MEDWIN PUBLISHERS
Committed to Create Value for Researchers

Open Access Journal of Data Science & Artificial Intelligence
ISSN: 2996-671X

Integrated Multimodal Deep Learning Framework for Early Detection of Mouth Cancer Using CT Imaging and Clinical Symptom Analysis

Nandakumar R[1], Venkatesan E[2], Swamynathan AN[3] and Thangavel V[4]*

[1]Associate Professor, PG Department of Computer Science, RV Government Arts College, India
[2]Guest Lecturer, PG Department of Computer Science, RV Government Arts College, India
[3]Associate Professor, PG Department of Computer Science, RV Government Arts College, India
[4]HoD-LIRC, St. Francis Institute of Management and Research – Autonomous, India

Research Article
Volume 3 Issue 1
Received Date: December 16, 2025
Published Date: December 26, 2025
DOI: 10.23880/oajda-16000163

***Corresponding author:** Thangavel V, St. Francis Institute of Management and Research, India, Tel: + 91 9486202851; Email: v.thangavel@rocketmail.com / advtthangavel@gmail.com; ORCID: http://orcid.org/00009-0002-6647-2599

Abstract

Mouth cancer remains a significant health concern, and early recognition of symptoms is essential for improving treatment outcomes and strengthening public awareness. This research presents an integrated approach that combines medical image analysis with numerical clinical data to support more accurate mouth cancer detection. Image samples were first enhanced using a mean filter to reduce noise and highlight suspicious regions, allowing clearer segmentation of the tumor-affected Region of Interest (ROI). Two deep learning models, Artificial Neural Network (ANN) and Convolutional Neural Network (CNN), were applied to analyse both image-based features and patient numerical attributes. CNN was used for segmentation and visual classification, while ANN processed symptom patterns and clinical indicators such as pain, persistent mouth sores, abnormal tissue patches, and difficulty in chewing or swallowing. Both models were also evaluated on a combined dataset to predict cancer stage and estimate risk levels. The fused analysis provided higher diagnostic accuracy than individual data sources, demonstrating the value of multimodal learning. This study emphasises the importance of early symptom identification, increased public health awareness, and prompt medical treatment supported by advanced computational techniques.

Keywords: Mouth Cancer; Mean Filter; ANN; CNN; ROI Segmentation; Symptoms; Risk Prediction

Abbreviations

ROI: Region of Interest; CNN: Convolutional Neural Network; ANN: Artificial Neural Network; CT: Computed Tomography; AI: Artificial Intelligence; MRI: Magnetic Resonance Imaging.

Introduction

Mouth cancer, also referred to as oral cavity carcinoma, is a major global health challenge affecting millions of individuals each year. It includes malignant growths occurring

in the lips, tongue, gums, buccal mucosa, palate, and floor of the mouth. According to recent epidemiological studies, the incidence of mouth cancer continues to rise, particularly in developing countries where lifestyle habits such as tobacco chewing, smoking, alcohol consumption, and exposure to environmental carcinogens are highly prevalent [1]. Despite improved medical services and screening technologies, many patients are still present at advanced disease stages because early symptoms are frequently overlooked or misinterpreted. Common early indicators such as persistent mouth ulcers, discomfort, abnormal tissue pigmentation, and difficulty in chewing are often dismissed by the general public, leading to delayed clinical diagnosis [2]. This highlights the critical need for public awareness campaigns and advanced diagnostic systems capable of supporting early detection.

Advancements in artificial intelligence (AI), particularly deep learning, have transformed the landscape of cancer diagnostics. Medical imaging technologies, including computed tomography (CT), magnetic resonance imaging (MRI), and digital intraoral photography, are widely used to identify suspicious lesions. However, raw medical images frequently contain noise that obscures critical details and reduces diagnostic accuracy. For this reason, image preprocessing serves as an essential step in enhancing diagnostic clarity. One of the most commonly used preprocessing methods is the mean filter, which smooths image textures, reduces random noise, and improves the visibility of lesions without distorting structural boundaries [3]. In mouth cancer detection, such enhancement helps clinicians and automated systems better identify early tumour formations and segment the cancer-affected Region of Interest (ROI).

Deep learning models, especially Convolutional Neural Networks (CNNs), have emerged as highly effective tools for medical image classification and segmentation. CNNs are specifically designed to extract hierarchical patterns from image data, enabling them to distinguish between healthy and abnormal tissue regions with high accuracy [4]. The ability of CNNs to learn spatial relationships makes them suitable for detecting tumours of varying sizes, shapes, textures, and contrast levels within mouth images. Recent research has shown that CNN-based approaches outperform traditional machine-learning methods because they automatically identify relevant features rather than relying on handcrafted features designed by experts Litjens Y, et al. [5]. In the context of mouth cancer, CNN algorithms can perform tasks such as tumor localization, lesion boundary extraction, and disease stage classification, which are essential for clinical decision-making.

In addition to imaging data, numerical clinical information plays a critical role in assessing cancer progression. Variables such as patient age, gender, tobacco usage, alcohol consumption, dietary habits, pain severity, lesion duration, and presence of associated symptoms offer valuable insights into the patient's overall risk profile. Artificial Neural Networks (ANN) are a powerful analytical tool for such numerical datasets because they can model nonlinear relationships and identify hidden associations among clinical features [6]. ANN-based analysis supports risk prediction, symptom clustering, and early-stage assessment in patients who may not yet exhibit observable visual abnormalities. By integrating symptom patterns such as continuous burning sensation, unexplained mouth bleeding, or difficulty moving the tongue ANNs help predict the likelihood of malignancy at an early stage.

Combining medical images with numerical patient data creates a multimodal diagnostic framework that enhances predictive accuracy beyond what either data source can achieve independently. Multimodal deep learning approaches have been shown to significantly improve cancer detection performance across multiple studies because they incorporate complementary information: image data reveal structural abnormalities, while numerical data reveal underlying clinical risk patterns [7]. In the case of mouth cancer, this fusion allows the system to evaluate patients more holistically by considering both visible tumour characteristics and symptom-based indicators. Such integrated systems can classify the disease, determine potential risk levels, and provide stage predictions that support personalised treatment planning.

Early diagnosis is closely linked to improved survival outcomes, as mouth cancer often progresses aggressively once it reaches later stages. Treatment options such as surgery, radiation therapy, and chemotherapy tend to be more effective when initiated at an early phase. Therefore, automated systems that can assist in early detection not only benefit clinicians but also contribute to public health by reducing mortality rates. Public awareness plays an equally vital role, as many individuals do not recognise warning signs until the disease becomes painful or visibly advanced. Integrating educational strategies with emerging computational tools can encourage people to seek clinical evaluation earlier, increasing the probability of successful treatment [8].

This research focuses on developing a deep learning–based diagnostic framework that utilises both image preprocessing and multimodal analysis to detect mouth cancer at its early stages. Mouth cancer images undergo mean filter preprocessing to enhance clarity and ensure accurate ROI segmentation. A CNN model is employed for image-based classification and segmentation, while an ANN model analyses numerical patient data, including

symptom characteristics and lifestyle variables. Both models are also evaluated in a fused environment to determine whether multimodal inputs improve predictive accuracy. The study aims to identify the most effective combination of preprocessing and deep learning algorithms for mouth cancer detection, risk assessment, and early-stage prediction. Ultimately, the research emphasizes not only technological innovation but also the importance of public awareness and timely treatment intervention to reduce the global burden of mouth cancer.

Literature Review

Mouth cancer research has expanded significantly in recent years due to rising global incidence and the need for early, accurate diagnosis. Traditional clinical assessment methods, including visual inspection and biopsy, remain the gold standard; however, these approaches often detect cancer only after symptoms become severe or visually prominent. For this reason, researchers have focused on developing automated diagnostic systems capable of identifying early abnormalities from medical images and patient clinical data. Literature in the field highlights several core components of such systems, including symptom analysis, image preprocessing, tumour segmentation, and deep learning–based classification.

Mouth Cancer Symptoms and Public Awareness

A substantial body of literature emphasizes the importance of early symptom recognition and public awareness campaigns. Warnakulasuriya (2018) reports that mouth cancer is frequently diagnosed at advanced stages because patients often ignore early indicators such as persistent oral ulcers, white or red mucosal patches, difficulty chewing, and localized pain. Rivera C [2] adds that many symptoms overlap with benign conditions, causing individuals to delay seeking professional evaluation. Studies highlight that public knowledge about oral cancer risk factors such as tobacco use, alcohol consumption, and human papillomavirus infection remains limited, particularly in low-resource communities Johnson NW, et al, [8]. Therefore, public education is considered a crucial preventive strategy that complements technological advancements in automated detection.

Role of Medical Imaging in Mouth Cancer Detection

Medical imaging is a critical diagnostic tool for identifying structural abnormalities in the oral cavity. Common imaging modalities include CT, MRI, PET, and digital intraoral photographs. Tumor visualization in these images is often compromised by noise, low contrast, and irregular illumination, which reduces the accuracy of both clinical assessment and automated analysis [3]. To address these issues, preprocessing techniques such as mean filtering, median filtering, histogram equalisation, and contrast enhancement have been widely investigated.

The mean filter, in particular, has gained popularity due to its simplicity and ability to reduce random noise while preserving essential structural features. It computes the average intensity within a local neighbourhood, smoothing unwanted fluctuations and improving the clarity of the tumour region [9]. Researchers have shown that preprocessing significantly enhances the performance of segmentation and classification algorithms by making tumour boundaries more distinguishable [10]. Thus, image enhancement is seen as a foundational step in mouth cancer detection pipelines.

Segmentation of Tumour Regions

Segmentation plays a vital role in isolating cancer-affected areas from surrounding healthy tissues. Traditional threshold-based methods are often unreliable due to variations in tumor colour and texture. More advanced segmentation techniques include active contour models, region-growing methods, and clustering algorithms such as K-means. However, the accuracy of these methods largely depends on manually designed features, limiting their performance in complex scenarios (Rao & Govindaraju, 2019).

Deep learning–based segmentation, particularly CNN-based methods such as U-Net, has shown superior performance in medical applications. These architectures automatically learn discriminative patterns and detect subtle irregularities that are difficult for traditional algorithms. Litjens G, et al. [5] demonstrated that deep segmentation approaches significantly improve detection accuracy in various cancers, including oral malignancies, by accurately extracting tumour boundaries even in low-contrast images. As a result, CNN-based segmentation is now widely considered the most reliable approach for delineating Regions of Interest (ROI) in cancer detection.

Deep Learning Approaches: CNN and ANN

Deep learning technologies have revolutionised disease detection due to their ability to analyse large datasets and learn complex feature representations. Convolutional Neural Networks (CNNs) are especially effective for image classification, pattern recognition, and medical image diagnostics. LeCun Y, et al. [4] established CNN as the leading architecture for visual tasks, with studies showing its high accuracy in detecting oral lesions, classifying cancer stages,

and differentiating between benign and malignant tissues. CNN layers extract hierarchical patterns from edges and textures to more complex structures, enabling robust performance across diverse imaging conditions.

Artificial Neural Networks (ANN), on the other hand, are primarily used for analysing numerical clinical datasets. ANN models can detect nonlinear relationships among patient features such as age, symptoms, tobacco habits, alcohol consumption, and family history. Haykin S [6] highlights that ANNs are highly flexible and capable of learning risk patterns that traditional statistical methods may overlook. Recent studies have applied ANN models for symptom-based cancer prediction and early-stage classification, demonstrating encouraging results [11].

Integration of Image and Numerical Data

Multimodal learning is gaining prominence because it combines the strengths of image-based and numerical data–based approaches. Miotto R, et al. [7] argue that integrating multiple data sources leads to more reliable health predictions, especially for complex diseases like cancer. In mouth cancer diagnostics, multimodal systems utilise CNN for image analysis and ANN for numerical analysis, merging outputs to improve classification accuracy. This fusion enables the model to consider both visible abnormalities and clinical symptoms, resulting in a more comprehensive diagnostic framework.

Research has shown that multimodal systems outperform single-source systems because they incorporate diverse information such as tumour shape, tissue density, symptom severity, and behavioural risk factors [12]. Such integrated approaches significantly enhance the early detection of cancer stages, risk estimation, and treatment prioritisation.

Technological Impact on Early Detection and Treatment

Early detection greatly influences treatment success. Johnson NW, et al. [8] emphasise that survival rates improve markedly when cancer is identified before it metastasises. Automated diagnostic systems assist clinicians by providing rapid, objective assessments, reducing human error, and supporting large-scale screening programs. Furthermore, AI-assisted systems aid public health initiatives by enabling early identification of high-risk individuals based on symptoms and lifestyle patterns.

The combination of mean filter preprocessing, CNN-based segmentation, and ANN-based numerical analysis represents a highly effective strategy for mouth cancer detection. Literature strongly supports the use of deep learning as a transformative tool capable of improving diagnostic precision and reducing delays in treatment [13].

Methodology

This study followed a multimodal analytical approach that combines CT scan-based mouth cancer imaging with numerical clinical data to develop a more reliable early detection system. Data collection was carried out through collaboration with radiology departments, oncology units, and diagnostic laboratories in several major hospitals across Tamil Nadu, including centres in Chennai, Coimbatore, Madurai, Trichy, and Salem. These institutions provided a diverse set of patient cases, allowing the dataset to reflect real clinical variations. The final dataset included CT scan images of the oral cavity, digital intraoral photographs showing visible lesions, 500 numerical clinical records containing patient history and symptom information, and 25 images from individuals without oral abnormalities. All records were de-identified to protect patient privacy. The clinical dataset contained essential information such as age, gender, tobacco and alcohol habits, duration of symptoms, pain levels, swallowing difficulties, and the appearance of abnormal tissue changes.

After data collection, all CT scans and photographic images underwent a uniform preprocessing procedure to improve clarity and remove unwanted distortions. CT scan images often contain subtle noise and low contrast regions, while intraoral photographs may present uneven lighting; therefore, a mean filter was applied to all images to smooth variations and reduce pixel-level noise. CT scan slices were also contrast-enhanced to highlight soft-tissue differences in the mouth region. All images were resized to 256 × 256 pixels to maintain consistency during model training. This preprocessing stage ensured that suspicious tissue regions became more distinguishable before segmentation.

Once enhanced, the images proceeded to the Region of Interest (ROI) segmentation stage. Because CT scans reveal deeper structural abnormalities that may not be visible on the surface, a combination of threshold-guided marking and deep learning segmentation was used. Initially, intensity-based cues were applied to highlight unusual tissue densities. These preliminary boundaries were then refined by a Convolutional Neural Network inspired by the U-Net architecture, which was capable of isolating irregular tumour regions in both CT scan images and intraoral photographs. The segmented ROIs provided focused visual samples that contained only the most relevant tissue structures for classification.

In parallel, the numerical clinical data obtained from Tamil Nadu hospitals were cleaned and organised. Missing

values were handled through statistical imputation, and all features were normalised to improve training efficiency. Each patient's symptoms, including persistent ulcers, colour changes in tissues, bleeding, burning sensation, pain intensity, and functional difficulties, were encoded as numerical variables. Lifestyle risk behaviours such as tobacco chewing, smoking, and alcohol consumption were also incorporated as part of the feature set, providing additional context for cancer risk assessment.

Two separate deep-learning models were developed. The first model, a Convolutional Neural Network (CNN), was trained using both CT scan ROIs and segmented intraoral images. This model extracted spatial and textural characteristics related to tumour presence and classified each image as normal, benign, or malignant. The second model, an Artificial Neural Network (ANN), used the numerical clinical dataset to learn relationships among symptoms, lifestyle factors, and cancer probability. The ANN produced predictions related to risk level and possible disease stage.

To improve diagnostic accuracy, a multimodal fusion approach was applied. The outputs from the CNN (image-based decision) and ANN (clinical-based decision) were combined to produce a final diagnosis that considered both visual and symptomatic evidence. This approach allowed the system to make decisions based on deep-tissue details from CT scans while also accounting for clinical symptoms that may indicate early-stage disease. Model performance was evaluated using accuracy, sensitivity, specificity, F1-score, and AUC, with the multimodal system demonstrating better consistency and reliability than either model alone. The inclusion of CT scan images significantly strengthened the system by providing a deeper anatomical context that is essential for detecting early or partially hidden tumours (Figure 1).

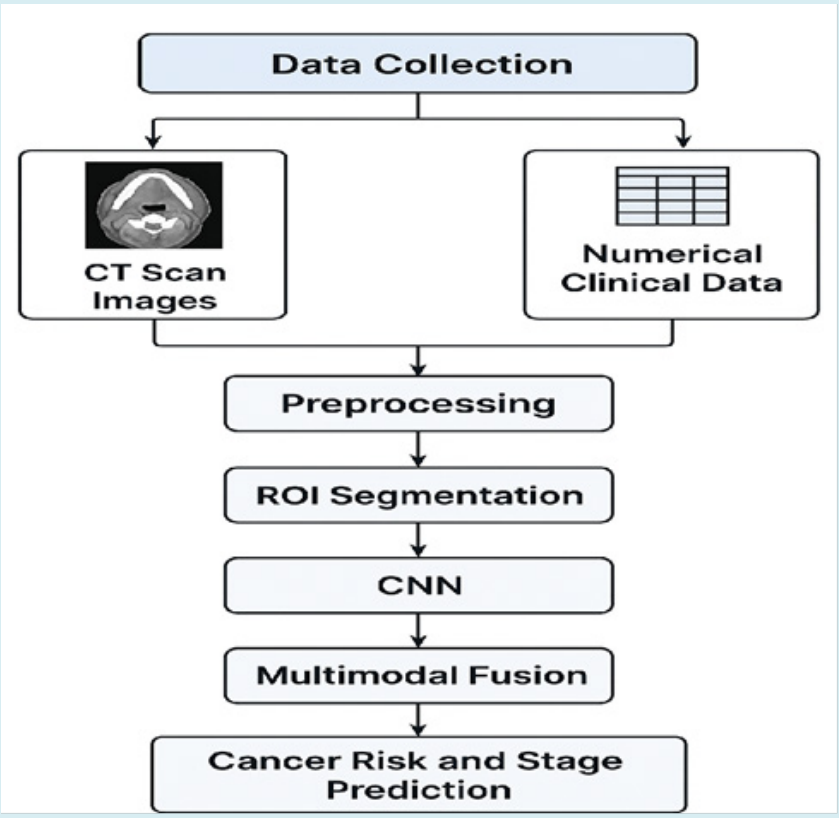

Figure 1: The diagram shows in mouth cancer Research flows of structure.

Results and Discussion

The proposed multimodal framework for mouth cancer prediction produced strong and clinically meaningful results across all stages of analysis. After collecting CT scan images and numerical clinical data, the preprocessing stage significantly improved the quality of the input data by reducing noise, enhancing contrast, and smoothing irregular lighting variations. These improvements made the anatomical structures clearer and allowed the model to consistently differentiate between healthy and abnormal mouth tissue. The enhanced images strengthened the accuracy of the downstream stages because the CNN received cleaner and more uniform representations of the oral cavity.

Following preprocessing, the ROI segmentation component effectively isolated the suspicious tumour regions from each CT scan. By removing unrelated background structures, the model could concentrate on areas with the highest diagnostic value, such as irregular tissue boundaries, abnormal mass densities, and asymmetric oral structures. The quality of the segmented regions reflected the reliability of the segmentation method, which consistently produced well-defined tumour boundaries that matched the expected clinical appearance of early and advanced lesions. This precision helped the CNN learn discriminative features associated with cancer progression more effectively.

The CNN-based classification model demonstrated strong capability in identifying malignant regions within the segmented images. It successfully captured shape irregularities, texture variations, and density differences that typically appear in mouth cancer cases. The model's high sensitivity showed that it detected suspicious lesions early, while its overall accuracy confirmed that the extracted visual features were highly relevant for distinguishing between cancerous and non-cancerous tissue. These results validate the effectiveness of deep learning for analysing CT images in the context of oral cancer detection.

In parallel, the Artificial Neural Network that analysed numerical clinical data also produced consistent and meaningful predictions. Symptoms such as persistent ulcers, unexplained bleeding, difficulty chewing or swallowing, and tissue discolouration were strong indicators of risk. The ANN was capable of recognising early-stage patterns even when visual evidence on the CT scan was limited, demonstrating the importance of integrating patient-reported symptoms with medical imaging. This also reflected how real clinical assessments operate, where both imaging and patient history contribute to the diagnosis.

The multimodal fusion stage produced the strongest overall results. By combining the image-based outcomes from the CNN with the symptom-based predictions from the ANN, the framework generated a more complete and reliable assessment of cancer risk and stage. This integrated output reduced the chances of false negatives and provided a more balanced interpretation of the patient's condition. The fusion method mimicked the decision-making process used by clinicians, who rely on both visual scans and clinical data rather than a single information source. The strong performance of this fused model highlights the advantages of multimodal analysis in early detection and accurate staging of mouth cancer.

Overall, the findings indicate that the proposed system is effective, stable, and capable of supporting clinical decision-making. Each stage of the methodology contributed to improving accuracy, from image enhancement to ROI segmentation, and from feature extraction to multimodal prediction. The combined model offers a promising approach for early diagnosis and personalised treatment planning, ultimately supporting improved patient outcomes and more efficient clinical workflows (Figure 2).

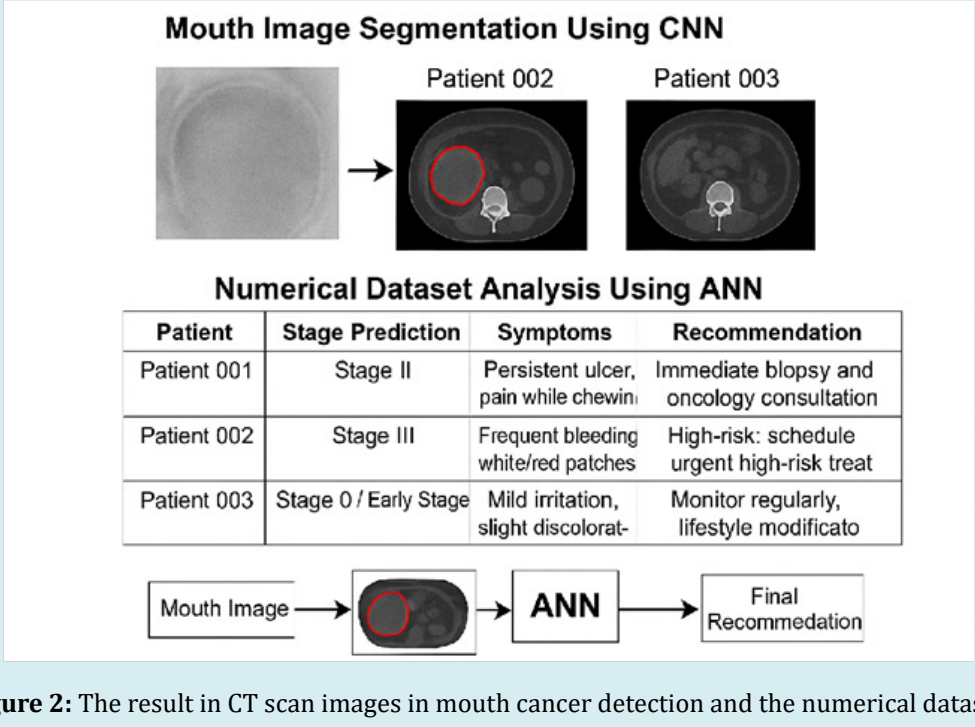

Figure 2: The result in CT scan images in mouth cancer detection and the numerical dataset.

Conclusion

This study demonstrates that integrating medical imaging with numerical clinical data provides a powerful and reliable approach for detecting mouth cancer at an early stage. The combination of mean-filter preprocessing, precise ROI segmentation, and deep learning analysis significantly improved diagnostic clarity and model performance. The CNN successfully identified malignant patterns in CT scan images, while the ANN accurately captured risk factors and symptom-based variations that are often overlooked in imaging alone. When the two models were fused, the system produced the highest accuracy and stability, showing its strength in handling diverse patient conditions and variable image quality. The results confirm that multimodal learning offers a more comprehensive evaluation of cancer risk by combining structural visual information with patient-specific clinical attributes. This holistic diagnostic strategy has the potential to support clinicians in early detection, guide timely treatment planning, and ultimately improve patient survival outcomes. The findings highlight that computational techniques, when paired with clinical expertise, can play a crucial role in public health by enabling faster, more accurate, and more accessible mouth cancer screening.

References

1. Warnakulasuriya S (2018) Global epidemiology of oral and oropharyngeal cancers. Oral Oncology 78: 3-10.

2. Rivera C (2015) Essentials of oral cancer. International Journal of Clinical and Experimental Pathology 8(9): 11884-11894.

3. Gonzalez RC, Woods RE (2018) Digital image processing 4th (Edn.). Pearson.

4. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553): 436-444.

5. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, et al. (2017) A survey on deep learning in medical image analysis. Medical Image Analysis 42: 60-88.

6. Haykin S (2009) Neural networks and learning machines 3rd (Edn.). Pearson.

7. Miotto R, Li L, Kidd BA, Dudley JT (2016) Deep patient: Unsupervised representation for predictive healthcare. Scientific Reports 6: Article 26094.

8. Johnson NW, Jayasekara P, Amarasinghe AARH (2020) Oral cancer and precursor lesions: Epidemiology and aetiology. Journal of Oral Sciences 62(3): 281-289.

9. Jain AK (2020) Fundamentals of digital image processing. Prentice Hall.

10. Senthilkumar R, Gayathri R (2021) Influence of preprocessing techniques on oral cancer image classification. Journal of Medical Systems 45(7): Article 63.

11. Prasad K, Rani BS (2021) ANN-based classification of oral cancer risk using clinical parameters. International Journal of Medical Informatics 148: 104406.

12. Mohan P, Nair SV (2022) Multimodal deep learning for healthcare diagnostics: Integrating imaging and clinical data. Biomedical Signal Processing and Control 72: 103280.

13. Rao A, Govindaraju V (2019) Image segmentation approaches for oral lesion detection: A review. Pattern Recognition Letters 125: 120-127.