**ADVANCEMENT IN EARLY DETECTION OF ORAL CANCER**

**AND METASTASIS CONTROL**

|  |  |  |  |
| --- | --- | --- | --- |
| **Mrs. Rimsy Dua** | **Dr. Santosh Kumar Singh** | **Ms. Pooja Kadam** | **Ms. Shreya Patil** |
| Asst. Professor | Head of the Department (I.T) | Student- MSc.I.T. | Student- MSc.I.T. |
| Thakur College of Science and Commerce,  Kandivali (E), Mumbai-101 | Thakur College of Science and Commerce,  Kandivali (E), Mumbai-101 | Thakur College of Science and Commerce,  Kandivali (E), Mumbai-101 | Thakur College of Science and Commerce,  Kandivali (E), Mumbai-101 |
| Email id.:  rimsydua051292@gmail.com | Email id.:  sksingh@tcsc.edu.in | Email id.: kadam2001pooja@gmail.com | Email id.:  shreyarpatil15@gmail.com |

***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.***

# **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.

# **ResNet-50**

The advance ResNet-50 introduces benefits for detection through its role as an advanced version of CNN. Old diagnostic techniques that include visual inspection and histopathological examination take too much time and contain room for human mistakes. ResNet-50 with deep learning represents an advanced ML technique that enables fast accurate non-invasive diagnostic tools. The Residual Network (ResNet) architecture produced ResNet-50 as an advanced variant that specializes in extracting elaborate features from image data. The concept was created to solve deep neural networks' gradient vanishing problem through introduction of residual path connections. The ability to achieve superior representation learning exceeds prior limits.

ResNet-50 stands vital in both metastasis control and early oral cancer detection because it performs advanced pattern analysis for feature extraction. The ResNet-50 model receives high-resolution medical images consisting of biopsy slides CT scans MRI scans and intraoral photos within oral cancer diagnosis. The 50-layer architectural structure of this model consists of convolutional layers together with pooling layers and residual connections directing gradients smoothly while training to prevent gradient explosions and vanishings. The model obtains information from both basic image characteristics (such as textures and edges) and meaningful patterns (including tumor borders and distortions of tissue structure). Weak structural distinctions that human observers cannot view become detectable with ResNet-50 because this model demonstrates exceptional performance in benign and malignant lesion classification. Due to its depth and ability to learn residual patterns ResNet-50 demonstrates strong potential for identifying cancer early which results in more successful diagnosis.

The analysis performed by ResNet-50 on both histopathology images and immunohistochemistry slides investigates cellular-level indications of metastasis development. The deep structure of this technology allows it to detect essential markers of metastasis which include microvascular invasion together with lymph node involvement and perineural invasion. Through transfer learning a ResNet-50 model becomes sensitive to distinct metastatic patterns by taking ImageNet pre-training and applying it on oral cancer-specific information. This model identifies tumor cells clusters along with recognizing invasive margins and measuring tissue differentiation levels. The system evaluates new blood vessel formation patterns during angiogenesis because these patterns serve as aggressive tumor indicators along with predicting metastatic potential. Medical staff can use the enhanced accuracy of metastasis predictions to develop purpose-built therapies that stop cancer from spreading and enhance treatment results.

# **III. Random Forest**

The survival of cancer patients together with their treatment results depend heavily on detecting cancer early through precise metastasis prediction. The diagnostic approaches traditional methods face problems when trying to deal with complex medical information that emerges from histopathology images and multiple patient historical elements. Random Forest solves the analytical difficulties by linking prediction abilities from several decision trees to guarantee improved accuracy while minimizing overfitting problems. Random Forest demonstrates peak utility in medical applications because its capabilities enable the handling of numerical and categorical data and the management of missing values and the prevention of overfitting through ensemble learning. His capacity to handle complex high-dimensional data enables Random Forest to become an efficient solution for cancer detection and metastasis risk assessment with reliable results.

Random Forest proves to be highly effective at oral cancer early diagnosis through its analysis of imaging data along with clinical information. Detailed features involving texture and color fluctuations along with cellular arrangements are extracted from histopathology images by the algorithm even though these patterns are outside human perception capabilities. Patient data combining both diagnostic information with individual variables such as age, tobacco usage and alcohol use and medical history and genetic components and personal characteristics. The system constructs various decision trees that use separate parts of the provided data for training purposes. The diagnosis of cancer tissue requires each tree to determine malignant or benign status before final results emerge through majority voting. The combination method raises both sensitivity for cancer detection and specificity in avoiding erroneous results thereby boosting early diagnosis precision and reliability. The early discovery of cancer enables doctors to provide treatment at an appropriate time which enhances both positive therapy results and decreased death rates.

The spreading of cancer to foreign body areas through Metastasis strongly affects how oral cancer patients will fare. For predicting metastasis risk through Random Forest depends on its analysis of both clinical indicators and histopathological indicators. The algorithm analyzes minute tissue features to assess cell geometry and tissue consistency along with nuclear form abnormalities since these indicate intense cancer behavior. Physical details about the tumor combined with information about lymph nodes plus genetic abnormalities and individual health-related information are integrated with imaging results for analysis. Random subsets of cancer-related features are used to create multiple decision trees by the algorithm which helps identify complicated metastasis patterns. When assessing metastasis risk during prediction, individual trees determine metastasis risk independently and the overall prediction combines their responses. Appropriate risk evaluations become possible when you go through the oncology procedure which guides physicians to develop individualized treatment strategies and deploy specific therapies. Random Forest contributes to better clinical choices and better results for patients through improved metastasis risk estimation.

# **IV. Support Vector Machines**

Support Vector Machine is an effective supervised learning method that medical image analysis uses to process high-dimensional data while detecting the complicated patterns. SVM helps themalignant from benign tissue abnormalities when examining histopathological images together with patient information in oral cancer detection. The algorithm produces its best results through margins optimized by hyperplane generation which helps achieve high accuracy with minimal data requirements. The integration of image attributes (\_texture\_, \_shape\_ and \_color\_) with patient data collection points (biological age, smoking records and genetic patterns) allows SVM to improve diagnostic accuracy which enables early medical action and better medical results.

The MATLAB application provides fitcsvm as well as other functions which enable users to train their model when analyzing combined image and patient history data. Through image processing methods involving noise reduction and edge detection alongside segmentation the system extracts essential features from the images. The normalization process combine the CSV data and extracted image features in a single dataset. The SVM model obtains instruction for malignant and benign classifications using these features throughout the training process. Model performance optimization uses cross-validation and hyperparameter tuning that includes selecting a kernel type from linear, polynomial or RBF options. MATLAB provides native visualization software which enables users to check classification precision together with ROC curve evaluation for model diagnostic attributes.

SVM develops predictions about metastasis through examinations of historical patient data in combination with image-derived biomarkers. A control model for metastasis requires training with datasets which include both diagnosis and follow-up data to establish predictive indicators of metastasis. SVM maintains reliable classification accuracy regardless of underrepresentation of metastatic cases in data samples. Risk assessments tailored to individual patients become possible through the combination of patient timeline information and imaging pattern analysis using SVM models which suggest treatment methods accordingly. The ability of MATLAB through multidimensional data handling ensures continuous refinement and update capacity for SVM-based metastasis prediction models with new patient datasets to enhance long-term treatment outcomes.

# **VI. Litreature Review**

**Vineet Vinay, Praveen Jodalli, Mahesh S. Chavan, Chaitanya. S. Buddhiko in research work titled “***Artificial Intelligence in Oral Cancer Diagnostics and Prognostics***”** from Diagnostics 2025 assesses artificial intelligence applications in oral cancer diagnosis and prognosis to achieve better early detection and improved sensitiveness in diagnostics. DCNN along with ANN algorithms demonstrate superior early detection performance because of their exceptional accuracy. The research draws information from three data types which consist of imaging data and histopathology findings and clinical data records. The research shows that diagnostic abilities and predictive power have been improved through the findings. The analysis addresses data diversity problems along with ethical matters and emphasizes both standardized datasets alongside robust clinical validation protocols. [1]

**Shintaro Sukegawa, Sawako Ono, Futa Tanaka, Yuta Inoue in reserach work titled in , “***Efectiveness of deep learning classifers in histopathological diagnosis of oral squamous cell carcinoma by pathologists***”** from Nature Portfolio in 2023 investigates the proper use of deep learning classifiers with VGG16 and ResNet50 models for porphological diagnosis of oral squamous cell carcinoma. Pathologists have prepared the data through their histopathological image collection process. The combination of VGG16 with SAM optimizer yielded the best results with 0.8622 accuracy along with 0.9602 AUC. The diagnostic performance of pathologists improved substantially when using CNN assistance according to the research findings (p = 0.031). The paper identifies the need of expanding the diverse datasets and conduct external validations while examining various CNN models to enhance their widespread use. ​[2]

**Olutomilayo Olayemi Petinrin, Faisal Saeed, Muhammad Toseef, Zhe Liu, Shadi Basurra in research work titled in, “***Machine learning in metastatic cancer research: Potentials, possibilities, and prospects***”** Computational and Structural Biotechnology Journal 2023 uses machine learning along with deep learning to enhance the early identification and treatment methods for patients with metastatic cancer. The analysis of PET/CT and MRI image data through CNN algorithms serves to detect metastasis while making survival outcome predictions. Two types of data sources include clinical trial datasets together with imaging records. The application of ML models produces better outcomes for diagnosis along with prognosis prediction. The discussion emerges with three main obstacles such as unbalanced datasets in addition to DL’s obfuscation and collection disparities. [3]

**Xiaoshuai Xu, Linlin Xi, Lili Wei, Luping Wu, Yuming Xu research work titled in, “***Deep learning-assisted contrast-enhanced CT–based diagnosis of cervical lymph node metastasis of oral cancer: a retrospective study of 1466 cases***”** from European Radiology 2023 investigation attempted to develop deep learning algorithms to boost the identification of oral cancer lymph node metastases. Image segmentation tasks required the employment of Mask R-CNN as the main algorithm. A total of 1466 CECT images from patient cases formed the data sources. The model achieved detection results that were comparable to radiological teams and better than outcomes from surgeons and students. The study discussed how the model's diagnostic accuracy improved with treatment planning as well as the advantage transfer learning brings to model operation performance enhancement. [4]

**Shriniket Dixit, Anant Kumar, and Kathiravan Srinivasan research work titled in, “***A Current Review of Machine Learning and Deep Learning Models in Oral Cancer Diagnosis: Recent Technologies, OpenChallenges, and Future Research Directions***” f**rom Diagnostics 2023 mainly focuses on improving the medical identification of oral cancer with artificial intelligence technology. The algorithm with a combination of machine learning with deep learning models. Medical images together with patient dataset information serve as the primary information sources. AI tools give better results in diagnostic performance according to study findings. AI demonstrates its diagnostic advantages in solving medical challenges such as remote cancer detection and benign versus malignant tissue differentiation yet there remains a necessity to follow ethical guidelines during AI system implementation in healthcare facilities. [5]

**Xaviera A. López-Cortés, Felipe Matamala, Bernardo Venegas, and César Rivera research work titled in,** “Machine-Learning Applications in Oral Cancer: A Systematic Review” from Applied Science 2022 focuses on using machine learning to improve processes of oral cancer diagnosis. The proposed algorithm depends on SVM along with ANN and LR models for operation. The data collection includes genomic information together with histopathological results and medical imaging. The research data indicates better diagnostics along with better prognosis prediction. The research shows that machine learning's ability to detect the diseases early together with treatment planning capabilities which should become part of clinical workflows to improve diagnosis effectiveness. [6]

**Carlos M. Chiesa-Estomba, Manuel Graña, Alfonso Medela, Jon A. Sistiaga-Suarez, Jerome R. Lechien** in their research work titled in, “Machine Learning Algorithms as a Computer- Assisted Decision Tool for Oral Cancer Prognosis and Management Decisions: A Systematic Review” from ORL 2022 evaluates how machine learning (ML) technology contributes to oral cancer projection prediction. The analysis of algorithms concentrated on research that implemented ML methods with ANN as well as SVM. The research utilized PubMed and Google Scholar as well as SciELO and Scopus databases to collect data and consulted studies that satisfied predefined selection requirements. A total of eight research papers showed that ML demonstrates strong capability to enhance prognosis prediction. The examined discussion explains how continuous learning capabilities of ML enable better cancer diagnosis and management worldwide. [7]

**María García-Pola, Eduardo Pons-Fuster, Carlota Suárez-Fernández,** in the research work titled in, “Role of Artificial Intelligence in the Early Diagnosis of Oral Cancer. A Scoping Review” from Cancers 2021 investigates how Artificial Intelligence contributes to the diagnosis of early oral cancers. The research algorithm used PubMed along with Web of Science and Embase and Google Scholar databases starting from 2000 up to 2020. The analysis gathered research material from studies which investigated the implementation of AI technology for non-invasive examinations of oral cancer. The research of 36 studies confirmed the prospective capability of AI for predicting oral cancer development. The researchers emphasize that methodological requirements must be resolved before implementing AI on a population scale for detection protocols. [8]

**Amr Bugshan , Imran Farooq** in their research work titled in, “Cancer Stem Cells in Oral Cavity Squamous Cell Carcinoma: A Review” from F1000Research 2020 explores oral squamous cell carcinoma (OSCC) metastasis and its related malignant disorders as well as their causes and diagnostic advances. The search process used the relevant keywords to get Google Scholar and PubMed. This research used a variety of articles that explored OSCC and potentially malignant disorders along with their etiological factors and diagnostic methods and salivary biomarkers. Research data confirmed that OSCC has the ability to metastasize in addition to demonstrating the malignant characteristics of different disorders while indicating how salivary biomarkers facilitate early detection. The necessity of traditional biopsy methods was confirmed during discussion but experts acknowledged that salivary biomarkers have potential to enhance early detection of OSCC. [9]

**Yi-Ju Tseng, Hsin-Yao Wang, Ting-Wei Lin, Jang-Jih Lu, Chia-Hsun Hsieh, Chun-Ta Liao** research work titled in “Development of a Machine Learning Model for Survival Risk Stratification of Patients With Advanced Oral Cancer” from JAMA Network Open 2020 Research sought to build a machine learning approach which identified survival risk categories for advanced oral cancer patients. The algorithm applied elastic net penalized Cox regression to analyze both clinical-pathological details and genetic information. Stage III/IV oral squamous cell carcinoma patients from 1996 to 2011 contributed data for this study. When genetic information was included in the model it surpassed models containing clinicopathologic information in accurately predicting cancer-specific and locoregional recurrence-free survival. An online calculator shows promise to customize postoperative patient care according to the model recommendations. [10]

# **V. Research 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.

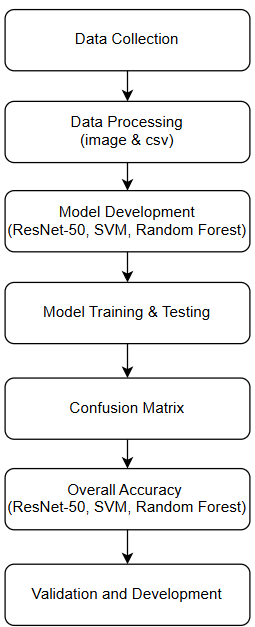


Figure1. Flowchart of Research Methodologies required for Early detection of oral cancer

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.

*Data Processing*

Data processing begins by collecting unstructured data before reorganizing its structure to create analyzable insights from the information. Through data processing procedures noise together with inconsistencies are eliminated from CSV patient history records and medical images to enhance early oral cancer diagnosis and metastasis management processes. Professional data cleaning produces a standard format that enables prediction patterns to be detected by Random Forest, SVM, and ResNet algorithms. High-quality repetitive input through data processing leads to increased model efficiency as well as more precise outcomes.

*Data Preprocessing*

Data preprocessing of CSV patient files requires sequential execution of cleaning followed by transformation before structural organization of the information. The data specialists must execute three essential operations which include handling missing values alongside converting categories into numbers and making a normalization process for numerical data. Through feature selection the learning process of the algorithm becomes more effective because it discards insufficient data. After preprocessing the CSV data it acquires a standardized format that allows Random Forest and SVM algorithms to discover significant oral cancer and metastasis connections. The speed of computational processing along with better model interpretability come from organized data structure.

*Image Enhancemnet*

Interface improvement remains the fundamental component for studying medical oral cancer images. The three fundamental techniques of contrast adjustment and noise reduction and image sharpening create improved visibility of cancer-related areas. The deep learning model ResNet improves its capabilities to detect elaborate cancer patterns and structures by analyzing enhanced images. The predictive success of cancer detection and metastasis assessment improves when improved image analysis is paired with the model because such images better display important features. By completing this process the algorithm gains stronger capacities to find cancer signs during their early stages.

*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.

*Model Development*

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 treatment designs.

Testing demonstrated ResNet50 performs at a high level for tumor detection purposes whereas SVM stands 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.

*Model Training and Testing*

Training the model requires the use of ResNet50 for image fine-tuning followed by implementing SVM to perform lesion classification. Metastasis risk assessments from Random Forest include both medical image data and patient historical information. The MATLAB platform divides the data collection into training sections occupying 70% of the total set and testing sections using 30% of the data. A model optimization step through cross-validation approaches enables maximum training accuracy but testing verifies both early identification and risk categorization performances to deliver reliable solutions in clinical settings.

*Confusion Matrix*

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.

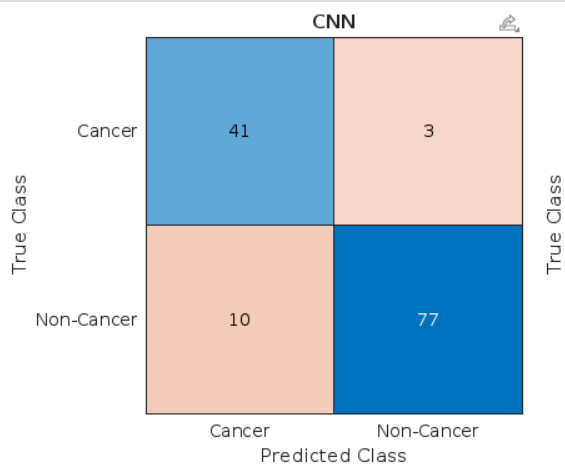


Figure 2. Confusion matrix of CNN’s ResNet-50 analysis for oral cancer detection

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.

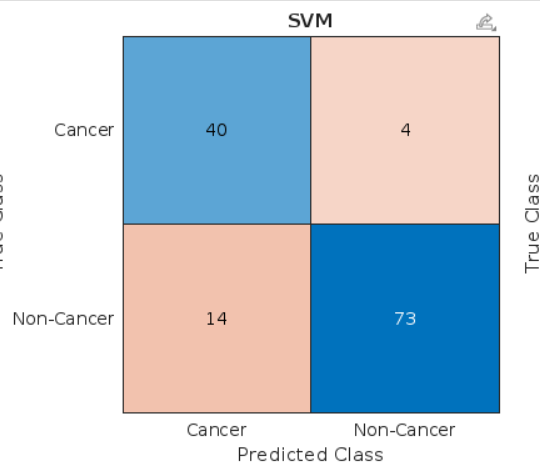
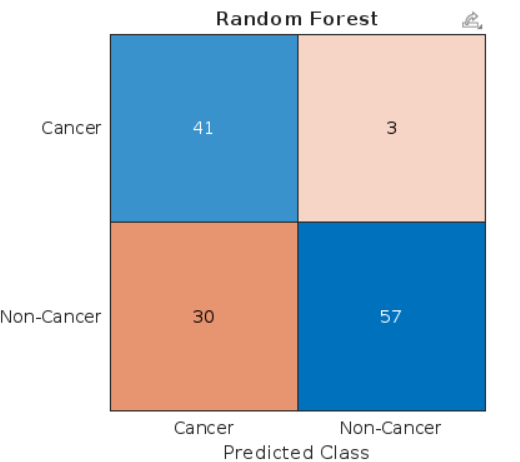


Figure 3. Confusion matrix of SVM analysis for oral cancer detection

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 4. Confusion matrix of Random Forest analysis for oral cancer detection

*G. Overall Accuracy*

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.

*H. Validation and Optimization*

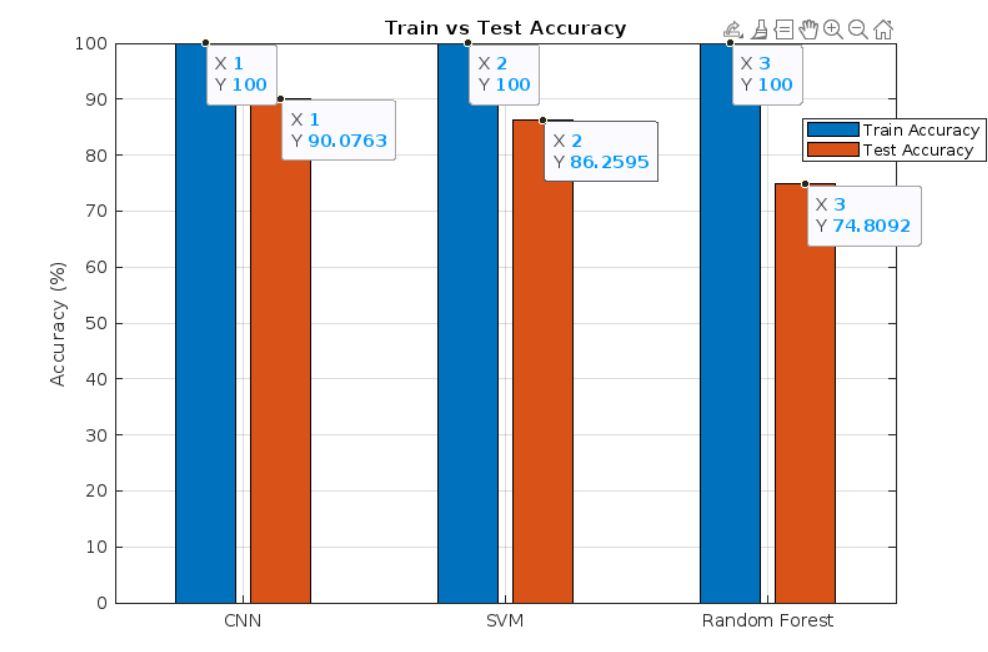
A splitting operation of MATLAB code data into training and testing sections is performed for oral cancer detection to avoid overfitting. The cross-validation method helps evaluate model performance as one of several evaluation techniques. The optimization process using Bayesian optimization and gradient descent method helps enhance the accuracy of ResNet50, SVM, and Random Forest through fine-tuning their hyperparameters. The implementation method provides reliable performance for clinical use.

# **VI. Data Accquired**

In this classification, we refer to the different types of research papers for applying Machine Learning Algorithm ResNet-50 , SVM, and Random Forest. And the some cancerous and Non Cancerous images were gathered from Kaggle, some renowned Oncologist and one which was made freely available online. The first dataset contains the 2157 images as training samples and 2157 Excel .

# **VII. Result**

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

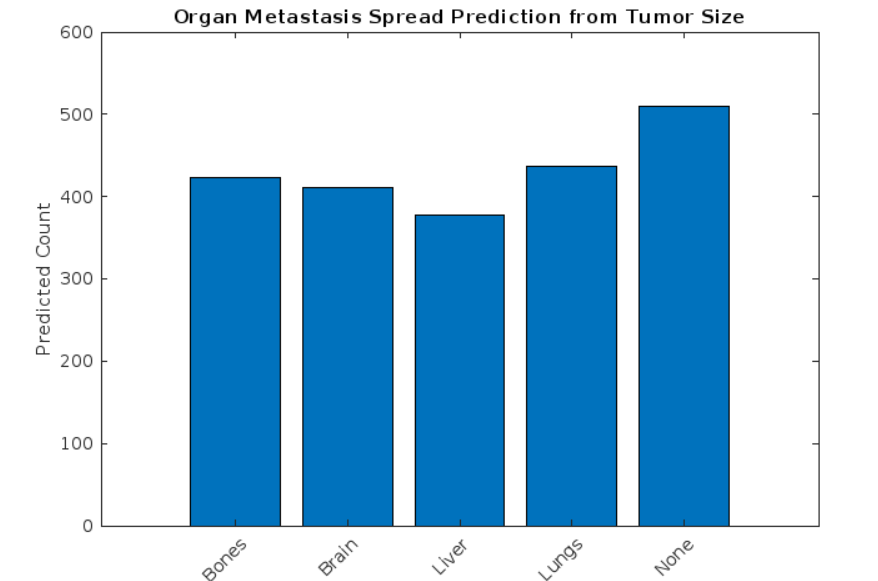


Figure 6. Bargraph analyzing the Organ Affected due to metastasizing of the cancer

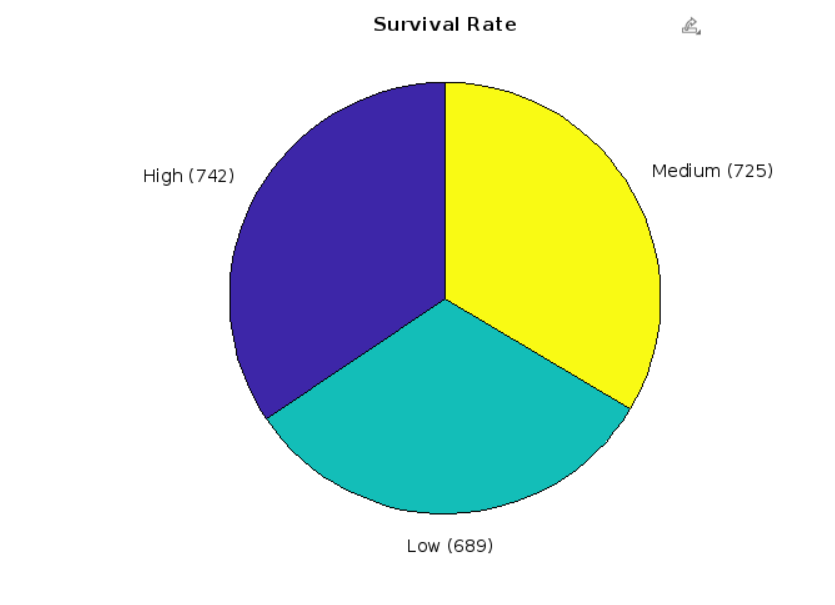


Figure 7. Piechart depicting the Survival rates on the basis of stages of oral cancer.

# **VIII . 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.

# **IX. Future Work**

Researchers should focus on evaluating new detection algorithms for oral cancer in MATLAB systems through investigation of transformers because of their enhanced photo extraction and graph neural networks for tracking tumor progression and attention-based methodology to direct analysis focus on high-risk image areas. Two relatively uncommon approaches involving evolutionary algorithms for automatic parameter selection and fuzzy logic for managing diagnostic uncertainty would allow the system to tackle decision-making and data variability problems. The combination of generative adversarial networks (GANs) for rare tumor image synthesis with swarm intelligence for feature selection would lead to superior dataset quality while improving model robustness. Unresearched techniques hold potential to enhance both the accuracy along with scaling potential and doctors' ability to use AI-based oral cancer detection systems.

# **X. Acknowledgment**

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 my 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.

# **XII. Reference**

1. Artificial Intelligence in Oral Cancer: A Comprehensive Scoping Review of Diagnostic and Prognostic Applications by Vineet Vinay, Praveen Jodalli, Mahesh S. Chavan, Chaitanya. S. Buddhikot 2025
2. Compares ResNet-50 models (VGG16 and ResNet50) and examines their impact on improving diagnostic accuracy when used by oral pathologists by Shintaro Sukegawa, Sawako Ono, Futa Tanaka, Yuta Inoue , 2023
3. Machine learning in metastatic cancer research: Potentials, possibilities,and prospects by Olutomilayo Olayemi Petinrin, Faisal Saeed, Muhammad Toseef, Zhe Liu, Shadi 2023
4. Deep learning-assisted contrast-enhanced CT–based diagnosis of cervical lymph node metastasis of oral cancer: a retrospective study of 1466 case by Xiaoshuai Xu, Linlin Xi, Lili Wei, Luping Wu, Yuming Xu, 2023

.

1. A Current Review of Machine Learning and Deep Learning Models in Oral Cancer Diagnosis: Recent Technologies, OpenChallenges, and Future Research Directions by Shriniket Dixit, Anant Kumar, and Kathiravan Srinivasan, 2023.
2. Machine-Learning Applications in Oral Cancer: A Systematic Review by Xaviera A. López-Cortés, Felipe Matamala, Bernardo Venegas, and César Rivera, 2022
3. Machine Learning Algorithms as a Computer-Assisted Decision Tool for Oral Cancer Prognosis and Management Decisions: A Systematic Review by Carlos M. Chiesa-Estomba, Manuel Graña, Alfonso Medela, Jon A. Sistiaga-Suarez, Jerome R. Lechien, 2022
4. Role of Artificial Intelligence in the Early Diagnosis of Oral Cancer. A Scoping Review by María García-Pola, Eduardo Pons-Fuster, Carlota Suárez-Fernández, 2021
5. Cancer Stem Cells in Oral Cavity Squamous Cell Carcinoma: A Review by Amr Bugshan , Imran Farooq , 2020
6. Development of a Machine Learning Model for Survival Risk Stratification of Patients With Advanced Oral Cancer by Yi-Ju Tseng, Hsin-Yao Wang, Ting-Wei Lin, Jang-Jih Lu, Chia-Hsun Hsieh, Chun-Ta Liao, 2020.