**ADVANCEMENT IN EARLY DETECTION OF ORAL CANCER**

**AND METASTASIS CONTROL**

**Rimsy Dua1, Santosh Kumar Singh2, Pooja Kadam3 ,Shreya Patil3**

1Assistant Professor, Information Technology, Thakur College of Science and Commerce, Mumbai, Maharashtra, India

2Head of the Department, Information Technology, Thakur College of Science and Commerce, Mumbai, Maharashtra, India

3Student, Information Technology, Thakur College of Science and Commerce, Mumbai, Maharashtra, India

4Student, Information Technology, Thakur College of Science and Commerce, Mumbai, Maharashtra, India

**ABSTRACT**

This research investigates machine learning-based disability detection of oral cancer combined with metastasis management strategies. The research evaluates the diagnostic capabilities of ResNet-50 CNN, SVM, and Random Forest algorithms through assessment of image-histopathological data and patient CSV information. The dataset contains two measurement variables which include metastasis status and organ dimensions.The outcomes demonstrated that CNN(ResNet-50) scored the best identification rate of 90.08% compared to SVM 86.26% and Random Forest 74.81%. Machine learning techniques show effective potential in oral cancer diagnosis according to the obtained results that lead to early detection and better healthcare results.

**Keywords:**CNN, Early Detection, MATLAB, Metastasis, Oral cancer, Random Forest, ResNet-50, SVM.

1. **INTRODUCTION**

The global concern with oral cancer needs immediate early detection to increase survival rates while lowering death figures. Modern diagnosis mainly depends on visual examination and tissue analysis under the microscope but these methods take time before rendering inaccurate human-based results. Machine learning (ML) approaches present themselves as a transformative diagnostic option by creating fast still precise and touch-free assessment solutions.

Current diagnosis practices continue to need enhancement because they depend on human observers along with their natural probability of making mistakes. Through machine learning algorithms medical practitioners expect to access diagnosis tools that provide faster and more precise services. These algorithms help medical practitioners boost their ability to detect diseases at an early stage and lower the number of incorrect test results and improve how they plan treatments. The research sets out to build and validate the efficiency of different machine learning methods for oral cancer detection along with metastasis prediction.

A total of 2157 histopathological oral cancer images form the dataset with additional patient history information contained in a CSV file which includes cancer-stage and metastasis data as features. The machine learning algorithms used in this project combine ResNet-50s with ResNet-50 architecture together with Support Vector Machines and Random Forest.ResNet-50 fulfills all requirements to classify images through its discovery of specific histopathological analysis elements. During meta-analysis of clinical data performed by Random Forest and SVM users can make metastasis occurrence predictions using cancer stage data and population metadata.

The data processing system alongside algorithm development runs on MATLAB for efficient execution of the project. The platform operates efficiently while providing a comparison feature among ML models. The diagnostic system under development integrates imaging processing methods with clinical data because its focus is early detection of oral cancer alongside metastasis control.Comparing ResNet-50 with SVM and Random Forest algorithms on the given dataset helps determine which ML technique works best for this application.

This study s AI-based oncology diagnostic advancement which will result in better patient outcomes and care improvements. Research uses ResNet-50 for image classification tasks because of its success rate alongside SVM and Random Forest for testing different machine learning methods with clinical data diversity. The proposed work shows great potential to transform the diagnosis process of oral cancer through improved detection methods which healthcare practitioners can use for early diagnosis and treatment decisions.

Existing research in oral cancer provides useful information regarding disease factors and progression in spite of the preventability of the disease. Early diagnosis of oral cancer remains critical in order to reduce mortality from the disease since the majority of oral cancers appear in visible and palpable sites. AI and machine learning approaches have made important contributions in oral cancer diagnosis as they improve early detection and risk prediction and provide treatment predictive estimates. The study is a contribution towards larger efforts in advancing oral cancer control through the use of advanced tools for enhancing diagnostic precision as well as speed of delivery.

The diagnosis is more precise as the project integrates medical image analysis and clinical information in creating more efficient treatment that results in better patient outcomes. With the use of machine learning algorithms, medical practitioners obtain speedy and non-invasive testing methods that outshine traditional diagnostic techniques in speed and operational efficiency. Future studies that integrate AI in oral cancer diagnosis will leverage these results in creating improved application methods and establishing standardized procedures. The project contributes value in the current studies in leveraging machine learning in order to improve healthcare and medical diagnostics.

1. **METHODOLOGY**

MATLAB program development methods which aim to detect oral cancer and metastasis at an early stage require research to enhance diagnostic accuracy along with treatment approaches. Through a combination of ResNet-50 and SVM and Random Forest machine learning models the approach uses image analysis together with clinical data for better detection purposes. MATLAB provides advantages for data management and algorithm execution which leads to the development of precise diagnostic systems. The method fills a requirement for timely diagnosis which leads to better patient survival statistics and decreased mortality rates from oral cancer while promoting superior clinical care.

For the research, we have acquired the image from the internet. And the data used in this study was collected via the image of the area affected by oral cancer. The data is stored in the Excel file.csv it known as the training set. It is consisting of four Column from which there are the twelve Column for input. The input Colum consists of the Cancer type, Metastasis, Stage, Size, Spread, Survival rate, Cause, and Metastasizing affected organ. The class will be on the base of the image which differentiate the cancerous and non-cancerous from the image. To acquire the data here, we have used ResNet-50, SVM and Random Forest classifier.

**2.1 Data Collection and Data Preprocessing**

It has 2157 images of number of cancer-affected parts and patient background data exist in a CSV file for early oral cancer detection. This CSV document contains information about patient ages with their metastasis status together with tumor size stages and affected organs and survival statistics. The preprocessing step includes image resizing to 224x224 pixels together with pixel normalization from 0 to 1 followed by ResNet50 feature extraction. The CSV contains missing data that is interpolated and the dataset becomes balanced through SMOTE application. We distributed our data into training and testing components (80:20). The model's accuracy capabilities are tested through metrics analysis to measure its efficiency in cancer and metastasis control prediction.

**2.2 Feature Extraction**

Early detection of oral cancer with machine learning requires the importance extraction from medical images and patient records to increase model operation effectiveness. ResNet-50 exists as a deep convolutional neural network (CNN) through its implementation of residual connections to facilitate gradient flow optimization in deep learning operations. The extracted features are processed by Support Vector Machine (SVM) to identify a hyperplane that delivers maximum distinction between cancerous and non-cancerous samples. The random forest model executes its functionality by applying multiple decision trees to different datasets for producing more accurate predictions and avoiding overfitting problems. The system uses SVM together with Random Forest to process clinical data and analyze ResNet-50 image features to diagnose early-stage cancer effectively by extracting complex patterns from combined visual and structured data types which leads to better diagnosis accuracy.

**2.3 Model Developement**

Early detection of oral cancer with metastasis control depends on the integration of ResNet50 along with SVM and Random Forest in MATLAB. The medical images go through ResNet50 for tumor detection while SVM analyzes lesions by their texture attributes. Random Forest merges visual examination data with clinical histories from patients to forecast metastasis risks. The developed algorithm executes a two-step approach that first trains ResNet50 for oral cancer recognition through fine-tuning then applies SVM to classify lesions from images. Risk assessment performed by Random Forest uses both medical imaging outputs together with clinical data inputs. Automated feature extraction implementing MATLAB scripts together with model training processes aids both the improvement of diagnostic precision and creation of individualized treat out for its excellence at lesion identification while Random Forest produces better metastasis predictions. The model demonstrates potential for clinical implementation in early detection and risk stratification but its pathway faces barriers from dataset quality concerns as well as ethical considerations.

**2.3.1 ResNet-50**

ResNet-50 improves the detection of oral cancer through its ability to deliver fast precise diagnosis without invasive procedures while surpassing visual inspection along with histopathological examination. The advanced version of CNNs solves image extraction issues by applying residual links to resolve the gradient vanishing problem in complex medical feature acquisition. The diagnostic system examines high-resolution images that human observers would not easily detect by performing automated analysis of biopsy slides and CT and MRI scan data. ResNet-50 shows excellent performance in medical diagnosis by its ability to identify the critical metastasis indicators which include microvascular invasion and lymph node involvement. By using transfer learning techniques the model accepts pre-trained models to discover particular cancer signatures in oral tissues. Angiogenic assessment capabilities of the model help predict possible metastatic potential through its analysis of new blood vessel development.

**2.3.2 Support Vector Machine**

In medical image analysis technicians employ Support Vector Machine (SVM) as a supervised learning technique to identify intricate patterns within high-dimensional datasets when detecting oral cancer in tissue samples between malignant and benign conditions. SVM gets excellent accuracy results while working with minimal amounts of data by making optimized margin calculations through its hyperplane generation system. The system uses image characteristics such as texture and shape and patient information including age and smoking history together with genetic patterns to provide precise diagnosis along with early medical assistance.The tool effectively predicts metastasis behavior through combining historical patient information with image biomarkers that manages accurate results even within scarcity of metastatic data. The system generates individualized risk evaluation protocols together with customized treatment protocols. SVM models in MATLAB gain efficiency for extended treatment success because they can incorporate fresh patient information to enhance their performance persistently.

**2.3.3 Random Forest**

Random Forest improves the better prediction quality and lowers overfitting by its ability to generate combined forecasts from various decision trees. The algorithm has been successfully evaluating quantitative and qualitative data alongside proficient handling of absent data points. Random Forest uses imaging data features alongside clinical information which comprises patient age together with lifestyle factors and genetic data for oral cancer diagnosis. A majority voting process among multiple decision trees enables tissue classification into benign and malignant categories which increases the detection sensitivity and specificity. The system evaluates cancer spread potential through combined analysis of tissue characteristics and clinical characteristics for creating tailored medical interventions. The combination of multiple decision trees in Random Forest allows improved risk evaluation for cancer which supports clinical decisions to yield better patient success and lower death statistics.

1. **MODELING AND ANALYSIS**

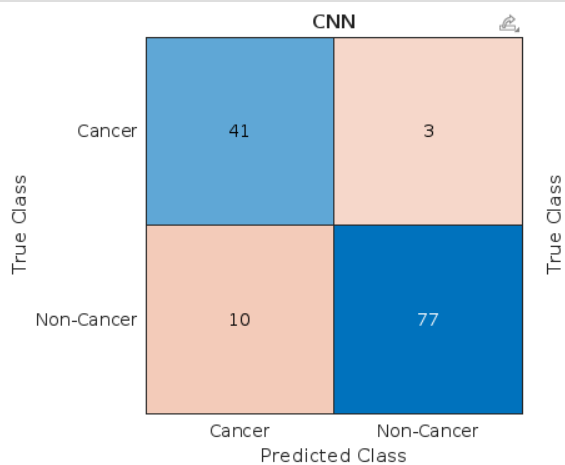
**3.1 Model Evaluation Metrics**

The model performance evaluation relied on accuracy results alongside confusion matrices which studied the classification accuracy. ResNet50 demonstrated the best performance in model accuracy by reaching 90.08% which exceeded the results from SVM (86.26%) and Random Forest (74.81%). The confusion matrices show how ResNet50 committed less misclassification errors than Random Forest and its confusion rates were highest between cancer and non-cancer cases. ResNet50 wrongfully diagnosed ten healthy tissue samples as cancerous tissue samples whereas SVM went wrong with fourteen healthy tissue sample classifications. The greater number of misclassifications within Random Forest resulted in degraded overall accuracy levels. The performance metrics demonstrate how ResNet50 achieves the best accuracy in classifying cancer from non-cancer tissues.

**3.2 Confusion Matrix**

**3.2.1 ResNet-50**

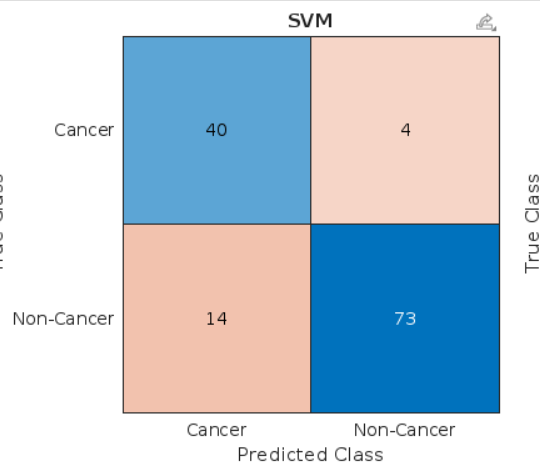
A confusion matrix provides essential performance information about a CNN model used to classify oral cancer. The model performed 41 precise identifications (TP) between oral cancer cases and successfully recognized 77 cases (TN) without cancer. The model failed to detect three oral cancer cases in addition to incorrectly classifying a total of 10 non-cancer cases as cancer and 3 case of being canceras non cancerous. Early detection of oral cancer is essential for successful treatment so FN cases must be minimized. Early oral cancer diagnosis requires a reduction of false negative readings as the main priority.



**Figure 1**. Confusion matrix of CNN’s ResNet-50 analysis for oral cancer detection

**3.2.2 Support Vector Machine**

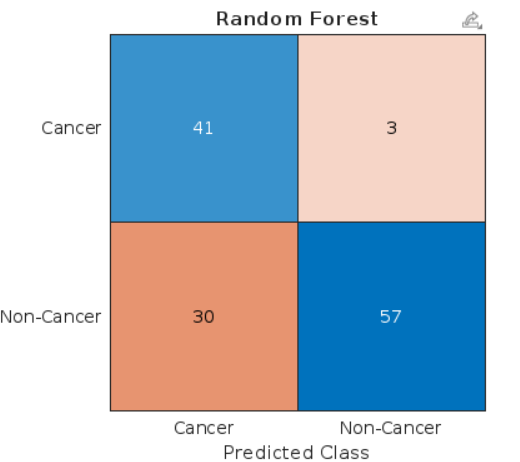
A confusion matrix demonstrates the evaluation results of an SVM model which performs oral cancer diagnosis. In the developed system there were 40 correctly identified oral cancer cases (TP) alongside 73 correctly identified non-cancer cases (TN). The model registers 14 wrong positive results suggesting healthy tissues classified as cancerous tissue leading to potential unnecessary medical interventions. The detection of 4 false negatives represents situations where the model failed to identify oral cancer while 4 false positives resulted in cases wrongly classified as cancer. Both FN and FP outcomes must be minimized because prompt cancer identification stands as crucial for successful treatments. The correct diagnosis of oral cancer depends on minimizing false negative results.



**Figure 2**. Confusion matrix of SVM analysis for oral cancer detection

**3.2.3 Random Forest**

A Random Forest model achieves its classification performance accuracy by utilizing the confusion matrix for oral cancer diagnosis. The confusion matrix demonstrates true positive identification of 41 oral cancer cases along with 57 correctly identified true negative instances. Thirty cases of misidentification by the model indicate oral cancer classifications made on non-cancer cases which could lead to unnecessary treatment procedures and patient stress. Early detection becomes crucial because the use of 3 false negative results allows oral cancer cases to go undetected.



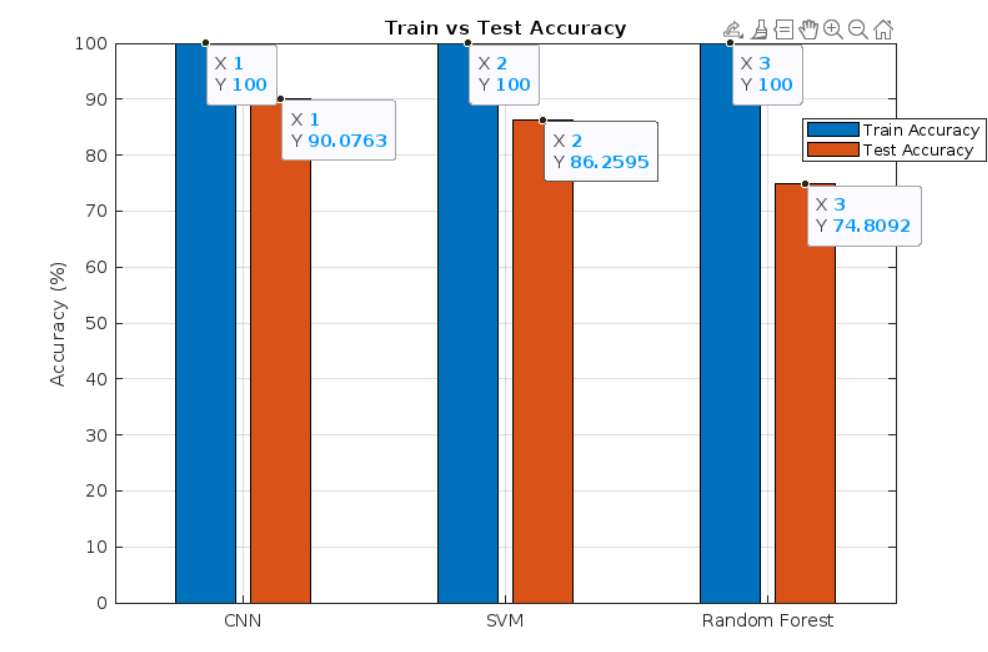
**Figure 3.** Confusion matrix of Random Forest analysis for oral cancer detection

**3.3 Overall Acurracy**

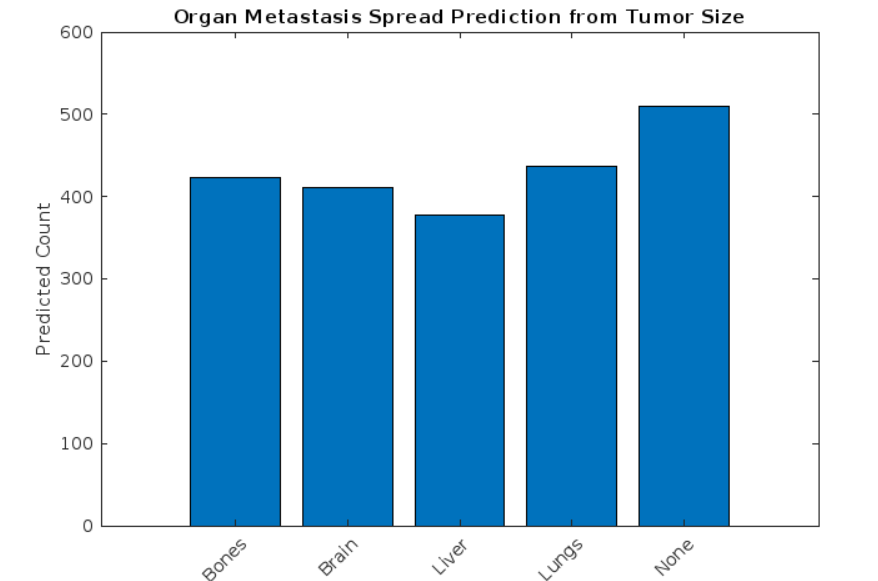
Oral cancer detection and metastasis control become possible through this MATLAB code which unites ResNet50 with SVM and Random Forest algorithms. ResNet50 demonstrates an image analysis capability to achieve 90.08% accuracy in testing. The performance accuracy of SVM is 86.26% while Random Forest achieves 74.81%. The combined use of these models increases general diagnostic accuracy because it combines their respective strengths to generate better clinical diagnostic results. By integrating these features researchers attain better accuracy when detecting cancers together with risk assessments.

1. **RESULTS AND DISCUSSION**

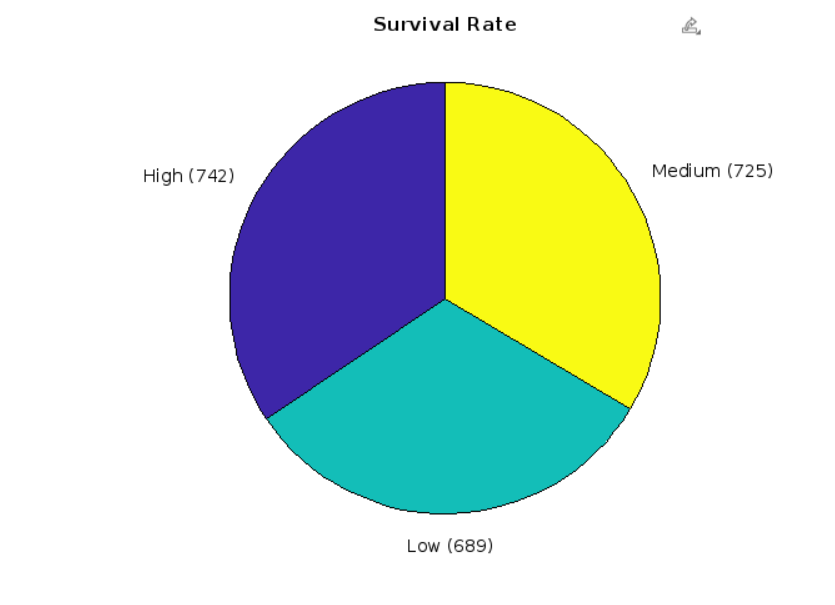
The machine learning technologies achieved notable performance metrics in oral cancer and metastasis control applications. ResNet-50 operating on CNN reached the best identification performance with 90.08% accuracy which exceeded the analysis from SVM reaching 86.26% and from Random Forest reaching 74.81%. By employing its deep architecture structure ResNet-50 identified multiple complex patterns in medical images resulting in better accuracy levels. The structured data performance of SVM was high but Random Forest achieved moderate accuracy because of its ensemble learning process. The study demonstrates how deep learning approaches led most efficiently by ResNet-50 serve to enhance both early diagnosis and treatment strategies in oral cancer cases.



**Figure 4.** Bargraph explaining the comparison between the accuracy obtained by training and testing the model developed for the ResNet-50 , SVM, and Random Forest Algorithm



**Figure 5.** Bargraph analyzing the Organ Affected due to metastasizing of the cancer



**Figure 6.** Piechart depicting the Survival rates on the basis of stages of oral cancer.

1. **CONCLUSION**

The combination of ResNet50 with SVM and Random Forest within MATLAB code delivers encouraging findings for early detection of oral cancer and metastasis regulation. The performance of ResNet50 reaches 90.08% accuracy which surpasses both SVM with 86.26% and Random Forest with 74.81%. By combining deep learning with traditional machine learning the diagnostic system can perform with higher precision and reliability. The system could benefit from future developments which would merge transformers along with graph neural networks to build a performance-enhanced system for clinical environments. The research demonstrates that AI technology represents a promising approach to enhance both diagnosis quality and management of oral cancer diseases.

**ACKNOWLEDGEMENTS**

We are deeply indebted to the numerous individuals whose support and encouragement were instrumental in the success of this research paper. Our sincere appreciation goes to Dr. (Mrs.) C.T. Chakraborty, Principal of Thakur College of Science and Commerce, and the teaching staff, whose unwavering assistance ensured timely completion of our work.

We extend collective gratitude to the IT Department for providing access to the lab facilities and their invaluable technical expertise, which significantly facilitated my research endeavors.

Special thanks are due to my guide, Prof. Rimsy Dua, whose insightful guidance was pivotal in shaping this project. Additionally, we are grateful to our family members, who provided unwavering support and shared their expertise throughout this journey.

1. **REFERENCES**
2. Vineet Vinay, Praveen Jodalli, Mahesh S. Chavan, Chaitanya. S. Buddhiko, “Artificial Intelligence in Oral Cancer Diagnostics and Prognostics” ,Diagnostics, 2025
3. Shintaro Sukegawa, Sawako Ono, Futa Tanaka, Yuta Inoue, “Efectiveness of deep learning classifers in histopathological diagnosis of oral squamous cell carcinoma by pathologists” , Nature Portfolio, 2023
4. Olutomilayo Olayemi Petinrin, Faisal Saeed, Muhammad Toseef, Zhe Liu, Shadi Basurra, “Machine learning in metastatic cancer research: Potentials, possibilities, and prospects” ,Computational and Structural Biotechnology Journal, 2023
5. Xiaoshuai Xu, Linlin Xi, Lili Wei, Luping Wu, Yuming Xu, “Deep learning-assisted contrast-enhanced CT–based diagnosis of cervical lymph node metastasis of oral cancer: a retrospective study of 1466 cases”, European Radiology, 2023
6. Shriniket Dixit, Anant Kumar, and Kathiravan Srinivasan, “A Current Review of Machine Learning and Deep Learning Models in Oral Cancer Diagnosis: Recent Technologies, OpenChallenges, and Future Research Directions” , Diagnostics, 2023
7. Xaviera A. López-Cortés, Felipe Matamala, Bernardo Venegas, and César Rivera, “Machine-Learning Applications in Oral Cancer: A Systematic Review” , Applied Science, 2022
8. Carlos M. Chiesa-Estomba, Manuel Graña, Alfonso Medela, Jon A. Sistiaga-Suarez, Jerome R. Lechien, “Machine Learning Algorithms as a Computer- Assisted Decision Tool for Oral Cancer Prognosis and Management Decisions: A Systematic Review”, ORL, 2022
9. María García-Pola, Eduardo Pons-Fuster, Carlota Suárez-Fernández, “Role of Artificial Intelligence in the Early Diagnosis of Oral Cancer. A Scoping Review” , Cancers, 2021
10. Amr Bugshan, Imran Farooq, “Cancer Stem Cells in Oral Cavity Squamous Cell Carcinoma: A Review” , F1000Research, 2020
11. Yi-Ju Tseng, Hsin-Yao Wang, Ting-Wei Lin, Jang-Jih Lu, Chia-Hsun Hsieh, Chun-Ta Liao, “Development of a Machine Learning Model for Survival Risk Stratification of Patients With Advanced Oral Cancer”, JAMA Network Open, 2020