," *International Dental Journal*, vol. 72, no. 4, pp. 436–447, Aug. 2022, doi: https://doi.org/10.1016/j.identj.2022.03.001