© APR 2026 | IRE Journals | Volume 9 Issue 10 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I10-1716822

AI-Assisted Early Detection of Oral Cancer Using Attention-Guided Deep Learning and Automated Diagnostic Reporting

SNEHA SAWANT[1], LOKESH THAKARE[2], MUSHFIQ SHAIKH[3], SUHAS WAGHMARE[4]
[1,2,3,4] *Dept. of Artificial Intelligence and Data Science New Horizon Institute of Technology and Management Thane, India*

Abstract- ***Late-stage oral cancer diagnosis is a problem we have had the tools to address for years — the bottleneck has always been getting those tools into the right hands fast enough. Our work grew out of that frustration. We put together AGMSFFNet, a histopathology classifier that pairs a modified EfficientNet-B5 backbone with a dual-branch attention design we call HSCA, which handles both spatial and channel recalibration at once. Alongside that, multi-resolution LBP texture descriptors feed into the feature pipeline to capture detail that convolution alone tends to skip over. The part we are most invested in, though, is the LangGraph reporting layer — rather than stopping at a class label, the system drafts a structured clinical summary automatically, flags anything it is uncertain about, and hands a readable report to the pathologist. Testing on ORCHID gave us 98.7% accuracy, 98.9% precision, 98.5% recall, and a 98.7% F1. External results on NDBUFES held up at 97.5%. ANOVA and Tukey tests confirmed the differences were statistically meaningful.***

Index Terms- ***— Oral Cancer Classification, Deep Learning, Efficientnet, Attention Mechanisms, Feature Fusion, Histopathological Image Analysis, Langgraph, Agentic AI, Interpretable AI.***

I. INTRODUCTION

Oral squamous cell carcinoma is one of those cancers where the statistics tell a frustratingly avoidable story. Around 377,000 new cases every year, 177,000 deaths, and a five-year survival rate hovering at 50% — not because treatment options are limited, but because most patients walk in when the disease has already progressed. Push the detection window earlier and that survival figure climbs to 80-90%. That gap between what is possible and what actually happens in practice is what motivated this project.

Histopathological biopsy is still the diagnostic gold standard, and it is not going away anytime soon. But it carries real limitations — the interpretation is subjective, the process is time-intensive, and the outcome depends heavily on how experienced the pathologist reading the slide happens to be. In hospitals and clinics where specialist expertise is thin on the ground, those limitations compound quickly.

We started this work knowing that deep learning had already shown genuine promise on this problem. What we found in the existing literature, though, was that most systems were solving only part of it. They classified well enough, but then handed a probability score to a clinician and stopped there. Writing the clinical note, interpreting what the result means in context, deciding what to do next — all of that still fell on the pathologist. Closing that gap was the reason we brought LangGraph into the design.

To summarise what we set out to do and what we built:

- A classification architecture combining EfficientNet-B5 with our HSCA dual-branch attention module and multi-level LBP texture fusion.
- A LangGraph-driven reporting pipeline that takes model outputs and produces structured, human-readable diagnostic summaries with uncertainty flagging.
- A Relief-based feature selection step that trims the fused representation down to what actually discriminates, rather than carrying redundant features through to the classifier.
- Validation on two independent datasets — ORCHID for primary evaluation, NDBUFES for

cross-institutional generalisability — backed by formal statistical testing.

II. REVIEW OF LITERATURE

Putting AGMSFFNet together meant pulling from three fairly distinct bodies of work: transfer learning approaches to oral cancer classification, attention mechanism design in medical imaging, and the newer territory of agentic AI in clinical workflows.

A. Deep Learning for Oral Cancer Detection

The earlier efforts in this space were largely about proving that automated classification was feasible at all. Panigrahi and colleagues ran VGG16, ResNet50, and MobileNet through careful fine-tuning on oral histopathology slides and topped out at 96.6% — solid results that established a useful baseline. Das et al. took a different approach and built a ten-layer CNN from the ground up, reaching 97.82%, though without any attention component or multi-scale feature handling. Both papers opened the door; neither fully walked through it.

B. Attention Mechanisms and Feature Fusion

Attention has earned its place in medical image analysis because tissue slides are not uniformly informative — the diagnostically meaningful content is concentrated in specific regions and specific feature channels, and models that treat everything equally tend to get distracted. Dharani and Danesh demonstrated this for oral cancer specifically, applying spatial-channel attention on EfficientNet-B5 and reaching 98% on ORCHID. Their work stopped short of multi-level fusion and did not attempt automated reporting, which left room for what we built here. On the texture side, Krishnan et al. established that LBP descriptors add value precisely where deep features are weakest — at fine-grained cellular structure — and that combination informed our fusion strategy directly.

C. Agentic AI in Clinical Applications

LangGraph is a stateful graph-based framework that lets you compose multi-step AI workflows where each node does a discrete job and the transitions between nodes follow conditional logic. For a diagnostic pipeline, that structure maps well onto how clinical reasoning actually works — you gather information, identify what is missing, retrieve what you need, then produce a coherent output. The ability to short-circuit unnecessary steps when confidence is high, and to pull in supporting literature when it is not, is exactly what makes it a better fit here than a simple linear inference chain.

III. SYSTEM ARCHITECTURE

The full system runs as a four-stage pipeline: an attention-enhanced CNN backbone, a multi-resolution texture extraction module, a feature fusion and selection stage, and the LangGraph reporting layer. Each component has a specific responsibility and feeds into the next.

A. Enhanced CNN Backbone with HSCA Attention

EfficientNet-B5 serves as the feature extraction base. On top of it we placed the Hybrid Spatial-Channel Attention module, which runs a spatial branch and a channel branch in parallel and combines their outputs to reweight the backbone's feature maps. The spatial branch identifies which regions of the slide are worth focusing on; the channel branch decides which feature dimensions carry the most signal for the current input. In practice this means the model tends to home in on things like chromatin irregularity and disrupted epithelial layer structure — the same things a trained pathologist would be looking at.

B. Texture Feature Extraction and Fusion

We ran LBP descriptor extraction at three radii (1, 2, and 3) to get texture information at multiple granularities, ending up with a 1,024-dimensional texture vector. That gets concatenated with the 2,048-dimensional CNN output to form a 3,072-feature combined representation. Relief-based selection then cuts that down to 1,536 features — the half that actually contributes to class discrimination. The fact that dropping half the features improved accuracy rather than hurt it confirmed our suspicion that the raw concatenation carried a lot of noise.

C. LangGraph-Based Agentic Reporting

Five agents handle the reporting pipeline in sequence. The Prediction Intake Agent structures the raw model output — probabilities, confidence scores, Grad-CAM maps — into a standardised record. The Gap

© APR 2026 | IRE Journals | Volume 9 Issue 10 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I10-1716822

Identification Agent examines that record for low-confidence or clinically inconsistent predictions and flags them. When a flag is raised, the Context Retrieval Agent queries a clinical knowledge base for relevant literature and grading criteria. The Report Generation Agent then assembles a complete pathology report from all available information. Finally the Review Agent checks the draft for internal consistency before it reaches the pathologist. If no gap is flagged, the pipeline skips directly from intake to report generation, which keeps latency reasonable.

IV. EXPERIMENTAL SETUP AND RESULTS

A. Implementation Details

Training ran on an NVIDIA A100 with 40 GB VRAM under TensorFlow 2.10 and Keras, with FP16 mixed-precision throughout. Our primary evaluation used a stratified 50,000-image subset of ORCHID, balanced across normal tissue, dysplastic tissue, and OSCC. For external validation we selected 1,000 images from NDBUFES. All images were resized to 224x224 and normalised to [0,1], with Macenko stain normalisation applied beforehand to reduce the colour variability that comes from different institutional staining protocols.

B. Evaluation Metrics

We tracked accuracy, precision, recall, and F1-score at both per-class and macro-averaged levels. Statistical validation used one-way ANOVA across five independent training runs, followed by post-hoc Tukey HSD tests to confirm that AGMSFFNet's improvements over each comparison model were not attributable to random variation.

C. Results

Table I shows the per-class breakdown on the ORCHID test set. Dysplastic tissue was the hardest category, as expected — 97.9% recall there compared to 98.7-98.9% on the other two classes — but even that number holds up well against prior work.

TABLE I
Classification Performance on ORCHID Test Set

Class	Precision	Recall	F1
Normal	0.991	0.989	0.990
Dysplastic	0.984	0.979	0.981
Malignant	0.992	0.987	0.989
Macro Avg.	0.989	0.985	0.987

Table II shows where AGMSFFNet sits relative to the methods we compared against. The accuracy gap over the next-best approach is 0.7 percentage points, and it comes with a parameter count of 67.8M against the TSA Ensemble's 103.9M — a better result from a leaner model.

TABLE II
Comparison with State-of-the-Art Methods

Model	Accuracy	Precision	Recall
VGG19	0.778	0.782	0.778
ResNet152V2	0.700	0.739	0.700
ViT-16	0.792	0.794	0.792
DeepPatchNet	0.867	0.868	0.867
TSA Ensemble	0.980	0.981	0.979
AGMSFFNet (Ours)	0.987	0.989	0.985

D. Discussion and Limitations

Looking at where the gains came from — HSCA attention steered the network toward diagnostically relevant structures rather than letting it spread attention uniformly across the slide. The LBP fusion picked up texture patterns at a scale that convolution alone tends to flatten out. Relief selection kept the feature space focused. On NDBUFES, accuracy dropped to 97.5%, a 1.2 point gap from ORCHID that we think is acceptable evidence of cross-institutional robustness. The reporting layer is harder to quantify in a table, but in practical terms it means a pathologist gets a structured clinical summary rather than a raw probability — which changes how useful the system actually is in a real workflow.

On limitations: our training data likely underrepresents geographic and demographic variation in OSCC presentation. Three-class grading is a simplification of clinical reality. All evaluation was retrospective. GPU availability is a real constraint in the low-resource settings where this tool would arguably be most valuable.

V. CONCLUSION

What we wanted to build was a system that does not just classify oral cancer tissue images but actually produces something a clinician can use. AGMSFFNet gets to 98.7% accuracy on ORCHID and 97.5% on NDBUFES while running on fewer parameters than the previous best ensemble approach. The LangGraph pipeline turns those classification results into structured diagnostic reports, handles uncertainty explicitly, and reduces the interpretive work that would otherwise fall on the pathologist. The obvious next steps are finer-grained classification, integration of clinical and genomic data, and prospective testing in real hospital settings — but as a foundation for that work, we think this holds up.

VI. ACKNOWLEDGMENT

We thank the teams who built and shared the ORCHID and NDBUFES datasets — this work would not have been possible without them. Computational support came from Mohammed VI University of Sciences and Health and the Saveetha Institute of Medical and Technical Sciences.

REFERENCES

[1] H. Sung et al., "Global cancer statistics 2020: GLOBOCAN estimates," CA: A Cancer Journal for Clinicians, vol. 71, no. 3, pp. 209-249, 2021.

[2] S. Warnakulasuriya, "Global epidemiology of oral and oropharyngeal cancer," Oral Oncology, vol. 45, no. 4-5, pp. 309-316, 2009.

[3] S. Panigrahi et al., "Classifying histopathological images of OSCC using deep transfer learning," Heliyon, vol. 9, no. 3, p. e13444, 2023.

[4] R. Dharani and K. Danesh, "Optimized deep learning ensemble for accurate oral cancer detection," Intelligence-Based Medicine, vol. 11, p. 100258, 2025.

[5] I. Tafala et al., "DeepPatchNet: A deep learning model for enhanced screening of oral cancer," Informatics in Medicine Unlocked, vol. 53, p. 101658, 2025.

[6] M. M. R. Krishnan et al., "Automated diagnosis of oral cancer using higher order spectra features and LBP," Technology in Cancer Research & Treatment, vol. 10, no. 5, pp. 443-455, 2011.