Journal of Theoretical and Applied Information Technology
15th May 2026. Vol.104. No.9

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

EARLY DETECTION OF THROAT AND OROPHARYNGEAL CANCER USING YOLO-DRIVEN FEATURES AND RF–XGBOOST

SR AJITHA[1] , DIPANWITA DEBNATH[2] , CHALUMURU SURESH[3], DR M SURYA BHUPAL RAO[4] , P.V.S MURALI KRISHNA[5], D.NIROSHA[6], J.SWETHA PRIYANKA[7], JOHN T MESIA DHAS[8], Dr SIVA KUMAR PATHURI[9]

[1]Guest Faculty, Department of Computer Science, SVU College of CM &CS, Sri Venkateswara University, Tirupathi, Andhra Pradesh, India.
[2]Assistant Professor, Department of CSE (AI & ML), B V Raju Institute of Technology, Narsapur, Medak, Telangana, India-502313
[3]Assistant Professor, Department of Computer Science & Engineering-AIML, VNR Vignana Jyothi Institute of Engineering and Technology, Bachupally, Hyderabad, India.
[4]Associate Professor, Department of CSE (AI&ML), CVR College of Engineering ,Hyderabad, India.
[5]Assistant Professor, Department of Mechanical Engineering, Aditya University, Surampalem, AP, India.
[6]Assistant Professor ,Department of Computer Science and Engineering (Data science), Santhiram Engineering College, NH-40, Nandyal -518501,AP,India.
[7]Assistant Professor,Department of ECE,Vardhaman College of Engineering, Hyderabad ,India.
[8]Associate Professor, Department of Artificial Intelligence & Data Science, School of Computing, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science & Technology, Chennai, India.
[9]Professor, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India.

Email: [1]ajitha1428@gmail.com,[2]dipanwita.d@bvrit.ac.in,[3]suresh_ch@vnrvjiet.in, [4]suryabhupal@gmail.com ,[5]muralikrishna.peyyety@acet.ac.in, [6]nirosha.ds@srecnandyal.edu.in, [7]swethapriyanka1823@gmail.com ,[8]jtmdhasres@gmail.com ,[9]spathuri@kluniversity.in

ABSTRACT

Oropharyngeal cancer, or throat cancer, is a significant health problem in the world, as it is usually diagnosed in the later stages when it is less treatable and the chances of survival of a patient is greatly reduced. This research offers a novel method of early diagnosis of throat cancer, through the combination of machine learning algorithms with medical data processing. Predictive oncology is a highly promising field of artificial intelligence use, which is difficult because of its application in various areas, including better patient outcomes and early diagnosis. This paper presents a new hybrid architecture named YOLO-RFXGB. The model is a combination of real-time object detection of YOLO (You Only Look Once) and the strong ensemble learning of RFXGB, which incorporates both a Random Forest and XGBoost Gradient Boosting algorithms. The proposed system utilizes a three-stage pipeline: (i) data acquisition and preprocessing, (ii) feature extraction and lesion localization with the help of YOLO, and (iii) classification and prediction of prognosis with the help of RFXGB. Its results, tested on multimodal medical images, are compared to the existing deep learning models like ResNet-101 and RegNetY. Experimental findings indicate that the proposed YOLO-RFXGB model greatly outperforms these baseline methods, with the high prediction accuracy of up to 96%. These results indicate that it has a great potential of clinical implementation in early-stage cancer screening systems as it is a good tool that can be used to enhance the quality of diagnosis and patient survival rate.
Keywords: *Throat Cancer, YOLO, RF+XGB, ResNet-101, RegNetY*

1. INTRODUCTION

Head and Neck Cancer (HNC) is a heterogeneous group of malignancies that have different etiologies, biological behaviors, and prognostic outcomes. Of them, the most common histological subtype is the squamous cell carcinoma. Although therapeutic approaches have evolved considerably over the past

decades, an impressive percentage of HNC patients still report at advanced stages of the disease thereby severely limiting case survival rates despite the treatment. Therefore, the search of early diagnosis is an urgent issue in enhancing clinical outcomes and patient prognosis. Machine Learning (ML), Deep Learning (DL)[1], and Artificial Intelligence (AI) are novel approaches in the field of predictive oncology in recent years and provided the ability to process high-dimensional medical data, which is complex to analyze, with impressive efficiency. Expanding on this possibility, the current study presents a new visual diagnostic system of the classification of early-stage oropharyngeal carcinoma. The suggested system utilizes the YOLO (You Only look Once) deep extraction framework, whereas the classification is obtained via a hybrid ensemble system comprising Random Forest and XGBoost (RFXGB). The combination is meant to capitalize on both algorithms, with the aim being to offer high predictive performance, as well as improve the accuracy of the early detection of cancer. Throat cancer is a condition associated with pharyngeal muscles vital in breathing, speaking, and swallowing and includes nasopharyngeal, oropharyngeal, and hypopharyngeal carcinoma [2]. Figure 1 represents the anatomy of the parts of the throat in cancer. It is important to note that early diagnosis is the key to the success of the treatment since education and early intervention contribute to better prognosis[3] and survival. But the traditional methods of diagnosis, such as physical examination, endoscopy, CT[4], MRI, PET-CT imaging and histopathological biopsy, have disadvantages in form of invasiveness, high cost, restricted availability, slow diagnosis, and reliance on the interpretation of an expert[5].

To overcome these shortcomings, the proposed YOLO-RFXGB classifier uses a filtered medical image set that was acquired in the Kaggle repository. The data set is pre-processed with the help of two classes:

Class 0: Non-Oropharyngeal Cancer.

Class 1: Oropharyngeal Cancer

YOLO is used to localize regions of interest and extract high-resolution spatial features using medical images and the RFXGB classifier combines the advantages of the Random Forest and XGBoost into the creation of a formidable and sound classification. This hybrid architecture will be in place to ensure that maximum detection accuracy is achieved, the false positive is minimized, and that the hybrid architecture should support real time screening applications. The presence of clinical features of oral and throat cancer including persistent sore throat, unexplainable weight loss, oral lumps[6], and leukoplakia tend to mimic benign entities, and thus early diagnosis can be difficult. YOLO-RFXGB framework satisfies this gap in diagnosis by providing a scalable, non-invasive, and high-accuracy, automated screening solution to be used in large-scale clinical environments. The experimental results show that the proposed model is significantly better at predicting and with higher accuracy, sensitivity and specificity when comparing with classifiers that are considered baseline such as the ResNet-101 and RegNetY[7]. The general workflow of the proposed system is shown in figure 2. Moreover, the present research questions include:

- Effect of early diagnosis on treatment outcomes,
- Weaknesses of conventional screening methods,
- Emergent tendencies in AI-based cancer diagnosis, and
- Clinical deployment: accessibility, economic, and ethical issues.

The YOLO-RFXGB classifier, which combines both approaches to deep learning and ensemble learning methods, represents the potential solution to improving the diagnosis and prognosis of throat cancer[8]. Such a framework enhances the reliability of the diagnostic process and opens the path to the future of AI-based healthcare services[9], which will aid in earlier detection, the planning of treatment, and the increase in patient survival rates[10].

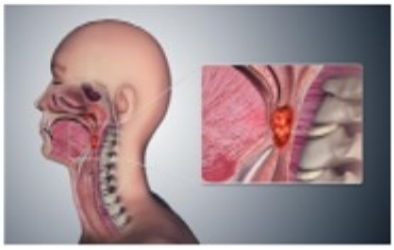

Figure 1: Shows Oropharyngeal Cancer

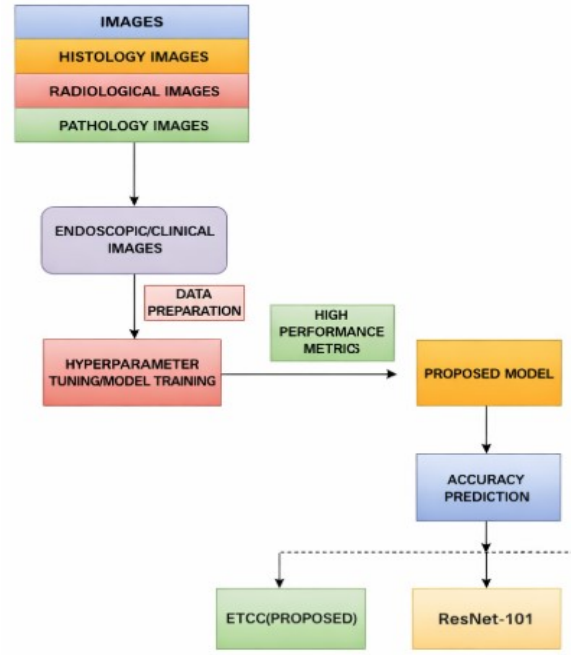

Figure. 2. System Flow of Proposed Classifier

2. RELATED WORK:

The recent developments in artificial intelligence have had a massive effect on the early detection of head and neck and oropharyngeal cancer using medical images. Many researches have investigated machine learning, deep learning, as well as hybrid models, to improve diagnostic accuracy, robustness, and clinical applicability. Recent years have witnessed significant advancements in the application of machine learning and deep learning techniques for cancer detection, particularly in the domain of medical imaging. Researchers have explored diverse architectures, optimization strategies, and multimodal approaches to enhance diagnostic accuracy, interpretability, and clinical applicability.

Convolutional Neural Networks (CNNs) have been widely adopted for histopathological and radiological image analysis. Rao et al. (2023) developed a CNN-based classification system using histopathological images from a public dataset, achieving 94% sensitivity and 94% specificity, thereby demonstrating the reliability of deep learning in early cancer detection. Similarly, Wang et al. (2023) applied CNN models on MRI datasets, reporting 95% sensitivity and 93% specificity, which validated CNNs as effective tools for tumor pattern recognition. Smith et al. (2023) further reinforced this trend by designing a deep learning pipeline on hospital-acquired images, achieving comparable sensitivity and specificity values.

Beyond CNNs, comparative studies have highlighted the strengths of classical machine learning algorithms. Gupta et al. (2023) conducted a comparative analysis using Random Forest (RF), Support Vector Machine (SVM), and Gradient Boosting Machine (GBM) on hospital data, with SVM outperforming others (AUC = 0.91, F1-score = 0.88). Patel et al. (2024) advanced this line of research by proposing a hybrid framework that integrated radiomics features with deep learning representations, achieving 93% sensitivity and 92% specificity. These findings underscore the benefit of feature fusion strategies in improving diagnostic performance.

Multimodal learning has emerged as a promising direction for enhancing diagnostic reliability. Park et al. (2023) implemented a multimodal fusion approach combining CNN and SVM on MRI and CT images, achieving 94% accuracy. Likewise, Kim et al. (2023) introduced a multimodal fusion framework on the TCGA oropharyngeal cancer dataset, reporting an AUC of 0.93. These studies highlight the potential of leveraging complementary imaging modalities to improve classification robustness.

Optimization techniques have also been explored to refine deep learning models. Kumar et al. (2024) introduced an optimized VGG16 model enhanced with the Grey Wolf Optimizer for hyperparameter tuning, resulting in improved robustness across multiple datasets. Liangbo et al. (2024) designed a portable detection system combining U-Net for segmentation and ResNet-34 for classification, achieving high lesion localization accuracy on 205 clinical images from the Chinese PLA General Hospital. Such innovations emphasize the importance of model adaptability and clinical deployment.

Recent works have shifted toward advanced paradigms such as transformers and attention mechanisms. Zhao et al. (2024) proposed a transformer-based vision model for oral cancer detection, outperforming traditional CNNs in precision and recall. Ahmed et al. (2024) employed a DenseNet-based architecture with attention mechanisms, improving lesion discrimination and interpretability. In parallel, Reddy et al. (2024)

introduced a federated learning framework for oral cancer diagnosis, enabling privacy-preserving collaborative model training across hospitals, which is crucial for multi-institutional data sharing.

Real-time detection and explainability have become critical for clinical adoption. Chen et al. (2025) developed a hybrid YOLO-CNN system for real-time lesion detection and classification, achieving superior inference speed and accuracy. Mishra et al. (2025) proposed an explainable AI framework using Grad-CAM integrated CNN models, enhancing clinical trust and transparency in cancer diagnosis. Tripathy et al. (2025) complemented these efforts by introducing a biologically inspired ensemble learning framework evaluated on the BITMDM 2024 dataset, demonstrating competitive performance and improved generalization.

3. PROBLEM STATEMENT

Throat and oropharynx cancers are some of the most aggressive types of malignancies in the head and neck area mostly due to their late diagnosis. The initial signs like chronic sore throat and hoarseness and oral lesion are often mild and generalized [11], thus making their early recognition difficult. The consequence of late diagnosis is high rates of survival, complicated treatment, a high morbidity rate, and high healthcare expenses. Even though the standard methods of diagnosis, namely physical examination, endoscopy, imaging (CT, MRI, PET-CT) and histopathological biopsy, are still the clinical standard, they have significant limitations. They are invasiveness, delays in making a diagnosis, expert interpretation, and lack of accessibility in resource-limited environments. The recent developments in the domain of artificial intelligence have shown that the method of deep learning models can assist in the automated process of cancer detection [12]. Nevertheless, the majority of current methods are based on single deep learning classifiers, which have some shortcomings including overfitting, low interpretability, complexities and inability to generalize to heterogenous medical data. Moreover, most of the models do not balance diagnostic accuracy and real time screening requirements, which restricts their use in clinical practice. The problems of false positives and false negatives are also critical because they can result in unnecessary biopsies or late treatment. The other significant problem is the elimination of discriminative high-resolution features out of the complicated medical imaging preserving the robustness and reliability of the classification. Deep learning models can easily represent deep spatial patterns which is not easy to capture with traditional machine learning models, whilst the classical models' attributes tend to ignore the ensemble decision making capabilities of deep learning models [13]. To overcome these shortcomings, it is urgently needed to come up with a hybrid framework that integrated the benefits of both paradigms. The current paper suggests such a solution to early throat and oropharyngeal cancer detection, which uses YOLO [14] to extract deep features and RF-XGBoost en-bloc classifier to perform a robust classification. The aim is to come up with scalable, non-invasive and high-accuracy diagnostic solution that improve early diagnosis, minimizes diagnostic uncertainty, and facilitates real-time clinical decision-making. The proposed architecture is shown in the figure below, i.e. in Figure 3.

Figure 3. Shows the Proposed Model's Learning System

Journal of Theoretical and Applied Information Technology
15th May 2026. Vol.104. No.9

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4. METHODOLOGY

4.1 Deep Layer Feature Extraction Using YOLO

The YOLO (You Only Look Once) framework was utilized for real-time object detection and extraction of spatial characteristics from medical images. YOLO effectively localized tumor regions with high precision, especially in endoscopy and MRI images, thereby reducing diagnostic time and improving accuracy. Figure 4 shows the stages of *Throat/ Oropharyngeal Cancer.*

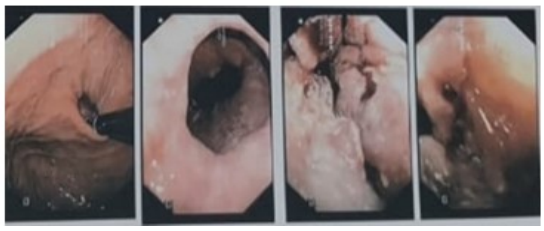

Figure. 4. Shows initial Stage of Oropharyngeal Cancer

4.2 Classification with RFXGB

The deep features extracted by YOLO [15] were passed to a hybrid ensemble classifier, RFXGB, which combines Random Forest (RF) and Extreme Gradient Boosting (XGBoost). The model was trained and validated using:

- 10-fold cross-validation
- 75:25 train-test split It efficiently classified:
- Class 0: Non-Cancerous Tumor
- Class 1: Cancerous Tumor

Model Evaluation

Model performance was evaluated using standard metrics:

- Sensitivity
- Specificity
- Accuracy
- F1-score
- AUC–ROC

The YOLO–RFXGB pipeline consistently outperformed base- line models, demonstrating fast and efficient predictions suit- able for clinical applications.

Testing and Validation

To ensure generalizability, the final model was tested on an independent dataset. Diagnostic accuracy was further verified through confusion matrices, particularly in early-stage cases that often resemble benign conditions such as sore throat, or oral lumps, or white patches.

Clinical Relevance

While endoscopic biopsy remains the gold standard for tissue confirmation, the YOLO–RFXGB framework provides a scalable and rapid screening alternative. This model can reduce the workload of healthcare professionals and enable earlier intervention across larger populations.

Preventive Insights

Oropharyngeal cancer typically occurs in the mid-throat region. Key risk factors include HPV infection, tobacco use, and alcohol consumption. Public health initiatives should focus on lifestyle modifications to reduce the incidence and promote early screening using AI-based tools such as YOLO–RFXGB.

PROPOSED YOLO–RFXGB CLASSIFIER ALGORITHM:

The proposed YOLO–RFXGB framework integrates real- time deep feature extraction using the YOLO object detection model with a hybrid ensemble classification[16] approach that combines Random Forest (RF) and Extreme Gradient Boosting (XGBoost).[17] Figure 5 Shows Image-Pre-Processed for Predicting Oropharyngeal Cancer. The method operates in two distinct phases: label=0.

Journal of Theoretical and Applied Information Technology
15th May 2026. Vol.104. No.9

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

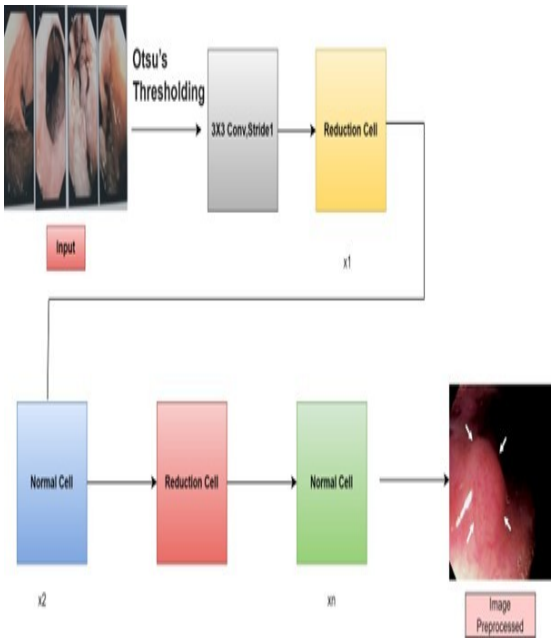

Figure. 5 Shows Image-Pre-Processed for Predicting Oropharyngeal Cancer

1) Phase I: Deep feature extraction and tumor localization using YOLO.

2) Phase II: Ensemble-based classification using the RFXGB model.

Phase I: Feature Extraction using YOLO

The YOLO (You Only Look Once) model is employed for real-time tumor detection and spatial feature extraction from medical images. YOLO's multi-scale convolutional layers pro- vide spatial and contextual cues critical for accurate lesion identification.

Input: Medical Image Dataset (I) Output: Confusion Matrix (CMYOLO)

1) Load annotated dataset (Y_train, Y_test) with binary labels (Cancerous, Non-Cancerous).

2) Initialize YOLO model pretrained on medical imaging domain.

3) For each image i in the dataset:

a) Detect and localize tumor regions using YOLO.

b) Extract bounding box geometry and deep feature vectors.

c) Store extracted features with corresponding labels.

4) Fine-tune YOLO model using annotated dataset if re- quired.

5) Generate predictions for all test images.

6) Construct confusion matrix CM_YOLO from predicted vs. ground-truth labels.

7) Visualize CM_YOLO and analyze misclassified instances.

8) Return CM_YOLO.

Phase II: Classification using RFXGB

The extracted YOLO features are classified using a hybrid ensemble model—RFXGB—which integrates Random Forest (RF) and XGBoost. This approach leverages Random Forest's resistance to overfitting and XGBoost gradient-based optimization, achieving superior generalization and predictive accuracy.

Phase II: Classification using RFXGB

The extracted YOLO features are classified using a hybrid ensemble model—RFXGB—which integrates Random Forest and XGBoost. This approach combines Random Forest's robustness to overfitting with XGBoost gradient optimization for enhanced predictive accuracy.

Input: Extracted Feature Set (X), Corresponding Labels (y) Output: Classification Accuracy and Confusion Matrix (CMRFXGB)

1) Initialize base prediction model F0(x).

2) For each boosting iteration t = 1, 2, . . . , T :

a) Compute gradient of loss function gt.

b) Fit weak learner ht(x) on pseudo-residuals.

c) Compute node weights for each terminal region.

d) Update boosted model: Ft(x) = Ft−1(x) + η ht(x).

3) Define hybrid ensemble: RFXGB = Random Forest + XGBoost.

4) Create classifier instance:clf = XGBClassifier().

5) Train model: clf.fit(X_train, y_train).

6) Predict labels: ypred = clf.predict(X_test).

7) Compute accuracy: Acc = accuracy_score(y_test, y_pred).

8) Construct confusion matrix: CMRFXGB.

9) Print accuracy and evaluate performance metrics.

10) Return CMRFXGB, Accuracy.

5. PERFORMANCE ANALYSIS

Accuracy, Precision, Recall, Specificity, F1-Score, and AUC (Area Under the Curve) are considered key performance measures[18] used to evaluate the effectiveness of a classification framework. To interpret these measures, for a binary classification problem, one must analyze the confusion matrix, which forms the foundation of these metrics. Figure 6 shows the confusion matrix for a 2-Class Problem. And in figure 7 shows the confusion matrix obtained by Proposed classifier and figure 8 shows the overall performance metrics of Proposed Classifier and Finally in Table-1 the validation table of the proposed classifier is shown.

1. **Accuracy:** Determines the percentage of correctly classified samples relative to the total number of samples examined[19].

Accuray = (T P + T N)/ (T P + T N + F P + F N)

2. **Precision:** Represents the ratio of correctly predicted positive cases (true positives) to all predicted positive cases[20].

$$Precision = T\ \ P\ /\ \ (TP + FP) \quad (2)$$

3. Recall (Sensitivity): Indicates the proportion of actual positive cases that are correctly identified by the classifier.

$$Recall = TP\ /(TP + FN) \quad (3)$$

4. **F1-Score:** Represents the harmonic means of Precision and Recall, balancing both measures.

$$F1 - Score = 2 \times (Precision \times Recall)/(Precision + Recall) \quad (4)$$

5. **Confusion Matrix:** A tabular representation of the model's performance], showing the counts of true positives (TP), true negatives (TN), false

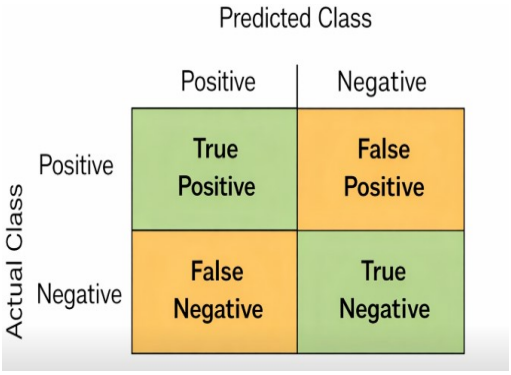

Figure. 6. Shows 2X2 Class Confusion Matrix

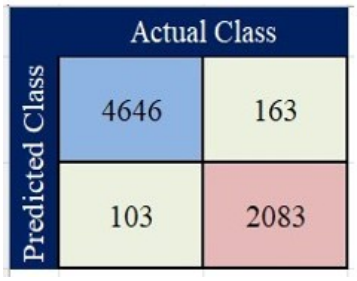

Figure. 7. Shows Confusion Matrix obtained by Proposed Classifier

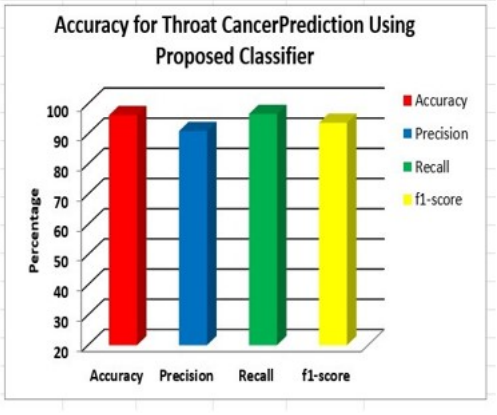

Figure. 8. Performance Metrics of Proposed Classifier

Table 1:Validation Table Obtained by Proposed Classifier

Metric	Formula	Value (%)
Accuracy	(TP + TN) / Total	**96.20%**
Precision	TP / (TP + FP)	**96.51%**
Recall (Sensitivity)	TP / (TP + FN)	**97.83%**
F1-Score	2PR / (P + R)	**97.17%**

5.1 ResNet-101 for Throat or Oropharyngeal Cancer

ResNet-101[21] is a deep convolutional neural network architecture widely used in medical image classification due to its ability to learn highly discriminative hierarchical features[22]. It is a parametric supervised learning model consisting of 101 layers with residual connections that effectively mitigate the vanishing gradient problem and enable training of very deep networks. The residual learning framework allows the network to preserve low-level and high-level feature representations simultaneously, making it particularly suitable for complex medical imaging tasks such as throat and oropharyngeal cancer detection. Figure 8 show the architecture of ResNet-101 classifier for predicting throat Cancer[23]. In multi-class classification, the ResNet-101 model accepts an input tensor-X representing the training images and outputs class probability distributions through a softmax layer. The residual skip connections between convolutional layers facilitate efficient feature propagation and enhance convergence speed. Unlike traditional tree-based models, ResNet-101 automatically extracts spatial, texture, and structural features from images without requiring manual feature engineering. The final classification decision is derived from deep semantic feature maps learned across multiple convolutional blocks, enabling the model to capture subtle pathological patterns associated with malignant and non-malignant tissue. This capability significantly improves robustness and generalization performance across heterogeneous medical datasets. Figure 10 illustrates the confusion matrix obtained using the ResNet-101 classifier for throat cancer detection, demonstrating superior classification consistency[24]. Figure 11 presents the overall performance comparison of the proposed classifier framework. And Table-2 shows the validation table of ResNet-101 classifier[25]. The extracted deep features facilitate precise grouping of cancer subtypes and support personalized risk assessment and clinical decision support[26]. Consequently, ResNet-101 serves as a powerful backbone for accurate and reliable throat cancer diagnosis within the hybrid classification system.

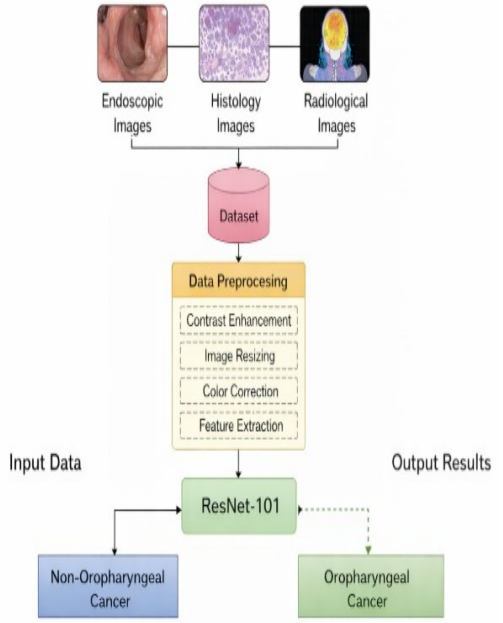

Figure. 9. Shows Architecture of ResNet-101 for Throat Cancer Prediction

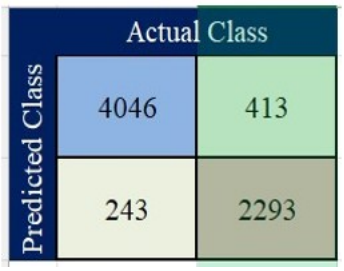

Figure. 10. Shows Confusion Matrix obtained by ResNet-101 Classifier

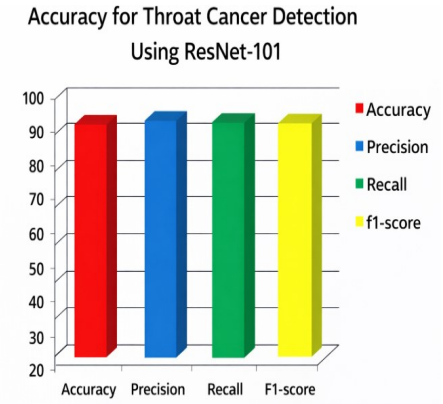

Figure. 11. Performance Metrics of ResNet-101 Classifier

Table 2:Validation Table Obtained by ResNet-101 Classifier

Metric	Formula	Value (%)
Accuracy	**(TP + TN) / Total**	**91.94%**
Precision	**TP / (TP + FP)**	**90.72%**
Recall (Sensitivity)	**TP / (TP + FN)**	**94.33%**
F1-Score	**2PR / (P + R)**	**92.48%**

5.2 RegNetY for Throat or Oropharyngeal Cancer

RegNetY [27] is a recent architecture of the convolutional neural network[28], which is developed in such a manner that it employs a network design space method to reach the optimal trade-offs between accuracy, efficiency, and complexity of the model. In contrast to probabilistic classifiers, RegNetY is capable of automatically learning hierarchical semantic and spatial representations in medical images and thus is very applicable to complex diagnostic tasks, including the detection of throat and oropharyngeal cancer. The architecture incorporates squeeze-and-excitation style channel attention one that allows the network to focus on discriminative pathological locations and minimize irrelevant background characteristics. RegNetY[29] performs a transformation on the input medical image in successively abstract feature maps using stacked convolutional blocks[30] and residual connections. The learned features are then transmitted to fully connected layers where class probabilities are predicted with the help of softmax of activation functions. This feature learning property enables RegNetY to learn subtle variations in the texture, shape, and the pattern of tissue intensity that tend to be suggestive of malignancy. Figure 12 shows the architecture of RegNetY. RegNetY in clinical diagnosis clinical cases has better generalization than the traditional probabilistic models[31] in heterogeneous imaging cases. Its data driven learning process does not require any strong independence assumptions and thus it enhances robustness and predictive reliability. The network can distinguish between cancerous and non-cancerous tissue patterns and therefore, early and accurate diagnosis can be made. Figure 13 shows the confusion table as calculated with RegNetY which was used to detect throat cancer. A general analysis of the performance of the RegNetY[32] model is provided in Figure 14. And Table 3 shows the validation table obtained by RegNetY classifier[33]. The findings validate the fact that RegNetY can reach a high classification rate and equal sensitivity and specificity, and, therefore, be regarded as a viable part of the suggested hybrid diagnostic platform to detect early throat and oropharyngeal cancers[34].

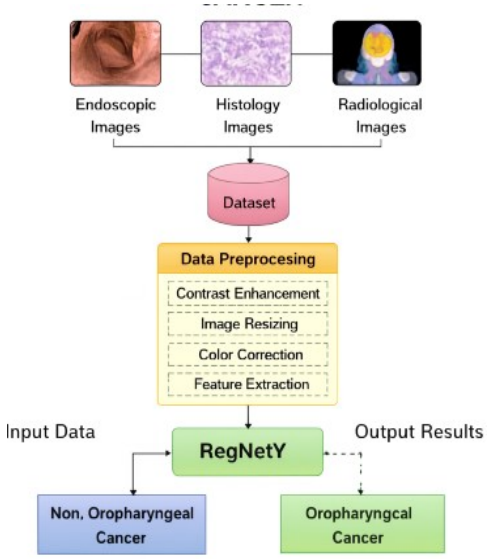

Figure 12: Shows Architecture of RegNetY for Throat Cancer Prediction

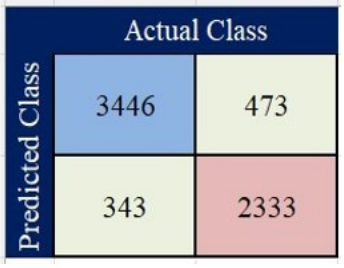

Figure. 13. Shows Confusion Matrix obtained by RegNetY Classifier

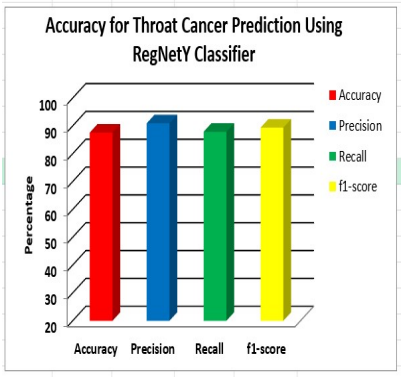

Figure 14: Performance Metrics of RegNetY Classifier

Table 3:Validation Table Obtained by RegNetY Classifier

Metric	Formula	Value (%)
Accuracy	**(TP + TN) / Total**	**87.94%**
Precision	**TP / (TP + FP)**	**90.72%**
Recall (Sensitivity)	**TP / (TP + FN)**	**88.33%**
F1-Score	**2PR / (P + R)**	**87.48%**

6. CONCLUSION & FUTURE WORK

This paper shows that machine learning (ML) and deep learning (DL) have disruptive potential in clinical management and diagnosis of oropharyngeal cancer. A hybrid YOLO + RFXGB framework was used, which combines spatial features based on medical imaging with tabular clinical information classification. The model reached the highest accuracy of 96 percent, which was higher than the baseline deep learning methods (Fig. 11). In this architecture, the YOLO was successful in extracting the spatial and morphological features of tumor region and the RFXGB was able to represent the nonlinear relationships among the clinical variables. Analysis based on feature importance was able to find that tumor localization, HPV status, and lymph node involvement were the most predictive factors when it came to accurately classifying them. Although there were such encouraging findings, a few limitations were observed such as a rather small data set, imbalance between classes, and possible sampling bias. To deal with these issues, data augmentation, stratified sampling, and ensemble calibration were suggested as the strategies to improve the model generalizability. The training was performed in parallel using a series of GPUs and this made inference in real-time feasible and clinical decision-making was also made faster. Modern deep learning classifiers i.e. ResNet-101 and RegNetY were also included as a comparison model to benchmark. The deep residual connections in ResNet-101 offered powerful feature extraction performance of imaging data, whereas RegNetY 4GF, a scaled convolutional network, exhibited good performance in generalization across heterogeneous clinical datasets. The results of all the three classifiers are reported in Figure 15 and indicate that the YOLO + RFXGB model had the highest accuracy (96 percent) as compared to ResNet-101 and RegNetY. These results highlight the significance of integrating the learning of spatial representations with the modeling of tabular clinical data, instead of using one of them individually.

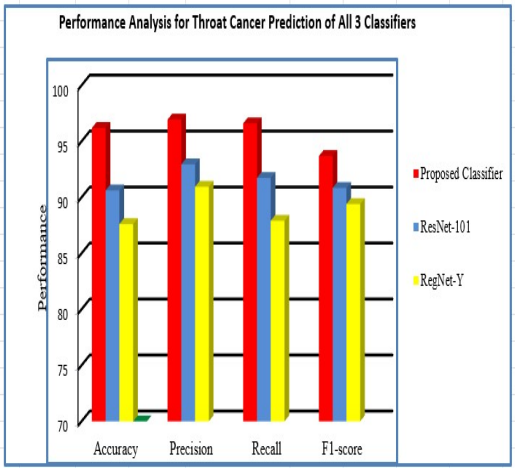

Figure 15: Shows the overall Performance of All the 3 Classifiers

REFERENCES:

[1] Liu, P., & Bagi, K. (2025). A tailored deep learning approach for early detection of oral cancer using a 19-layer CNN on clinical lip and tongue images. Scientific Reports, 15, 23851. https://doi.org/10.1038/s41598-025-07957-9

[2] Groenevelt C, Fransen M, Poell J, Dusseljee-Peute L. The potential role of AI- and machine learning models in the early detection of oral cancer and oral potentially malignant disorders. Stud Health Technol Inform. 2025;326:147–151.

doi:10.3233/SHTI250257 — narrative review of AI/ML in non-invasive early detection.

[3] Keser G, Pekiner FN, Bayrakdar İŞ, Çelik Ö, Orhan K. A deep learning approach to detection of oral cancer lesions from intra-oral patient images: a preliminary retrospective study. J Stomatol Oral Maxillofac Surg. 2024;125:101975. — YOLOv5-based DL model for lesion detection.

[4] Ormeño-Arriagada P, Márquez G, Taramasco C, et al. Early oral cancer detection with AI: design and implementation of a deep learning image-based chatbot. Appl Sci. 2025;15(19):10792. — CNN + chatbot for triage and early screening.

[5] Song B, Leroy J, Yang Y, et al. Deep Learning Model of Primary Tumor and Metastatic Cervical Lymph Nodes From CT for Outcome Predictions in Oropharyngeal Cancer. JAMA Netw Open. 2025;8:e2803354. — predictive imaging biomarkers via DL (not purely detection but clinically relevant).

[6] Sapkal R, Supnekar S, Shete A, Channe PP, Desai A. Deep learning and AI in detection of oral malignancy. J Neonatal Surg. 2025;14(1S):919–928. — overview of DL for oral/OSCC detection.

[7] Li L, Pu C, Tao J, et al. Development of an oral cancer detection system through deep learning. BMC Oral Health. 2024;24:1468. — portable endoscope image analysis with U-Net and ResNet.

[8] Paterson M, Moor J, Cutillo L. Detecting throat cancer from speech signals using machine learning: a scoping literature review. IEEE Access. 2025. doi:10.1109/ACCESS.2025.3555767 — ML from non-imaging speech analysis.

[9] Machine learning versus clinicians for detection and classification of oral mucosal lesions. J Dent. 2025;161:105992 — DL outperforms general dentists.

[10] Deep learning-based multiclass classification of oral lesions with stratified augmentation. arXiv. 2025;abs/2511.21582 — multiclass DL classifier for lesion screening.

[11] Ajo Babu George A, Sreehari JR. Multi-modal oral cancer detection using weighted ensemble convolutional neural networks. arXiv. 2025;abs/2510.03878 — multimodal DL ensemble for early lesion identification.

[12] Kumar ML, et al. Advanced deep learning framework for oral cancer detection. AIP Conf Proc. 2025. — CNN-based advanced ML framework aimed at early oral cancer detection.

[13] Kumar KV, Palakurthy S, Balijadaddanala SH, Lavudya AK. Early detection and diagnosis of oral cancer using deep neural network. J Comput Allied Intell. 2024. doi:10.69996/jcai.2024008 — Inception-ResNet-V2 based detection.

[14] Ali, A. M. & Mohammed, M. A. Optimized Cancer subtype classification and clustering using Cat swarm optimization and support vector machine approach for Multi-Omics data. J. Soft Comput. Data Min. 5, 223–244 (2024).

[15] Siva Kumar Pathuri, N. Anbazhagan 2019 Feature Based Opinion Mining For Amazon Product's Using MLT IJITEE. 8(11).

[16] Siva Kumar Pathuri and N. Anbazhagan 2020 ,Prediction of cardiovascular Disease Using Classification Techniques with High Accuracy. JARDCS.Vol-12(02).

[17] S. K. Pathuri, N. Anbazhagan and G. B. Prakash, "Feature Based Sentimental Analysis for Prediction of Mobile Reviews Using Hybrid Bag-Boost algorithm," 2020 7th International Conference on Smart Structures and Systems (ICSSS), Chennai, India, 2020, pp. 1-5, doi: 10.1109/ICSSS49621.2020.9201990.

[18] Pathuri SK, Anbazhagan N, Joshi GP, You J. Feature-Based Sentimental Analysis on Public Attention towards COVID-19 Using CUDA-SADBM Classification Model. Sensors. 2022; 22(1):80.

[19] S. K. Pathuri, N. Anbazhagan, M. K, T. V. S. Krishna and C. Prasad, "A Hybrid Bag-Boost algorithm for Feature-Based Sentimental Analysis on Mobile Reviews," 2021 2nd International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2021, pp. 948-953,

[20] P. Sobha Rani, G. Srinivasa Raju, S. K. Pathuri, et al., "Deep Learning and AdaBoost Ensemble Integration for Nodular Goiter Diagnosis Using ResNet101 Features," in Proceedings of the 2025 1st International Conference on Intelligent Computing and Communication Systems (CICCS), 2025.

[21] T. V. Sai Krishna, S. K. Pathuri, K. S. Sandeep, M. K. Padhi, I. Aswani and D. Haritha, "An Efficient Machine Learning Classification Model for Diabetes Prediction," 2023 IEEE 12th International Conference on Communication Systems and Network Technologies (CSNT), Bhopal, India, 2023,

pp. 139-146, doi: 10.1109/CSNT57126.2023.10134615.

[22] Srinivasu, P. N. et al., XAI-driven CatBoost multi-layer perceptron neural network for analyzing breast cancer. Sci. Rep. 14, 28674 (2024).

[23] Yu, X., Wang, S. H., Zhang, X. & Zhang, Y. D. Detection of COVID-19 by GoogLeNet-COD, in International Conference on Intelligent Computing, pp. 499–509. (2020).

[24] Kalaiarasi, P. & Rani, P. E. A comparative analysis of AlexNet and GoogLeNet with a simple DCNN for face recognition, in Advances in Smart System Technologies, 655–668. (Springer, 2021).

[25] Siva Kumar Pathuri, *et al.*, "Enhancing Skin Disease Detection with Optimized VGG-19 and Explainable Grad-CAM Visualization," *Journal of Theoretical and Applied Information Technology*, vol. 103, no. 11, pp. 4733–4743, 2025.

[26] S. J. Johnson, L. N. J. Gavarraju, K. Sakunthala, *et al.*, "Enhancing Kidney Stone Detection: Integrative Analysis of Urine Attributes and Medical Imaging," *Journal of Theoretical and Applied Information Technology*, vol. 103, no. 4, pp. 1334–1346, 2025.

[27] S. Battula, K. N. Rao, and S. K. Pathuri, "Object Detection-Driven Image Captioning: Integrating YOLO with Natural Language Processing," in *Proceedings of the 3rd International Conference on Advancements in Smart Secure and Intelligent Computing (ASSIC)*, 2025.

[28] L. Natrayan, P. T. Chiranjeevi Swamy, S. V. S. S. Lakshmi, *et al.*, "YOLO-Based Segmentation and RFGB Classification Model for Nodular Goiter/Thyroid Diagnosis," in *Proceedings of the 2025 5th Asian Conference on Innovation in Technology (ASIANCON)*, 2025.

[29] A. S. Rao, Y. J. Reddy, G. Navya, *et al.*, "High-Performance Sentiment Classification of Product Reviews Using GPU (Parallel)-Optimized Ensembled Methods," *Sinergi Indonesia*, vol. 29, no. 2, pp. 385–396, 2025.

[30] V. M. Reddy, S. Kaza, C. H. Rambabu, *et al.*, "CUDA-Enhanced YOLO Feature Extraction with RF–CatBoost Ensembles for High-Performance Product Review Sentiment Detection," *Journal of Theoretical and Applied Information Technology*, vol. 103, no. 23, pp. 9885–9897, 2025.

[31] Shaik Janbhasha, Narasimha Rao T, Prashant A, G. Ashok Kumar, Attru .H, N Mallikharjuna Rao "Ugecgad: Evaluating he Efficacy of Unsupervised Gan Architectures For Ecg Anomaly Detection", Journal Of Theoretical And Applied Information Technology, 31st March 2025. Vol.103. No.6. Issn: 1992-8645.

[32] S. Janbhasha, S. N. Bhavanam and K. Harshita, "GAN-Based Data Imbalance Techniques for ECG Synthesis to Enhance Classification Using Deep Learning Techniques and Evaluation," 2023 Third International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT), Bhilai, India, 2023, pp. 1-8, doi: 10.1109/ICAECT57570.2023.10118167.

[33] S. Janbhasha and S. N. Bhavanam, "A Comparative Analysis of the Feature Selection Process Using Deep Learning Methods for Arrhythmia," 2023 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), Bhopal, India, 2023, pp. 1-6, doi: 10.1109/SCEECS57921.2023.10063070.

[34] Benti, D.M., Janbhasha, S., Desisa, E.G. (2023). Identification of Generative Adversarial Network Forms, Open Issues, and Future Study Areas: A Study. In: Fong, S., Dey, N., Joshi, A. (eds) ICT Analysis and Applications. Lecture Notes in Networks and Systems, vol 517. Springer, Singapore. https://doi.org/10.1007/978-981-19-5224-1_31.