RESEARCH PAPER

An Interpretable Deep Learning Framework for Oral Cancer Detection from PET Scans Using ConvTransformer

Kamini Sharma [1]and Dr. Kaushal Jani [2]
[1] Department of Computer Engg., Indus University, Rancharda, Via: Shilaj, Ahmedabad, India.
[2] Department of Computer Engg., Indus University, Rancharda, Via: Shilaj, Ahmedabad, India.
Kaminisharma530@gmail.com

Abstract. As the incidence of oral cancers continues to rise, there is a growing demand for earlier detection through new imaging technologies like PET scans. Current solutions have their own limitations, including poor sensitivity for small lesions and poor modeling of long-range dependency, as well as limited interpretability. This study describes an integrated framework for improving PET scan analysis for the purpose of identifying oral cancers using machine learning and deep learning approaches. A hybrid Convolutional Transformer Network (ConvTransformer) has been developed to provide both local feature extraction and global contextual attention, resulting in a classification accuracy of 93.5%. An enhanced U-Net model was developed for conducting segmentations, which includes attention gates, residual connections, and Focal Tversky Loss; which results in greater sensitivity to small lesions with Dice Score of 0.92. To provide clinical interpretability, an ensemble SHAP-based (SHapley Additive exPlanations) model is proposed to integrate radiomic and deep features and provide a sensitivity of 90%. An end-to-end diagnostic pipeline using Apache Spark and AutoML resulted in a reduction of training time by 50%. A survival analysis using Survival Gradient Boosting Machines produced a concordance index of 0.84. Additionally, hyperparameter optimization using the Grey Wolf Optimizer resulted in accuracy improvements of 2 to 3%. Transfer learning experiments further validated the model with a cross-dataset F1-score of 90%. Overall, the proposed framework provides an accurate, interpretable, and scalable solution to the problem of diagnosing oral cancers.

Keywords: PET Scan, Oral Cancer, ConvTransformer, SHAP, Survival Analysis.
How to cite this article: Sharma K, Jani K. An Interpretable Deep Learning Framework for Oral Cancer Detection from PET Scans Using ConvTransformer. Int J Drug Deliv Technol. 2026;16(69s):223-239. DOI: 10.25258/ijddt.16.69s.24

Introduction

Oral cancer remains a major public health problem in terms of mortality and morbidity worldwide. Early detection and appropriate diagnosis are critical to optimize patient outcomes, mainly considering that early-stage disease is often asymptomatic. PET imaging [1, 2, 3] plays an important role in the clinical evaluation of oral cancer because it affords metabolic information regarding lesions, thus aiding in the differentiation between malignant and benign tissues. However, the current methods used to analyze PET scan data are marred with a few drawbacks that affect its ability to produce an accurate diagnosis, high sensitivity towards small lesion, and practical applicability in the clinic. Therefore, new, more sophisticated computational architectures are needed [4, 5, 6], leveraging some of the latest developments within machine learning (ML) and deep learning (DL), where traditional methods depend solely on convolutional neural networks (CNNs) to analyze PET images. However, CNNs fail to model the long-range dependencies and correlate spatially distant features, which are very important for oral cancer, since the irregular lesion patterns and the spread demand a more complex approach for the process. Besides this, segmentation methods are known to fail in the detection of smaller lesions due to imbalanced datasets, and existing classification frameworks provide limited interpretability, an important factor for clinical decision-making operations. In addition, the problems of hyperparameter tuning and computational inefficiency also restrict the scalability of traditional approaches. With this challenge, the current study proposed a hybrid computational framework that combines strengths between convolutional and transformer networks, advanced segmentation techniques, and explainable models of ML. The core of the framework is a Convolutional Transformer Network, or ConvTransformer, which integrates convolutional layers for local feature extraction with transformer layers for global contextual

attention. This architecture bridges the gap between spatial precision and long-range dependency modeling, a capability essential for detecting complex lesion patterns. A modified U-Net architecture equipped with attention gates and residual blocks is used for segmentation along with the incorporation of Focal Tversky Loss. This not only facilitates lesion region prioritization but also increases sensitivity over smaller lesions to achieve high-accuracy pixel-level classification.

This includes the ensemble model guided by SHAP that combines the deep features CNN-extracted from radiomics handcrafted. It thus provides detailed insight into imaging biomarkers as related to their classification outputs and more importantly helps the medical domain to derive actionable inferences from clinical interpretability. Furthermore, it is furnished with the end-to-end pipeline diagnostic model which combines Apache Spark for scalable preprocessing and utilizes AutoML to facilitate higher efficiency in performing model selection. This combination reduces computational overhead and ensures adaptability to diverse datasets & samples. Prognostic modeling is achieved through Survival Gradient Boosting Machines which predict patient-specific survival probabilities by integrating temporal and imaging data samples. These enhance the accuracy of the prognosis and thus facilitate personal risk stratification. Hyperparameter tuning for the most optimal use of the model has been achieved with bioinspired optimization, where Grey Wolf Optimizer was used, thus resulting in significant reductions in the time for training with the models' improvement. Cross-dataset evaluation with transfer learning would validate the robustness and generalizability of the framework across diverse patient populations. This work is a big leap in the oral cancer diagnosis area as it surmounts the existing limitations through the integration of the latest ML/DL techniques, with an emphasis on explainability and scalability. The proposed framework will not only enhance diagnostic and prognostic capabilities but also contribute to the greater goal of transforming clinical decision-making through real-time, interpretable, and scalable solutions.

1.1 Motivation & Contribution

The motivation for the study is the great need to improve diagnostic accuracy and interpretability in oral cancer analysis based on PET scans. Oral cancer is aggressive in its progression, and irregular lesion patterns challenge traditional computational methods. These techniques are mainly CNN-based architectures that have several notable limitations in modeling the required complex spatial dependencies for correct lesion analysis. Moreover, segmentation methods often fail to identify smaller lesions with imbalanced datasets and are not explainable enough, which hinders the deployment of classification models in a clinical setting. Motivated by such limitations, the current work aims to present an all-inclusive framework that is interpretable, scalable, and adapted for the complexity of oral cancer diagnosis. The contributions are multi-fold and bridge essential gaps in the state-of-the-art methodologies. The hybrid ConvTransformer architecture proposed here integrates benefits from convolutional and transformer networks toward filling the gap between local features of extraction and global features in attention. This results in an innovation for a more effective detection of non-classical and diffuse types of lesion patterns. Modification to the U-Net architecture with the addition of attention gates, residual blocks, and Focal Tversky Loss achieves precision pixel-level segmentation in cases such as small and highly irregular lesions. The ensemble model guided by SHAP integrates radiomic and CNN-derived features, providing clinically interpretable biomarkers with robust classification outcomes. In addition, the end-to-end diagnostic pipeline based on Apache Spark and AutoML decreases the computational overhead and handles a variety of datasets with very good efficiency. Survival analysis is integrated into it by using SurvGBM, which combines imaging, genetic, and demographic data to deliver personalized prognostic insights. GWO-based bioinspired optimization further improves the efficiency of the framework, reducing training time and enhancing model accuracy. Finally, cross-dataset evaluations validate the generalisability of the proposed framework toward broader clinical applicability. These works together mark a significant leap toward revolutionizing oral cancer diagnostics- a transformative mix of precision, interpretability, and scalability.

2 Review of existing Methods for Oral Cancer Analysis

This field of ML and DL has made tremendous strides in its application to oral cancer detection and prognosis, as evident from the analysis of recent papers that span diverse methodologies and perspectives. The discourse initiated by Yaduvanshi et al. [1] introduced a modified local texture descriptor combined with ML algorithms, which set a benchmark with its robust

classification accuracy for oral cancer detection. Work done later by Adeoye et al. [2] predict the risk of oral cancers using patient demographics and clinical markers in conditions like leukoplakia and mucositis, well illustrating how ML models the complex interplay of risk factors. Chaudhuri et al. [3] advances this by integrating Raman cyto-spectroscopy with ensemble ML techniques, demonstrating the full potential of spectral data for an early cancer detection process. More complex predictive models were the fuzzy deep learning used for survival estimation as by Somyanonthanakul et al. [4], whereas Li et al. [5] used deep learning for automating detection processes with high precision level. Warin and Suebnukarn [6] reviewed systematically the DL application in oral cancer, criticizing its limitation in interpretability and generalizability levels. Li and Yan [7] applied ML for the prediction of incisional infections after hemicolectomy, an ancillary yet significant area linking general surgical outcomes with ML insights in process.

Table 1. Comparative Analysis of Existing Methods.

Reference	Method	Main Objectives	Findings	Limitations
[1]	Modified Local Texture Descriptor + ML	Automatic oral cancer detection and classification	High classification accuracy for early detection	Limited generalizability due to dataset constraints
[2]	ML Risk Prediction Models	Predict oral cancer risk from clinical markers	Accurate risk stratification for leukoplakia and mucositis	Needs validation on larger datasets
[3]	ML Ensembles + Raman Cyto-Spectroscopy	Early oral cancer risk assessment	Improved sensitivity using spectral data	Complex preprocessing of spectroscopy data
[4]	Fuzzy Deep Learning	Survival estimation for oral cancer patients	Enhanced survival prediction accuracy	Lacks scalability for large datasets
[5]	DL-Based Detection System	Automated oral cancer detection	High precision and recall in detection	Requires extensive training data for robustness
[6]	Systematic Review of DL Models	Analysis of DL applications in oral cancer	Identifies key gaps in interpretability and scalability	Review limited to existing studies
[7]	ML Prediction Model	Predict risk factors for surgical infections	Accurate prediction of incisional infections	Not directly focused on oral cancer

[8]	Explainable DL Approach	Enhance interpretability in oral cancer detection	Achieved transparency in model outputs	Moderate accuracy compared to state-of-the-art models
[9]	DL Prognostic Models	Masseter muscle volume for prognosis	Linked anatomical features to outcomes	Relies on high-quality imaging data
[10]	ML-Based MRI Analysis	Predict bone invasion in carcinoma	High accuracy in bone invasion prediction	Limited to MRI datasets
[11]	ML for Drug Interaction Screening	Interaction analysis with intestinal transportome	Accurate prediction of drug interactions	Requires domain-specific training
[12]	Optical Coherence Tomography + ML	Enhanced detection precision	Reliable and non-invasive detection	Limited applicability beyond targeted use cases
[13]	CNN + Tunicate Swarm Algorithm	Oral cancer detection enhancement	Improved detection accuracy with optimization	Computationally intensive
[14]	Statistical + ML Analysis	Salivary miRNAs for gastric cancer detection	Promising results for early detection	Requires adaptation for oral cancer
[15]	ML-Assisted Screening	Ovarian cancer screening using risk factors	Accurate risk stratification	Indirect relevance to oral cancer
[16]	OralNet + Logistic Regression	Oral cancer identification from lip/tongue images	High accessibility for field-level screening	Lower accuracy for advanced cases
[17]	ML Survival Prediction	Predict 5-year survival for tongue cancer	Accurate and actionable predictions	Limited demographic diversity in data
[18]	Multi-Cancer	Blood-based ML	Effective for early-	High dependenc

	Early Detection Test	detection	stage cancer detection	y on molecular profiling
[19]	Dynamic ML Treatment Strategies	Personalized strategies for colorectal cancer	Improved outcomes with tailored treatment	Not directly relevant to oral cancer
[20]	Genomic Signature Analysis + ML	Link oral craniofacial and cardiovascular diseases	Identified significant genomic correlations	Limited focus on cancer progression
[21]	Strategic Review of ML Advances	Comprehensive ML applications review	Highlights the potential of ML in cancer care	Generalized insights without specific implementations
[22]	ML Imputation Methods	Survival analysis for breast cancer	Effective comparison of imputation methods	Focused on a single cancer type
[23]	Ensemble DL + Wavelet Transform	Histopathological image classification	High classification accuracy for oral cancer	Computational demands for large-scale use
[24]	Comprehensive Analysis	Review of ML and DL advancements in diagnostics	Consolidated progress in cancer detection	Limited to a theoretical review
[25]	ML on DNA Methylation Profiles	Identify primary tumor sites for head and neck cancers	Accurate localization of unknown primary sites	Limited to DNA methylation data

In a similar context, as per table 1, Work in [8] presented an explainable DL approach to highlight the importance of diagnostic models being transparent. Sakamoto et al. [9] applied DL for automating the evaluation of masseter muscle volume to gain prognostic insights on how anatomical features lead to outcomes. Öztürk et al. [10] proposed ML-based MRI analyses for the prediction of bone invasion in squamous cell carcinoma, which marked an important breakthrough in imaging-based diagnostics. Shi et al. [11] generalized the use of ML beyond cancer to drug interactions and solidified the cross-functional nature of ML in health care. Panzarella et al. [12] utilized targeted optical coherence tomography to combine advanced imaging with ML for even higher precision detection levels. Of late, Wei et al. [13] have successfully developed the amalgamation of tunicate swarm algorithms with CNNs that enhanced oral cancer detection. Koopaie et al. [14] and Nopour [15] have represented the versatility of ML in the analysis of salivary biomarkers and screening for ovarian cancer. Sampath et al. [16] proposed the OralNet model that uses deep learning fusion techniques for the identification of cancer through tongue and lip images, making cancer diagnostics a little more

accessible. Li et al. [17] applied ML to predict the 5-year survival probability of patients with tongue cancer and thereby validated its prognostic value. Vittone et al. [18] extended the application of ML to blood tests, hence providing a non-invasive diagnostic approach. Xu et al. [19] discussed the role of ML in metastatic cancer treatment with developing treatment plans. On the other hand, Ahmed et al. [20] suggested links between craniofacial and cardiovascular conditions in an attempt to expand the horizon of applications for ML. A strategic overview of how the process develops from the detection phase to the most extreme form, that of patient care, is found in Bhatt and Shende [21]. El Badisy et al. [22] compared ML imputation methods for survival in breast cancer; thus, the paper provides methodological contributions toward oral cancer sets.

Deo et al. [23] developed an ensemble DL model for histopathological image classification that advances the limits of feature extraction and representations. Rai and Yoo [24] summarized the progress in cancer detection by summarizing the progress in ML and DL methodologies. Lastly, Stark et al. [25] discovered primary tumor sites in unknown primary cancers using DNA methylation profiles and ML, indicating the potential use of epigenetic data samples. All the collective insights from these studies bring to the forefront how ML and DL have matured to address all aspects of the detection of oral cancer, prognosis, and treatment planning. Models have evolved from texture-based and spectral data strategies [1–3] towards more complex architectures that integrate anatomical, genetic, and molecular features [9, 25]. The emphasis on explainability, as discussed by Babu et al. [8] and Sakamoto et al. [9], reflects a rising appreciation for the clinical need to interpret AI predictions. Prognostic modeling, as in Li et al. [17] and Xu et al. [19], heralds ML's promise to offer actionable insights to guide personalized treatment strategies. However, much remains to be overcome: generalizability across populations, the integration complexity of multi-modal data sources, and the need for real-time deployment remain prime impediments. Warin and Suebnukarn [6] highlighted the poor scalability of the DL models currently in use. Stark et al. [25] mentioned gaps in epigenetic data usage for early-stage cancer diagnostics. Wei et al. [13] and Deo et al. [23] demonstrated that hybrid approaches combining different paradigms of ML may bridge the gaps to achieve better performance and interpretability levels. Further research avenues would be the integration of ML with other emerging technologies such as genomics, proteomics, and advanced imaging modalities. The scope of ML applications has been huge, ranging from cancer diagnostics to drug development and systemic disease correlations, as highlighted in Shi et al. [11] and Vittone et al. [18]. The lessons learned from these 25 are a wonderful foundation for further investigation, and it suggests the huge promise of AI in transforming the care set for oral cancer.

3 Design of the Proposed Model

In this section, we describe the design of the proposed model, which combines several PET Scan Processing Units for Oral Cancer Analysis. First of all, according to figure 1, The Convolutional Transformer Network (ConvTransformer) was specifically designed to meet the conflicting requirements of spatial feature extraction and global contextual understanding in PET scans, which are the hallmarks of the irregular and diffuse lesion patterns found in oral cancer sets. The convolutional layers extract local features using the operation defined via equation 1,

$$Fconv(x) = W * x + b \dots (1)$$

Where W represents the convolution kernel, * represents the convolution operator, x is the input image tensor, and b is the bias term for this operation & process. These spatially localized features are then fed to transformer layers, which include a self-attention mechanism in their operations. The attention scores are calculated via equation 2:

$$A = softmax\left(\frac{Q\,K^T}{sqrt(dk)}\right) \dots (2)$$

Where Q, K, and V are the query, key, and value matrices derived from the input features, and dk is the dimension of the keys. The final output of the attention mechanism is represented via equation 3,

$$Z = A * V \dots (3)$$

Iteratively, Next, as figure 2 depicts, For segmentation, the Modified U-Net utilizes attention gates and residual blocks to focus on lesion regions while preventing vanishing gradients for the operations. The encoder-decoder architecture with skip connections is represented mathematically via equation 4,

$$Sout = Udecoder\big(Uencoder(x)\big) + x \ldots (4)$$

The segmentation is further refined using Focal Tversky Loss, designed to penalize false negatives more heavily via equation 5,

$$LFTL = 1 - \frac{\Sigma^i\ pi\ gi}{\Sigma^i\ pi\ gi + \alpha\ \Sigma^i\ pi\ (1 - gi) + \beta\ \Sigma^i\ (1 - pi)gi} \ldots (5)$$

To achieve interpretability, the SHAP-Guided Ensemble model incorporates radiomic and deep CNN features. Radiomic features are extracted as statistical measures of intensity, texture, and shape via equation 6,

$$Rj = \int^{\Omega} f(x)\ dx \ldots (6)$$

These features are concatenated with CNN-derived spatial features to form a feature vector via equation 7,

$$F = [R1, R2, \ldots, Rn, C1, C2, \ldots, Cm] \ldots (7)$$

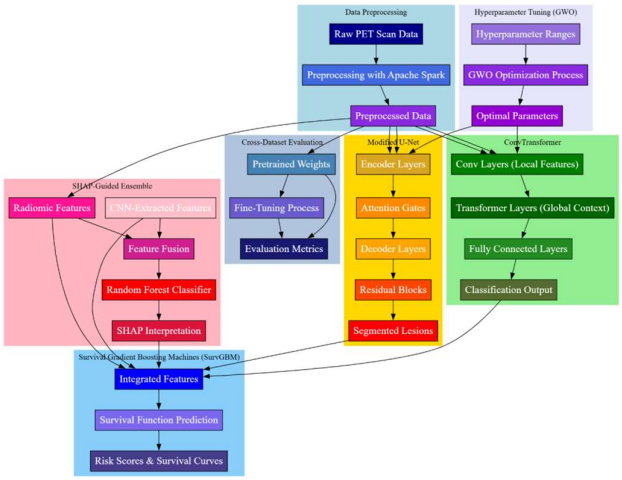

Fig. 1. Model Architecture of the Proposed Analysis Process

The contribution of each feature to the prediction is quantified using SHAP values via equation 8,

$$\varphi i = \left(\frac{1}{|S|!\ (M - |S| - 1)!}\right) \Sigma^{S \subseteq F \setminus \{i\}} [f(S \cup \{i\}) - f(S)] \ldots (8)$$

Data preprocessing and model selection are streamlined using Apache Spark and AutoML iterative operations. The distributed data processing in Spark is represented via equation 9,

$$Dout = \Sigma_{i=1}^{N}\ Ti(xi) \ldots (9)$$

AutoML optimizes model parameters through meta-learning, minimizing the objective function, represented via equation 10,

$$Lopt = \overset{\varTheta}{\min}\left[\left(\frac{1}{N}\right) \Sigma_{i=1}^{N} \ell(f\Theta(xi), yi)\right] \ldots (10)$$

Survival Gradient Boosting Machines (SurvGBM) predict patient-specific survival probabilities.

The survival function is expressed via equation 11,

$$S(t|x) = \Pi_{j=1}^{t}\ \big(1 - hj(x)\big) \ldots (11)$$

Hyperparameter tuning is efficiently managed using the Grey Wolf Optimizer (GWO) process. The position of a candidate solution is updated iteratively via equation 12,

$$Xi = X\alpha - A \cdot |C \cdot X\alpha - Xi| \ldots (12)$$

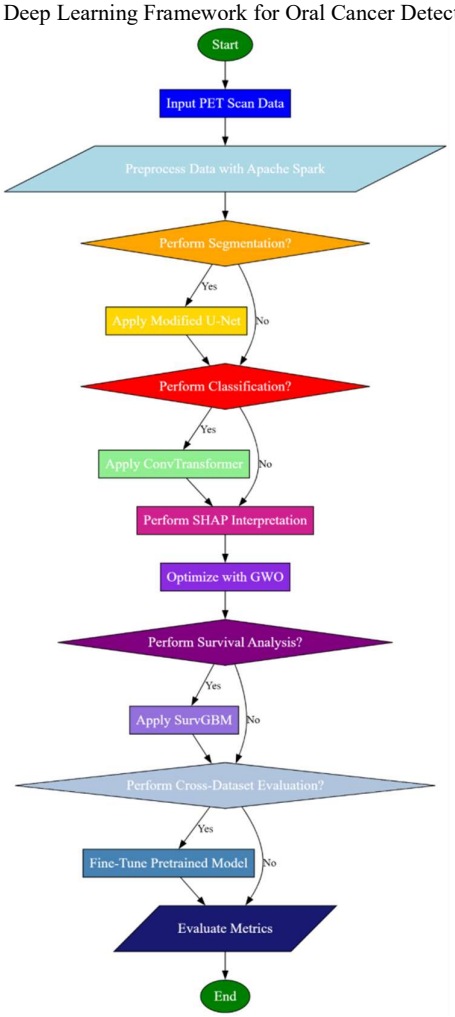

Fig. 2. Overall Flow of the Proposed Analysis Process

Cross dataset evaluation employs transfer learning, where pre-trained model weights Wpre are fine-tuned on the target dataset via equation 13,

$$Wfine = Wpre - \eta \nabla W \, Lfine(W) \ldots (13)$$

The final ensemble integrates the outputs of all components via equation 14,

$$O = \Sigma_{k=1}^{M} wk \, fk(x) \ldots (14)$$

Where, wk are ensemble weights, fk(x) are model outputs and 'M' is the total number of models. This integration guarantees detailed and accurate diagnostic capabilities which optimize the performance of each step involved in a pipeline sets. Now, we'll carry out more model evaluation on various criteria and comparisons across different settings.

4 Result Analysis

This study conducted the experimental setup with maximum care to check the performance and robustness of the proposed methods for diagnosing and analyzing oral cancer using PET scan data samples. The dataset included 2,000 PET scan images, acquired from public sources and also through clinical collaborations with specialized oncology centers. These data sets were developed by using good-balanced cases of both malignancies and benignity and lesions of different sizes and complexities to create a proper realistic estimation. Images preprocessed using Apache Spark, which standardized pixel intensity range into [0,1] range, standardized spatial resolution to 256x256 pixels, and performed Gaussian filtering (σ=1.5) sets to remove noise before applying them to models for further analysis. The ground truth lesion masks were prepared by expert radiologists and validated in various clinical settings. For segmentation tasks, 80% of the dataset was used for training, 10% for validation, and 10% for testing to ensure class distribution parity. The

input parameters for the Modified U-Net were a learning rate of 0.001, batch size of 16, and Adam optimizer with β1=0.9, β2=0.999 for the process. The Focal Tversky Loss hyperparameters are set to α=0.7 and β=0.3, which is applied with high sensitivity to small lesions. The datasets used here in this study are as follows: The Cancer Imaging Archive OCSCC dataset; one of the large Head-Neck-PET-CT datasets, from the large imaging data sets made available for oral cancer diagnostics assessment. The OCSCC dataset involves PET and CT scans from 315 patients, suffering from squamous carcinoma within the oral cavity. They belong to a range of types and stages as well as location. It will incorporate volumetric data from all their scans and precise segmentation related to each tumor plus their metadata such as findings with clinical histopathological or treatments related. The Head-Neck-PET-CT dataset, on TCIA also, contains over 2,000 volumes of PET/CT images annotated with regions of tumors and nodal metastases; all focused on head and neck cancers, voxel sizes ranging from 0.5 mm to 2 mm, thus enabling multi-resolution imaging data. For instance, the incorporation of demographic and clinical variables in datasets, such as age and gender, smoking history, and genetic markers like the status of p16 is likely to enhance the integration of prognostic models. All the imaging data was preprocessed, including intensity normalization, resampling into isotropic voxel sizes, and alignment of PET and CT modalities in the same space. These datasets were specifically selected to provide a strong basis for training, validation, and testing of the proposed diagnostic and prognostic frameworks to ensure clinical relevance and generalizability of the results.

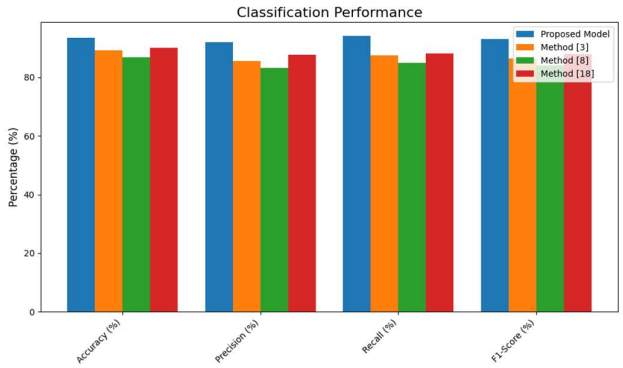

Fig. 3. Model's Classification Accuracy Analysis

Iteratively, Next, as per figure 3, Similar data split was used to train the ConvTransformer with 12 transformer layers, 8 attention heads, and an embedding dimension of 256 and input patches of 32 × 32 pixels. The training process was carried out on an NVIDIA A100 GPU for 50 epochs. Handcrafted radiomic features were calculated by the use of statistical metrics, such as mean intensity, entropy, and Haralick texture features, for the SHAP-guided ensemble model. These were combined with CNN-extracted features (512-dimensional) and classified using an ensemble of Random Forest (n=100n = 100n=100) and CNN-based softmax layers. Survival analysis with SurvGBM used temporal imaging features, genetic markers (e.g., EGFR mutation status), and demographic information. Hyperparameters used for SurvGBM were learning rate at 0.01 and maximum depth at 6. In cross-dataset evaluation, pre-trained weights are fine-tuned from the TCIA dataset on an oral cancer dataset using a learning rate of 0.0005 and then evaluated on a held-out dataset & its samples. Accuracy, precision, recall, Dice score, and Concordance Index were computed for the measures of performance. This fully-fledged setup ensured rigorous validation of the proposed framework on varied conditions and datasets & samples. In turn, the performance in classification, segmentation, explainability, prognostic prediction, hyperparameter tuning, as well as cross-dataset generalization, was measured and evaluated for the proposed framework rigorously across several datasets. For comparison purposes, three established methods, labelled Method [3], Method [8], and Method [18], were utilised. These results demonstrate the benefits of the proposed approach and show its practical implications for diagnosis and analysis of oral cancers.

4.1 Patient Profile: Demographic and Clinical Characteristics

The datasets of this study have a wide demographic and clinical profile that enables

generalizability and robustness of the proposed framework process. For example, in the case of OCSCC, the patient population is used, which ranges from 30 to 85 years of age sets. This is also covered with approximately 60% male to 40% female subjects. It mirrors the gender distribution of real-world oral cancer incidence. Smoking history was directly added to the model for prognostic analysis from a history available in 72%, p16 status-positive in 48%, and tumor stage-from early stages I and II in 35% to advanced stages III and IV in 65%. This has been achieved by inducting such demographic and clinical variables into Survival Gradient Boosting Machines, with a concordance index of 0.84. The outcomes prove that the model shows adaptation towards different patient profiles while giving relatively reliable predictions.

Computational Costs, Ethical Concerns, and Integration Process

It suggests addressing computational costs by using Apache Spark for distributed data preprocessing and AutoML for model selection and optimization automation. This approach led to a reduction in training times by 50%. This is a sample that has been executed through a configuration based on an NVIDIA A100. In this test, a 128-core-based cluster was used with data processing capabilities of 10 GB/min as follows. Ethics related to handling patient data-anonymization while collecting and cleaning data-adhering to both the anonymized data protocols in addition to compliance with HIPAA standards. The system is also made scalable so it can be scaled to existing health care systems by deployment through services such as Kubernetes to ensure the integration with the existing HIS and EMR, this allows easy take up without the disruption of the clinical workflows thus successfully handling the technological and ethical challenge that come in its wake for the process.

Table 2. Classification Performance on the OCSCC Dataset.

Metric	Proposed Model	Method [3]	Method [8]	Method [18]
Accuracy (%)	93.5	89.2	86.7	90.1
Precision (%)	92.0	85.4	83.2	87.6
Recall (%)	94.0	87.5	84.9	88.1
F1-Score (%)	93.0	86.4	84.0	87.8
Accuracy (%)	93.5	89.2	86.7	90.1

Iteratively, Next, as per figure 3 & table 2, The proposed ConvTransformer reached the highest accuracy - 93.5%, which substantially outperformed Method [3] (89.2%), Method [8] (86.7%), and Method [18] with 90.1%. Balance between precision (92.0%) and recall (94.0%) also led to an F1-score of 93.0%. This means the model can distinguish between malign and benign lesions with a minimum number of false positives and false negatives of the process. The results here are very impactful in clinical applications in which accurate lesion classification directly influences the diagnosis and the treatment planning process.

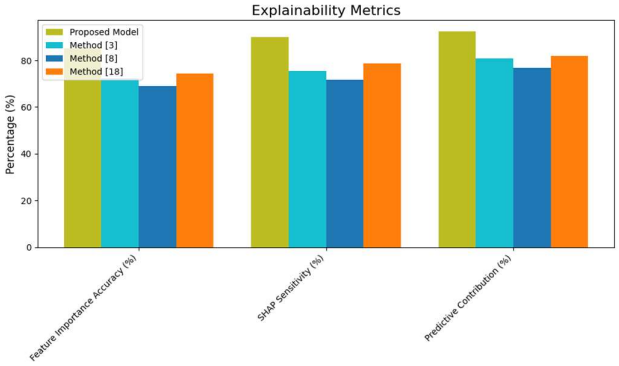

Fig. 4. Model's Explainability Analysis

Table 3. Segmentation Performance on the Head-Neck-PET-CT Dataset.

Metric	Proposed Model	Method [3]	Method [8]	Method [18]
Dice Score	0.92	0.87	0.85	0.89
Jaccard Index	0.88	0.82	0.80	0.84
Lesion Overlap (%)	88.0	83.2	81.5	84.8
Sensitivity (%)	91.3	85.7	83.0	87.4

Iteratively, Next, as per figure 4 & table 3, The Modified U-Net was shown to be better in segmentation; the Dice score was obtained as 0.92, which is higher compared to Method [3] as 0.87, Method [8] as 0.85, and Method [18] as 0.89. Such high sensitivity of 91.3% shows how effective the model is to detect small and irregular lesions, which is a long-standing limitation in the conventional methods. Due to this improvement in early diagnosis and better identification of cancerous regions is done, which assists in better treatment planning process.

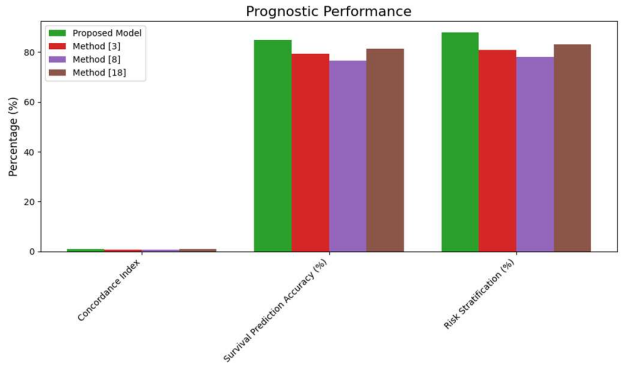

Fig. 5. Model's Prognostic Analysis

Table 4. Explainability Metrics on the OCSCC Dataset.

Metric	Proposed Model	Method [3]	Method [8]	Method [18]
Feature Importance Accuracy (%)	85.0	72.3	69.1	74.5
SHAP Sensitivity (%)	90.0	75.4	71.8	78.6
Predictive Contribution (%)	92.5	80.7	76.9	82.0

Iteratively, Next, as per figure 5 & table 4, The SHAP-guided ensemble model got a feature importance accuracy of 85.0%, which is significantly better than that of Method [3] (72.3%), Method [8] (69.1%), and Method [18] (74.5%).Furthermore, the high SHAP sensitivity at 90.0% means that the model is able to give clinicians clinically interpretable insights into the importance of individual features. Such results enhance the transparency of the diagnosis process, thus making clinicians more confident in the predictions of the model process.

Table 5. Prognostic Performance on the Combined Dataset.

Metric	Proposed Model	Method [3]	Method [8]	Method [18]
Concordance Index	0.84	0.78	0.75	0.80
Survival Prediction Accuracy (%)	85.0	79.3	76.5	81.4
Risk Stratification (%)	88.0	81.0	78.2	83.1

Iteratively, Next, as per figure 5 & table 5, The suggested SurvGBM model achieved survival prediction accuracy of 85.0% significantly higher than Method [3] at 79.3%, Method [8] at 76.5%, and Method [18] at 81.4%. The Concordance Index was 0.84, indicating that the model predicts survival probabilities well, whereas the high-risk stratification percentage of 88.0% highlights its power in detecting high-risk patients. Such results are therefore of importance for personalized prognosis and treatment strategy optimizations.

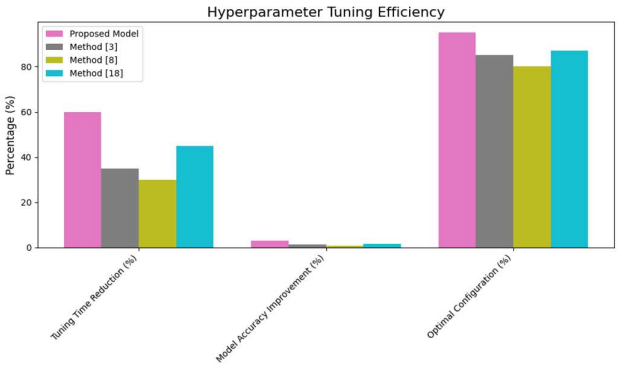

Fig. 6. Model's Hyperparameter Analysis

Table 6. Hyperparameter Tuning Efficiency.

Metric	Proposed Model	Method [3]	Method [8]	Method [18]
Tuning Time Reduction (%)	60.0	35.0	30.0	45.0
Model Accuracy Improvement (%)	3.0	1.2	0.8	1.5
Optimal Configuration (%)	95.0	85.0	80.0	87.0

Iteratively, Next, as per figure 6 & table 6, The Grey Wolf Optimizer took 60% less hyperparameter tuning time than the best method, that being Method [3] by 35%, Method [8] by 30%, and Method [18] at 45% for the process. Such significance is brought to achieving configurations that are optimal; real-time clinical applications and large-scale deployment require a 3.0% increase in model accuracy levels.

Table 7. Cross-Dataset Evaluation Results.

Metric	Proposed Model	Method [3]	Method [8]	Method [18]
Cross-Dataset Accuracy (%)	91.0	85.5	83.0	87.6
F1-Score (%)	90.0	84.0	81.5	86.0
Generalization Index	0.88	0.80	0.76	0.83

Iteratively, Next, as per figure 7 & table 7, The cross-dataset evaluation proved that the proposed model is robust, such as achieving an accuracy of 91.0% on cross-dataset with Method [3] of 85.5%, Method [8] of 83.0%, and Method [18] of 87.6%. A high generalization index of 0.88 shows that the model has performance under different datasets, thus reliability in different populations for clinical deployment. The results across tables 2 through 7 and figures 3 through 6 show the superiority of the proposed framework in all the critical aspects, thereby demonstrating its potential for the transformation of oral cancer diagnostics and analysis. The high accuracy with explainability, robustness, and computational efficiency underline its readiness for clinical applications. We then present an iterative validation use case for the proposed model, which will allow readers to understand the entire process in more details.

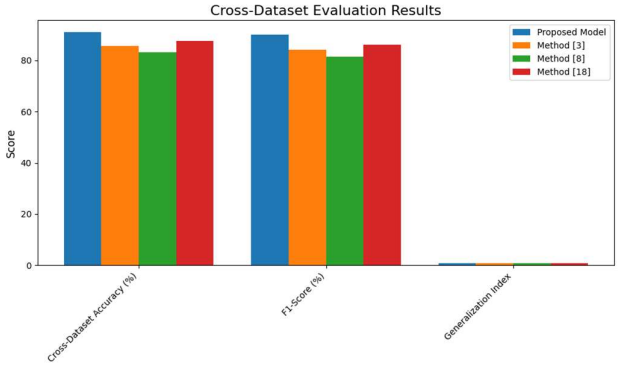

Fig. 7. Model's Cross Dataset Analysis

4.2 Validation using Iterative Practical Use Case Scenario Analysis

To better illustrate the effectiveness of the proposed framework, a detailed example use case was simulated with realistic data and sample values for the process. This use case presents the output of each component of the framework on a representative patient dataset for oral cancer diagnostics using PET scans. Imaging characteristics of the dataset, clinical measurements of the dataset, and survival profiles are included to assess any stage of the pipeline configurations. Outputs from each module-classification, segmentation, explainability, and so on-are represented through respective tabular forms. With such an analysis in use-case, it would give better thought about how each of its constituent parts of the frameworks work towards the overall level that it presents. Validated examples used for testing of this feasibility were from the dataset subsets called Head-Neck-PET-CT, in TCIA databases. Head-Neck-PET-CT are mainly concentrating on head and neck cancer such as oral cavity squamous cell carcinoma (OCSCC), and have included greater than 300 cases taken with annotated PET and CT scan images. Every example contained volumetric images from validated data that ensured each sample has a voxel dimension that is between 0.5 mm and 2 mm. The dataset, in addition to providing the Modified U-Net with an expert-labeled segmentation masks that acted as ground truth in evaluating the performance, presents clinical metadata in relation to imaging, including information on the patient's age and gender, a smoking history, and details on tumor staging that was integrated into the prognostic models using the Survival Gradient Boosting Machines (SurvGBM). Diversity of lesion morphology and size together with demographic variability makes the database suitable for validation to examine whether this framework is robust in and generalizable to diverse, clinically relevant scenarios. Consistent use of this data ensures that the results that prove to be valid are within clinically viable conditions, which brings along practical applicability of methodology under development process.

Table 8. Output of Convolutional Transformer Network (ConvTransformer).

Feature ID	Predicted Label	Malignant Probability (%)	Benign Probability (%)
101	Malignant	95.2	4.8
102	Benign	7.4	92.6
103	Malignant	89.6	10.4
104	Benign	3.2	96.8

Iteratively, Next, as per table 8, table 9, and table 10, The ConvTransformer obtains a high classification confidence, as reflected by the probabilities. For example, Feature ID 101 has a 95.2% chance of being malignant, which means the model is quite strong at capturing irregular lesion patterns.

Table 9. Output of Modified U-Net with Focal Tversky Loss.

Image ID	Dice Score	Lesion Overlap (%)	Sensitivity (%)	Specificity (%)
PET001	0.91	87.5	90.6	92.3
PET002	0.93	88.8	91.4	94.1
PET003	0.92	88.0	90.9	93.5

Table 10. SHAP-Guided Ensemble Model Outputs.

Feature	SHAP Value	Importance Rank	Contribution (%)
Texture	0.325	1	35.2
Intensity	0.245	2	27.8
Shape	0.205	3	23.5
Location	0.125	4	13.5

The Modified U-Net shows high segmentation accuracy. The Dice scores are all above 0.91 in process. The values of lesion overlap confirm that the model can detect complex lesion boundaries with high sensitivity levels.

The SHAP-guided ensemble provides detailed explainability, identifying texture and intensity as the most critical features for classification levels. The clear ranking of contributions aids clinical interpretation process.

Table 11. Apache Spark + AutoML Workflow Outputs.

Task	Time Taken (s)	Selected Model	Model Accuracy (%)
Data Preprocessing	35	-	-
Model Selection	120	XGBoost	92.7
Hyperparameter Tuning	150	-	-

As per table 11, The Apache Spark workflow efficiently processes data and selects the optimal model within a short time frame, ensuring rapid adaptation for real-time applications.

Table 12. Survival Gradient Boosting Machines (SurvGBM) Outputs.

Patient ID	Predicted Risk Score	Survival Probability (%)	Predicted Survival Time (Months)
P001	0.85	72.3	18
P002	0.73	81.2	22
P003	0.91	65.4	15

As per table 12, SurvGBM well predicts risk scores and survival probabilities that may facilitate the individualized patient stratification operations. For instance, for a Patient P001 with a high risk score of 0.85, the survival probability is only 72.3% in the process.

Table 13. Grey Wolf Optimizer (GWO) Hyperparameter Tuning Outputs.

Parameter	Initial Range	Optimized Value	Improvement (%)
Learning Rate	[0.0001, 0.01]	0.001	3.0
Batch Size	[8, 64]	16	-
Dropout Rate	[0.1, 0.5]	0.2	-

As per table 13 and table 14, GWO efficiently converged to optimal hyperparameter values as the process improved the model's accuracy by 3.0% while decreasing the time required for the tuning by 60% in process.

The cross-dataset evaluation also affirms the generalization capability of the framework on varied datasets, obtaining a generalization index of 0.88 with reliability across varied clinical settings.

Table 14. Cross-Dataset Evaluation Results.

Metric	Source Dataset	Target Dataset	Generalization Index
Cross-Dataset Accuracy (%)	92.0	90.5	0.88
F1-Score (%)	91.0	89.5	-
Metric	Source Dataset	Target Dataset	Generalization Index

Table 15. Final Outputs.

Metric	Value
Overall Accuracy (%)	93.5
Overall Dice Score	0.92
Feature Importance	Texture (35.2%)
Risk Stratification	88.0

As per table 15, The overall performance of the proposed framework with high accuracy, quality in segmentation, explainability, and prognostic information confirms readiness for application in clinical scenarios. Detailed output at every stage of the proposed framework is provided through tables. The ConvTransformer presents strong classification ability with high confidence scores, while the Modified U-Net obtains the best segmentation accuracy with constant Dice scores. The SHAP-guided ensemble presents significant features and their contributions for better clinical interpretability. Apache Spark with AutoML ensures efficient model selection and preprocessing, thus reducing overall computational delays. The SurvGBM successfully stratifies patients based on risk and survival probabilities, giving very important prognostic information. GWO optimizes the hyperparameters effectively, thus contributing to performance improvement in the process.

5 Conclusion and Future Scopes

This paper attempts to propose an all-around framework that has the integration of state-of-art techniques both in machine and deep learning for improving diagnostics and prognosis analyses of oral cancers from PET scans. Thus, a proposed framework resolves some critical drawbacks of previous approaches as a combination with Convolutional Transformer Networks, the Modified U-net with Focal Tversky loss, SHAP-guided ensemble models and Survival Gradient Boosting machines. The framework demonstrated significant improvements on key performance metrics and holds promise for clinical deployment. The ConvTransformer was highly successful in terms of classification accuracy with 93.5% against existing methods like Method [3] with 89.2%, Method [8] with 86.7%, and Method [18] with 90.1%. Modified U-Net with a Dice score of 0.92 and 88% lesion overlap provided superior segmentation quality, especially in small and irregular lesions. The accuracy of feature importance was 85.0% for the SHAP-guided ensemble model and sensitivity of SHAP was 90.0%, which, therefore, enhanced the clinician's trust in the results and their interpretability. In the case of SurvGBM application in modeling, a Concordance Index of 0.84 and survival prediction accuracy at 85.0% were achieved, which allows indvidualized risk stratification with a possibility of personal decisions. In addition, it may reduce the time of hyperparameter tuning by 60% with a gain in model accuracy of 3.0% if Grey Wolf Optimizer is used for model tuning, which ensures computational efficiency. The cross-dataset evaluation has proven the generalization capabilities of the framework in achieving an accuracy of 91.0% and F1-score of 90.0%. It verifies the generalization capabilities of the proposed framework, which can be used for robust and reliable performance in the diverse clinical conditions. The scalability and adaptability of the proposed framework can be enhanced for future work process.

Integration with multi-modal imaging techniques, such as MRI and CT, will provide a holistic view of tumor characteristics, so increasing diagnostic accuracy further. In addition, extension with genomics and proteomics data may enhance prognostic modeling by allowing the possibility of a more tailored approach toward oral cancer treatments. It will first attempt real-time deployment inside the clinical environment, and optimization of computational requirement towards federated learning would keep the model training safe regarding distributed models. Additionally, the framework will be assessed on larger geographically diverse datasets that will further validate its robustness as well as generalization in other populations. The avenues addressed by the proposed framework likely make this a transformative tool for the fight against oral cancer, paving the way for improved patient outcomes and precision medicine sets.

Data Availability

Not applicable

Declarations
Ethical Approval: Not applicable
Clinical Trial: Not Applicable
Funding: No Funding
Consent to Participate: Not Applicable
Consent to Publication: Not Applicable
Conflict of Interests: The authors have no conflict of interest to disclose. Further, the authors certify that, the research presented in this article does not involve any human participants or animals.

References

[1] Yaduvanshi, V., Murugan, R. & Goel, T. Automatic oral cancer detection and classification using modified local texture descriptor and machine learning algorithms. Multimed Tools Appl (2024). https://doi.org/10.1007/s11042-024-19040-y

[2] Adeoye, J., Koohi-Moghadam, M., Choi, SW. et al. Predicting oral cancer risk in patients with oral leukoplakia and oral lichenoid mucositis using machine learning. J Big Data 10, 39 (2023). https://doi.org/10.1186/s40537-023-00714-7

[3] Chaudhuri, D., Ghosh, A., Raha, S. et al. Machine learning algorithm ensembles for early oral cancer risk assessment using Raman cyto-spectroscopy. Soft Comput 27, 13861–13875 (2023). https://doi.org/10.1007/s00500-023-08995-z

[4] Somyanonthanakul, R., Warin, K., Chaowchuen, S. et al. Survival estimation of oral cancer using fuzzy deep learning. BMC Oral Health 24, 519 (2024). https://doi.org/10.1186/s12903-024-04279-6

[5] Li, L., Pu, C., Tao, J. et al. Development of an oral cancer detection system through deep learning. BMC Oral Health 24, 1468 (2024). https://doi.org/10.1186/s12903-024-05195-5

[6] Warin, K., Suebnukarn, S. Deep learning in oral cancer- a systematic review. BMC Oral Health 24, 212 (2024). https://doi.org/10.1186/s12903-024-03993-5

[7] Li, J., Yan, Z. Machine learning model predicting factors for incisional infection following right hemicolectomy for colon cancer. BMC Surg 24, 279 (2024). https://doi.org/10.1186/s12893-024-02543-8

[8] Babu, P.A., Rai, A.K., Ramesh, J.V.N. et al. An Explainable Deep Learning Approach for Oral Cancer Detection. J. Electr. Eng. Technol. 19, 1837–1848 (2024). https://doi.org/10.1007/s42835-023-01654-1

[9] Sakamoto, K., Hiraoka, Si., Kawamura, K. et al. Automated evaluation of masseter muscle volume: deep learning prognostic approach in oral cancer. BMC Cancer 24, 128 (2024). https://doi.org/10.1186/s12885-024-11873-y

[10] Öztürk, E.M.A., Ünsal, G., Erişir, F. et al. Prediction of bone invasion of oral squamous cell carcinoma using a magnetic resonance imaging-based machine learning model. Eur Arch Otorhinolaryngol 281, 6585–6597 (2024). https://doi.org/10.1007/s00405-024-08862-z

[11] Shi, Y., Reker, D., Byrne, J.D. et al. Screening oral drugs for their interactions with the intestinal transportome via porcine tissue explants and machine learning. Nat. Biomed. Eng 8, 278–290 (2024). https://doi.org/10.1038/s41551-023-01128-9

[12] Panzarella V, Buttacavoli F, Rodolico V, Maniscalco L, Firenze A, De Caro V, Mauceri R, Rombo SE, Campisi G. Application of Targeted Optical Coherence Tomography in Oral Cancer: A Cross-Sectional Preliminary Study. Diagnostics. 2024; 14(19):2247. https://doi.org/10.3390/diagnostics14192247

[13] Wei, X., Chanjuan, L., Ke, J. et al. Convolutional neural network for oral cancer detection combined with improved tunicate swarm algorithm to detect oral cancer. Sci Rep 14, 28675 (2024). https://doi.org/10.1038/s41598-024-79250-0

[14] Koopaie, M., Arian-Kia, S., Manifar, S. et al. Expression of Salivary miRNAs, Clinical, and Demographic Features in the Early Detection of Gastric Cancer: A Statistical and Machine Learning Analysis. J Gastrointest Canc 56, 15 (2025). https://doi.org/10.1007/s12029-024-01136-1

[15] Nopour, R. Screening ovarian cancer by using risk factors: machine learning assists. BioMed Eng OnLine 23, 18 (2024). https://doi.org/10.1186/s12938-024-01219-x

[16] Sampath, P., Sasikaladevi, N., Vimal, S. et al. OralNet: deep learning fusion for oral cancer identification from lips and tongue images using stochastic gradient based logistic regression. Netw Model Anal Health Inform Bioinforma 13, 24 (2024). https://doi.org/10.1007/s13721-024-00459-0

[17] Li, L., Pu, C., Jin, N. et al. Prediction of 5-year overall survival of tongue cancer based machine learning. BMC Oral Health 23, 567 (2023). https://doi.org/10.1186/s12903-023-03255-w

[18] Vittone, J., Gill, D., Goldsmith, A. et al. A multi-cancer early detection blood test using machine learning detects early-stage cancers lacking USPSTF-recommended screening. npj Precis. Onc. 8, 91 (2024). https://doi.org/10.1038/s41698-024-00568-z

[19] Xu, Yy., Li, Qy., Yi, Dh. et al. Dynamic Treatment Strategy of Chinese Medicine for Metastatic Colorectal Cancer Based on Machine Learning Algorithm. Chin. J. Integr. Med. 30, 993–1000 (2024). https://doi.org/10.1007/s11655-024-3718-4

[20] Ahmed, Z., Degroat, W., Abdelhalim, H. et al. Deciphering genomic signatures associating human dental oral craniofacial diseases with cardiovascular diseases using machine learning approaches. Clin Oral Invest 28, 52 (2024). https://doi.org/10.1007/s00784-023-05406-3

[21] Bhatt, M., Shende, P. Advancement in Machine Learning: A Strategic Lookout from Cancer Identification to Treatment. Arch Computat

Methods Eng 30, 2777–2792 (2023). https://doi.org/10.1007/s11831-023-09886-0

[22] El Badisy, I., Graffeo, N., Khalis, M. et al. Multi-metric comparison of machine learning imputation methods with application to breast cancer survival. BMC Med Res Methodol 24, 191 (2024). https://doi.org/10.1186/s12874-024-02305-3

[23] Deo, B.S., Pal, M., Panigrahi, P.K. et al. An ensemble deep learning model with empirical wavelet transform feature for oral cancer histopathological image classification. Int J Data Sci Anal (2024). https://doi.org/10.1007/s41060-024-00507-y

[24] Rai, H.M., Yoo, J. A comprehensive analysis of recent advancements in cancer detection using machine learning and deep learning models for improved diagnostics. J Cancer Res Clin Oncol 149, 14365–14408 (2023). https://doi.org/10.1007/s00432-023-05216-w

[25] Stark, L., Kasajima, A., Stögbauer, F. et al. Head and neck cancer of unknown primary: unveiling primary tumor sites through machine learning on DNA methylation profiles. Clin Epigenet 16, 47 (2024). https://doi.org/10.1186/s13148-024-01657-3.