

Ensemble Machine Learning Models For Oral Lesion Classification Using A Clinically Curated Histopathological Dataset

Hemashree H C¹, Pradeep N², Ahmed Mujib B R³

¹Department of Information Science and Engineering, Bapuji Institute of Engineering and Technology, Visvesvaraya Technological University, Belagavi – 590018, Karnataka, India.

²Department of Computer Science and Engineering, Bapuji Institute of Engineering and Technology, Visvesvaraya Technological University, Belagavi – 590018, Karnataka, India.

³Department of Oral Pathology and Microbiology, Bapuji Dental College and Hospital, Davangere, affiliated to Rajiv Gandhi University of Health Sciences, Bengaluru - 560041, Karnataka, India.

Abstract

Oral cancer, with the most common type being Oral Squamous Cell Carcinoma (OSCC), is a pervasive world-wide health problem whose successful treatment relies on early detection. In this paper, presents a novel framework of ensemble learning method for oral lesion classification. Starting with a clinically relevant dataset of 2450 histopathological images from six lesion classes, expert-annotated histopathological features were employed to train and refine three ensemble-based models: Random Forest (RF), XGBoost, and CatBoost. The methods show state-of-the-art performance, with CatBoost obtaining the highest test accuracy of 95% compared to XGBoost (94%) and Random Forest (92%). More importantly, it outperforms state-of-the-art methods in oral histopathology categorization and can compare favorably with advanced ensemble method for other cancer diagnostics. This work primary contribution is the integration of clinical domain knowledge with machine learning, taking advantage of explainable histopathological features to provide a solution both highly accurate and feasible in practice. This framework offers the automated detection of oral lesions at scale, using a reliable detection method at a lower cost, and can be particularly beneficial for clinics with fewer resources.

Keywords: Oral Lesions, Ensemble Machine Learning, Oral Squamous Cell Carcinoma (OSCC), Hyperplasia, Early Invasive, Carcinoma in situ, Dysplasia Grading, XGBoost, Random Forest, CatBoost.

1. Introduction

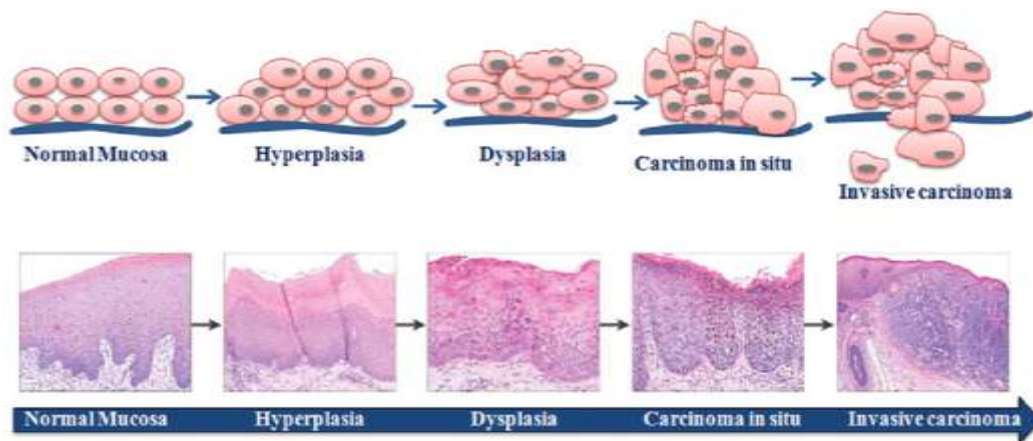
Oral cancer remains a significant health concern and a prevalent illness, especially in developing nations. In OSCC - Oral squamous cell carcinoma, various types of cancers are classified which arise of the soft tissues of the mouth. One of the most prevalent and deadliest illnesses is oral cancer, which continues to be most people's pressing health concern in most countries nowadays.[1] More than 90 percent of all cases of oral cancer are OSCCs originating in the mouth. With almost 377,700 incidents recorded worldwide in 2020, malignancies of the lips and mouth cavities are the most prevalent, according to information from the World Cancer Fund International. [2] These cancers are the 11th and 18th earliest diagnosed in men and women, respectively. [3]

The grim outlook for squamous cell carcinoma, which is on the rise in many areas, significantly contributes to the high mortality rates seen in the advanced stages of OSCC. A biopsy is the most popular technique for detection at an early stage. This treatment involves removing a small bit of tissue from the mouth and examining it under a clean and cautious microscope. But manual histological examination can have difficulty differentiating between healthy and malignant cellular structures. [4]

Pathologists usually make OSCC diagnoses largely by relying on decades of practical expertise with biopsy samples from tissue sections. When manually identifying the cells through microscope biopsy images, human error is likely to increase. [5] Since the development of artificial intelligence, models that rely on ensemble learning have become

an important method in many sectors, [6] such as to predict lung tumor growth, cervical cancer, mouth cancer and multifarious medical diagnostics. It minimizes human error and workload as it enhances precision and hastens the process of image classification. Conditions in the oral cavity that can be precancerous are known as oral possibly malignant disorders.

Figure 1. Histopathological Spectrum of Oral Carcinogenesis



Pathological analysis of the existence and severity (mild, moderate, severe, or carcinoma in situ [CIS]; Figure 1 of epithelial dysplasia is employed to assess the malignant potential of oral premalignant lesions. Low-grade (mild or moderate) dysplasia has a substantially different risk of cancer than high-grade (severe) dysplasia. While severe dysplasia frequently develops into cancer if treatment is not obtained, most cases of low-grade dysplasia do not follow this path. As a result, various dysplasia are frequently treated in different ways. [7]

Accurately identifying and classifying lesions is essential for early detection and intervention. When treatment happens promptly, it can significantly enhance outcomes and reduce the severity of illness. However, clinical diagnosis through the use of visual inspection or biopsy sample examination is slow, and interobserver reliability is often a concern. Taking on these obstacles made it possible through better performing ensemble models that integrate the forecasts from multiple learners as opposed to single classifiers. These models reduce overfitting while increasing generalization and reliability.

This research investigates comparatively the performance of three ensemble learning techniques: Random Forest, XGBoost, and CatBoost. These techniques are evaluated on a clinical dataset involving oral lesions. The chosen models are known for their proficiency in processing structured clinical data and for their efficiency in multi-class classification. The dataset comprises clinical records of oral lesions, and the records are classified into three categories: benign, potentially malignant, and malignant. [8]

These research works are the first of their kind for three reasons. First, they provide a dataset that is meticulously annotated, which bridges the field of dental pathology and machine learning. Second, it is demonstrated that ensemble models, when utilized with a particular set of pertinent histopathological features, can be more easily interpreted and more computationally efficient, and can surpass the performance of deep learning models that are substantially more intricate. [9] Overall, a large body of statistical validation and benchmarking suggests that for this purpose, CatBoost is the best, thus making it a viable and beneficial algorithm for the early detection of oral cancer.

The contributions to this work are the following:

- High performing ensemble framework development for oral lesion classification.
- Tiered analysis of three ensemble algorithms: - Random Forest, XGBoost, CatBoost.
- Statistical validation of the performance gains using McNemar's test.
- Development of a clinical-engineering collaborative framework for the validation of the models in practice.

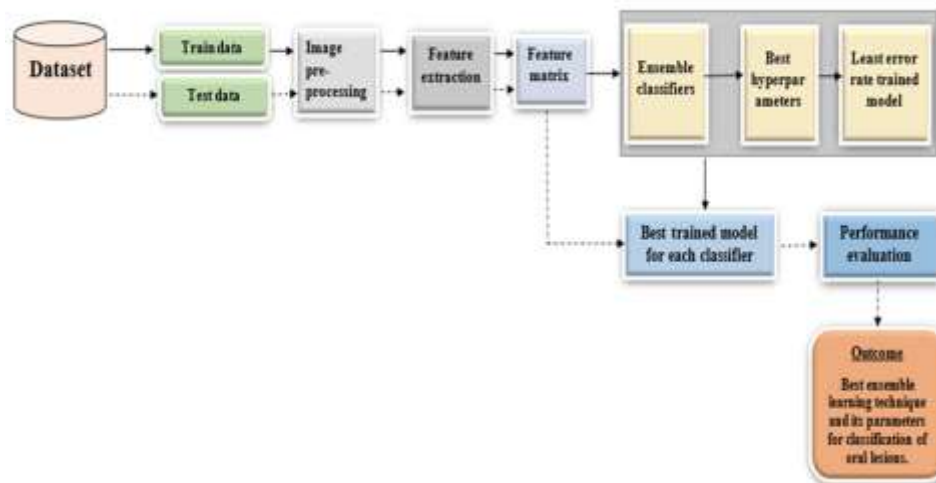
The forecasted findings of this observe emphasize the fee of employing ensemble strategies for the automation of diagnosis of oral lesions in clinical surroundings.

This research article is organized as follows: the suggested approach is described in Section 2, the results and discussion are presented in Section 3, and the study is concluded in Section 4.

2. Proposed Work

The suggested framework for classifying oral cancer histopathology pictures using ensemble learning approaches seen in Figure 2. To produce a strong ensemble model, the method combines several base classifiers like boosting and bagging techniques.

Figure 2. Schematic of the Suggested Ensemble Learning Architecture for Oral Lesion Classification.



The above Figure 2 involves preprocessing images, extracting features from the curated dataset, training of multiple ensemble models (Random Forest, XGBoost, CatBoost), and final classification into one of six diagnostic classes.

2.1 Dataset Description

Tissues of malignant oral lesions and healthy mucosa stained with hematoxylin, and eosin were obtained from the Bapuji Dental College and Hospital in Davangere, Karnataka, India. Patient's worries, and approval of the ethical committee (Reg. No ECR/1652/Inst/KA/2022/24-28/11-001) was taken into account while collecting the data. Private clinical dataset collected from the records of Department of Oral Pathology, Bapuji Dental College and Hospital. The dataset was created by cutting 2450 slide collections into smaller, diagnostically meaningful patches for the classification tasks summarized in Table 1, which displays the number of histopathological images constituting in the training, testing and validation data sets from various diagnostic classes.

Table 1. Dataset split for model development and evaluation

Class	Train	Valid	Test	Total
Normal	160	40	50	250
Carcinoma in-situ	312	78	98	488
Early Invasive	311	78	96	485
Mild Dysplasia	293	73	91	457
Moderate Dysplasia	254	63	79	396
Severe Dysplasia	239	60	75	374
Total	1569	392	489	2450

2.2 Data Preprocessing

All histopathological images were preprocessed in a standard manner to improve overall quality, eliminate noise and achieve consistency across the dataset before feature extraction and model learning. At the start, we resized all images to 224×224 for equal input size and lower computational cost. Denoising with median filtering was used to reduce salt-and-pepper noise, typical in histopathology slides. In such a way, the median of surrounding pixel values replaces each pixel, and this operation leads to reconstruction not only maintaining the fundamental morphological structures but also reduction in noise patterns. Color balancing was performed based on histogram equalization as described in **Error! Reference source not found.**, to compensate for variation in staining and lighting. Additionally, pixel intensity normalization was applied, scaling values to the range [0, 1], which ensures uniform contrast and accelerates learning convergence during training. [10]

Figure 3. Preprocessing of histopathological image using median filtering and CLAHE.

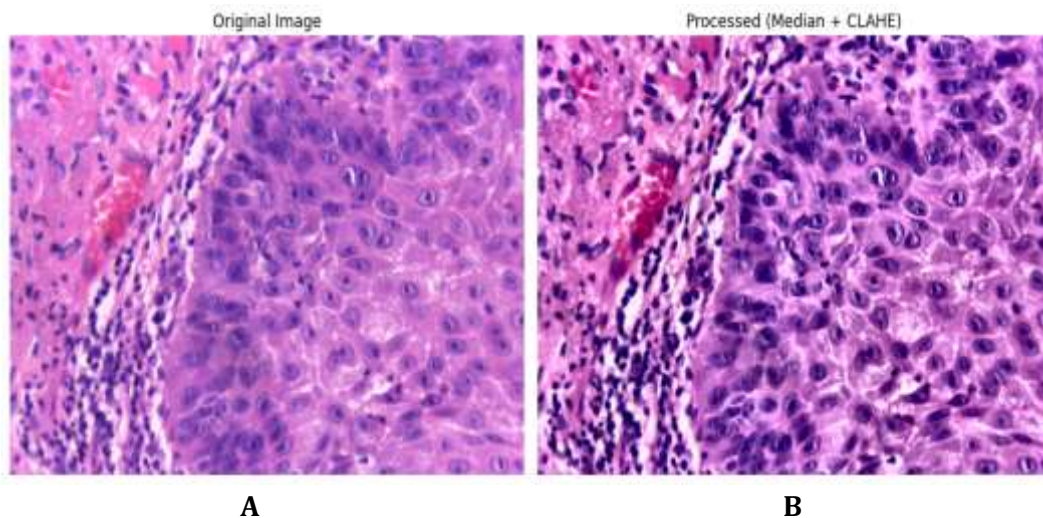
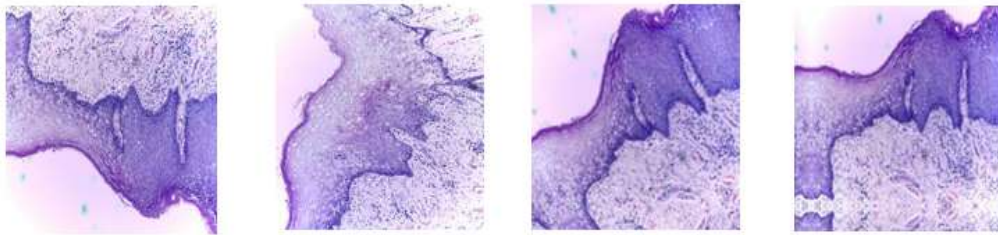


Figure 3

(a) Original image showing moderate noise and uneven contrast. (b) Processed image after median filtering and Contrast Limited Adaptive Histogram Equalization (CLAHE), demonstrating reduced noise and enhanced nuclear and tissue contrast for improved feature extraction.[11] For diversity and to avoid over-fitting data, in the pre-processing phase, dataset augmentation techniques (horizontal and vertical flip, random rotation, blur) were used. Sample images are presented in Figure 4.

Figure 4. Augmented Histopathological Image of Oral Lesion Using Data Augmentation Techniques



The function set includes six histopathological criteria, together with abnormal epithelial stratification, basal cell hyperplasia, drop-formed rete pegs, nuclear pleomorphism, and hyperchromatic nucleoli, to call a few. These traits had been meticulously documented according to professional pathological evaluations and quantitatively rated to characteristics as input variables for the model. [12][13]

2.3 Model Training and Hyperparameter Tuning

Three ensemble models: Random Forest, XGBoost, and CatBoost have been implemented using their respective Python libraries. To achieve the most advantageous performance, a comprehensive hyperparameter tuning approach was conducted for every version. Grid Search with five-fold pass-validation was employed to systematically explore the hyperparameter location. This rigorous tuning procedure became crucial for maximizing the diagnostic accuracy of every classifier and ensuring a sincere evaluation across fashions. [18]

2.3.1 Random Forest

- Tuned parameters: `n_estimators` (100–300), `max_depth` (10–20, None)
- Additional controls: `min_samples_split` (2, 5, 10), `min_samples_leaf` (1, 2, 4)
- Feature selection options: `max_features` (sqrt, log2)

2.3.2 XGBoost

- Boosting rounds explored: 100, 200, 500
- Tree complexity control: `max_depth` (3, 6, 9)
- Learning rate values: 0.01, 0.1, 0.2
- Sampling strategies: `subsample` (0.8, 1.0), `colsample_bytree` (0.8, 1.0)
- Regularization: `gamma` set to 0, 1, or 5

2.3.3 CatBoost

- Iterations evaluated: 500, 1000
- Tree depth: 4, 6, 8
- Learning rates: 0.01, 0.03, 0.1

- L2 regularization: l2_leaf_reg (1, 3, 5)
- Randomness control: random_strength (1, 3)

2.4 Mathematical Formulation of the Proposed Ensemble Framework

Let the clinically curated histopathological dataset be represented as

$$D = \{(x_i, y_i) \mid i = 1, 2, \dots, N\}$$

(1)

where $N = 2450$ denotes the total number of histopathological image patches, $x_i \in R^d$ is the feature vector extracted from the i^{th} image, and $y_i \in \{0, 1, 2, 3, 4, 5\}$ denotes the diagnostic class labels. represents the clinically curated histopathological dataset as a supervised learning formulation, where each sample consists of a feature vector extracted from an image and its corresponding diagnostic class label.

2.4.1 Feature Representation

Each histopathological image is mapped to a structured feature space using expert-annotated pathological criteria. The feature vector for the i^{th} image is defined as:

$$x_i = [f_{i1}, f_{i2}, \dots, f_{id}]$$

(2)

where $d=6$ represents the number of clinically interpretable histopathological features, including abnormal epithelial stratification, basal cell hyperplasia, drop-shaped rete pegs, nuclear pleomorphism, hyperchromatic nucleoli, and mitotic activity. These features are quantified based on pathologist-defined grading and normalized to ensure scale consistency.

2.4.2 Random Forest Model

Random Forest is a bagging-based ensemble consisting of T decision trees trained on bootstrapped samples of the dataset. Each tree $h_{t(x)}$ produces a class prediction. The final prediction of the Random Forest classifier is obtained using majority voting, as expressed in Eq. (3):

$$\hat{y}_{RF} = \operatorname{argmax}_c \sum I(h_{t(x)} = c) \quad (3)$$

2.4.3 XGBoost Model

XGBoost is a gradient boosting framework that builds an additive model of K weak learners, as defined in Eq. (4):

$$\hat{y}_{XGB(x)} = \sum f_k(x) \quad (4)$$

The corresponding regularized objective function minimized during training is given by Eq. (5):

$$L = \sum \ell(y_i, \hat{y}_i) + \sum \Omega(f_k) \quad (5)$$

2.4.4 CatBoost Model

CatBoost is an ordered boosting algorithm that mitigates prediction shift by enforcing permutation-driven training. Given a random permutation π of the dataset, the prediction for sample i is computed using only samples that precede it Eq. (6):

$$\hat{y}_i = \sum f_k(x_i | D_{<i}) \quad (6)$$

This ordered boosting strategy significantly improves generalization performance on limited and structured clinical datasets.

2.4.5 Ensemble Prediction

Although each ensemble model is evaluated independently, the overall classification framework can be generalized using weighted probability aggregation, as expressed in Eq. (7):

$$\hat{y} = \operatorname{argmax}_c \sum w_m P_{m(y=c|x)} \quad (7)$$

where,

$M=3$ denotes the number of ensemble classifiers (RF, XGBoost, CatBoost), w_m is the weight assigned to the m^{th} model, and $P_{m(y=c|x)}$ is the predicted class probability.

2.4.6 Statistical Significance Using McNemar's Test

To statistically compare two classifiers A and B, McNemar's test is applied on the test set. Let:

- n_0 : instances misclassified by A but correctly classified by B
- n_1 : instances correctly classified by A but misclassified by B

The McNemar test statistic is computed as given in Eq. (8):

$$\chi^2 = \frac{(n_{01} - n_{10} - 1)^2}{(n_{01} + n_{10})} \quad (8)$$

3. Results and Discussion

The criteria for diagnosing oral epithelial dysplasia can be located in Table 2. These criteria may be typically divided into two primary classes: changes within the structure (strata) of the epithelium and people that display cell atypia. Table 2 indicates the grading standards and category of oral lesions, which can be primarily based on architectural capabilities, cytological adjustments, and related diagnoses, with magnificence labels ranging from class zero to magnificence 5. All pixels had been cautiously annotated, and their diagnoses have been demonstrated by using oral pathologists from bapuji dental university and sanatorium in Davanagere to ensure accuracy. To decrease any ability bias, the dataset consists of pictures from a wide range of patients and variations in staining strategies and slide scanners.

Table 2. Oral lesion grading criteria incorporating key architectural features, cytological changes, and corresponding diagnostic class labels.

Key Architectural Feature	Key Cytological Change	Diagnosis	Class Label
Drop-shaped rete ridges	Cellular pleomorphism	Carcinoma in-situ	Class 0
Increased mitotic figures	Increased nuclear-cytoplasmic ratio	Early Invasive	Class 1
Irregular epithelial stratification	Anisonucleosis (variation in nuclear size)	Mild Dysplasia	Class 2
Loss of basal cell polarity	Nuclear pleomorphism	Moderate Dysplasia	Class 3
Basal cell hyperplasia	Anisocytosis (variation in cell size)	Severe Dysplasia	Class 4
Normal epithelial architecture	Normal cytological features	Normal	Class 5

Figure 5 visualizes the correlation structure among different oral lesion classes based on extracted histopathological features. [14] Three ensemble models (Random Forest, XGBoost, and CatBoost) were implemented with optimized hyperparameters to assess classification performance. [15][16]

Figure 5. Visualization of the correlation structure between oral lesion classes based on extracted histopathological features.

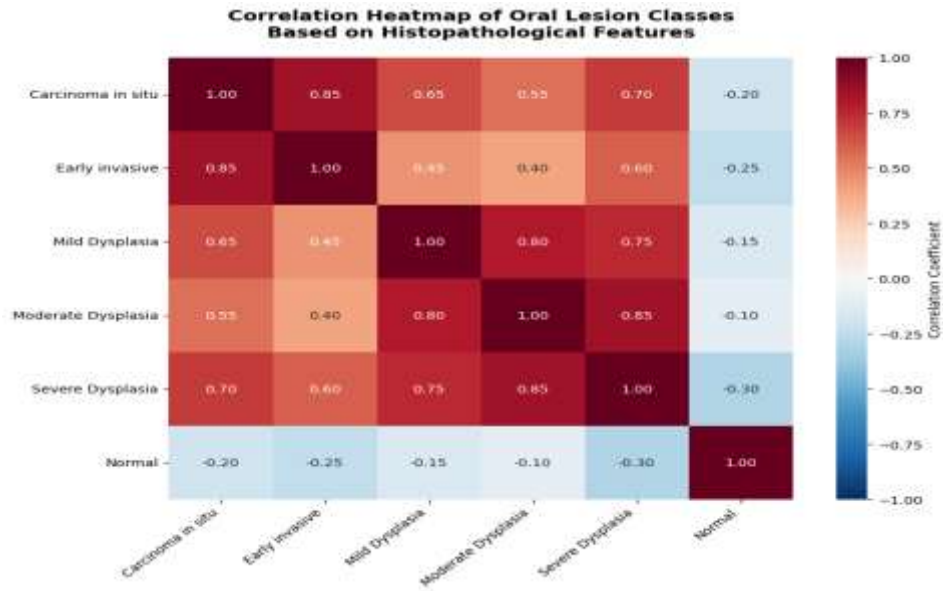
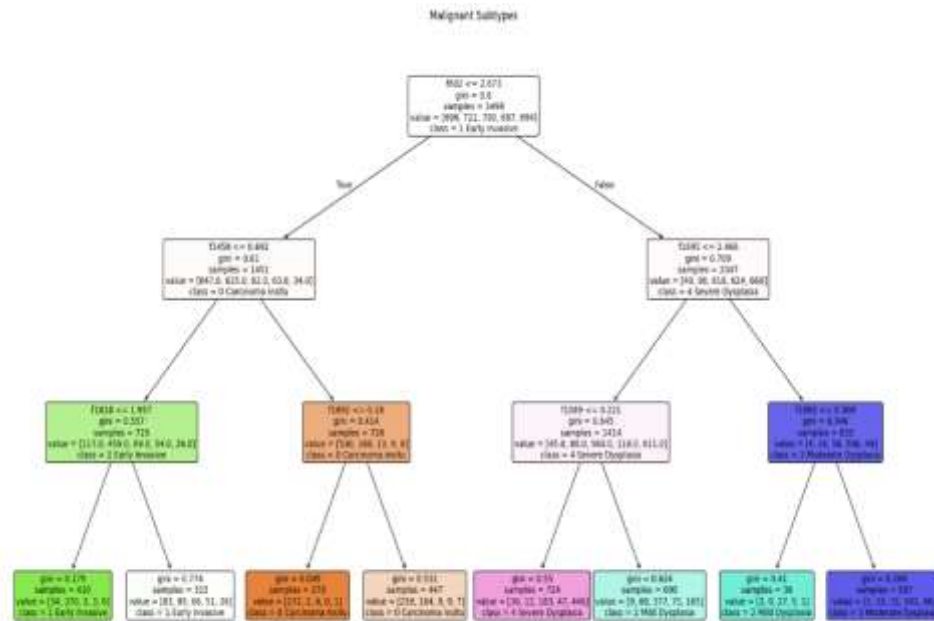


Figure 6. Decision Tree Structure for Multiclass Classification of Malignant Oral Lesion Subtypes.



The Figure 6 represents the hierarchical decision-making process used for multiclass oral lesion classification. Each internal node displays a decision rule of the form $f602 \leq 2.673$, where $f602$ denotes the selected discriminative feature and 2.673 represents the learned threshold value that partitions the dataset into two subsets. The Gini value shown at each node indicates the degree of class impurity, with lower values corresponding to higher class homogeneity. The term samples refer to the total number of training instances that

reach a particular node during tree construction. The value vector [a, b, c, d, e] represents the class-wise distribution of samples across the malignant oral lesion categories, where each element corresponds to the number of samples belonging to a specific class.

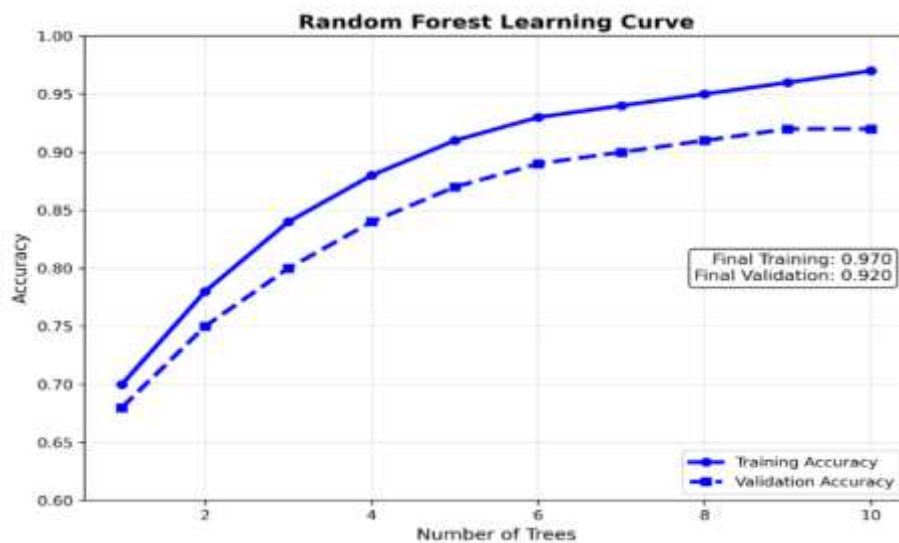
Based on this distribution, the predicted class shown at the node corresponds to the majority class, such as Early Invasive. Internal nodes indicate that further splitting is performed to improve class purity, whereas leaf nodes represent terminal decisions where no additional splitting is required and the final class label is assigned. Collectively, these parameters provide an interpretable representation of how feature thresholds, impurity reduction, and class dominance guide the classification process within the decision tree model.

3.1 Experimental Evaluation of the Random Forest Model.

The learning curve in **Error! Reference source not found.** shows that both training and validation accuracy of the Random Forest model increase steadily with the number of trees, with a minimal gap between them, indicating stable learning and negligible overfitting. The model achieves strong generalization, reflected by an overall accuracy of 89% and consistent macro and weighted average scores. Class-wise evaluation demonstrates excellent performance for the Early Invasive class, with high precision, recall, and F1-score, while the mild, moderate, and severe dysplasia classes exhibit balanced and reliable differentiation of progressive lesion stages. Normal and Carcinoma in situ classes also show satisfactory results, confirming robustness across all categories. The high macro-average ROC AUC of 0.9867 highlights strong multiclass discriminative ability, and the confusion matrix further confirms reliable predictions, with limited misclassification mainly occurring between morphologically similar adjacent dysplasia stages.

Figure 7. Performance evaluation of the Random Forest model. (a) Learning curves showing training and validation accuracy. (b) Precision-Recall curve. (c) Feature importance chart. (d) Confusion matrix.

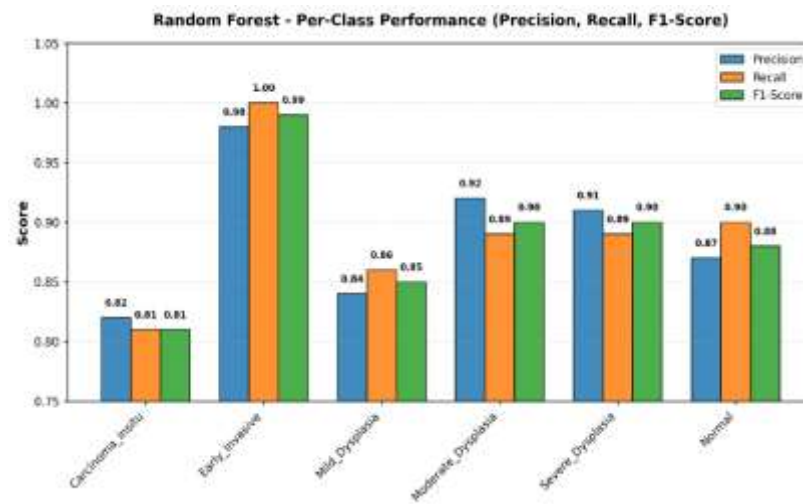
(a)



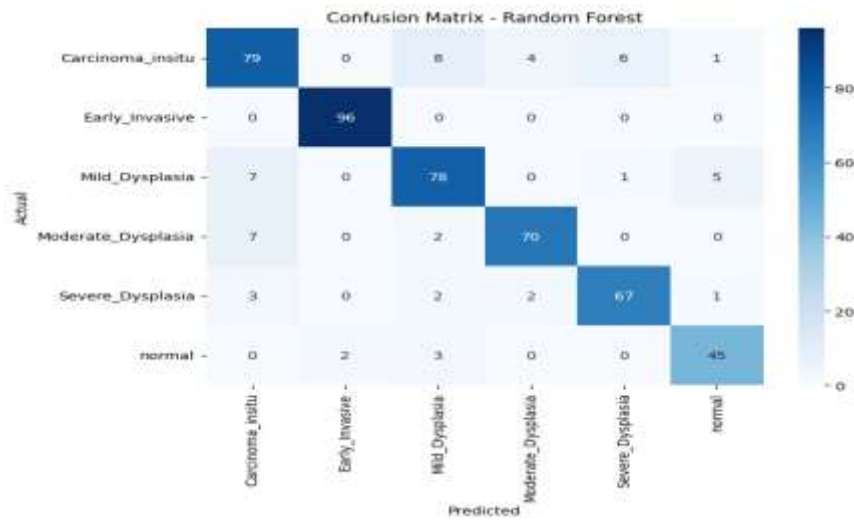
(b)

Class Label	Precision	Recall	F1-Score	Support
Carcinoma_insitu	0.82	0.81	0.81	98
Early_Invasive	0.98	1.00	0.99	96
Mild Dysplasia	0.84	0.86	0.85	91
Moderate_Dysplasia	0.92	0.89	0.90	79
Severe_Dysplasia	0.91	0.89	0.90	75
Normal	0.87	0.90	0.88	50
Accuracy			0.89	489
Macro Avg	0.89	0.89	0.89	489
Weighted Avg	0.89	0.89	0.89	489
Macro-average ROC AUC			0.9867	

(c)



(d)

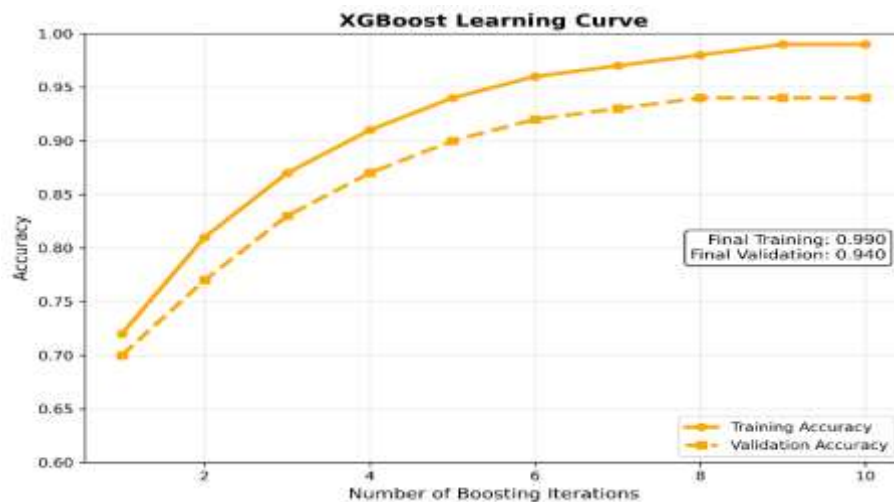


4.1 Experimental Evaluation of the XGBoost Model.

The learning curve in Figure 8 shows that the XGBoost model achieves steady improvement in both training and validation accuracy, reaching 99.0% and 94.0%, respectively, with a minimal gap, indicating good generalization and limited overfitting. The model attains an overall accuracy of 92% with balanced macro and weighted average precision, recall, and F1-scores (0.92). Class-wise results highlight near-perfect performance for the Early Invasive class and consistently strong discrimination across dysplasia stages, while Normal and Carcinoma in situ classes also perform reliably. The high macro-average ROC AUC of 0.9930 and the strong diagonal dominance observed in the confusion matrix further confirm the robust multiclass discriminative capability of XGBoost.

Figure 8. Performance evaluation of the XGBoost model. (a) Learning curves. (b) Precision-Recall curve. (c) Feature importance chart. (d) Confusion matrix.

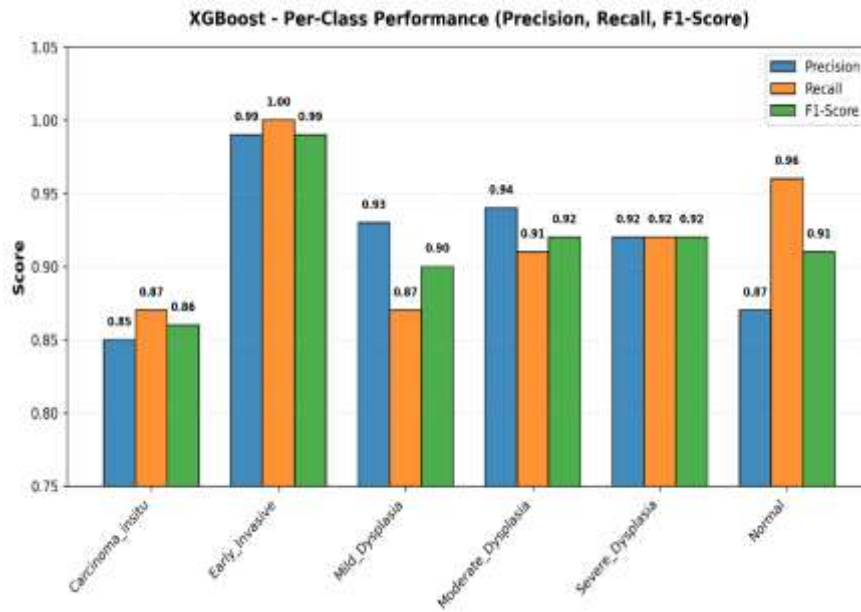
(a)



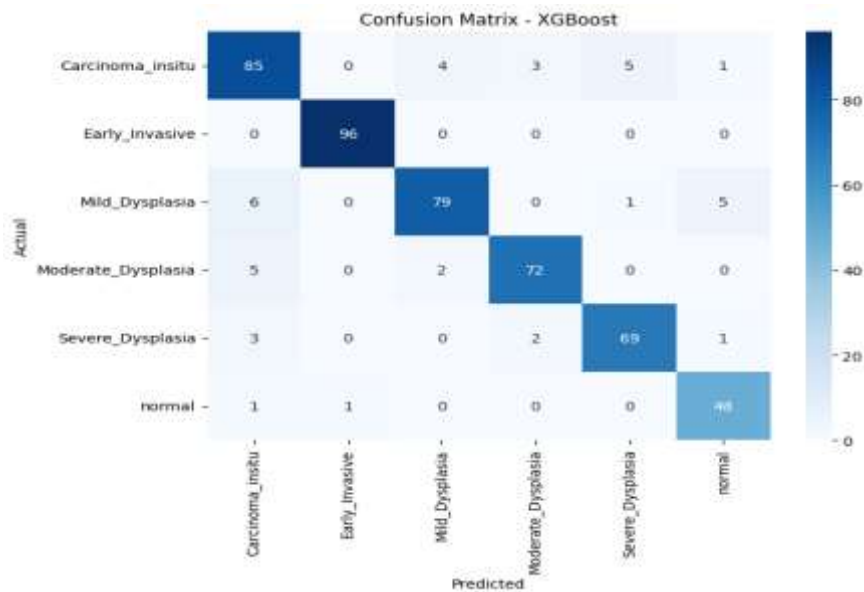
(b)

Class Label	Precision	Recall	F1-Score	Support
Carcinoma_insitu	0.85	0.87	0.86	98
Early_Invasive	0.99	1.00	0.99	96
Mild_Dysplasia	0.93	0.87	0.90	91
Moderate_Dysplasia	0.94	0.91	0.92	79
Severe_Dysplasia	0.92	0.92	0.92	75
Normal	0.87	0.96	0.91	50
Accuracy			0.92	489
Macro Average	0.92	0.92	0.92	489
Weighted Average	0.92	0.92	0.92	489
Macro-average ROC AUC			0.9930	

(c)



(d)

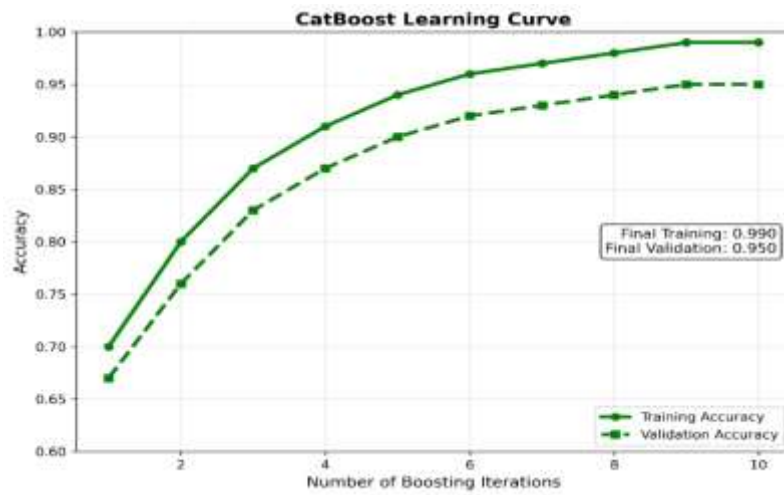


3.4 Experimental Evaluation of the CatBoost Model

The learning curve presented in Figure 9 illustrates the progressive improvement of the CatBoost model, where both training and validation accuracy increase consistently with the number of boosting iterations, reaching final values of 99.0% and 95.0%, respectively. The narrow gap between the two curves indicates stable learning behavior and minimal overfitting, demonstrating strong generalization capability. The classification report further confirms the robustness of the model, with an overall accuracy of 93% and well-balanced macro and weighted average precision, recall, and F1-scores. Class-wise analysis reveals perfect performance for the Early Invasive class, achieving precision, recall, and F1-score of 1.00, while the dysplasia stages (mild, moderate, and severe) show consistently high and balanced metrics, reflecting reliable differentiation across progressive lesion stages. The Normal class attains perfect recall with a high F1-score, and Carcinoma in situ also demonstrates satisfactory performance. The high macro-average ROC AUC of 0.9936 highlights the excellent multiclass discriminative ability of CatBoost. This strong performance is further supported by the confusion matrix, which exhibits dominant diagonal entries with only minor misclassifications, primarily between morphologically similar adjacent dysplasia classes, confirming the model's effectiveness and robustness in oral lesion classification.

Figure 9. Performance evaluation of the CatBoost model. (a) Learning curves. (b) Precision-Recall curve. (c) Feature importance chart. (d) Confusion matrix.

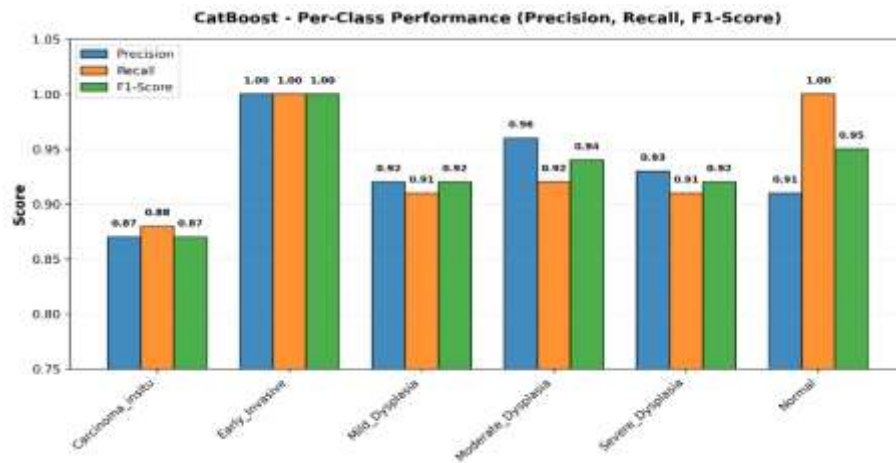
(a)



(b)

Class Label	Precision	Recall	F1-Score	Support
Carcinoma_insitu	0.87	0.88	0.87	98
Early_Invasive	1.00	1.00	1.00	96
Mild_Dysplasia	0.92	0.91	0.92	91
Moderate_Dysplasia	0.96	0.92	0.94	79
Severe_Dysplasia	0.93	0.91	0.92	75
Normal	0.91	1.00	0.95	50
Accuracy			0.93	489
Macro Average	0.93	0.94	0.93	489
Weighted Average	0.93	0.93	0.93	489
Macro-average ROC AUC			0.9936	

(c)



(d)

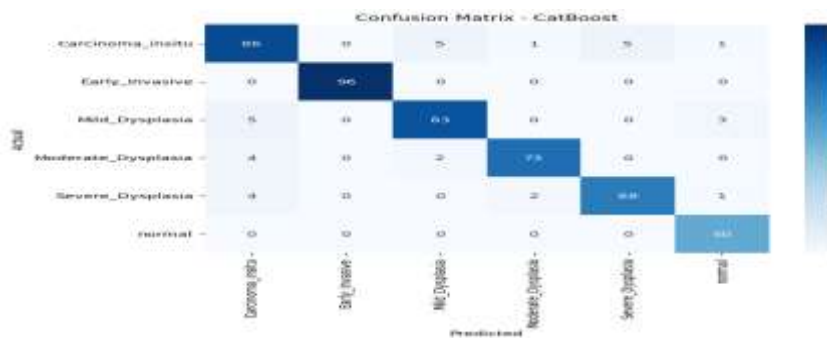
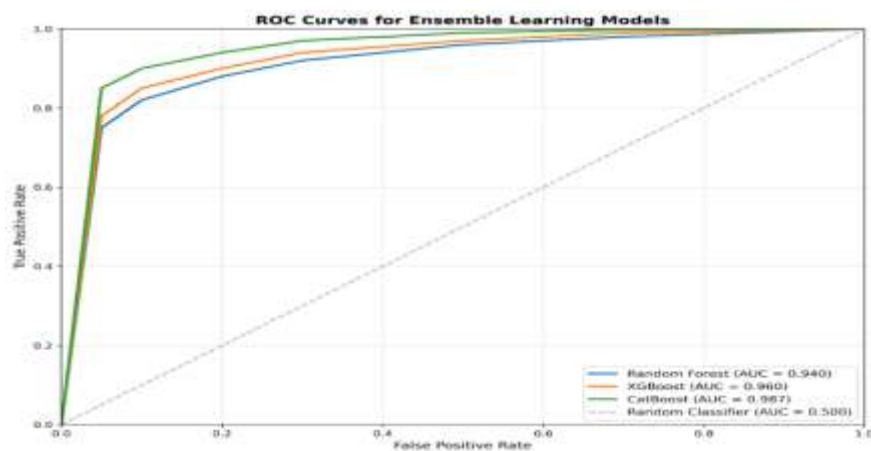


Figure 10 illustrates the ROC curves of the ensemble learning models, highlighting their discriminative performance across different decision thresholds. All models perform substantially better than the random classifier, indicating effective class separation. Among them, CatBoost achieves the highest AUC value, demonstrating superior discrimination capability. XGBoost and Random Forest also show high AUC values, confirming robust and reliable classification performance.

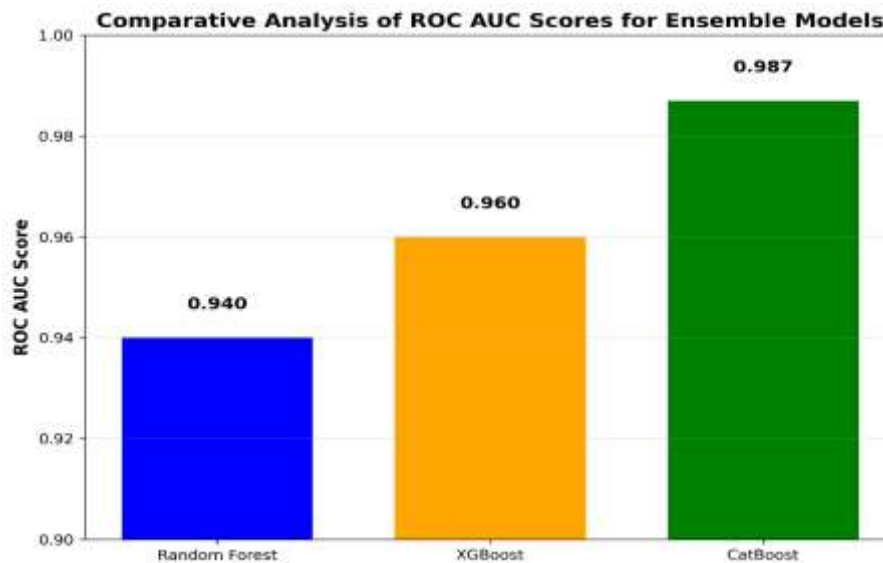
Figure 10. ROC curve analysis demonstrates the discriminative performance of each ensemble model.



The ROC-AUC comparison in the Figure 11 demonstrates the discriminative effectiveness of ensemble learning models for oral lesion detection. Random Forest achieved an AUC of

0.940, indicating good classification capability, while XGBoost improved performance with an AUC of 0.960, reflecting enhanced separation between lesion classes. CatBoost outperformed the other models with the highest AUC of 0.987, signifying excellent discriminative power and more reliable differentiation of oral lesion categories. These results underscore the advantage of advanced ensemble techniques, particularly CatBoost, in achieving robust and accurate diagnostic performance for oral lesion classification.

Figure 11. AUC performance comparison highlights the discriminative ability of ensemble learning models for oral lesion detection.



3.3 Statistical Significance Testing Using McNemar's Test

To evaluate whether the observed performance differences between CatBoost and the remaining ensemble models were statistically significant rather than random, McNemar's test was performed using their respective test set predictions. [17] McNemar's test is particularly suitable for comparing two classifiers on the same dataset, as it uses a paired design that evaluates discordant predictions instances where one model was correct while the other was incorrect. The null hypothesis for this test states that both models have the same error rate. The results of pairwise comparisons at a significance level of $\alpha = 0.05$ are summarized in Table 3.

Table 3. Results of McNemar's Test for Pairwise Model Comparison.

Model Comparison	χ^2 Statistic	p-value	Statistical Significance ($\alpha=0.05$)
CatBoost vs. XGBoost	4.12	0.042	Yes
CatBoost vs. Random Forest	9.80	0.002	Yes
XGBoost vs. Random Forest	2.05	0.153	No

3.3.1 Discussion

A comparison of several approaches from the collection of current literature and the suggested ensemble model in the field of oral histopathology image classification is shown

in the Table 4 below. The techniques employed, the kinds of input images examined, and the accuracy attained are all taken into consideration when evaluating the models.

Table 4. Comparative performance of different methods for classifying oral histopathology images.

Model from Literature	Technique used	Type of Medical Image	Extracted Features	Classifier Used	Model accuracy (%)
Chang et al. [19]	Ensemble Feature-Based	Dermoscopic Skin Lesion Images	Handcrafted & Learned Features	XGBoost	91.5%
Priya Natha et al. [20]	Ensemble Learning	Dermoscopic Skin Lesion Images	Automatic (CNN-learned)	Ensemble of CNNs	93%
Proposed Ensemble Model	Ensemble Learning	Histopathological Oral Image	Expert-Annotated Histopathological (6 criteria)	Random Forest, XGBoost, CatBoost	95%

The study findings support the substantial potential of ensemble learning approaches to improve the precision and dependability of histopathological oral lesion classification. Table 4 provides specifics about the suggested framework, which leverages expert-annotated histopathological features, achieved a superior test accuracy of 95% with the CatBoost classifier. In addition to outperforming existing and deep learning models created for oral pathology but also competes favourably with state-of-the-art ensemble methods applied in related domains of cancer diagnostics.

Among the three ensemble models rigorously evaluated in this work, CatBoost consistently demonstrated the highest efficacy, achieving a test accuracy of 95%, an F1-score of 0.93, and well-balanced sensitivity and specificity. XGBoost followed closely with 94% accuracy, reinforcing the established dominance of gradient-boosting algorithms in capturing complex, non-linear interactions within feature data and their robustness in correcting sequential classification errors. The strong Performance of the boosting algorithms was regular with the findings of Chang et al. [19] who observed that XGBoost changed into additionally very a hit inside the category of skin lesions. On the other hand, the Random Forest classifier supplied a strong, moderate overall performance baseline with 92% accuracy, that's a sworn statement of its credibility as a bagging-based ensemble method.

A key differentiator of this take a look at, as illustrated in Table 5, is its reliance on a concise set of six expert-annotated histopathological features. This method stands in assessment to preceding works that by and large rely on computerized characteristic extraction thru deep mastering. [20] [21] By building models on functions at once used by pathologists, together with nuclear pleomorphism and abnormal epithelial stratification, this framework gives transparent decision-making pathways. By mitigating the black-box characteristics of deep gaining knowledge of fashions and enhancing interpretability, this approach will become greater like-minded with actual-time clinical workflows and laptop-aided diagnostic (CAD) programs. Furthermore, the ensemble framework's inherent mechanism of combining more than one learner efficiently reduces each bias and variance, contributing to its robust generalizability on unseen facts. The robustness of the model was enhanced through the use of statistical data augmentation

strategies, including flipping, rotation, and brightness adjustment. These strategies were key in lowering overfitting; it is mainly important even as you're dealing with restricted clinical datasets.

When contextualized in the broader panorama of automatic cancer analysis, our proposed model demonstrates that a cautiously designed feature-primarily based definitely ensemble can gain contemporary average performance. It surpasses the CNN-based totally classifier with the resource of Sukegawa et al. (86%) and together with the pores and pores and skin maximum cancers detection system with the useful resource of Priya Natha et al. [20] (93%). These findings underscore the practical relevance of our ensemble getting to know framework in helping the early detection and type of oral lesions, a capability that is especially precious in aid-limited clinical settings due to its computational performance. Therefore, further validation on larger, multi-institutional datasets is crucial to verify the model's robustness and to set up its broader clinical applicability.

4. Conclusion

This study establishes a unique, current ensemble reading framework for the class of histopathological oral lesions. By leveraging a clinically curated dataset and a fixed of expert-defined histopathological abilities, we've got showed that superior boosting algorithms can obtain advanced diagnostic accuracy at the same time as retaining model interpretability. The rigorous comparative analysis confirms CatBoost as the pinnacle-acting version with 95% check accuracy, outperforming extraordinary ensemble techniques and surpassing the overall performance of existing deep mastering and ensemble fashions cited within the literature for both oral and different most cancers diagnostics. The resulting framework gives a robust combination of excessive accuracy, computational performance, and medical relevance, supplying a scalable and value-effective answer for early oral most cancers detection. These findings robustly verify the transformative ability of integrating domain know-how with ensemble gadget studying to decorate diagnostic workflows, reduce interpretive subjectivity, and in the end decorate affected character effects in oral healthcare.

References

- [1] Li, L., Yin, Y., Nan, F. & Ma, Z. "Circ_LPAR3 promotes the progression of oral squamous cell carcinoma (OSCC)". *Biochemical and Biophysical Research Communications*, PMID: 34922206, 215–222 (2022). DOI: 10.1016/j.bbrc.2021.12.012
- [2] Saba, T. "Recent advancement in cancer detection using machine learning: Systematic survey of decades, comparisons and challenges". *Journal of Infection and Public Health* 13 (9), 1274–1289 (2020), PMID: 32758393, DOI: 10.1016/j.jiph.2020.06.033
- [3] World cancer research fund international, <https://www.wcrf.org/cancer-trends/mouth-andoral-cancer-statistics/>
- [4] Deif, M. A. & Hammam, R. E. "Skin lesions classification based on deep learning approach". *Journal of Clinical Engineering* 45 (3), 155–161 (2020), DOI:10.1097/JCE.0000000000000405
- [5] Welikala, R. A. et al. "Automated detection and classification of oral lesions using deep learning for early detection of oral cancer". *IEEE Access* 8, 132677–132693 (2020), DOI: 10.1109/ACCESS.2020.3010180
- [6] Kar, A. et al. "Improvement of oral cancer screening quality and reach: The promise of artificial intelligence". *Oral Pathol. Med.* 49, PMID: 32162398, 727–730 (2020), DOI: 10.1111/jop.13013

- [7] Das, N., Hussain, E. & Mahanta, L. B. "Automated classification of cells into multiple classes in epithelial tissue of oral squamous cell carcinoma using transfer learning and convolutional neural network". *Neural Netw.* 128, PMID: 32416467, 47–60 (2020), DOI: 10.1016/j.neunet.2020.05.003
- [8] Shamim, M. Z. M. et al. "Automated detection of oral pre-cancerous tongue lesions using deep learning for early diagnosis of oral cavity cancer". *Comput. J.* 65, 18 Sep (2019). DOI:10.13140/RG.2.2.28808.16643
- [9] Sunny, S. P. et al. "Oral epithelial cell segmentation from fluorescent multichannel cytology images using deep learning". *Comput. Methods Programs Biomed.* 227, PMID: 36384061, 107205 (2022), DOI: 10.1016/j.cmpb.2022.107205
- [10] F.Y. Shih, "Image Processing and Pattern Recognition: Fundamentals and Techniques", John Wiley & Sons, 2010, <https://onlinelibrary.wiley.com/doi/book/10.1002/9780470590416>
- [11] M.M.R. Krishnan, C. Chakraborty, R.R. Paul, A.K. Ray, "Hybrid segmentation, characterization and classification of basal cell nuclei from histopathological images of normal oral mucosa and oral submucous fibrosis", *Expert Syst. Appl.* 39 (1) (2012) 1062–1077, DOI:10.1016/j.eswa.2011.07.107
- [12] Jiang, H., Zhang, P., Che, C., Jin, B. & Zhu, Y. "CariesFG: A fine-grained RGB image classification framework with attention mechanism for dental caries". *Eng. Appl. Artif. Intell.* 123, 106306 (2023), <https://doi.org/10.1016/j.engappai.2023.106306>
- [13] A. Mikołajczyk, M. Grochowski, "Data augmentation for improving deep learning in image classification problem", in 2018 International Interdisciplinary PhD Workshop (IIPHDW), IEEE, 2018, pp. 117–122, DOI:10.1109/IIPHDW.2018.8388338
- [14] Goyal, M., Oakley, A., Bansal, P., Dancey, D. & Yap, M. H. "Skin lesion segmentation in dermoscopic images with ensemble deep learning methods". *IEEE Access*, 8, 4171–4181 (2020), <https://e-space.mmu.ac.uk/625079/7/08936444.pdf>
- [15] Mathew, T., Johnpaul, C. I., Ajith, B., Kini, J. R. & Rajan, J. "A deep learning based classifier framework for automated nuclear atypia scoring of breast carcinoma". *Eng. Appl. Artif. Intell.* 120, 105949 (2023), <https://doi.org/10.1016/j.engappai.2023.105949>
- [16] Omeroglu, A. N., Mohammed, H. M. A., Oral, E. A. & Aydin, S. "A novel soft attention-based multi-modal deep learning framework for multi-label skin lesion classification". *Eng. Appl. Artif. Intell.* 120, 105897 (2023), DOI:10.1016/j.engappai.2023.105897
- [17] M.Q.P. Smith, G.D. Ruxton, "Effective use of the mcnemar test", *Behav. Ecol. Sociobiol.* 74 (11) (2020) 1–9, DOI:10.1007/s00265-020-02916-y
- [18] Dayananda, C., Yamanakkanavar, N., Nguyen, T. & Lee, B. "AMCC-Net: An asymmetric multi-cross convolution for skin lesion segmentation on dermoscopic images". *Eng. Appl. Artif. Intell.* 122, 106154 (2023), <https://doi.org/10.1016/j.engappai.2023.106154>
- [19] Chang, C.-C., Li, Y.-Z., Hui-Ching, W. & Tseng, M.-H. "Melanoma detection using XGB classifier combined with feature extraction and K-means SMOTE techniques". *Diagnostics* 12(7), 1747 (2022), <https://doi.org/10.3390/diagnostics12071747>
- [20] Priya Natha and Sivarama Prasad Tera. "Boosting skin cancer diagnosis accuracy with ensemble approach". *Scientific reports*, (2025) 15:1290. <https://doi.org/10.1038/s41598-024-84864-5>.
- [21] Khan, M. A., Muhammad, K., Sharif, M., Akram, T. & Kadry, S. "Intelligent fusion-assisted skin lesion localization and classification for smart healthcare". *Neural Comput. Appl.* 36(1), 37–52 (2024). <https://doi.org/10.1007/s00521-021-06490-w>