STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **x**, Month 202x, pp 1–17.
Published online in International Academic Press (www.IAPress.org)

A Multimodal Deep Learning Framework for Oral Cancer Prediction Using DenseNet-169 and Clinical Data Fusion

Nissreen A. El-Saber [1,2], Mahmoud I. Gaber [3], Hasnaa R. Hamed [1], Mahmoud A. Zaher [4]

[1] *Information Systems Department, Faculty of Computers and Informatics, Zagazig University, Egypt*
[2] *Software Engineering Department, Computer Science College, MIU, Egypt*
[3] *Data Science Department, Faculty of Artificial Intelligence, Egyptian Russian University (ERU), Egypt*
[4] *Data Science Department, Faculty of Artificial Intelligence and Information, Horus University (HUE), Egypt*

Abstract Oral squamous cell carcinoma (OSCC) and leukoplakia with or without dysplasia require careful histopathological assessment. We examine a DenseNet-169-based multimodal framework combining H&E tissue images with clinical metadata for three-class oral lesion classification, emphasizing patient-independent validation rather than architectural novelty. The NDB-UFES cohort contains 237 paired images from 69 patients. We partitioned 55 patients (189 images) for development and 14 disjoint patients (48 images) for a locked internal test. Patient-grouped five-fold cross-validation supported image-only, clinical-only, fusion ablations, and frozen ImageNet backbone comparisons (DenseNet-169, ResNet-50, EfficientNet-B0). Internal validation Grad-CAM visualizations and clinical permutation sensitivity were exploratory. In controlled development ablations, the DenseNet-169 image-only branch attained mean patient-level accuracy 67.3% and macro-F1 0.612, versus 60.0% and 0.488 for fusion with a reducer. On the prespecified independent-patient internal test, the multimodal fusion model obtained patient-level accuracy 64.29% (9/14; exact 95% CI 35.14–87.24%), macro-F1 0.600 and macro one-vs-rest AUC 0.860; the image-only comparator obtained 57.14% (8/14), macro-F1 0.470 and AUC 0.756. Exact paired McNemar p=1.000 did not establish superiority. Fusion is feasible, but benefits remain uncertain owing to the small cohort, minority-class errors and absent external validation. The original submission's 87.18% image-level result arose from a patient-overlapping split and is not presented as an independent-patient estimate.

Keywords oral squamous cell carcinoma; oral leukoplakia; DenseNet-169; histopathology; patient-disjoint validation; clinical data fusion.

DOI: 10.19139/soic-2310-5070-4746

1. Introduction

Oral squamous cell carcinoma (OSCC) is a major disease burden, and accurate characterization of malignant and potentially malignant oral lesions depends on histopathological assessment and clinical context. Computer-assisted image analysis is being investigated as an adjunct to expert judgment rather than a replacement for clinical diagnosis [1, 4, 9, 24].

Oral cancer research spans several distinct endpoints: diagnosis of tissue lesions, prediction of recurrence or survival, and identification of prognostic features. These endpoints should not be conflated when comparing model results. In this study the outcome is a three-class tissue-based diagnosis: leukoplakia with dysplasia, leukoplakia without dysplasia, or OSCC.

Structured-data algorithms can model patient-level factors, while convolutional neural networks capture morphological information from H&E images. In a paired dataset, simple feature concatenation offers a transparent multimodal reference model, although small samples and patient-correlated images demand strict grouping during evaluation.

Recent reviews of AI in oral cancer discuss promising imaging and prognostic applications alongside selection bias, dataset heterogeneity, explainability, and external validation challenges [12, 13, 21, 23]. Reported scores from different endpoints and cohorts must therefore be treated as contextual literature, not same-condition benchmarks.

ISSN 2310-5070 (online) ISSN 2311-004X (print)
Copyright © 202x International Academic Press

Earlier evaluations in this project emphasized image-level performance but did not establish independence between training and test patients. Re-examination of source identifiers revealed repeated images of the same patients across those subsets. The revised study therefore emphasizes group-aware evaluation and explicitly quantifies the impact of adding clinical variables.

Our application-level contribution is a reproducible assessment of a conventional DenseNet-169/MLP fusion architecture for paired NDB-UFES oral lesion records. We provide (i) patient-disjoint holdout selection and five-fold development cross-validation, (ii) controlled modality and fusion-component ablations, (iii) same-fold frozen-backbone comparisons, and (iv) exploratory interpretability and paired test uncertainty. We do not claim a novel fusion operator or clinical deployment readiness.

The key research question is whether limited clinical metadata improves the patient-level discrimination of three oral lesion categories over a strong image-only encoder. We evaluate that question with patient-independent comparisons, acknowledging that a beneficial effect of fusion cannot be assumed a priori.

The paper is organized as follows: Section 2 critically reviews related work; Section 3 describes source data and patient grouping; Section 4 details the two-stream architecture and the reproducible experimental protocol; Section 5 reports grouped development and locked internal-test results; and Section 6 discusses limitations and future directions.

2. Related Work

Relevant studies fall into three strands: clinical predictors of prognosis/recurrence [5, 22], deep visual analysis of oral or general pathology images [4, 9, 11, 16], and methods combining data sources or reviewing multimodal clinical AI [12, 13, 19, 23, 25]. The endpoints, sample units, imaging modalities and validation protocols differ substantially. Consequently, published accuracy and AUC values are not directly comparable to the present three-class, paired-sample experiment.

Wan-Ting Tseng et al. [1] adopted clinical data from a Southern Taiwanese oncology center, having a total of 673 cases after discarding cases with incomplete data. In their analysis, the decision tree (DT) and artificial neural network (ANN) models were found superior to conventional statistical methods for oral cancer diagnosis.

Rasheed Omobolaji Alabi et al. [2] analyzed 311 cases using predictors such as age, gender, tumor stage (AJCC 7th edition), and WHO grade. The postoperative tissue samples were examined histologically. ANN model was able to achieve 92.7% accuracy as opposed to logistic regression (86.5%). The web-based predictive model had 88.2% accuracy with sensitivity of 71.2% and specificity of 98.9%.

A retrospective analysis on 467 patients with OSCC was performed by Chui S. Chu et al. [3]. ML algorithms, such as linear regression, SVM, DTs, and KNN were assessed. It was the DT model that used 34 prognostic variables and showed the highest accuracy of 70.59% (AUC = 0.67). In their study, Roshan Alex Welikal et al. [4] used a Faster R-CNN model, trained on ResNet-101, to analyze more than 2,000 oral lesion images. The model demonstrated an excellent capability to classify lesions as measured by F1 score (87.07%).

According to Alkhadar et al. [5], analysis of data of 416 OSCC patients through five machine learning models revealed the DT algorithm as having the highest accuracy (76%), followed by logistic regression (60%). Unfortunately, there was no external validation. Kantharimuthu M. et al. [6] developed an early prediction model for OSCC that uses probabilistic neural networks (PNN). The accuracy of the model reached 80%, but failed to localize cancerous regions precisely.

Nanditha and Annegowda [14] discussed oral cancer detection using machine and deep learning methods. Their work is distinct from the global epidemiological review by Salehiniya and Raei [7]; the earlier text inadvertently assigned Nanditha to reference [7]. Chen et al. [8] investigated neural network-supported risk assessment of potentially malignant lesions, addressing a different outcome from the present histopathology three-class endpoint.

M. A. Rahman et al. [9] implemented AlexNet along with transfer learning to differentiate between cancerous and normal biopsy images. They attained training accuracy of 97.66%, but only 90.06% accuracy in the test data, while overfitting and imbalanced data problem was observed. Heli Shah et al. [10] utilized one-way ANOVA as a feature selection method and trained multiple machine learning models using 10-fold cross-validation approach.

The best results were obtained by the use of hybrid models and polynomial regression, indicating that advanced feature engineering can play an important role in model development.

Ibrahim Abdulrab Ahmed et al. [11] used OSCC-400x images and combined GoogLeNet, ResNet-101, VGG16, artificial neural network and XGBoost via ARG hashing technique. The hybrid model attained 99.3% accuracy, 98.2% sensitivity and 98.85% AUC score, making it one of the best-performing models in the area. Vinay et al. [12] presented a comprehensive literature review on AI techniques for oral cancer. The authors paid particular attention to CNN algorithms and discussed ethical issues associated with their clinical application. Dixit et al. [13] concluded that oral cancer is one of the biggest health concerns worldwide, with very high prevalence in developing countries.

Prior literature reviews cover oral cancer applications and broader deep-learning limitations, including small data, imbalance and risks of optimistic evaluation [12, 13, 15]. These concerns motivate patient-disjoint rather than solely image-random validation. Results from separate cohorts should not be interpreted as evidence that one particular architecture outperforms another under identical conditions.

Dong et al. [16] studied multi-view pathology learning. Yang et al. [17] reviewed medical image super-resolution, while Parisi and Manaog [18] addressed interpretable learning in several domains. Their contributions concern different technical questions and cannot substitute for direct, controlled oral lesion classification baselines.

A study carried out by Wei Yuan et al [19]. proposed a deep radiomics analysis model that was based on optical coherence tomography imaging and a 3D deep learning technique (MC3 network) together with inflammatory biomarkers such as systemic immune-inflammation index (SII), platelet-to-lymphocyte ratio (PLR), and neutrophil-to-lymphocyte ratio (NLR). The proposed model incorporated imaging features and associated them with immunological markers to give non-invasive predictions of the prognosis of tumors without the need for surgery. Despite the high potential of the model in improving disease prognosis, some difficulties like small sample size and complex models exist.

Recently, a research paper has been published in which Kumari et al. [20] proposed a novel model that utilized NASNet and ShuffleNet with deep transfer learning concatenation and hybrid optimization algorithms, and achieved a precision rate of 0.84786 for diagnosing oral cancer.

The recent systematic review and meta-analysis by [21] examined the application of artificial intelligence and radiomics for predicting lymph node metastasis and tumor grading among oral cancer patients. There were 40 studies that were used in the review and meta-analysis, while 33 were used in the meta-analysis. In relation to the predictive ability of artificial intelligence-based models in detecting lymph node metastasis, the sensitivity, specificity, and diagnostic odds ratio (DOR) were 0.86, 0.91, and 56.58, respectively. For tumor grading of oral squamous cell carcinoma, the sensitivity and specificity were 0.88 and 0.82, respectively.

In another study [22], a collaborative machine learning (ML) approach was developed to predict the overall survival rate among patients with oral squamous cell carcinoma based on a large dataset collected from the Surveillance, Epidemiology, and End Results (SEER) database involving 9,439 patients. Various machine learning techniques such as voting ensemble, stacked ensemble, extreme gradient boosting, light boosting, and logistic regression were applied to predict the survival rate among patients in the dataset with an accuracy rate of 69.1% to 70.2%. With the

By employing collaborative learning between various models used in the ML process, the cML system was able to perform equally well as the best-performing model, ensuring reliability and robustness compared to other models. Some of the factors affecting prognosis, such as age of diagnosis, tumor stage, grade, gender, type of therapy, and metastasis stage, were determined.

[23] is an article that reflects the results of research carried out by the authors on the subject of the latest tendencies in the implementation of AI technologies as an important step in precision medicine aimed at diagnosing oral cancer. As one could understand from the discussion in the article, oral squamous cell carcinoma poses an issue as patients are diagnosed too late and survival rates from the disease are quite low. According to the authors, modern equipment used for diagnosing requires manual sample evaluation which might be not only subjective but also inefficient. On the contrary, using AI, specifically machine learning methods and deep learning algorithms such as convolutional neural networks, advanced imaging such as fluorescent and hyperspectral imaging can be evaluated. Moreover, according to the authors, smartphones equipped with AI technologies are capable of detecting oral lesions just as

precisely as experts. Nonetheless, several issues associated with bias and patient privacy and black box model of AI were identified.

Table 1. Related Work Summary

Ref.	Year	Focus Area	Technique/Model Used	Data type	Key Outcome	Limitations
[1] Tseng et al.	2015	Prognosis prediction	Data Mining / ML(C4.5, Naive Bayes, SVM)	Clinical (Large Taiwanese medical dataset)	Mined patterns for prognostic prediction	Outdated dataset; does not include recent medical advances
[2] Alabi et al.	2019	Recurrence prediction	Machine Learning (SVM, Logistic Regression)	Clinical (Early-stage Tongue cancer data from web tool source)	Forecasted recurrence of tongue cancer	Limited to oral tongue cancer; may not generalize
[3] Chu et al.	2020	Treatment outcome prediction	Machine Learning (Random Forest, SVM)	Clinical (Medical records from oral cancer cases)	Predicted survival and response to treatment	Clinical variables differ across cohorts; limited interpretability of model decisions
[4] Welikala et al	2020	Lesion detection	Deep Neural Networks (Deep CNN)	lesion images	Detected and classified oral lesions with high accuracy	Model needs high-quality annotated images; lacks generalizability
[5]Alkhadar et al.	2021	Survival prediction	Machine Learning (Random Forest (RF), Gradient Boosting (GBoost))	Clinical (5-year OSCC follow-up data from hospital database)	Predicted 5-year survival for OSCC patients	No external testing; biased by single-center data
[6]Kantharimuthu et al.	2023	Oral cancer prediction	Probability Neural Network (PNN)	Clinical (Patient records hospital-based)	Identified presence of oral cancer using probabilistic NN	Limited dataset size; lacks external validation
[7]Salehiniya & Raei	2020	Global epidemiology	Statistical review	Epidemiological	Descriptive epidemiology of oral cavity/lip cancer (not a diagnostic model)	Epidemiological review; no classifier evaluated
[8]Chen et al.	2022	Cancer risk prediction	Artificial Neural Network (ANN)	Clinicalhistory/Lesion records	Predicted precancerous lesion risks	Model explainability low; dataset size small
[9]Rahman et al.	2022	Histopath Diagnosis	Deep Learning (Transfer Learning InceptionV3)	Biopsy images (OSCC biopsy image dataset, augmented)	High-accuracy diagnosis of oral squamous carcinoma	Dataset imbalance; potential overfitting in transfer learning
[10]Shah et al.	2023	Treatment planning	Machine Learning (Decision Trees, Naive Bayes)	Clinical (Institutional patient treatment data)	Suggested treatment options via ML models	Dataset lacks feature diversity; early-stage model
[11] Ahmed et al.	2023	Early diagnosis	Hybrid Deep Learning (CNN Fusion) (VGG, ResNet)	Histopathology (digital histopathology slides)	Enabled early OSCC detection through image fusion	Computationally intensive; lacks multicenter data diversity
[12] Vinay et al.	2025	AI applications review Diagnostic/Prognostic	Mixed Review	Mixed (Discusses common datasets (e.g., TCGA, ImageNet)	Broad scoping of AI in oral cancer Reviewed diagnostic and prognostic tools	Review lacks performance meta-analysis
[13] Dixit et al.	2023	Review of ML/DL in oral cancer (Diagnostic Models)	Mixed Review	Mixed (Surveys datasets across studies)	Surveyed ML/DL capabilities in diagnosis	Lists gaps but lacks dataset benchmarking or metrics
[14] Nanditha & Annegowda	2022	Detection systems	Hybrid ML & DL (SVM, CNN)	Mixed (Combined image + numerical datasets)	Created hybrid detection model for accuracy	Model comparison superficial; lacks clinical testing
[15] Talaei Khoei et al.	2023	DL methods review	Deep Learning Review (ResNet, DenseNet, VGG)	General	DL model comparison & trends (across domains not cancer-specific)	General review not cancer-specific; lacks clinical relevance
[16] Dong et al.	2023	Pathology image diagnosis	Deep Learning (Multi-view CNN)	Pathology Images (Multi-view image dataset from pathology labs)	Advanced diagnosis using pathology image views	Requires high computing resources; scalability issues
[17] yang et al.	2023	Medical image super-resolution	Deep Learning Review (GANs, SRNet)	Imaging (High-resolution datasets for SR training)	Overview of enhancement techniques Image resolution enhancement for medical images	Super-resolution gains not tested on oral datasets
[18]Parisi & Manaog	2023	Cross-domain AI	Feature-driven ML/DL	Mixed *Public and institutional AI datasets for cross-dom (Health, finance, and education)	Applied AI models across healthcare domains	road focus; healthcare application segment limited
Wei Yuan1, Et al [19]	2024	Prognostic prediction of oral cancer	Deep radiomics model integrating Optical Coherence Tomography with 3D deep learning (MC3 network with residual layers) and multi-view radiomics fusion guided by inflammatory biomarkers	3D OCT imaging + peripheral blood inflammatory indicators (SII, PLR, NLR)	Achieved improved non-invasive prognostic prediction by combining imaging and inflammatory features, enabling better risk stratification of oral cancer patients	Limited dataset size (68 patients); high model complexity; requires multi-modal data acquisition and computational resources
Alka Kumari et al [20]	2025	Early detection and classification of oral cancer	Concatenated Deep Transfer Learning using NASNet and ShuffleNet with hybrid optimization algorithms	Histopathological images (OSCC)	Achieved precision of 0.84786, demonstrating improved diagnostic performance for oral cancer detection	High computational complexity; requires intensive preprocessing and parameter tuning
Khadijah Mohideen et al [21]	2026	Prediction of lymph node metastasis and tumour grading	Systematic Review and Meta-analysis of AI and Radiomics Models	CT, PET/CT Imaging Studies	Pooled sensitivity/specificity of 0.86/0.91 for lymph node metastasis prediction; demonstrated strong potential of radiomics-based AI	Significant heterogeneity among included studies; limited standardization across imaging protocols

Ref.	Year	Focus Area	Technique/Model Used	Data type	Key Outcome	Limitations
Alabi et al [22]	2026	Overall survival prediction in OSCC	Collaborative Machine Learning (Voting, Stacking, XGBoost, LightBoost, Logistic Regression)	SEER clinical database (9439 patients)	Achieved ˜70% accuracy; identified key prognostic factors for OSCC survival prediction	Moderate accuracy; limited interpretability of ensemble/collaborative models
Sinha Kumari,et al [23]	2026	AI-driven precision medicine and early detection in oral cancer	Review of ML, DL (CNN, AI imaging systems, smartphone-based AI)	Multimodal imaging and clinical data	Demonstrated potential of AI in early detection and precision medicine with near expert-level performance	Lack of transparency, data bias, and ethical/privacy concerns

Overall, the literature supports the feasibility of medical image learning and multimodal analysis but exposes gaps in patient-level leakage control, matched ablations, model comparability, and uncertainty estimates. Three relevant studies in Statistics, Optimization & Information Computing also provide context: Omran et al. [25] on medical imaging and metadata fusion, Ali and Sadeeq [26] on multiclass dental imaging and interpretability, and Abdulkareem and Abdulazeez [27] on imbalanced multiclass learning. Our revised contribution is an evaluation of whether straightforward fusion adds value under patient-independent oral-lesion testing; no between-paper numerical ranking is asserted.

3. Dataset

The publicly described NDB-UFES oral cancer/leukoplakia dataset provides H&E histopathological images and linked patient metadata [24]. The paired analytical subset comprises 237 image–clinical records from 69 unique patients and 77 lesions. Counts by three-class diagnosis were 89 images of leukoplakia with dysplasia, 57 of leukoplakia without dysplasia, and 91 of OSCC. Multiple images can originate from the same patient; thus, neither images nor slides are interchangeable with independent patient observations. Class membership is derived from tissue diagnosis rather than from a binary cancer/non-cancer proxy.

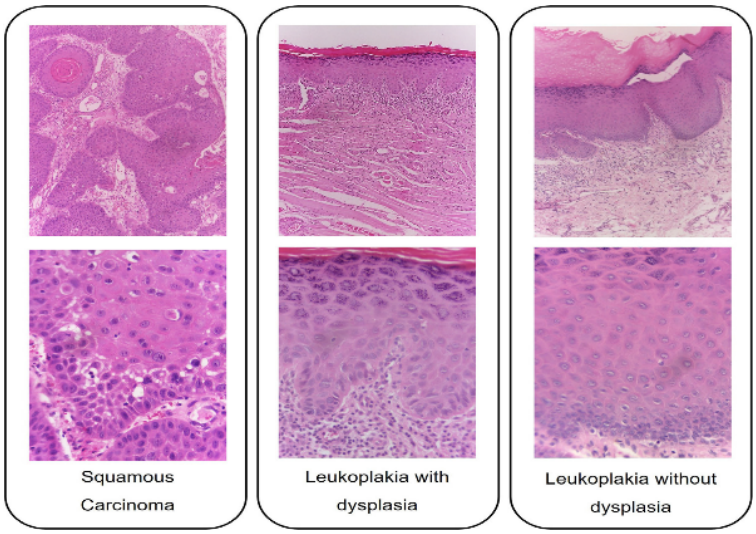


Figure 1. Comparison between Squamous Carcinoma, Leukoplakia with Dysplasia, and Leukoplakia without Dysplasia

Figure 1 depicts representative H&E morphology for the three classes in the original NDB-UFES illustration. It is retained unchanged from the submitted article. The present task predicts the histopathological diagnosis, not a separately evaluated binary clinical outcome. Representative morphology does not by itself establish diagnostic reliability of any algorithm [24].

3.1. Image Modality

The linked images are hematoxylin-and-eosin-stained microscopy photographs, not necessarily full digital whole-slide images. Each of the 237 analytical records was matched to its existing PNG path; images from the larger source folder that lacked a paired record were excluded from this experiment.

3.2. Clinical and Demographic Modality

The accompanying clinical records supply lesion localization, tobacco use, alcohol consumption, patient gender, and a lesion-size variable. These five inputs were used in the fusion experiments, with no target-defining TaskIII/TaskIV labels, diagnosis or dysplasia-severity variables allowed as clinical predictors. Table 2 preserves the source-variable overview.

Table 2. Clinical Features Included in the Dataset

Feature	Type	Categories / Range
Gender	Categorical	Male, Female
Lesion Location	Categorical	Buccal mucosa, Buccal sulcus, Floor of mouth, Labial mucosa, Lip, Palate, Tongue
Cigarette Use	Categorical	Yes, No, Former, Not informed
Alcohol Use	Categorical	Yes, No, Former, Not informed
Lesion Size	Continuous / Categorical	Measured in mm or classified by grade

3.3. Data Challenges and Considerations

Two principal challenges were the limited number of patients and the imbalance of diagnostic classes. There were 237 analytical images but only 69 patients; repeated photographs of one patient may share staining, preparation or morphology. We therefore grouped all splits by patient. Categorical values were handled with fold-specific one-hot encoding, while missing lesion-size values were imputed and scaled using training patients only. Frozen ImageNet image encoders reduced trainable parameters in the controlled comparisons; they do not eliminate overfitting or guarantee generalization. We did not establish a benefit from focal loss or synthetic oversampling and do not report these as implemented.

3.4. Dataset Splitting

Using a fixed random seed (42), a patient-stratified partition allocated 55 patients (189 images) to model development and 14 independent patients (48 images) to a locked internal holdout, with no patient overlap. Within development, five shuffled stratified patient folds were constructed (44 training and 11 validation patients per fold; fold-level image counts varied). Clinical imputation, one-hot encoding, and scaling were refitted separately on each training fold. The 14 holdout patients were excluded from model selection and evaluated only once after the two final configurations and training durations were fixed from development folds. This is internal patient-independent validation, not external or multicenter validation.

Table 3. Summary of the dataset splitting strategy

Set	Proportion	Purpose
Training Set	55 development patients; 189 images; five grouped folds	44-patient training subset per CV fold; full 55 patients for final fitting
Validation Set	11 distinct patients per fold; images vary	Training-only tuning, early stopping and reporting fold scores
Test Set	14 independent patients; 48 images	Single locked internal test; no overlap and no test-driven tuning

Table 3 distinguishes the locked internal test from development folds. Grouping was enforced by patient identity; splitting individual images independently would leak patient-specific information across subsets.

4. The Proposed Methodology

The original schematic in Figure 2 outlines the two-stream system. The final reproducible experiments used a frozen ImageNet-pretrained CNN followed by trainable image and clinical heads, late concatenation where applicable, and three-way softmax classification. The source schematic is preserved; dimensions shown illustratively in its original labels should be interpreted using the executable dimensions documented below.

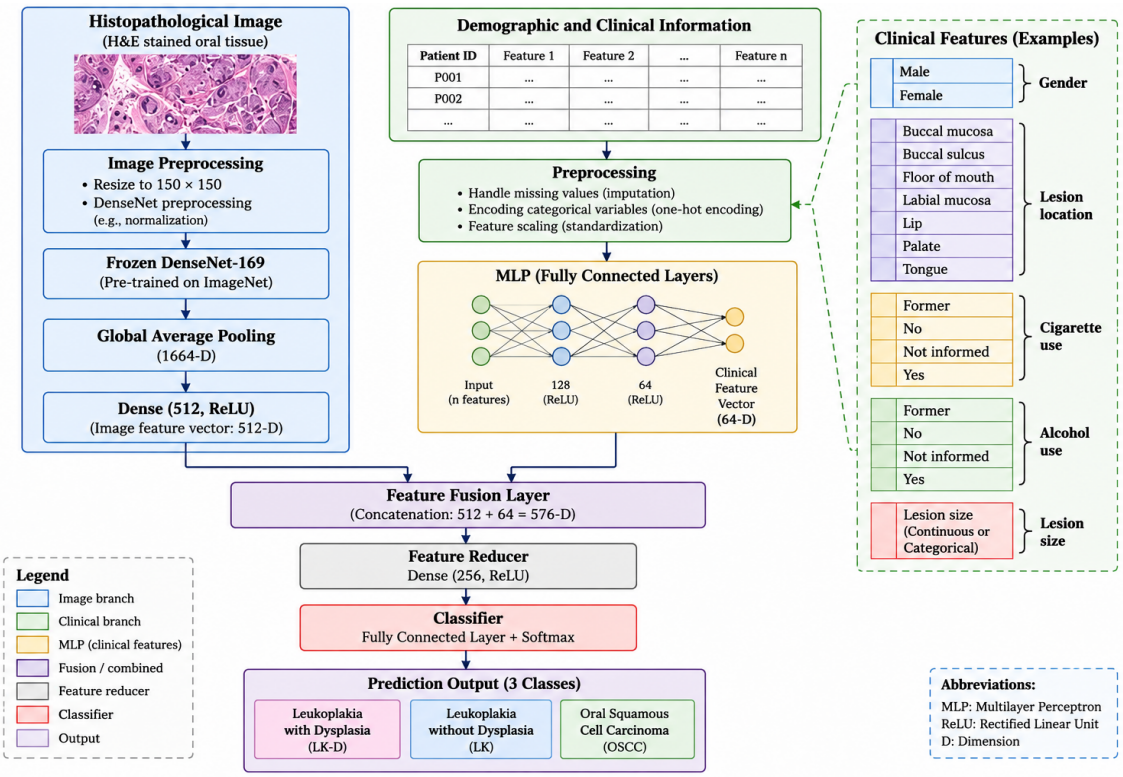


Figure 2. Original submitted conceptual two-stream DenseNet-169 + clinical MLP architecture (retained). Executed branch dimensions and class semantics are specified in Sections 4.2–4.7.

4.1. Dataset Description

The analytic cohort consists of 237 linked H&E image/metadata records from 69 patients. Diagnostic targets were encoded as 0 = leukoplakia with dysplasia, 1 = leukoplakia without dysplasia, and 2 = OSCC. All experiments addressed this three-class task; there was no separately trained or independently evaluated binary cancer classifier.

4.2. Image Branch: Histopathological Image Processing

For the revised controlled experiments, histopathology images were encoded with an ImageNet-pretrained DenseNet-169 network without its original classification top. To prevent image-level information leakage, the encoder was

frozen during the comparative feature-extraction protocol, and only classification heads were trained. The original training notebook used additional augmentation/Mixup settings, but those results were not treated as the corrected patient-independent benchmark.

4.2.1. Image Preprocessing Paired RGB images were read from their verified PNG paths and resized to 150 $\times$ 150 pixels using bilinear interpolation. Pixel arrays in their original $[0, 255]$ range passed once through tf.keras.applications.densenet.preprocess_input before the frozen encoder. There was no duplicate divide-by-255 operation in the corrected comparisons. No test-set augmentation was performed; the new feature-caching protocol did not employ Mixup or label smoothing. The old 224 $\times$ 224 narrative described a different/incorrect configuration and does not characterize these measured experiments.

4.2.2. Feature Extraction via DenseNet-169 DenseNet-169 (include_top=False; ImageNet weights) was kept non-trainable and pooled with global average pooling to yield a 1664-dimensional visual vector. A trainable Dense(512, ReLU) projected this vector, followed by normalization/dropout operations in the classification head. Images were resized to 150 $\times$ 150; the frozen convolutional feature maps for the explainability experiment were 4 $\times$ 4 $\times$ 1664, limiting localization resolution.

4.3. Clinical Branch: Structured Data Processing

The clinical branch was trained only on variables available at prediction time. The feature table was paired to the correct patient/image records. Categorical coding and continuous scaling were fit exclusively to the current training patients, then applied unchanged to validation or holdout patients.

4.3.1. Clinical Data Preprocessing The categorical features (localization, tobacco_use, alcohol_consumption, gender) were imputed using training-set most-frequent values and one-hot encoded with unseen categories ignored. The numeric larger_size variable underwent training-only median imputation and standard scaling. Clinical preprocessing was learned independently per fold and fitted once on all 55 development patients for the final locked evaluation; no diagnoses, class indicators or patient IDs were used as predictors.

4.3.2. Clinical Feature Extraction via Multilayer Perceptron (MLP) The clinical multilayer perceptron mapped its encoded input through Dense(128, ReLU) and Dense(64, ReLU) layers; the resulting 64-dimensional embedding formed the clinical contribution to the fused network. L2 penalties and the model' s dropout/normalization operations are described below. Table 4 is updated to reflect the actually executed 128-to-64 architecture.

Table 4. Executed two-layer clinical MLP in the corrected fusion study.

Layer	Units	Activation
Input	n features	—
Hidden Layer 1	128	ReLU
Hidden layer 2 / embedding	64	ReLU
Not used in revised runs	—	—
Clinical output	64-D	Embedding

The executed clinical embedding has 64 components. The originally described extra Dense(32) layer and 128-dimensional output did not correspond to the controlled experiments; the corrected architecture is provided in Table 4.

4.4. Multimodal Fusion and Classification

The multimodal classifier applies a simple late-fusion strategy: a DenseNet-derived 512-dimensional image embedding is concatenated with the 64-dimensional clinical embedding (576 total). The combined vector is optionally passed through a trainable 256-unit reducer, before a three-output softmax. This approach does not

introduce attention or a learned inter-modality weighting mechanism and is reported as a conventional reproducible baseline.

4.4.1. Fusion Layer Fusion uses feature concatenation, not a novel trainable attention or gating layer. Separate image and clinical representations are first learned by trainable dense heads and merged as a 576-dimensional vector; the direct-fusion ablation omits the learned reducer, while the full model includes its Dense(256) layer.

4.4.2. Feature Reducer The full fusion/reducer head uses dropout, normalization and a 256-unit ReLU reduction layer, with the precise executable layout preserved in the accompanying training notebook. The ablation compares its predictions against direct concatenation without this reducer. In this small cohort, the reducer cannot be claimed to improve performance consistently without reporting paired patient-level results.

4.4.3. Classifier A trainable three-neuron dense layer with softmax maps the selected image-only or fused representation to probabilities for the three histopathology classes. Categorical cross-entropy with sparse labels and Keras kernel L2 penalties was optimized using Adam; no second-stage end-to-end CNN fine-tuning was undertaken in the controlled backbone study.

4.4.4. Prediction Output The prediction is the argmax of three probabilities: 0 = leukoplakia with dysplasia, 1 = leukoplakia without dysplasia, and 2 = OSCC. The labels were not collapsed into a binary decision. At patient level, all image probability vectors from the same patient were averaged before taking the argmax; this aggregation yields one patient prediction per held-out patient.

4.5. Training hyperparameters and safeguards

The controlled ablation/backbone protocol used a fixed seed of 42, 150×150 RGB images, a batch size of 8, Adam optimizer (learning rate 1×10^{-4}), maximum 30 epochs, early stopping patience 7 with restoration of the best validation cross-entropy weights, and L2 kernel coefficient 0.01. The loss comprised sparse categorical cross-entropy and Keras L2 penalties. No Mixup, label smoothing, or mixed-precision claim is made for these measured controlled runs. Cross-validation selected each fold checkpoint using validation classification cross-entropy rather than the architecture-dependent total regularized loss.

Table 5. Reproducible settings of the corrected patient-disjoint experiments.

Item	**Verified setting**
Dataset	NDB-UFES: 237 linked images / 69 patients
Target	Three classes: dysplastic leukoplakia / nondysplastic leukoplakia / OSCC
Group split	55 development patients (189 images); 14 holdout patients (48 images); seed 42
Cross-validation	Five patient-stratified development folds (44 train, 11 validation patients each)
Image processing	150×150 RGB; model-specific Keras ImageNet preprocess_input once
CNN baseline	DenseNet-169 frozen (ImageNet); 1664-D pooled embedding
Clinical predictor set	localization; tobacco_use; alcohol_consumption; gender; larger_size
Clinical transform	Training-fold-only most-frequent/mode, median imputation, one-hot/scaling
Fusion head	Image Dense512; clinical Dense128 $\rightarrow$ Dense64; concatenate; optional Dense256 reducer; softmax 3
Optimizer	Adam, lr 1e-4, batch 8; cross-entropy + L2(0.01)
Model selection	Maximum 30 epochs; patience 7; best validation classification CE
Final fixed epochs	Fusion/reducer 11; image-only 4; development-fold median best epochs
Software / hardware	TensorFlow 2.20.0; Kaggle Tesla T4 GPUs reported (2 available)

4.6. Controlled experiments and statistical analysis

Four DenseNet-169 variants (image only, clinical only, concatenation without reducer, concatenation plus 256-unit reducer) were compared within the same five patient folds. Three frozen CNN encoders—DenseNet-169, ResNet-50 and EfficientNet-B0—were assessed with image-only and fusion heads on matched development splits. All image feature caches were generated with frozen pretrained encoders and encoder-specific preprocessing. For patient k, the mean of image-level three-class probability vectors was classified with argmax. Development means and sample standard deviations were calculated across five folds; pooled out-of-fold patient predictions comprised one prediction per development patient. Exact two-sided binomial confidence intervals were calculated for held-out patient accuracy and class-wise recall. Exploratory paired patient-level comparisons used exact McNemar testing and a patient-resampling bootstrap; they are not evidence of multicenter generalization.

4.7. Exploratory explanation and final locked test

Grad-CAM was generated for nine internal validation images per model (18 overlays total) using a final feature map of approximately 4×4 positions; region saliency is coarse and cannot be interpreted as validated tumor localization without specialist review. A small-cohort patient-level permutation analysis assessed sensitivity to each clinical variable. Before any final holdout predictions, model roles and epochs were fixed from development evidence: primary DenseNet-169 fusion with a reducer for 11 epochs, and image-only comparator for 4 epochs. Both heads were trained from scratch on the 55 development patients using the frozen ImageNet encoder; the patient-independent 14-patient internal test was then evaluated once. No external dataset was tested.

5. Results and Discussion

The revised experimental evaluation included: (i) patient-grouped five-fold development cross-validation, (ii) a four-arm modality/reducer ablation using DenseNet-169, (iii) a three-backbone matched patient-fold study, (iv) exploratory Grad-CAM and clinical permutation sensitivity, and (v) one prespecified locked internal holdout. Performance was calculated separately per image and per independent patient using accuracy, balanced accuracy, macro-F1 and multiclass one-vs-rest macro AUC. The original submission' s 39-image result is archived only as a historical image-split finding because all 31 test patients also appeared in training.

The original submission reported 87.18% image accuracy and approximately 0.981 AUC on 39 images, but audit of patient identifiers showed that all 31 patient identities in that test subset were also present in the 198-image training subset. This violated patient-independence and could inflate the score. Accordingly, the revised evidence and conclusions use the new 14-patient untouched internal holdout and the patient-disjoint development folds; historical Figures 3–5 below are retained solely to document the originally submitted non-independent evaluation.

In patient-grouped five-fold validation, a preliminary multimodal run with L2=0.01 yielded mean patient accuracy $70.91\% \pm 9.96\%$, patient macro-F1 0.683 ± 0.115, and patient macro OvR AUC 0.806 ± 0.084. A separate L2=0.0001 sensitivity run yielded mean patient accuracy $63.64\% \pm 6.43\%$, F1 0.609 ± 0.078 and AUC 0.825 ± 0.061. Reducing regularization lowered total reported loss but did not improve averaged accuracy or macro-F1, so the controlled ablation and backbone comparisons retained L2=0.01.

The original ROC and precision-recall plots and their numerical interpretations described the patient-overlapping 39-image split, and therefore do not provide evidence of performance on previously unseen patients. For the revised experiment, patient-disjoint results and uncertainty are reported in Tables 8–11 and Figure 6. The source figures are preserved for traceability and clearly marked as superseded.

The controlled ablation showed that the inclusion of clinical variables did not reliably improve validation performance. On the locked internal holdout, fusion performed better overall than the image-only comparator on this particular group (9/14 vs. 8/14 correct patients), but its OSCC recall was lower (4/6 vs. 5/6). The small cohort and wide intervals prevent strong conclusions about improved clinical performance.

Original Figures 3–5 reproduce the originally submitted ROC/precision-recall curves, per-class measures and confusion matrix. These plots correspond exclusively to the earlier patient-overlapping evaluation (39 images),

not to the revised patient-independent split or locked test. They are retained solely so that none of the original submission' s figure content is lost and must not be cited as corrected validation evidence.

Archived analyses from the original submission

The following two historical tables and three historical figures document the image-level, patient-overlapping evaluation in the original submission. They are included for traceability and are not evidence of performance on independent patients.

Table 6. ORIGINAL SUBMISSION ONLY — three-class image-level precision, recall and F1 on a patient-overlapping 39-image split (superseded; NOT the independent internal-test results).

Class	Precision	Recall	F1-Score
0	0.88	0.93	0.90
1	1.00	0.56	0.71
2	0.83	1.00	0.91
Macro Average	**0.90**	**0.83**	**0.84**

The original submission's Table 6 compared estimates from different datasets and nonmatched protocols and cannot substantiate superiority over Decision Trees or earlier fusion architectures. It is retained as background from the first submission only. The direct, controlled comparisons on identical patient-disjoint folds are now summarized in Tables 8 and 9.

Table 7. HISTORICAL original nonmatched literature-style AUC comparison (not a controlled benchmark).

Model	Archived AUC	Not matched / not patient-independent
Decision Tree	67%	Different clinical dataset / different endpoint
DenseNet-169	98.18%	Earlier image-split estimate; patient overlap
MetaBlock fusion & ResNetV2 backbone	83.24%	Provenance/protocol not validated as fair comparison

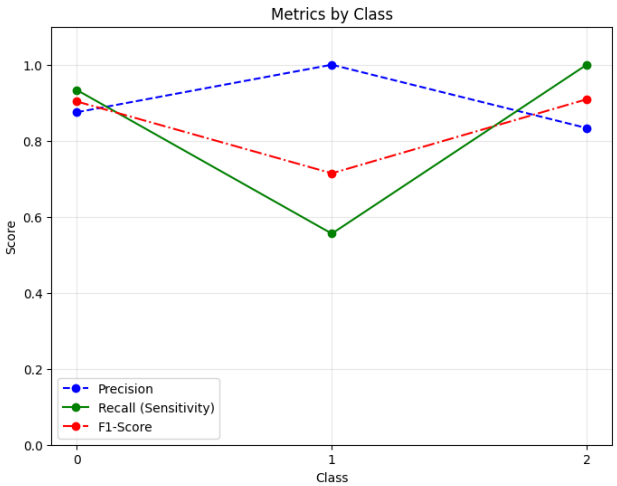


Figure 3. HISTORICAL — originally submitted per-class precision, recall and F1 visualization from the patient-overlapping image-split evaluation (superseded).

The numerical differences in the original submission's Table 6 must not be interpreted as causal architectural advantages. The corrected same-fold evaluation of DenseNet-169, ResNet-50, and EfficientNet-B0 with harmonized frozen encoders and matched heads is presented in the new results tables. Comparisons across different endpoints, cohorts or image sets are not valid evidence of architectural superiority.

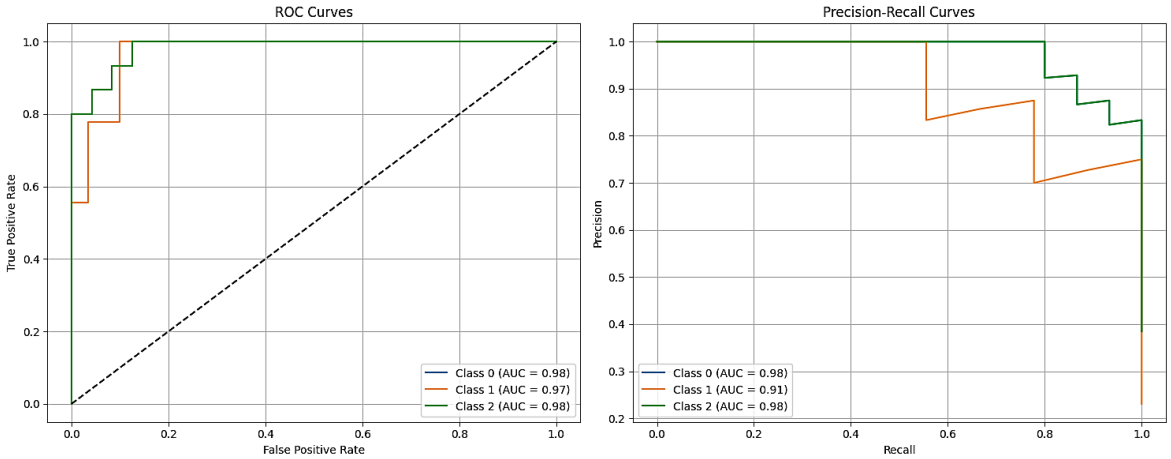


Figure 4. HISTORICAL — originally submitted ROC and precision-recall curves based on the patient-overlapping image split (superseded).

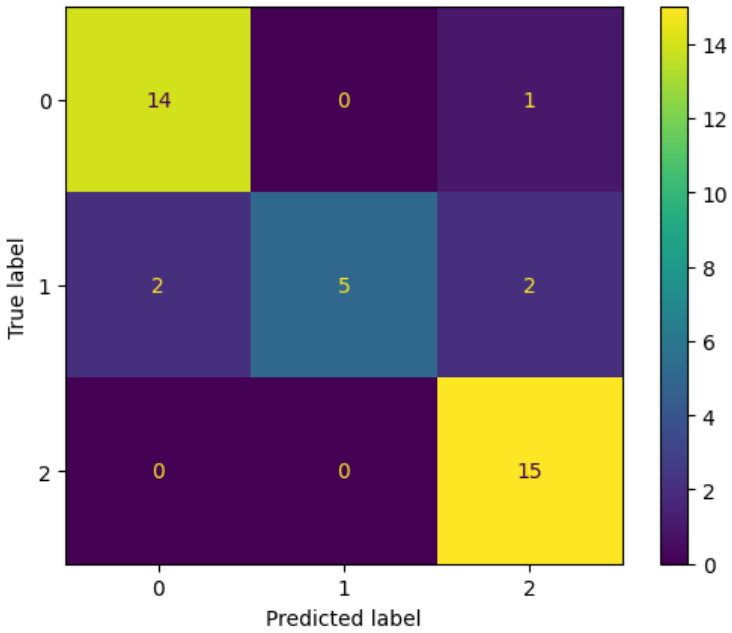


Figure 5. HISTORICAL — originally submitted confusion-matrix heatmap for patient-overlapping test (superseded).

5.1. Five-fold development validation and L2 sensitivity

The five-fold patient-disjoint DenseNet fusion pilot with L2=0.01 achieved mean image accuracy 59.88% ± 3.36%, patient accuracy 70.91% ± 9.96%, patient macro-F1 0.683 ± 0.115, and patient macro OvR AUC 0.806 ± 0.084.

Reducing L2 to 0.0001 reduced the reported regularized loss but resulted in patient accuracy 63.64% $\pm$ 6.43% and patient macro-F1 0.609 $\pm$ 0.078. Because the controlled ablation and backbone experiments share L2=0.01, their estimates are analyzed separately from this preliminary sensitivity study.

5.2. Matched modality and reducer ablation

Table 8. DenseNet-169 ablation: 5-fold development mean (patient units; image-level accuracy also shown).

Variant	Image acc.	Patient acc.	Patient macro-F1
Image only	57.8%	67.3%	0.612
Clinical only	54.3%	50.9%	0.471
Direct fusion	52.1%	52.7%	0.448
Fusion + reducer	57.4%	60.0%	0.488

In these development folds, the image-only DenseNet head attained the highest mean patient accuracy and macro-F1. Fusion with reducer did not establish incremental benefit from clinical metadata. This result is an important qualification of the original motivation: application of a modality does not prove that the modality improves performance. The analysis is exploratory and based on 55 development patients.

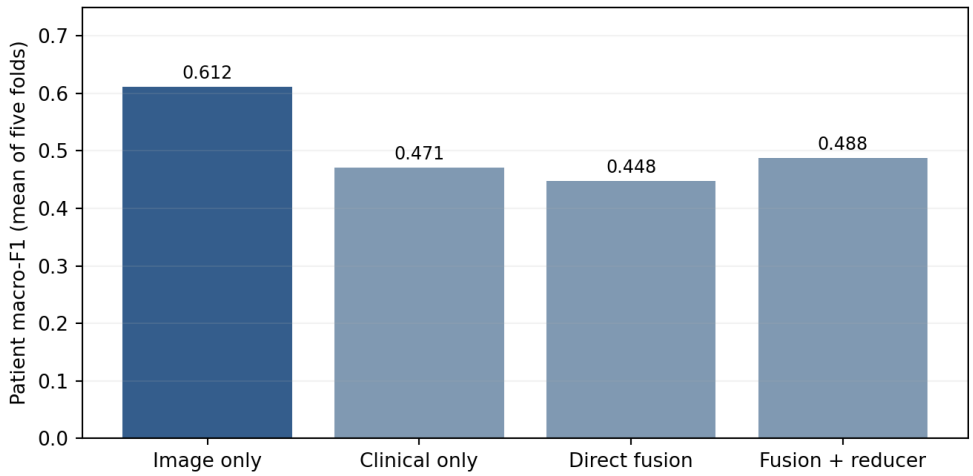


Figure 6. Controlled DenseNet-169 ablation (mean patient-level macro-F1 over five development folds).

5.3. Matched frozen-CNN backbone comparisons

Table 9. Five-fold matched-backbone comparison; mean $\pm$ sample SD on patient-level outcomes.

Backbone	Variant	Patient acc.	Patient macro-F1
DenseNet-169	Image only	0.673 $\pm$ 0.122	0.612 $\pm$ 0.118
DenseNet-169	Fusion + reducer	0.600 $\pm$ 0.104	0.488 $\pm$ 0.058
ResNet-50	Image only	0.618 $\pm$ 0.041	0.541 $\pm$ 0.058
ResNet-50	Fusion + reducer	0.582 $\pm$ 0.081	0.508 $\pm$ 0.057
EfficientNet-B0	Image only	0.582 $\pm$ 0.081	0.547 $\pm$ 0.086
EfficientNet-B0	Fusion + reducer	0.636 $\pm$ 0.064	0.581 $\pm$ 0.103

Across these same patient folds, DenseNet-169 image-only had the highest mean patient macro-F1 (0.612), whereas EfficientNet-B0 fusion attained 0.581. Encoders were frozen with ImageNet weights; these findings do not support claims about end-to-end fine-tuning or universal superiority of a backbone. Comparisons from unrelated published datasets are not interpreted as controlled baselines.

5.4. Prespecified internal holdout (not external validation)

Table 10. Single locked patient-independent internal test (14 patients; 48 images).

Outcome	DenseNet-169 fusion + reducer	DenseNet-169 image only
Image accuracy	58.33% (28/48)	47.92% (23/48)
Image macro-F1	0.568	0.454
Image macro OvR AUC	0.828	0.718
Patient accuracy	64.29% (9/14)	57.14% (8/14)
Patient macro-F1	0.600	0.470
Patient macro OvR AUC	0.860	0.756
95% exact patient-accuracy CI	35.14%–87.24%	28.86%–82.34%

The two final models were selected and trained without examining the locked test predictions. The fusion model correctly classified nine of 14 patients versus eight of 14 for image-only. The absolute difference (7.14 percentage points) represents just one patient. Exact paired McNemar testing returned p=1.000 (2 discordant patients favored fusion and 1 favored image-only). An exploratory stratified patient-bootstrap 95% percentile interval for the patient macro-F1 difference (fusion minus image) spanned -0.101 to +0.356; the data do not establish statistical superiority or clinical utility.

Table 11. Locked internal-test patient-level recall by diagnostic class.

True diagnosis (patients)	Fusion	Image only
Leukoplakia with dysplasia (n=5)	4/5 (80.0%)	3/5 (60.0%)
Leukoplakia without dysplasia (n=3)	1/3 (33.3%)	0/3 (0.0%)
OSCC (n=6)	4/6 (66.7%)	5/6 (83.3%)

Importantly, fusion misclassified two of six OSCC patients versus one of six for image-only. Nondysplastic leukoplakia recall was especially uncertain because only three such patients were included. On the one-time test, patient-level fusion confusion-matrix rows (true class 0/1/2) were $[4, 1, 0]$, $[2, 1, 0]$, and $[1, 1, 4]$; image-only rows were $[3, 2, 0]$, $[3, 0, 0]$, and $[1, 0, 5]$. No post-hoc model retuning was conducted.

5.5. Explainability and computational context

Nine internal validation images per model yielded 18 Grad-CAM overlays (not external or test examples). Coarse 4×4 class-activation grids often extended beyond obvious lesion areas, precluding histopathology localization claims without specialist assessment. In exploratory patient-level permutation sensitivity, lesion localization caused a small mean patient macro-F1 reduction (0.0113, SD 0.0248); tobacco use, alcohol consumption, gender and lesion size had measured reductions of 0 under this narrow protocol. Zero observed sensitivity must not be interpreted as absence of medical relevance. The Kaggle environment reported TensorFlow 2.20.0 and two Tesla T4 GPUs; the final frozen head-training logs recorded approximately 13.7 s for 11-epoch fusion and 4.2 s for 4-epoch image-only training (excluding image feature extraction). Holdout encoder inference for all 48 images took approximately 9.37 s in the recorded run; head prediction added approximately 0.64 s for fusion and 0.32 s for image-only. These hardware-specific measurements are not validated clinical throughput benchmarks.

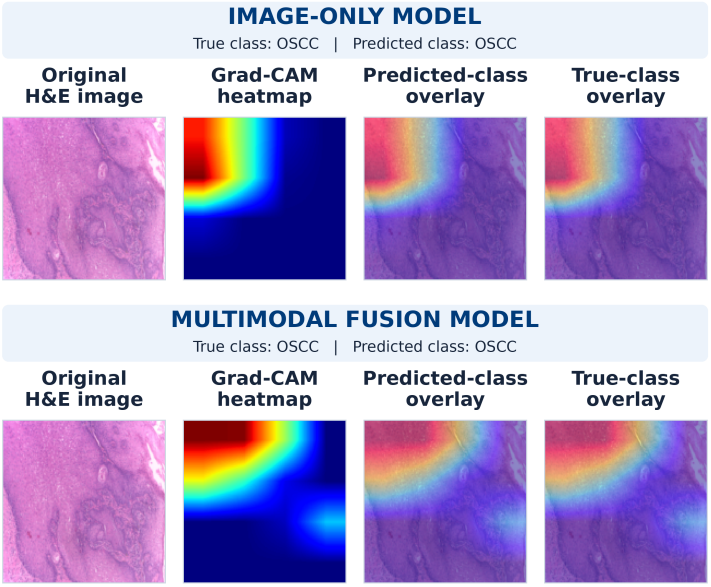


Figure 7. Internal validation Grad-CAM comparison: image-only model (top) and multimodal fusion model (bottom). Original H&E images, heatmaps, and class-specific overlays are presented at an enlarged, readable scale. Exploratory visualizations only; not clinically verified tumor maps.

5.6. Interpretation, caveats and clinical implications

Development ablations did not support a stable contribution from the clinical branch, while the single independent-patient internal holdout numerically favored fusion overall. This discrepancy underscores variance under small-sample sampling and cautions against threshold-independent claims of improved decision making. In particular, the fusion model missed two of six OSCC patients and performed weakly on nondysplastic leukoplakia. Clinical application would require larger patient-independent samples, external multisite data, specialist review of interpretability outputs, prospective evaluation, bias monitoring, appropriate consent and patient privacy governance. The reported analysis is research evidence, not a deployable diagnostic device.

6. Conclusion and Future Directions

This work assesses the feasibility and limitations of a conventional DenseNet-169 plus clinical-MLP feature-fusion system for three-class NDB-UFES oral lesion classification. The major methodological improvement is patient-grouped development and a locked patient-independent internal test. On 14 held-out patients, fusion achieved 64.29% accuracy and macro-F1 0.600 versus 57.14% and 0.470 for image-only; the paired contrast was not statistically significant (exact McNemar p=1.000). Although the fixed fusion model improved overall holdout accuracy in this small sample, its OSCC recall was 4/6 rather than 5/6 in image-only, and performance on nondysplastic leukoplakia was limited (1/3). Development ablations did not demonstrate a consistent gain from clinical fusion. External multicenter validation, larger independent patient cohorts, calibrated predictions, pathology review of Grad-CAM maps, and prespecified clinical decision thresholds remain necessary before any deployment claim. No clinical decision benefit is established by this study.

Author Contributions

H.R. and M.G. conceived the study and contributed to the research design. M.G. and M.Z. collected and analyzed the data, developed the methodology, prepared the figures and tables, and wrote the main manuscript. M.Z. and N.E. supervised the research, validated the methodology, and critically reviewed the manuscript. N.E. and H.R. contributed to the interpretation of the results and revised the manuscript. All authors reviewed and approved the final version of the manuscript.

REFERENCES

1. Tseng, W. T., Chiang, W. F., Liu, S. Y., Roan, J., & Lin, C. N. (2015). The application of data mining techniques to oral cancer prognosis. Journal of Medical Systems, 39(1), 1–7.
2. Alabi, R. O., Elmusrati, M., Sawazaki-Calone, I., Kowalski, L. P., Haglund, C., Coletta, R. D., & Almangush, A. (2019). Machine learning application for prediction of locoregional recurrences in early oral tongue cancer: A web-based prognostic tool. Virchows Archiv, 475(4), 489–497.
3. Chu, C. S., Lee, N. P., Adeoye, J., Thomson, P., & Choi, S. W. (2020). Machine learning and treatment outcome prediction for oral cancer. Journal of Oral Pathology & Medicine, 49(10), 977–985.
4. Welikala, R. A., Remagnino, P., Lim, J. H., Chan, C. S., Rajendran, S., Kallarakkal, T. G., & Barman, S. A. (2020). Automated detection and classification of oral lesions using deep learning for early detection of oral cancer. IEEE Access, 8, 132677–132693.
5. Alkhadar, H., Macluskey, M., White, S., Ellis, I., & Gardner, A. (2021). Comparison of machine learning algorithms for the prediction of five-year survival in oral squamous cell carcinoma. Journal of Oral Pathology & Medicine, 50(4), 378–384.
6. Kantharimuthu, M., Malachi, M., & Cynthia, P. (2023). Oral cancer prediction using a probability neural network (PNN). Asian Pacific Journal of Cancer Prevention (APJCP), 24(9), 2991.
7. Salehiniya, H., & Raei, M. (2020). Oral cavity and lip cancer in the world: An epidemiological review. Biomedical Research and Therapy, 7(8), 3898–3905.
8. Chen, W., Zeng, R., Jin, Y., Sun, X., Zhou, Z., & Zhu, C. (2022). Artificial neural network-assisted cancer risk prediction of oral precancerous lesions. BioMed Research International, 2022, 7352489.
9. Rahman, A. U., Alqahtani, A., Aldhafferi, N., Nasir, M. U., Khan, M. F., Khan, M. A., & Mosavi, A. (2022). Histopathologic oral cancer prediction using oral squamous cell carcinoma biopsy empowered with transfer learning. Sensors, 22(10), 3833.
10. Shah, H., Patil, Y., & Patel, R. (2023). Prediction of oral cancer treatment plan using machine learning. International Journal of Computer Information Systems and Industrial Management Applications, 15, 10.
11. Ahmed, I. A., Senan, E. M., & Shatnawi, H. S. A. (2023). Analysis of histopathological images for early diagnosis of oral squamous cell carcinoma by hybrid systems based on CNN fusion features. International Journal of Intelligent Systems, 2023(1), 2662719.
12. Vinay, V., Jodalli, P., Chavan, M. S., Buddhikot, C. S., Luke, A. M., Ingafou, M. S. H., ... & Testarelli, L. (2025). Artificial intelligence in oral cancer: A comprehensive scoping review of diagnostic and prognostic applications. Diagnostics, 15(3), 280.
13. Dixit, S., Kumar, A., & Srinivasan, K. (2023). A current review of machine learning and deep learning models in oral cancer diagnosis: Recent technologies, open challenges, and future research directions. Diagnostics, 13(7), 1353.
14. Nanditha, B. R., & Annegowda, G. K. (2022). Oral cancer detection using machine learning and deep learning techniques. International Journal of Current Research and Review, 14(1), 64–70. https://doi.org/10.31782/IJCRR.2021.14104
15. Talaei Khoei, T., Ould Slimane, H., & Kaabouch, N. (2023). Deep learning: Systematic review, models, challenges, and research directions. Neural Computing and Applications, 35(31), 23103–23124.
16. Dong, W., Sun, S., & Yin, M. (2023). A multi-view deep learning model for pathology image diagnosis. Applied Intelligence, 53(6), 7186–7200.
17. Yang, H., Wang, Z., Liu, X., Li, C., Xin, J., & Wang, Z. (2023). Deep learning in medical image super resolution: A review. Applied Intelligence, 53(18), 20891–20916.
18. Parisi, L., & Manaog, M. L. (2023). Innovative feature-driven machine learning and deep learning for finance, education, and healthcare. Neural Computing and Applications, 35(16), 11477–11480.
19. Yuan, W., Rao, J., Liu, Y. et al. Deep radiomics-based prognostic prediction of oral cancer using optical coherence tomography. BMC Oral Health 24, 1117 (2024). https://doi.org/10.1186/s12903-024-04849-8
20. Kumari, A., Mehta, A.K. & Khan, D.A. Concatenated Deep Transfer Learning with Hybridized Optimization Algorithm for Early Prediction and Diagnosis of Oral Cancer Disease. J. Electr. Eng. Technol. 21, 1073–1082 (2026). https://doi.org/10.1007/s42835-025-02462-5
21. Mohideen, K., Ghosh, S., Krithika, C., Mulk, B. S., Chole, R., Chatterjee, J., & Dhungel, S. (2026). Application of artificial intelligence and radiomics in the prediction of lymph node metastasis and tumour grading of oral cancer-a systematic review and meta analysis. BMC Oral Health.
22. Alabi, R. O., Elmusrati, M., Leivo, I., Almangush, A., & Mäkitie, A. A. (2026). Collaborative machine learning-guided overall survival prediction of oral squamous cell carcinoma. Acta Oto-Laryngologica, 146(2), 133-140.
23. Kumari, S., Kumar, N., Khan, M. S., Sultana, S., & Khalid, M. (2026). Early detection in oral cancer: are we ready for AI-driven precision?. Annals of Medicine and Surgery, 88(1), 32-33.
24. Ribeiro-de-Assis, M. C. F., Soares, J. P., de Lima, L. M., de Barros, L. A. P., Grão-Velloso, T. R., Krohling, R. A., & Camisasca, D. R. (2023). NDB-UFES: An oral cancer and leukoplakia dataset composed of histopathological images and patient data. Data in Brief, 48, 109128.

25. Omran, A. S. A. E. A., Mohamed El-Gazzar, M. M., & Monir Saeid, M. (2025). A Hybrid Fusion Approach for Skin Cancer Detection using Deep Learning on Clinical Images and Machine Learning on Patient Metadata. Statistics, Optimization & Information Computing, 14(4), 2091–2103. https://doi.org/10.19139/soic-2310-5070-2811
26. Ali, D. A., & Sadeeq, H. T. (2025). An Interpretable Deep Learning Framework for Multi-Class Dental Disease Classification from Intraoral RGB Images. Statistics, Optimization & Information Computing, 14(6), 3380–3397. https://doi.org/10.19139/soic-2310-5070-2880
27. Abdulkareem, R.-zaki, & Abdulazeez, A. M. (2025, online publication). A Comparative Study of Multi-Class Classification based on Imbalanced Data. Statistics, Optimization & Information Computing, 15(5), 4337–4357. https://doi.org/10.19139/soic-2310-5070-2731