A Project Work – Phase I on

Explainable Artificial Intelligence (XAI) in Predictive Diagnostics of Oral Cancer using Neural Networks

Submitted in partial fulfillment of the requirements for the award of the

Bachelor of Technology

in

Department of Computer Science and Engineering

(Artificial Intelligence and Machine Learning)

by

G. Akhila	**21241A66F0**
E. Sathwika	**21241A66E8**
G. Anushka	**21241A66E9**

Under the Esteemed guidance of

Ms. P. Shirisha

Assistant Professor

Department of Computer Science and Engineering
(Artificial Intelligence and Machine Learning)

GOKARAJU RANGARAJU INSTITUTE OF ENGINEERING AND TECHNOLOGY
(Approved by AICTE, Autonomous under JNTUH, Hyderabad)

Bachupally, Kukatpally, Hyderabad-500090

GOKARAJU RANGARAJU INSTITUTE OF ENGINEERING AND TECHNOLOGY

(Autonomous)

Hyderabad-500090

CERTIFICATE

This is to certify that the major project entitled "**Explainable Artificial Intelligence (XAI) in Predictive Diagnostics of Oral Cancer using Neural Networks**" is submitted by **G Akhila (21241A66F0), E Sathwika(21241A66E8) and G Anushka (21241A66E9)** in partial fulfilment of the award of degree in BACHELOR OF TECHNOLOGY in Computer Science and Engineering (Artificial Intelligence and Machine Learning) during Academic year 2024-2025.

Internal Guide	**Head of the Department**
Ms. P. Shirisha	**Dr. G. Karuna**

External Examiner

ACKNOWLEDGEMENT

There are many people who helped us directly and indirectly to complete our project successfully. We would like to take this opportunity to thank one and all. First, we would like to express our deep gratitude towards our internal guide **Ms. P. Shirisha, Assistant Professor,** Department of Computer Science and Engineering (Artificial Intelligence and Machine Learning), for his/her support in the completion of our dissertation. We wish to express our sincere thanks to **Dr. G. Karuna,** Head of the Department**,** and to our principal, **Dr. J. PRAVEEN,** for providing the facilities to complete the dissertation. We would like to thank all our faculty and friends for their help and constructive criticism during the project period. Finally, we are very much indebted to our parents for their moral support and encouragement to achieve goals.

G Akhila (21241A66F0)

E Sathwika (21241A66E8)

G Anushka (21241A66E9)

DECLARATION

We hereby declare that the major project titled **“Explainable Artificial Intelligence(XAI) in Predictive Diagnostics of Oral Cancer using Neural Networks”** is the work done during the period from **02nd August 2024 to 29th November 2024** and is submitted in the partial fulfillment of the requirements for the award of degree of Bachelor of Technology in Computer Science and Engineering (Artificial Intelligence and Machine Learning) from Gokaraju Rangaraju Institute of Engineering and Technology (Autonomous under Jawaharlal Nehru Technology University, Hyderabad). The results embodied in this project have not been submitted to any other University or Institution for the award of any degree or diploma.

G Akhila (21241A66F0)

E Sathwika (21241A66E8)

G Anushka (21241A66E9)

ABSTRACT

Oral cancer is a viable global public health problem, and early diagnosis is important to increase the survival rate of cancer patients. The importance of inventiveness is highlighted by the fact that the conventional diagnostic techniques involve huge drawbacks concerning accessibility, accuracy, and dependability. This research applies deep learning for improving the diagnostic performance of oral cancer, by using two datasets containing images of lips and tongue and HPS images. The Proposed work uses a standard benchmark Convolutional Neural Network (CNN) that is compared with an optimized CNN+VGG16. Propagation through the CNN model reveals an accuracy of about 84.0% for clinical images and a slightly lower of 72.0% for the histopathological images. Notably, the CNN+VGG16 model outperforms, with accuracy rates of 92.0% and 96.0% respectively, proving a greater ability to develop patterns from within the datasets. SHAP and LIME XAI tools are also inculcated within the system to enable the user to understand the patterns and projected outcome. These visualizations increase the level of trust and openness as clinician and consumers gain insights into the trends likely to affect the diagnosis.

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

XAI	Explainable AI
CNN	Convolutional Neural Network
HPV	Human Papillomavirus
SHAP	Shapely Additive Explanations
LIME	Local Interpretable Model Agnostic Explanations
PDP	Partial Dependence Plots
GRAD-CAM	Gradient-Weighted Class Activation Mapping
AUC-ROC	Area Under the Receiver Operating Characteristic curve
DL	Deep Learning
OSCC	Oral squamous cell carcinoma

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

The following section explains the rationale behind the project work,impact of oral cancer,its detection and the usage of XAI in oral cancer visualizations.

1.1. Introduction to the Project Work

Oral cancer is defined as malignancies that occur in any part of the mouth or oral cavity, including the lips, tongue, cheeks, the floor of the mouth, the hard and soft palates, and the oropharynx-example as shown in Figure 1.1. Squamous cell carcinomas, which start in the thin, flat cells lining the lips and the inside of the mouth, are the most common type of oral cancer. This type of cancer is particularly dangerous because it often goes undetected until it has progressed, which can negatively impact a person's health and quality of life.Oral cancers are a subset of head and neck cancers and constitute a large portion of all new cancer cases globally.

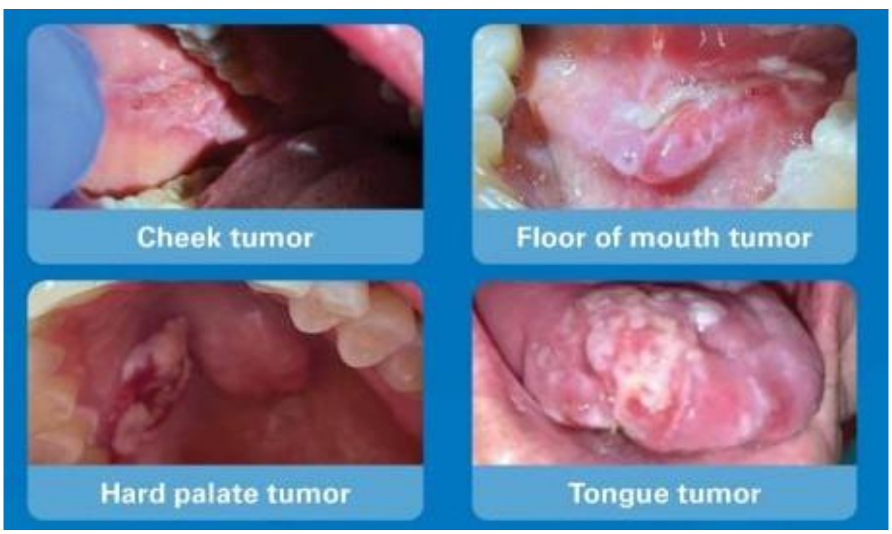

Figure 1.1. Oral Tumors(Courtesy: Source [26])

It affects 377,700 people annually, making it a major global health issue as shown in Figure 1.2. It is the sixth most common type of cancer globally, and it is particularly prevalent in areas with high rates of alcohol and tobacco use, such as South and Southeast Asia and Eastern Europe.The two most important characteristics of oral cancer are its high morbidity and aggressiveness. It involves malignant lesions in the mouth, lips, tongue, and throat. Oral cancer leads to 177,700 deaths annually and poses special difficulties since it is frequently detected in its advanced stages due to a lack of early warning signs or awareness.

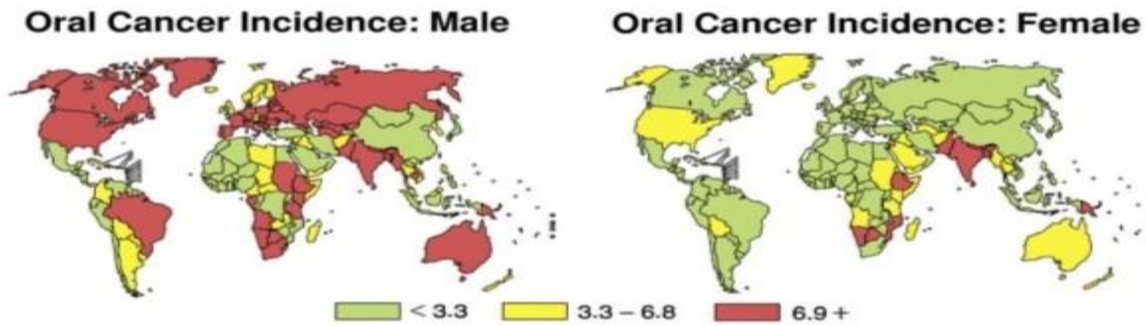

Figure 1.2. The Global oral Cancer rates by variation of gender(Courtesy: Source [27])

Human papillomavirus (HPV) exposure, alcohol use, and smoking are important risk factors. HPV-related cases have increased recently, especially among younger people who are less likely to have risk factors including alcohol or tobacco use.With 119,992 new cases reported annually, India is one of the countries with the highest frequency of oral cancer, accounting for almost one-third of all cases globally.About 119,992 new cases of oral cancer are reported annually in India, one of the countries with the highest prevalence of the disease, making up over one-third of all cases worldwide. The extensive usage of tobacco and betel quid, a popular chewable stimulant made of tobacco, betel leaf, and areca nut, is mostly to blame for this high prevalence.After five years, the survival percentage for early-stage diagnosis is over 85%, but the

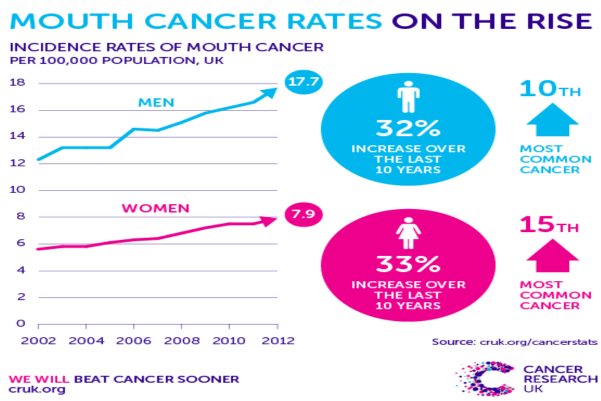

Figure 1.3. Mouth Cancer rates(Courtesy: Source [28])

survival rate for advanced-stage finding is roughly 40% as shown in Figure 1.3. This illustrates how increasing survival rates requires routine testing, early interventions, and public awareness. The increasing number of oral cancer cases worldwide, especially in high-risk areas, requires both awareness efforts and improvements in diagnostic technologies.

Staging

The size of the lesion and whether it has migrated to adjacent lymph nodes or other body areas indicate the stage of oral cancer.

Stage 0: Known as carcinoma, this stage has been defined by the presence of aberrant cells that have not yet disseminated.

Stage I: No lymph nodes have been affected, and the malignancy is 2 cm or less.

Stage II: The tumor has not spread and is 2-4 cm in size.

Stage III: The tumor has grown to one lymph node on the same side of the neck or is more than 4 cm in size.

Stage IV: A more advanced stage where there has been widespread local invasion or organ spread.

Risk Factors

- **Tobacco Use**: Smoking cigarettes, cigars, or pipes, and using smokeless tobacco products (like chewing tobacco) significantly increases the risk.
- **Alcohol Consumption**: Heavy alcohol use is another major risk factor, especially in conjunction with tobacco.
- **HPV Infection**: Certain strains of HPV, especially HPV-16, are associated with oropharyngeal cancers.
- **Sun Exposure**: For lip cancer, prolonged exposure to sunlight is a significant risk factor.
- **Diet and Nutrition**: A diet low in fruits and vegetables may increase susceptibility to oral cancer.

Role of AI & ML in Predictive Diagnostics of Oral Cancer

AI and ML also transform predictive diagnostics, allowing for early detection of diseases that could prevent them from deteriorating to serious conditions. This type of predictive diagnostics uses AI/ML algorithms which study an extensive pile of healthcare databases, like patient international histories, genetic information, lifestyle variables and real-time health measures. By understanding these signature patterns and such correlations, it allows us to intervene at biochemical levels where the disease may occur before any symptoms arise. Models can detect abnormalities, like tumors or fractures, with high precision, often surpassing human diagnostic accuracy.

AIML also enables personalized medicine, tailoring diagnostic and treatment approaches to individual patients. The predictive models can estimate the probability of a patient developing a chronic disease such as diabetes, heart disease, or oral cancer and suggest ways to avoid the disease. Furthermore, AI wearable devices monitor constant health information, like heart rate or glucose level, and offer diagnosis on the spot. These systems help patients and doctors to identify problems that can lead to hospitalizations and high healthcare expenses. However, the implementation of predictive diagnostics is not without its problems, including data privacy, algorithm bias, and workflow integration. The integration of AI/ML with healthcare has the potential of delivering a future of better , faster , and predictive healthcare.

These devices offer alarms to indicate when something is wrong and the patient and the physician can take action immediately. Further, using genetic, lifestyle, and environmental data, the models also enable the prognosis of disease severity, as well as the outcome of various therapies, thus creating the foundation for precision medicine. For example, convolutional neural networks (CNNs) can analyze X-rays, CT scans, and MRIs to a very high degree of accuracy. These models identify even nuisance-level structures such as microcalcifications in a mammogram or the earliest stage of a tumor, at times surpassing the ability of a human radiologist.Through the use of AI and ML, predictive diagnostics is changing the healthcare industry from a curative model to a preventive one, creating better patient outcomes, and at the same time, cutting costs and increasing the effectiveness of delivery systems.

What is Explainable AI(XAI)

XAI stands for eXplainable Artificial Intelligence, and Explainability is an emergent subfield of AI dedicated to explain ML models to end-users. For large and especially complex models, like deep neural networks, the models are considered “black box” because it is hard to interpret the results of the model in terms of how it came to make such predictions or decisions. To achieve these goals, XAI aims at opening the said black boxes through offering a good account of model outputs in a manner that end-users can comprehend to facilitate trust, validation and deeper comprehension of AI’s actions.

Importance of XAI

Building Trust: The practice of XAI allows users to see how an AI algorithm arrived at a particular conclusion and more importantly why they arrived at that conclusion thus improving the users trust in the system. If users have trust for an AI model, more will depend on it in various sophisticated applications.

Ensuring Fairness and Reducing Bias: XAI enables developers to assess and correct such biases by revealing to them which features contribute to what results. People must understand that they are choosing between concrete options when making a decision because this is ethical for specific professions such as employment, financial, legal, and others.

Compliance with Regulations: Several industries are required to provide compliance relating to their operations mainly in the healthcare, finance, and legal industries. XAI supports organizations to align them to these standards and hence lessen legal ramifications as well as promote responsible use of artificial intelligence.

Supporting Better Human-AI Collaboration: XAI makes it easier for humans to be able to combine forces with AI because the conclusions derived are easily understandable hence the users are in a position to make wise choices regarding the recommendations given by the AI. I’m aware of this especially in the health sector and the financial institutions.

Improving Model Reliability and Performance: XAI illuminates exactly how models make sense of the data.This can be used to pinpoint deficiencies that then allows the developers to fix the problem which otherwise would lead to increased inaccuracies and reduced robustness. That leads to more reliable AI.

XAI Techniques

SHAP (SHapley Additive exPlanations):

SHAP defines an important value of a feature in a prediction using Shapley values from the field of cooperative game theory. This method also provides localization and global attribution by depicting the contribution of each feature towards the output of a model.

Use Case: Often applied in finance and healthcare where interpretation of the feature importance for concrete predictions is required, for example, credit scoring.

LIME (Local Interpretable Model-agnostic Explanations):

LIME approximates a model locally by building an easier explanation, based on a locally interpretable model around the prediction. Thus, how the model behaves when some values are modified determines those features that are most important for a given instance in LIME.

Use Case: OPT used quite often in recommendation systems and fraud detection to know why this recommendation was made or why this flag was raised

Partial Dependence Plots (PDPs):

PDPs demonstrate the variation of the predicted value with regards to one feature at a time while holding all other features constant. This gives some information regarding contribution of certain features to the output given by the model.

Use Case: Applied in insurance or financial industries to see the impact some independent variables (such as age or income) have on the predictions and risk rate models.

Grad-CAM (Gradient-weighted Class Activation Mapping):

While Grad-CAM is developed for creating class-discriminator features, which is based on gradient information extracted from the final layer of a popular CNN architecture, Grad-CAM produces heatmap over an input image pointing the areas most important to a prediction.

Use Case: Often seen in image categorization problems especially in that of medical images where it may point to areas of interest in an x-ray or MRI scan as data for analysis**.**

Saliency Maps:

The Heat maps have arrows pointing to pixels/regions that are highly salient; this heavily influenced the models' decision. Saliency maps show what parts of the image the model 'zoomed in' to make its decision, by computing the derivative of the model's output regarding the input image.

Use Case: Typically a part of computer vision, especially in such fields like medicine or automotive industry, to give visual representation of model choices. For instance in cancer detection, saliency maps can indicate the exact parts of the X-ray that were used in arriving at a "positive" decision to help the radiologists on parts of the images to concentrate on.

XAI in oral cancer detection

In the context of oral cancer screening XAI systems are most valuable to increase the credibility and reliability of the created diagnostic classifiers. Practical studies of deep learning, especially of the CNN type, demonstrated an ability to detect cancerous lesions in images of the oral cavity. Nevertheless, these models impose challenges that result in the creation of models that function like "black boxes," which makes it hard for healthcare providers to know how specific predictions are made. This is especially so since explainability makes it possible for clinicians to trust and verify the predictions given by AI systems, which is crucial when dealing with patients. We found that using multiple XAI techniques together can greatly enhance the ability to explain the model. For the images of oral cavities, visual representation of focus points through the CAM which offers heat maps that identify aspects of the image for a specific diagnosis serves such a purpose well. However, CAM alone may not capture the degree of importance of differential features within the architecture of the proposed model at different levels.

In order to provide a better and broader perception of how CAM is integrated with other model-agnostic approaches such as the **SHAP** and the **LIME**, a better understanding has to be made.

Closely related to CAM, SHAP and LIME provide additional insights to the model predictions by showcasing how color, texture, and shape patterns in the regions of interest are implicated in the final assessment. SHAP values originating from game theory give us the framework of interpreting the extent of feature importance, while LIME gives us local interpretations by allowing examination of the effect of variations around images to learn about feature sensitivity. Combined with CAM's heatmaps of the model's reasoning, these approaches provide a complementary and comprehensive view of the model's decision-making process and, consequently, improve the reliability and interpretability of oral cancer detection. The utilization of multiple techniques ensures a global and local understanding of the results, which will help to build trust in the diagnostic capability of AI systems and support clinicians in decision making.

1.2. Objectives of project

- To develop a CNN model that is accurate for predicting oral cancer.
- To enhance Explainability of Models with XAI Visualisations
- To incorporate the visualizations with SHAP and LIME in order to maximize explainability.
- to further enhance the model's credibility by confirming that the areas found by the explainability techniques match clinically significant features.

The objective is to create a highly effective CNN model for predicting the likelihood of oral cancer, with an emphasis on improving model interpretability and encouraging clinician trust. In order to deploy XAI tools and increase the transparency of the model's decision-making process, this project will make use of advanced visualization approaches such as CAM, SHAP, and LIME.The ability to visually assess and understand the key factors driving the model's predictions would increase healthcare professionals confidence in its reliability. Furthermore, by ensuring that the sites suggested by the model prediction are accurate and applicable, the study aims to boost clinician trust.

1.3. Methodology Adapted

The approach taken by the project involves developing a **CNN model** and **a pretrained VGG16 model** to predict oral cancer across 2 distinct datasets: histopathological images and images of human mouths. While the VGG16 model, pretrained on a sizable image database, improves classification accuracy by transferring learnt features, the CNN model is specially designed to extract key features from the images. We utilize the deep visualizations of features that the VGG16 model already has by using transfer learning with VGG16, which cuts down on the time and resources required to train the model from scratch.A detailed comparison of the classification performance between microscopic histopathological images and macroscopic human mouth images is made possible by the training and evaluation of both models on each dataset.The models' ability to distinguish between cancerous and non-cancerous samples in each dataset is assessed using key metrics such as accuracy, precision, recall, and F1 score.

After the model is developed, the decision-making process is interpreted using XAI techniques.The technique seeks to pinpoint the regions or features within each kind of image that influence the model's predictions by using visualization tools such as CAM, SHAP, and LIME. CAM highlights the regions that the models deem most suggestive of cancer, whereas SHAP and LIME provide additional details regarding the contribution of each pixel or feature.

Convolutional Neural Networks(CNN)

CNN is a subset or a type of deep learning which is suitable for dealing with information presented in image format. The architecture of CNNs is inspired by how the brain cortex that is related to vision perceives objects from the natural environment. They are intended to learn spatially hierarchical features from the images in an automated manner and in the manner that can be adjusted if they are programmed, to perform various functions such as image categorization, object identification, and face recognition. CNNs are composed of several important layers as shown in Figure 1.4 that are capable of analyzing and extracting features from the admitted images.

Convolutional Layer: The Convolutional Layer is a set of vertically organized filters that scroll across the picture picking features such as edges and texture. Some are:

Size of the filters: The size controls the area of the image that will be captured Let’s denote Filter size by F.

Activation Function: The Activation Function for example ReLU adds non-linearity into the system in order to enable the model to learn the complex patterns. This makes it possible for the CNN to capture other features apart from the linear relationship of the features.

Pooling Layer: As a result of operations, the Pooling Layer compresses the feature maps making the model fast and less likely to be overfitting. The process of down sampling is accomplished by either Max pooling or Average pooling which loans important features for representation.

Fully Connected Layer: To be more precise, Fully Connected Layers connect every neuron from the previous layer, deals with high-level thinking, and makes conclusions – final forecasts.

In this case, activation functions that can either be softmax for problems or sigmoid based on the problem type are utilized here.

The study has demonstrated that CNNs are very usable and efficient in oral cancer detection tasks. They are particularly good at identifying hierarchical features from medical images, such as histopathological slide images, oral tissue images, or radiologic images. Being trained by a large and diverse set of labeled, medical images, CNNs can learn to differentiate between various gray scale intensity patterns, which reflect the tissue structure, cell formations and suspicious lesions that would indicate the presence of cancer. They are extracted in such a manner that

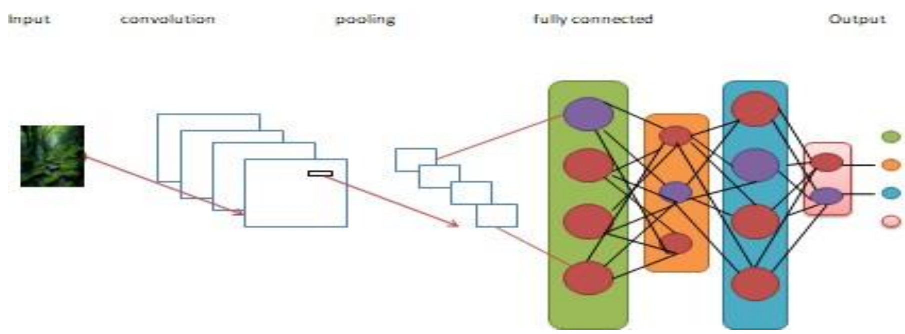

Figure 1.4. Working of Convolutional neural Networks

emphasizes the more relevant spatial characteristics for separating healthy tissue from cancerous tissue.

Pretrained VGG16

VGG16 is an implementation of deep Convolutional Neural Network, or CNN for short, which is widely used in image recognition tasks. It was developed at the University of Oxford's Visual Geometry Group and gained much recognition for its efficiency in the ILSVRC.

Layers and Structure: VGG16 model has 16 layers with weights: 13 layers of converging and 3 layers of pooling. It operates with filters of 3x3 and the stride of 1 with a padding so that the spatial dimensions of the input stay the same. This approach makes it possible to have deeper architectures than typical networks of the same number of parameters.

Pooling: Intermittently, the max pooling layer with a 2 x 2 window size and a stride of 2 is applied to decrease the spatial size of the feature maps, so as to optimize the calculation and alleviate the over-learning effect.

Activation Function: Batch normalization is not used in this model because it has been shown to perform poorly when it comes after a convolution operation Same with the drop-out layer which does not follow a convolution layer In order to introduce non-linearity into the model, the ReLU activation function is applied after each convolution operation.

Fully Connected Layers: The network also has three completely connected layers; the first two are 4,096 nodes and the last one is 1,000 nodes (for ImageNet classification) with softmax activation function.

CNNs integrated with VGG16 are very efficient in the diagnosis of oral cancer due to their capability of coming up with hierarchical features from the medical images which include histopathological slides, or oral tissue scans. Since VGG16 has a very deep structure, perhaps the subtle changes to the cell and tissue arrangements that announce cancer can be discerned. Using generated heatmaps, saliency maps, and class activation maps (CAMs), clinicians can see which portions of images the network pays attention to when making a cancer diagnosis and can enhance the interpretability.

XAI techniques used:

1. CAMClass Activation Mapping:CAM is an activation mapping visualization method employed principally in the Convolutional Neural Networks (CNNs) to point to the parts of the image that are most relevant to class activation. The enhanced visual representation of which parts of the input data affect a particular class decision, CAM offers meaningful insights into the model's areas of emphasis. The most sensible case is when you need to obtain specific information about the object, like in object detection or in medical imagery, where it is critical to know the reason behind the decision made by the model. CAM visualizations can assist in confirming that the model pays attention to relevant areas or it simply spots noise.

2. SHAP (SHapley Additive exPlanations) :It is a model-agnostic technique based on the axiomatic characterization of a solution concept in cooperative game theory, which aims at quantifying feature-wise importance to the model. It measures importance in a unified way to each feature so that it complements the Shapley value for the explanation of individual predictions. In this sense, one of the strongest claims of SHAP is that its theoretical framework ensures general reliability and equity of feature contribution attribution. They are not only useful to describe the behavior of global models but also include detailed diagnostics for individual prediction, thus suitable for either debugging or analyzing complicated models.

3. LIME (Local Interpretable Model-agnostic Explanations) LIME explaining the predictions of any classifier by fitting a generic, simple, interpretable model around the classifier in the vicinity of the instance. This involves modifying the input data a little and observing the behaviors of the model to learn a linear region around an instance. LIME is most useful in a scenario when the author cannot explain the global behavior of the model but local explanations are valuable. It excels in such areas of employment that decision making must be justified and accountable for, for example in fiscal or medical fields.

The integration of CAM with another two methodologies SHAP and LIME in a project would provide maximum explainability by utilizing the advantages of these techniques. CAM gives understanding of the spatial perspective of how CNNs make decisions on image data, whereas SHAP provides feature attribution and LIME gives local instance explanation. Such integration is useful because both general and specific knowledge can be obtained by employing those methods, and this is especially important in domains where a detailed understanding of model behavior is needed.

Concept of Early Stopping:

Early stopping is one of the best regularization techniques used by practitioners in machine learning and deep learning for preventing the model from overfitting. This is done through the use of a validation dataset, through which the model’s comprehension is checked and the training process is ceased once the performance in the comprehension checked measure elicits little or no enhancements. This is to avoid the model repeating what was learned during the training and even picking on incremental noises or patterns that are not relevant to the future new data.

The concept behind early stopping is therefore to achieve a better balance between under-and overfitting. Often during the training the error on the training data is gradually decreasing, while the performance on the validation data increases at some point and may even slightly decrease if the model becomes too complex and starts to overfit the data. This turning point is identified by early stopping and the model is restored at this stage since this provides the best validation metric with the added advantage of saving computer resources.

1.4. Architecture diagram with very brief description

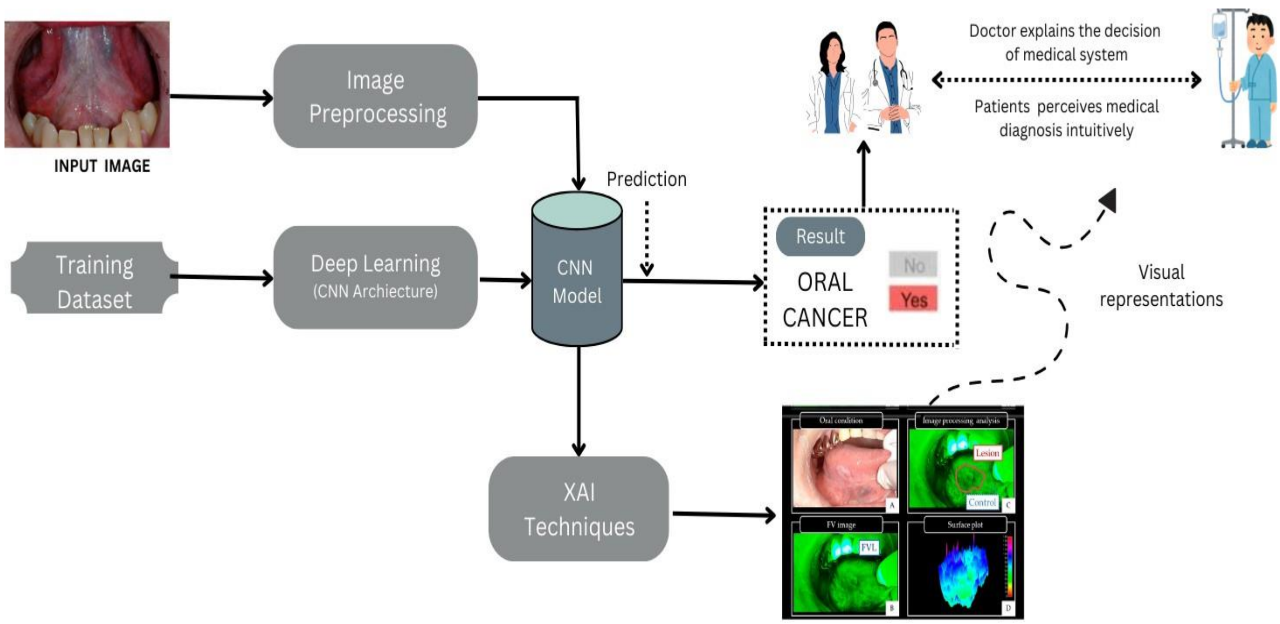

Figure 1.5. Architecture diagram

The architecture as shown in Figure 1.5 of the oral cancer prediction framework is outlined to require an input picture, regularly a restorative picture of the human mouth, and handle it through different stages for prediction and interpretability. The method starts with picture preprocessing, where the input picture experiences upgrade methods to plan it for investigation. This step includes resizing the picture, normalizing pixel values. These preprocessing strategies guarantee that the input information is steady, which in turn moves forward the execution and precision of the show when it analyzes unused pictures.

After preprocessing, the picture is passed into a profound learning demonstration that combines a custom CNN engineering and a pretrained VGG16 demonstration. This show, which is prepared employing a dataset of labeled pictures, is outlined to extricate noteworthy highlights that are characteristic of cancerous or non-cancerous locales. The CNN demonstrates these highlights through iterative preparing, permitting it to recognize designs that recognize solid tissue from possibly cancerous ranges. Based on these learned highlights, the show produces an expectation on whether or not the input image demonstrates the nearness of oral cancer. This expectation is a binary classification, with a “Yes” or “No” yield that communicates whether oral cancer is identified within the image.

To increase the interpretability of the model's expectations, Explainable AI (XAI) techniques are connected to imagine the decision-making handles. These strategies incorporate class Activation Mappings (CAM), SHAP and LIME which permit the show to highlight the particular ranges of the picture that impacted its determination. The XAI visualizations give a layered breakdown of the forecast, making it clear which locales were most demonstrative of cancer. By advertising these bits of knowledge, the framework increments straightforwardness, empowering clinicians to approve that the model's center ranges adjust with clinically important highlights, eventually boosting the model's validity.

At long last, the interpretability aspect of the show underpins the interaction between specialists and patients. The visualizations produced by XAI strategies make it less demanding for clinicians to clarify the conclusion in a reasonable way. The highlighted regions serve as visual helps, permitting specialists to naturally illustrate to patients why the AI show comes to its conclusion. This approach makes the model persistent believe in AI-assisted therapeutic conclusion but moreover helps clinicians in making educated choices based on the model's discoveries and visualizations provided. The combination of exact expectations and clear, interpretable conclusions comes about in this way makes an all encompassing framework that supports viable and straightforward care.

1.5. Organization of the Report

The organization of this study reports centers on the aforementioned components in detail. The study is structured into informational chapters that are given logically and with a thorough comprehension of the subject matter to preserve coherence. Every chapter concentrates on a different aspect of the work, and the information is presented in a straightforward and succinct way so that the reader can follow along with ease.

1.5.1 Introduction: A brief overview explaining the purpose of the study is given. In addition, the purpose of the research is stated clearly, the methodology used to achieve it is stated, and an outline of the schematic design is given.

1.5.2. Literature Survey: A literature survey is an in-depth review of the body of research, literature, and other sources of knowledge on a topic. where relevant information from journals, conference proceedings, and research articles was thoroughly examined in order to find oral tumor regions and predictions are made. An overview of current efforts for better explainability of model decisions have been provided, along with their benefits, outcomes, and shortcomings of decision understanding.

1.5.3. Proposed Method: It involves stating the problem statement and objectives of the project. The architecture diagram and modules are explained in-depth. To add software and hardware requirements were also specified.

1.5.4. Results and Discussions: After the implementation of the proposed method, results were analyzed and subsequently, the discussion about the proposed method and results were discussed.

1.5.5. Conclusion and Future Enhancements: Based on the results obtained and through analysis, conclusion is presented along with some future enhancements related to the proposed method.

1.5.6. Appendices: The source code of the project consisting of the main model architecture is attached in this section.

1.5.7. References: Proper referencing is given to the mentioned reference works.

CHAPTER 2

LITERATURE SURVEY

This chapter includes the description and summary of our project's current approaches, their advantages, results and their shortcomings.

2.1. Summary Of Existing Approaches

Antonio Ferrer-Sánchez and team [1] presented a deep learning technique involving U-Net for image segmentation and multi-task CNNs for classification of oral leukoplakia lesions. This work analyzed 261 lesions and obtained an average APR of 0.560 and multi-task models achieved significantly better sensitivity and AUC for cancer risk prediction and high-risk dysplasia. Insights gained using the LIME method helped to better understand how the model reached a certain decision.

The paper by Kahraman and co-authors [2] enhance the Interpretability of AI in Medical Diagnostics explains a methodology involved to make medical diagnostics using AI in a format understandable by the healthcare staff through fuzzy logic. In particular, using explainable AI models as an approach, the focus is placed on enhancing the reliability of diagnostic results and their compliance with clinical knowledge when making decisions.

P. Ashok Babu and team [3] suggested a deep learning technique, using transfer learning to improve diagnosis precision. The method employed deep neural networks to perform the classification of the oral cancer lesions and the Inception-V3 algorithm was recommended for use. This approach also makes an attractive and less invasive and less costly detection approach that the authors highlighted some limitations on biopsy specificity and that the necessity for larger scale testing was still required.

The research work [4] was entitled "A fully automated and explainable algorithm for predicting malignant transformation in oral epithelial dysplasia" and authored by Adam J Shephard and colleagues. The identified research methodology comprises a deep learning workflow, for predicting an OMT risk score from whole slide images stained with Haematoxylin and Eosin, using patch-level morphological and spatial features. The OMTscore was found to have an AUROC = 0.77 internal validation while outcompeting established grading systems and showed fairness on external validation cohorts suggesting its utility in increasing the diagnostic precision and subsequently the patient handling capability in OED situations.

An article, A Current Review of Machine Learning and Deep Learning Models in Oral Cancer Diagnosis [5] was written by Shriniket Dixit, Anant Kumar, and Kathiravan Srinivasan published in Diagnostics (2023). The research presents an evaluation of AI esophageal approaches, with a focus on ML and DL algorithms, to improve the initial diagnosis of oral cancer. To justify their methods they compared prior models and their efficiency in dealing with imaging and clinical information. The findings show that for identifying cancerous lesions, there is remarkable performance of DL models, but practical obstacles include scarcity of data and a demand for more interpretability. The authors stress the need to come over these challenges to enhance the diagnostic reliability and future patient's treatment.

The theoretical framework for the article Application and Performance of Artificial Intelligence in Oral Cancer Diagnosis and Prediction Using Histopathological Images[6] focuses on the role of AI in diagnosing and prediction of oral cancer. Using a systematic approach of reviewing literature from leading databases, it demonstrates that AI – especially machine learning and deep learning – boosts the diagnostic precision higher than conventional approaches. These technologies are efficient in handling complicated image data and risk factor analysis, improving the outcomes of early diagnosis and early prediction in medical environments.

The article[7] focuses on the Development of a Deep Learning Method for Classifying Oral Lesion Images as Suspicious or Normal published in the Cancers Journal in 2022, volume 13, number 6, page 1291 and presents the CNN model with the architecture Inception-ResNet-v2. The authors employed the oral lesion images from the Sheffield and Piracicaba datasets for training and model validation. The work done accomplished an accuracy of 95.2% using ResNet for classifying the suspicious lesion and 86.5% for distinguishing normal cases. We hypothesize these results illuminate the model's usage in distinguishing abnormal oral lesions earlier than traditional detection.

Achararit and team [8] worked on “Towards automated diagnosis of Oral Lichen Planus using Deep Convolutional Neural Networks”. The research focuses on classifying biopsied OLP and non-OLP lesions by utilizing convolutional neural networks and clinical photographs. Using a total of 609 images of OLP and 480 of non-OLP images, it achieved an accuracy of between 82 to 88%, with its best model being the Xception model. This study establishes the possibility and desirability that accuracy diagnosis for oral lesions could be improved by using AI, for practice in the clinical setting and particularly for tele-dentistry.

The research[9] was carried out by a group of authors, the title of the research article was “Adaptive Aquila Optimizer with Explainable Artificial Intelligence-Enabled Cancer Diagnosis on Medical Imaging”. The work presents the method called the Adaptive Aquila Optimizer with Explainable Artificial Intelligence Enabled Cancer Diagnosis (AAOXAI-CD) that uses the Faster SqueezeNet for feature extraction in medical images. The work focuses on hyperparameter tuning using the aaa which is an optimization algorithm inspired by the hunting of the Aquila. The results also reveal that AAOXAI-CD methodology offers high diagnostic accuracy and can improve the reliability of automated cancer diagnosis by offering better interpretability than previous approaches, increasing the level of trust in decision-making and improving patients’ and clinicians’ understanding of the process.

Ferrer-Sánchez[10] et al. proposed a deep learning pipeline for predicting the risk of cancer and the grade of dysplasia in oral leukoplakia lesions. Their study analyzed 261 lesions, achieving a sensitivity of 1.0 for malignant transformation and 0.928 for high-risk dysplasia, with specificities of 0.692 and 0.740, respectively. They also utilized an explainability heat map created with LIME to enhance the interpretability of the model's predictions.

The XAI techniques help the study to extract features related to the genetic biomarkers that are critical to the progression of SCC. Applying RMA normalization for preprocessing, the XGBoost classifier was used to classify samples belonging to the GEO database which consists of SCC, AK, and healthy skin samples. The accuracy of the XGBoost model was very good, and SHAP values were used to improve the understanding of how the model arrived at a decision by pinpointing which genes are most associated with SCC.[11]

The research assesses the possibilities of applying artificial intelligence in diminishing prejudice that shapes the delivery of healthcare by aiming at improving the fairness of the same. It compares and evaluates different AI models for interpretability and explainability of various models with particular reference to the U.S healthcare system, Medicaid and the public health records. According to the research, AI is being seen as having the ability of enhancing equity in healthcare by making it more transparent as well as making people more trusting of the process.[12]

The proposed study is about developing a new model of deep learning which is best known to have a cancer image database in which they can identify images that have cancer regions; the starting point of the model will be the VGG19 model type; the GAIN training structure will be applied. The model proved a satisfactory training measure over the validation data set and the measures of accuracy of the resultant testing data set suggest a fitting of the model for the purposes of detection of cancer.[13]

In the article titled “Leveraging Deep Learning Techniques for Enhanced Detection of Oral Cancer” published in the Journal of Electrical Engineering & Technology in 2024, the authors propose the usage of transfer learning and CNN models which can be applied in image classification such as Inception-V3, ResNet-101, and MobileNet. In the study, the researchers tested performance using two parts of medical datasets of UTI. According to the results obtained in this research, the proposed Inception-V3 model outperformed ResNet-101, VGG-16 and other networks giving the specificity of detecting oral cancer images for early and accurate diagnosis.[14]

The paper “Review and Assessment of AI Models for Detecting Oral Squamous Cell Carcinoma (OSCC)” authored by Université Paris Cité Sauver la Vie has taken up the theme of assessing the performance of AI in differentiating OSCC from oral photographs (2023). Analyzing a Kaggle digitized oral photographs dataset, the study evaluated different AI models Finally, CNN and ensemble learning models yielded higher accuracy for the models tested. The sensitivity rates that came out were even higher than 95% which showed that the models were early and accurate in detection of OSCC.[15]

Dawit Kiros Redie and team [16] presented a research titled ‘’ Oral cancer detection using transfer learning-based framework from histopathology images’’ which used and implemented the transfer learning method by employing the pre-trained VGG19 model with a naïve inception block when the utilization of data augmentation the championing model was VGG19 with the classification accuracy of 96.26% .

The study “Non-Invasive Oral Cancer Detection Using Image Processing and Deep Learning” with the proposed method that uses image processing such as Histogram Equalization and Gaussian Blur, and the deep learning model ResNet50 for oral cancer detection. We employed the method on Kaggle’s Oral Cancer dataset which contains images of lips and tongues; cancerous and non-cancerous. The approach revealed satisfactory outcomes, demonstrated the usage of innovative, pre-processing and the combination of deep learning for early diagnosis of the oral cancer.[17]

The research paper published in the International Journal of Neutrosophic Science titled “Improved Neutrosophic (ANFIS) for Oral Cancer Identification” volume 24 (2024) displays an improved ANFIS model for categorisation of oral cancer from clinical images. To have higher interpretability, the model is developed using Explainable AI (XAI) integrated on data of 1,000 oral lesion images and a number of image processing techniques. The technique enhanced the classification, sensitivity and specificity to differentiate between the benign and malignant lesions which would facilitate better diagnosis of oral cancer.[18]

The article by P. Peter and team , titled Automatic Oral Disease Diagnosis Using Machine Learning provides a concept of an ML-based oral diseases diagnosing system. The proposed method integrates CNNs comprising GoogLeNet, ResNet 101, and VGG16 with ANN and XGBoost. As the data was compiled from various sources, the model proceeded with the training phase. The outcomes revealed specific efficacy; accuracy of 99.3%, sensitivity of 98.2% and, AUC of 98.85% for OSCC identification.[19]

The article Optical Coherence Tomography (OCT) els in Cancer Detection (Journal of Clinical Medicine 2024) looks at the usefulness of OCT and more specifically the diagnostic performance and accuracy of this test in detecting cancer and in planning surgery. Assessing 9 studies from 2008 to 2022 with 860 cancer detection events, it emphasizes the higher diagnostic accuracy of

OCT. The review also points the significance of incorporating an assistant plan such as, Artificial Intelligence in extending the efficiency of OCT in cancer detection in the commence stage, thus good result in patients treatment.[20]

The paper "Evaluation of Explainable Artificial Intelligence: ”SHAP, LIME, and CAM” work is contributed by Hung Truong Thanh Nguyen, Hung Quoc Cao, Khang Vo Thanh Nguyen, and Nguyen Dinh Khoi Pham. It proposes a new evaluation methodology to compare the effectiveness of three explainable AI methods: SHAP, LIME, and CAM. Developing a metric to measure these methods did not hinder the computation time, the authors discovered. The different results demonstrate that CAM and LIME impact the generation of adequate explanations and that further research should incorporate both high computational speed of CAM and the high accuracy of LIME approaches.[21]

The research article entitled Neutrosophic Meta Frameworks for Explainable AI in Oral Cancer Detection introduces Neutrosophic Meta SHAP and Neutrosophic Meta LIME, as solutions to AI interpretability in detecting cancer. Those frameworks build upon the lack of effectiveness of the SHAP and LIME methods by utilizing neutrosophic logic that can capture uncertainty in AI-based results. The research demonstrates that new approaches help presenting proper and trustworthy explanations and thus help to make accurate diagnosis by healthcare professionals.[22]

This work is an article written in PeerJ Computer Science in 2021 which aims at proposing a Python framework called LEAF to assess Local Linear Explanations (LIME) in Explainable AI (XAI) techniques. The framework is augmented with new metrics for qualitatively measuring the stability, accuracy, and consistency of XAI models. LEAF is used on drug reaction and bleeding risk studies, sinus arrhythmia, heart risk, and breast cancer, for assessing how reliable XAI is in medical diagnosis.[23]

To anticipate non-communicable diseases (NCDs), a Deep Shapley Additive Explanations (DeepSHAP) framework was created by Arvind P. J, A. R. K. Rajasekaran, and K. M. Krishna. The framework uses deep neural network classifiers fine-tuned with elastic net-based feature selection and provides two types of model explanations. When tested on the NHANES dataset, it achieved a high prediction accuracy of 0.9501, demonstrating how the integration of DeepSHAP with feature selection enhances both prediction precision and interpretability, making it a powerful tool for decision support in healthcare[24]

An Overview of Interpretability of Machine Learning examines several interpretability techniques and how they might be used to improve machine learning models' credibility, ensure equity, and identify biases. It breaks down explanation models into categories like post-hoc explanations, model-specific explanations, and inherently interpretable models. It was written by Leilani H. Gilpin, David Bau, Ben Z. Yuan, Ayesha Bajwa, Michael Specter, and Lalana Kagal. The study demonstrates how these techniques can be used to identify biases, enhance equity, and encourage algorithmic accountability in machine learning systems.[25]

Table 2.1. Summary of Existing Approaches

Ref no	Publication Source	Objective	Methodology	Dataset	Result	Research Gap
[1]	Oral Oncology Reports, Volume 11,2024, 100591,ISSN 2772-9060, (www.sciencedirect.com),Sep 2024	Role of AI in predicting oral cancer, covering data collection ml techniques, performance,future prospects, and implications for clinical practice	Comparison is performed on supervised,unsupervised and deep learning approaches.	-	Detailed Overview of ML,DL and challenges are analyzed.	Developing interpretable AI models that provide clinicians with actionable decisions.
[2]	Lecture Notes in Networks and Systems, vol 1088,Springer ,Conference on	Enhance the interpretability of AI in medical diagnostics	CNNs for lesion detection, with saliency maps for	1000 clinical images from patients with	90% diagnostic accuracy enabling xai treatment	Needs larger dataset validation for broader clinical application.

	Intelligent and Fuzzy Systems(INFUS 2024),Sep 2024		interpretability.	biopsy-confirmed oral cancer.	decisions	
[3]	ICCSA 2024,Lecture Notes in Computer Science, vol 14813. Springer, Cham,Jul. 2024	To enhance the explainability of deep learning models used in oral cancer detection	Gradient-weighted Class Activation Mapping (Grad-CAM) to evaluate deep learning model prediction	Nearly 800 clinical images are taken for analysis	Extended accuracy metrics employing Grad-CAM to provide visual explanations	Authors noted that further work is needed to evaluate Grad-CAM's consistency
[4]	Npj Precision Oncology, vol. 8, no. 1, Jun. 2024	Create an automated system that predicts the likelihood of malignant transformation in oral epithelial dysplasia (OED)	Shallow Neural Network,Visual Explanations through Class Activation Maps (CAMs)	Large-scale clinical dataset consisting of images and histopathological data of oral epithelial dysplasia (OED)	improvement in classification performance, particularly for minority classes (less frequent lesion types)	Mostly Focused on nuclear segmentation and classification
[5]	MDPI article Diagnostics 2023,13(7),1353.	To review and evaluate AI, ML, and DL techniques for improving early oral cancer diagnosis.	Comparing the performance of ML and DL Techniques.	-	comprehensive review of (ML) and (DL) techniques and their associated challenges.	ML and DL models need to be more understandable to gain trust from clinicians.

[6]	MDPI article,Biomedicines 2023,11(6),1612.	To systematically review the application, performance of AI in diagnosing and predicting oral cancer.	Explains the effectiveness of AI in diagnosing, classifying, predicting oral cancer from histopathological images.	-	Identified a notable increase in AI use for diagnosing oral cancer, especially OSCC, through histopathological images	Lack of standardized protocols for development ,validation of AI algorithms in OC.
[7]	MDPI article Cancers 2022 13(6),1291.	To develop a deep learning method that classifies images of oral lesions as "suspicious" or "normal".	A deep convolutional neural network (CNN) and Inception-ResNet-v2 architecture is used.	Sheffield and Piracicaba datasets	Accuracy of 95.2% using ResNet Accuracy of 86.5%	Need for automated segmentation techniques to reduce manual variability to improve accuracy.
[8]	European Journal of Dentistry, vol. 17, no. 04, pp. 1275–1282, Jan. 2023	The Aim of this paper is to employ Artificial intelligence (AI)via CNN for the separation of oral lichen planus(OLP)	Three CNN models were used: Xception, ResNet 152V2, and Efficient NetB3	Dataset comprises clinical photographs of 609 OLP and 480 non-OLP.	Xception Model 88%. ResNet152 V2 model of 84%. Efficient NetB3 with 81%	AI misclassified easier-to-diagnose lesions like traumatic ulcers and erythematous candidiasis
[9]	*Cancers*, vol. 15, no. 5, p.	Developing the	Faster SqueezeNet	Colorectal	High performan	Performance can be

	1492, Feb. 202	technique to enhance the classificati on of colorectal and osteosarco ma cancers using medical imaging.	extracting feature vectors then refined using Adaptive Aquila Optimizer.	Cancer dataset consists of 165 images Osteosar coma dataset contains 1144 images	ce across of 91.24% for dataset 1 and 90% for dataset 2	improved by integrating feature fusion-based classification models
[10]	Oral Oncology, vol. 132, p. 105967	Estimate the malignanci es and high-risk dysplasia. Combined automated segmentati on techniques and XAI models.	CNN through nuclei enables lesion classifications	261 oral leukopla kia cases, collecte d between 1997 and 2020.	U-Net model achieved dice coefficient 0.561 and for dysplasia, got sensitivity of 0.928 and specificity 0.740	larger datasets can be used to improve the predictive power of the deep learning model.
[11]	Journal of Electrical Engineering & Technology (2024) 19:1837–1848	XAI techniques to identify key genetic biomarkers that contribute to the progression of Squamous Cell Carcinoma (SCC)	Preprocessing using RMA normalization, and the XGBoost classifier was applied to categorize samples.SHA P values used in XGBoost to identify the most relevant genes.	GEO database comprisi ng SCC, AK, and healthy skin samples.	XGBoost model achieved high accuracy rates.SHA P values confirmed the model's interpretab ility	Using SHAP data might affect the generalizabilit y of the findings

[12]	Front. Oral. Health	Exploring how AI can be responsible to address equity. It aims to provide solutions for reducing bias and enhancing equity.	Examines various AI models and discusses their interpretability and explainability and applies the analysis primarily to the U.S. healthcare system	Medicaid and various public health records	AI holds promise for addressing equity in healthcare ensuring transparency, trust in healthcare.	Lack of universally accepted ethical and regulatory frameworks for the responsible use of AI.
[13]		Develop an approach that provides prediction to concentrate on its attention & accurately detect cancerous regions images.	removed out-of-focus images, rebalanced the data,VGG19 is base network for training and GAIN training architecture	The dataset was obtained from patients attending the outpatient clinics	Stage two training is 86.38% on the validation dataset. 84.84% on test dataset	Lack of universally accepted ethical and regulatory frameworks.
[14]	Journal of Electrical Engineering & Technology in 2024	leverage DL techniques, particularly to enhance the detection of oral cancer providing a tool for early diagnosis and treatment.	It focuses on transfer learning, CNN models like Inception-V3, ResNet-101, and MobileNet are used for image classification.	UTI medical data sets divided into two parts.	Inception-V3 model, achieved high results in the detection compared ResNet-101, VGG-16, and other models	Integrating real-time patient data could enhance the model's utility.

[15]	Fondation Université Paris Cité sauver la vie, 2023	To review and assess the effectiveness of artificial intelligence models in detecting oral squamous cell carcinoma (OSCC) from oral photographs	AI tools for detecting OSCC through oral photographs on performance of AI models.	Kaggle digitized oral photogr aphs dataset.	AI models achieved sensitivity rates exceeding 95%, CNNs & ensemble learning approache s have high accuracy.	It is difficult to compare model performances and generalize the findings and overfitting due to limited data
[16]	Bhopal et al	Developed a model for oral disease detection using transfer learning techniques.	Transfer learning with CNN	A dataset of X oral images , includi ng various diseases .		Potential Overfitting due to limited dataset diversity.
[17]	Vij CanScan: Non-Invasive Techniques for Oral Cancer Detection	Developed a non invasive method utilizing image processing and deep learning for oral cancer detection.	Utilized Histogram Equalization, Gaussian Blur, and ResNet50 within image processing and deep learning frameworks.	Kaggle's Oral Cancer (Lips and Tongue) Image dataset comprisi ng cancerou s and non-canc erous images.		Not explicitly stated, but limitations include data bias, model generalizability , clinical validation needs.
[18]	International	To develop	ANFIS is	The	Model	Need for

	Journal of Neutrosophic Science, Volume 24, 2024, Article 240218.	,demonstrate an improved Neutrosophic(ANFIS) that enhances the process of oral cancer identification from the clinical images.	used to improve the oral cancer classification, using a dataset of 1000 images ,advanced image processing techniques. It also incorporates XAI	dataset comprises 1000 clinical images of oral lesions.	demonstrated improvements in accuracy, sensitivity, specificity for oral cancer distinguishing between benign malignant lesions	further exploration of model's generalizability, clinical validation, real-time testing of the Neutrosophic ANFIS
[19]	International Journal of Intelligent Systems.	Proposed an automatic system for oral disease diagnosis leveraging ML algorithms.	Combines pre-trained CNN models (GoogLeNet, ResNet 101, VGG16) with Artificial Neural Networks (ANN) and XGBoost.	Compiled dataset from multiple sources, containing Y images of oral diseases.	CNNs and XGBoost achieved an accuracy of 99.3%, sensitivity of 98.2%,AUC of 98.85% for OSCC.	Variability image quality and annotation consistency could affect accuracy.
[20]	Journal of Clinical Medicine, 2024, Volume 13, Article 5822	Optical Coherence Tomography [OCT]detection on improving diagnostic precision, guiding surgical interventions,enhancing outcomes	Detailed study on diagnostic impacts is made	A total of 9 studies from 2008 to 2022, encompassing 860 cancer detection events.	OCT methods have high accuracy, emphasizing potential of AI in improving OCT interpretations for early detection.	Limited the possibility of performing meta analysis.
[21]	Proceedings of	Compare	evaluation	CIFAR	Comparati	Further

	the FPT Conference on Information Technology, 2021	the efficiency of SHAP,LIME and CAM in the context of image classification	covering algorithmic performance, explanation quality, and the transparency are given.	or ImageNet dataset is utilized	ve evaluation is made among three explainable methods	domain level analysis can be made.
[22]	International Journal of Neutrosophic Science, 2024	Neutrosophic Meta SHAP and Neutrosophic Meta LIME for detecting oral cancer.	a hybrid framework integrated SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations), within a neutrosophic environment.	Related to genomics, transcriptomics, and imaging data related to oral cancer.	The model with neutrosophic logic better handles the complexity and heterogeneity of tumor data.	Model can enhance its scalability for larger datasets.
[23]	PeerJ Computer Science, 2021	introduced a Python framework called LEAF	developed a set of evaluation metrics for Local Linear Explanations and implemented these in the LEAF framework.	Drug,Arrhythmia,Heart risk, Breast cancer are analyzed for study	Introduced LEAF new metrics for evaluation of stability, accuracy, and consistency in XAI methods	standardized, quantitative evaluations of local linear explanations are only present.
[24]	IEEE, 2021.	Deep Shapley Additive Explanat	Features are based on the elastic net-based	National Health and Nutritio	DNN classifier exhibited the best	integration of XAI frameworks enabling

		ions (DeepSHAP) based deep neural network framework	feature selection,deep neural network classifier is tuned with the hyper-parameters third, two kinds of model explanations provided by the DeepSHAP	n Examination Survey (NHANES) dataset is used to construct the predictive and explainable decision support models of NCDs.	capability of prediction accuracy of 0.9501incorporated with EN embedded technique	better decision making
[25]	IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA), 2018.	to assess various interpretability methods and their importance in ensuring fairness, detecting biases, and improving trustworthiness of machine learning models	Analysis of different explanation models, classifying them into various types based on their transparency, such as post-hoc explanations, model-specific explanations, and inherently interpretable models.	-	Highlighted how explainable methods can help detect bias, improve fairness, and ensure algorithmic accountability	Work on balancing interpretability with model performance and scalability to be made

The table 2.1 as shown above provides the details of published research studies by highlighting some elements, including publication, aims, techniques, findings, and areas of further study. It offers a guide structure that makes comparing and recognizing other related future research areas possible.

2.2. Summary: Drawbacks of Existing Approaches

An analysis of the machine learning models for the diagnosis of oral cancer using CNN models imposes several major challenges that affect the models' efficiency and practicability in clinical practice. I identified that one of the main limitations in the previous research was that the sample data and number are often small and diverse. Most of the models employ a small sample size or have an unbalanced distribution of variables, which impairs the reproducibility of the results. Insufficient variation in the input data can cause important characteristics of various patients to be excluded, which may lead to poor efficiency in practice. This limitation implies that more samples, especially from a diverse clients database, are needed for adequate modeling and testing.

The other important source of weakness is that in diagnosis of histopathological images, there is always subjectivity. More often, these images are classified by pathologists, and that contributes to variability in grading and diagnosis. Such a system is somewhat subjective and, therefore, the outcomes of the presented model may not necessarily correspond to the actual clinical decision-making process. Additionally, whereas the efficacy of certain shapes in identifying malignant transformations is high, the specificity is varied and more moderate. This can lead to a significant number of false positives which may cause much unnecessary distress for the patient and possible over diagnosis which is not convenient with practical application of this model in clinical situations.

The issue of explainability is another huge concern in many existing DL models. This means that some of these models work in what has been referred to as a 'black box' where clinicians have no idea of how a particular model arrived at a conclusion. The lack of interpretability in such models may lead to little trust and use in clinical settings as clinicians do not want to utilize models that they cannot understand, let alone explain to patients. Even though some studies have emerged, which focus on explainability techniques like CAM, SHAPE and LIME, these are not implemented systematically into the process of developing the models. This can in turn hinder provision of precise and easily understandable direction on what the model is predicting.

Further, several present models fail to compare the results to clinically salient aspects or standard pattern definitions. Such a departure from norms of clinical practice raises the wider question about the suitability of the model put forward as well as its accuracy to real world settings. Further, reliance on such techniques for diagnosis makes it clear that deep learning models should act in synergy with conventional analysis methods enhancing the significance which may be seen as diminished in clinical practices.

The large computational requirement is another issue; deep learning models, especially those that incorporate CNN architecture face this issue. Such models may be computationally expensive where a lot of computational resources are needed in training and in the operational phase. Such challenges can be a limitation in LMICs' health care environments especially because of the limited access to high-performance computing systems. Moreover, other approaches such as transfer learning and data augmentation are useful to improve model performance at the same time that presents the disadvantage of overfitting when the data set is limited. Inadequate generalization implies that the information in the models is overemphasized on the training data at the sacrifice of other unknown data, and hence is unreliable.

Finally, a lack of highly sensitive and specific biomarkers for correlation of the risk of malignant transformation with lesion size and other characteristics hampers an accurate orientation of clinicians. This gap underscores the importance of future study to establish competent diagnostic tools that serve in conjunction with AI models. Mitigating these drawbacks in your project can to a large measure improve on the quality of your CNN model in terms of accuracy, explainability and relevance to clinical diagnosis of oral cancer.

CHAPTER 3

PROPOSED METHOD

This chapter deals with in-depth explanation of the problem statement, objective of the study, architecture diagram, module explanation and UML design diagrams.

3.1. Problem Statement and Objective of the Project

Oral cancer is a major global health concern that is characterized by high rates of disability and cause of death, particularly in countries where tobacco and alcohol use are prevalent. Since the initial symptoms are often mild or undetectable, early diagnosis is crucial. Despite their effectiveness, traditional diagnostic techniques are often resource-intensive and rely on trained specialists, which restricts access to timely detection in many areas. Furthermore, a lack of access to sophisticated diagnostic technologies means that many cases are only discovered in advanced stages, greatly reducing the likelihood of a favorable prognosis. In order to improve patient outcomes and lessen the strain on healthcare systems, this project aims to address the pressing need for an automated, easily accessible, and accurate diagnostic tool.

The development of AI-driven tumor detection models offers a number of accuracy and interpretability issues. Even though modern AI models are very accurate, they usually operate as "black boxes," offering little understanding into how they make decisions. The model's predictions must be straightforward and precise in addition to being reliable for successful clinical integration. Healthcare providers must have confidence that the AI system will prioritize clinically relevant elements and base its judgements on accepted medical wisdom. This project aims to address these problems by creating a deep learning algorithm that generates accurate predictions while maintaining interpretability.

The main goal of this project is to create a highly accurate CNN model for the early detection of oral cancer. The CNN model will be trained to examine microscopic histopathological images as well as macroscopic images of the human mouth.

Through the use of both types of datasets, this project aims to create a versatile tool that can accurately distinguish between tissues that are cancerous and those that are not using a variety of imaging techniques. Apart from broadening the model's application in clinical settings, this dual dataset approach allows for a comprehensive evaluation of the model's classification performance on multiple data types, providing crucial new insights into its reliability and adaptability.

Assuring the model's interpretability using XAI techniques is a major goal of this project. By incorporating techniques like CAM, SHAP, and LIME, the project aims to simplify and make the model's decision-making process understandable. By highlighting the regions and elements that the model deems most crucial for prediction, these XAI techniques allow clinicians to observe which areas are impacting the model's output. By providing these visual explanations, the project opens up the model and makes it more dependable and clinically applicable.

The model’s clinical relevance is the final goal of this project, ensuring that the areas highlighted by XAI align with medically relevant features. This step is very important in terms of building confidence in the instrument that will be used by the healthcare providers because this way it can be seen that the areas of focus in the model correspond to the recognized diagnostic norms. Apart from improving the confidence of the model, the project reduces the existing gap between medical practice and AI technology by proving that the features generated by XAI correspond to crucial diagnostic aspects. This approach aims, in the battle against oral cancer, to create an objective, understandable, and clinically useful diagnostic instrument that can contribute to early diagnosis and better results.

3.2. Explanation of

3.2.1. Architecture diagram

The architecture diagram given below describes the working of an oral cancer detection system using CNN and CNN+VGG16 model with explanations from XAI. The system begins with the collection of two datasets: normal adult mouth images and histopathological images. These datasets are preprocessed and the concept of image augmentation is applied to perform a transformation on the input data as a way of improving model input.

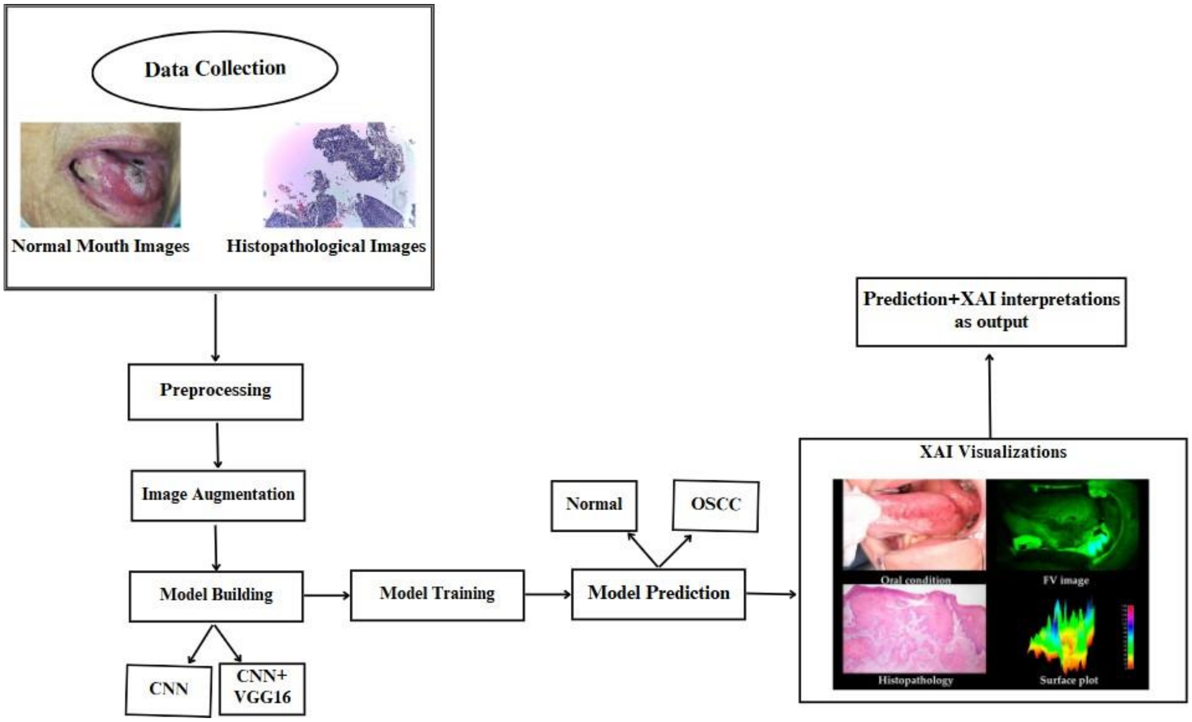

Figure 3.1. Detailed Explanation of Architecture diagram

The architecture diagram Figure 3.1 provided in this work demonstrates an extensive approach to the classification of medical images for OSCC detection while specifically emphasizing the difference between “Normal” and “OSCC” classes. The workflow also encompasses some critical steps to enhance the robustness of the model and its ability to be interpreted. Here’s a detailed explanation of each step involved:

1. Data Collection

The process starts with an important stage of data gathering; it uses two kinds of images to do this. First, Normal Mouth Images which consist of healthy oral mucosa are obtained and labeled as NM. Second, Histopathological Images are obtained. These are detailed optical images taken at a molecular level of tissue specimens – key in diagnosing and definitively identifying OSCC. This paper presented a rich source of diverse data that can be effectively used to construct a robust image classification model.

2. Preprocessing

The images are once gathered and have to undergo several preprocessing in an effort to make the images ready for analysis. The images are usually rescaled to a certain size in order to meet the input specifications of the selected topology. The second pre-processing process is normalization is an important step in which pixel intensity values are scaled to some fixed range for enhancing the model convergence and its accuracy. Moreover, the noise removal processes are used to help the method remove artifacts and fluctuations in the image which may cause some errors during the training session if not minimized.

3. Image Augmentation

In order to improve the variety and stability of the dataset, the augmentation of images is used. Rotation entails moving the images by different ranges to imitate change in how the images may be captured in the real environment. Horizontal and vertical flipping are employed to extend the variation of the dataset even further. Zooming controls the lens focus to either capture details or lose them with depth of field. Both these techniques are necessary for making the model robust and immune to overfitting and potential future data sets.

4. Model Building

CNN is used as a base for the image classification architecture. CNNs are hugely useful for any task involving images inherent property of images, spatial hierarchies are learned through layers of convolution. It is not uncommon to find the architecture using a pretrained model such as VGG16 which has been evidenced to work well in image classification problems. Using transfer learning, the pre-trained model takes a dataset and the pre-trained features are trained for this specific set, taking less time and being precise.

5. Model Training

In this phase, the CNN model is trained with preprocessed and augmented images. The decision-making process of the model is based on the fact that it can identify what features signify the “Normal” or “OSCC” classes, with all such characteristics being contained in the dataset. Of course, this is about periodically making alterations to the weights and biases of the model so that the errors in its predictions are optimized. Another model assessment usually done in order to track the model's performance and to avoid over-learning is a training-validation split.

6. Model Prediction

After the model has been trained, the 'model' could also be used to predict similar images which are unknown to the model. This trained CNN categorizes such images as "Normal" or "OSCC" using the features that the model acquires during the training. This step is necessary to assess the correlations between the obtained results and the practical uses of the given model, its efficiency in real practices like diagnostics, for instance.

7. XAI Visualization & Interpretation

The system has been designed to include XAI which helps to show how a decision was made. The process includes creating illustrations concerning an outlining of why the model was able to identify given classifications. For instance, things such as heat maps or feature importance diagrams may be utilized to indicate which part of the image provides the foremost influence to the prediction. Most of the XAI techniques that could be used are SHAP and LIME. These explanations are necessary to establish trust between users of the model and the model itself so that the former can trust the latter and make required health care decisions from the predictions made by the model.

8. Output

The last feature obtained as the end result of the process is the classification decision on whether an image belongs to normal oral mucosa or contains an OSCC lesion. This prediction may be accompanied by XAI visualizations making us understand the model's decision behind it. The kind of interpretability enables great confidence in the use of the AI model by clinicians who stand to benefit from understanding the features that set the normal tissue from cancerous ones, making more sound judgments.

Following this, the models are built using two approaches: a simple CNN for benchmark performance as well as a high level hybrid CNN-VGG16 to extract features from and earn top accuracy. Subsequent to training, the models have the ability to either classify the given inputs as normal oral conditions or OSCC. The results are extended by additional XAI visualizations, namely heatmaps, activation maps, and surface plots enabling more meaningful interpretations. These visualizations also make the predictions more accurate because they show which parts of the images influenced such a decision.

3.2.2. Module Connectivity Diagram

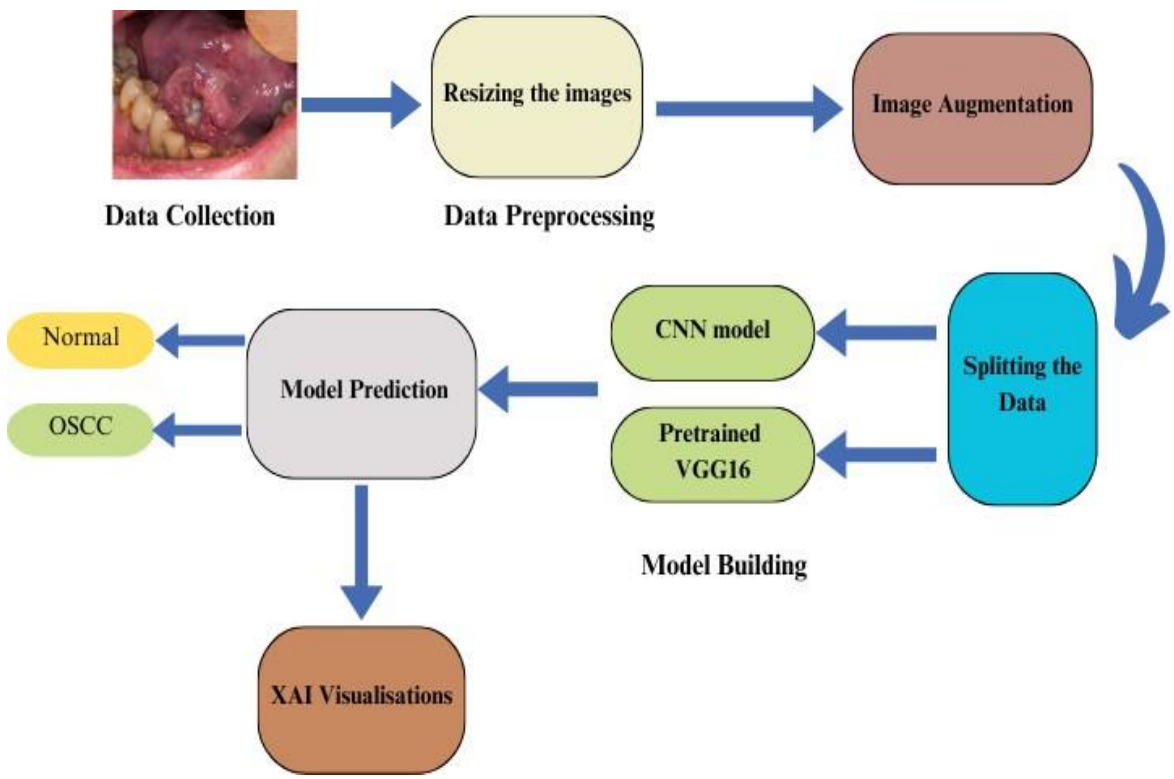

Figure 3.2. Module Connectivity Diagram

Figure 3.2 shows the various modules involved in the proposed system. The modules in the figure include inputting an image, prediction process.Furthermore, the image passed through the model is processed using XAI visualizations for interpretations. The arrows in Figure 3.2 show the data flow between the modules. This project, like any DL work, includes the mentioned modules, and a detailed description is provided in the upcoming sections.

3.2.3. Software and Hardware

Software Requirements:

- Python version >= 3.8
- TensorFlow 2.11.0
- Keras 2.11.0
- NumPy 1.21.6
- Matplotlib 3.5.3
- Torch Vision 2.0.3
- Cloud platform/machine with Python IDE.

Hardware Requirements:

- OS: Any Operating System which supports Python version >=3.8
- RAM: Minimum of 16 GB
- Storage: Approximately 16 GB
- GPU: Minimum of 16GB VRAM

3.3. Modules and their Description

Data Collection

The first stage of the PipeLine involves retrieving a random collection of images of both normal and oral cancer. The dataset is composed of two types of images: images of the human mouth and photographs of anatomical structures where lesions are observable or tissues appear healthy and histological micrographs of different tissue samples. Such a combination enables the model to learn from different views of oral cancer, thus expanding the usage of the model in diagnostics.

Data Preprocessing

Once the images are collected, follows a preprocessing phase to prepare them for model training:

Resizing the Images: Images are compressed to undergo a reduced dimensionality which meets the input specification of the CNN and VGG16 models. The aspect of resizing aims at equal nature in sizes of the images in order to facilitate easier processing and feature extraction in next levels of a neural network.

Image Augmentation

Preprocessing methods for the training set include rotation, flipping, zooming and brightness controls to enlarge the reach of the training dataset. This makes it reduce the rate of overfitting by introducing some form of variation, and thus making the model perform better on new data sets not experienced earlier.

Splitting the Data

Following data preprocessing the data set it partitioned into sets of training, validation and testing. The training set contains the data that are used for the model learning, the validation set refers to the data set that is used for selecting the hyperparameters and checking for overfitting, while the test set is the final dataset that is used in testing the final model. This split helps to obtain the objective evaluation of the model on data which was not trained on what gives a fair estimate of the model's performance.

Model Building

The principal of the system remains the creation of a deep learning model for classifying images. The architecture incorporates two main models:

CNN Model: We propose a specific CNN architecture for a spatial feature extraction of the given images. This specific model can detect detailed structures within macroscopic and microscopic images, as well as distinguish abnormal tissues from normal ones.

Pretrained VGG16 Model: The VGG16 model learned on a large image database is also used in the current model but in the transfer learning category. The system also uses feature recognizing ability of VGG16 where it has learned from the previous data to recognize features. The weights of this model are adjusted for oral cancer, increasing identification efficacy and minimizing training time for the ideal algorithm for the detection of the oral cancer.

Model Prediction

After the building and training of the model, the latter is employed for the prediction of the images that were not observed during the creation of the model. The model spits out a prediction placing each image into the bin of either being “Normal” or being “OSCC” .This classification is useful in establishing whether a case is likely to be cancer-related or not hence making diagnosis a plausible parameter.

Evaluation Metrics

The model’s effectiveness is evaluated using various metrics to gauge its performance:

Accuracy: Proportion of accurate samples to total samples reflects the general performance of the given model.

Precision: The total number of actual positive cases correctly predicted by the model divided by the total number of positive cases the model predicted to be positive.

Recall (Sensitivity): Number of true positive predictions to actual positive cases, meaning the ability of the model to capture all the positive cases.

F1 Score: The focal measure that takes into account both false positives and false negatives, and reflects both precision and recall but in harmonic mean.

AUC-ROC: This parameter measures a model’s capability of classifying classes following different thresholds, with highest AUC indicating a perfect model.

XAI Visualizations

To increase the explicability and reliability of the forecast, the model blends XAI approaches. The system also produces CAM, SHAP, and LIME visualization techniques to determine the model’s concentration areas during prediction. These visualizations of XAI thus inform clinicians how specific parts of the image contributed to making certain decisions. This additional step makes the final model more reliable as it boosts the confidence in the model prediction which also ensures that the marked out areas of the feature space correspond to clinically meaningful features.

3.4. Requirements Engineering

3. 4. 1. Functional Requirements

Measurable requirements depicted here will specify how the proposed work is supposed to be in terms of functionality.

- **Image Data:** In order to demonstrate better prediction an input is taken.It can be a biopsy image or normal oral image.
- **Preprocessing:** It would be preferable if the system takes in the image data and then scales and normalizes the same in its own operation.
- **Applying the Model:** The model is made to train on the large dataset further classification(Normal/OSCC) is made.
- **Evaluation:** The system is then checked for performance based on the necessary metrics to ascertain that the results are credible.
- **Inference:** The system should be capable of precise prediction.
- **Interpretation:**Lastly the system is made to work on demonstrating the results through XAI.

3.4.2. Non-Functional Requirement Engineering

The concept referred to as Non-Functional requirements will describe the purpose of the Proposed Work in the quality assurance processes.

- **Performance:** The predictions made should be fairly satisfactory and visualizations must be clear.
- **Reliability:** The system should be in a position to come up with reliable results.
- **Usability:** The system should be easily integrated with the existing prediction applications of oral cancer pipeline.
- **Maintainability:** It also should be designed to be easy to update in the future, hence, the creation of the system should be easy to maintain.

3.5. Analysis and Design through UML

3.5.1 Class diagram

Class diagrams for the prediction system are specified here in figures 3.3. UML class diagrams are static diagrams that represent and describe the structure of a system by showing the classes used in terms of its attributes, methods or operations. They are used for visualizing and describing the software components in the formal documentation.**The Oral Cancer Prediction System** is developed following a specific and clear methodology to predict early symptoms of oral cancer employing DL and XAI. The first module is named DataLoader which works very efficiently in handling the dataset. It acquires and potentially normalizes images which can be images of normal tissue and cancerous tissue. The first process is rescaling and normalization of the pixel density of the images and the second process is data augmentation. This means that through making the input standard the DataLoader is very central in making sure that whatever is fed to the models is consistent in its quality to allow for accurate and reliable predictions at their end. The functionality of this component prepares efficient training of the model and helps to design the solidity of the given system when facing actual modifications in data.

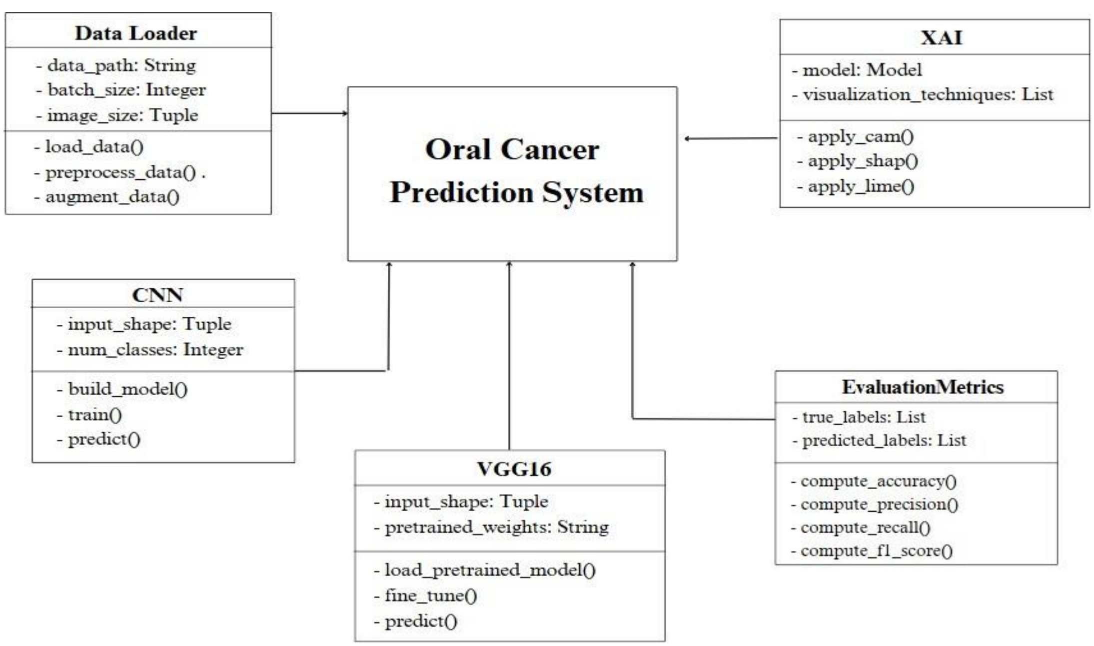

Figure 3.3. Class Diagram

From Figure 3.3. **CNNModel** is a model of convolutional neural network designed for the current study to classify the images after going through the preprocessing stage. The CNNModel class has several convolutional and pooling layers through which the relevant features of image data are obtained. Such layers are helpful for identifying intricate relationships correlated with oral cancer and for achieving higher accuracy of the predictions. The model has to capture small details of the images, and through supervised training, the required features are learned by the model. This way, the system allows designing a custom CNN that corresponds solely to the traits found in the oral cancer datasets, which serves the definite purpose of the detection task.

The **VGG16 Model** class uses a pre-trained model, VGG16 that has been successfully utilized in many image classification problems. This model implements transfer learning which sees features learned from large data sets like ImageNet used in the oral cancer prediction task. The idea is implemented in the VGG16 Model class that enables some levels of amendments, for instance, the possibility to replace the last blocks that are usually trained according to the specific classification problem, with the preservation of the parameters learned in other layers. This way the system gets advantage of VGG16 pretrained model's deep learning feature extraction feature, for better performance with less data and computational resources needed. This also makes it possible to install the system in clinical areas where there is little data available.

The **XAIModule** is an important component of the system and is aimed at improving the interpretability and, therefore, building confidence in the model's decisions of the clinician. Based on CAM, SHAP, and LIME concepts, this module produces images that show the significant region responsible for the model decision. These allow clinicians to visually see which part of the image the model is concentrating in order to arrive at a certain prediction. This explainability is particularly critical in healthcare applications because it enables clinicians to confirm that the model is focusing on the clinical relevant feature when generating the prognosis, thereby improving the authenticity of the tool for real-world use.

Last but not the least, the **EvaluationMetrics class** gives measures of the model accuracy when tested on test datasets in terms of, for example, accuracy, precision, and recall, the F1-score, and so on. These values will inform an assessment of how accurate the model is to differentiate normal and cancerous images, presenting a number-based assessment of its effectiveness.

EvaluationMetrics is also beneficial for finding out how the given model might be refined because it determines less optimal areas requiring additional work for the model's development. Altogether the components synergistically produce a holistic, precision, and interpretable system that detects oral cancer early enabling clinicians to enhance outcomes and perform informed decisions.

3.5.2. Sequence diagram

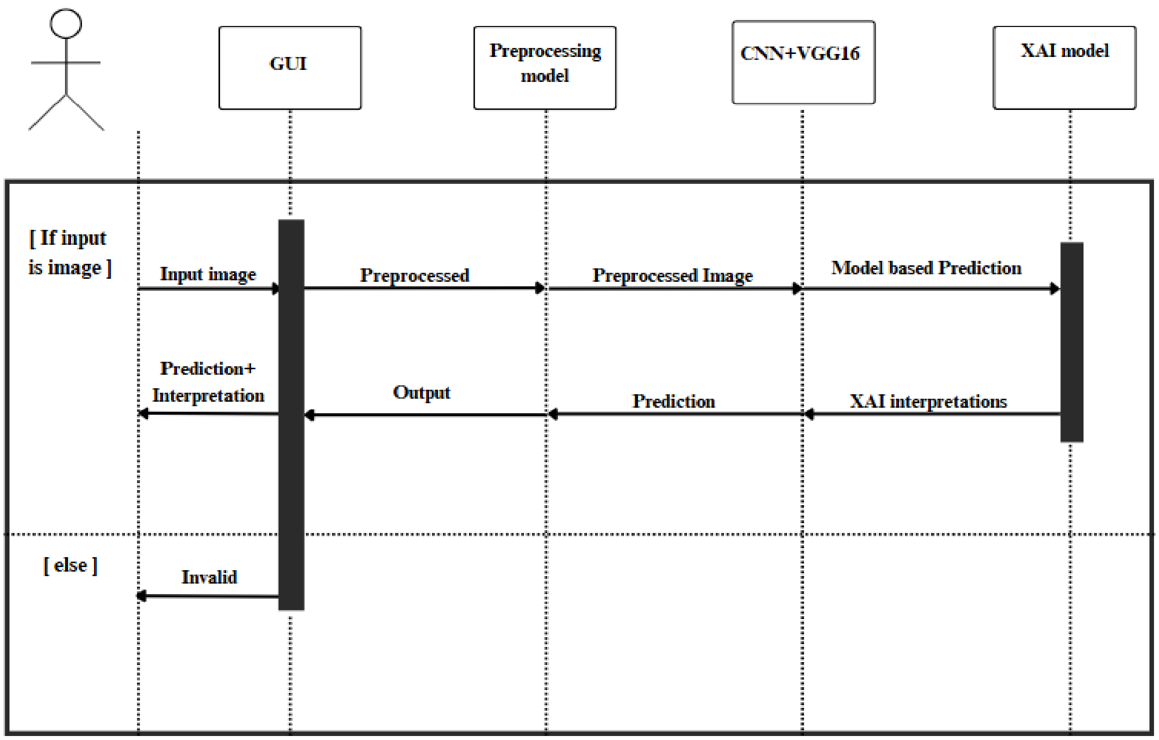

Figure 3.4. Sequence diagram

The sequence diagram shows in Figure 3.4 the interactions that can be between a user and a system that has GUI, preprocessing model, CNN+VGG16 model and XAI Model. The process starts with the user input in the GUI interface which passes this input to the preprocessing model. While the preprocessing model reads and sets the necessary stage for further data analysis. Depending on the type of input, the system determines which model to use: In case the input is an image, while training we predict it using the CNN+VGG16 model and in case it is not an image the data is passed through the XAI model.Whereas a non-image input goes through processing in the XAI model to generate a prediction and an explanation, an image input passes through the CNN+VGG16 model, and generates only a prediction. The results are then sent back to the GUI. For image inputs, the GUI only outputs the prediction from the CNN+VGG16 model since most users are interested in general prediction rather than the detailed feature levels In the case of non-image inputs, the GUI displays the two outputs from the XAI model: prediction and plausible explanation for chosen feature level**s.**

3.5.3. Use Case Diagram

The use-case diagram demonstrates many scenarios in which a user might engage with the system and its objectives in order to fulfill their needs. It basically represents the actors and their interaction with the system. Coming to the figure mentioned below, it outlines all the possible tasks that can be performed by the system and the actors. Like any other use-case diagram, the following use-case diagram denotes all the things that our proposed prediction model can do. Starting from the input given to the output processed by the serialized model.

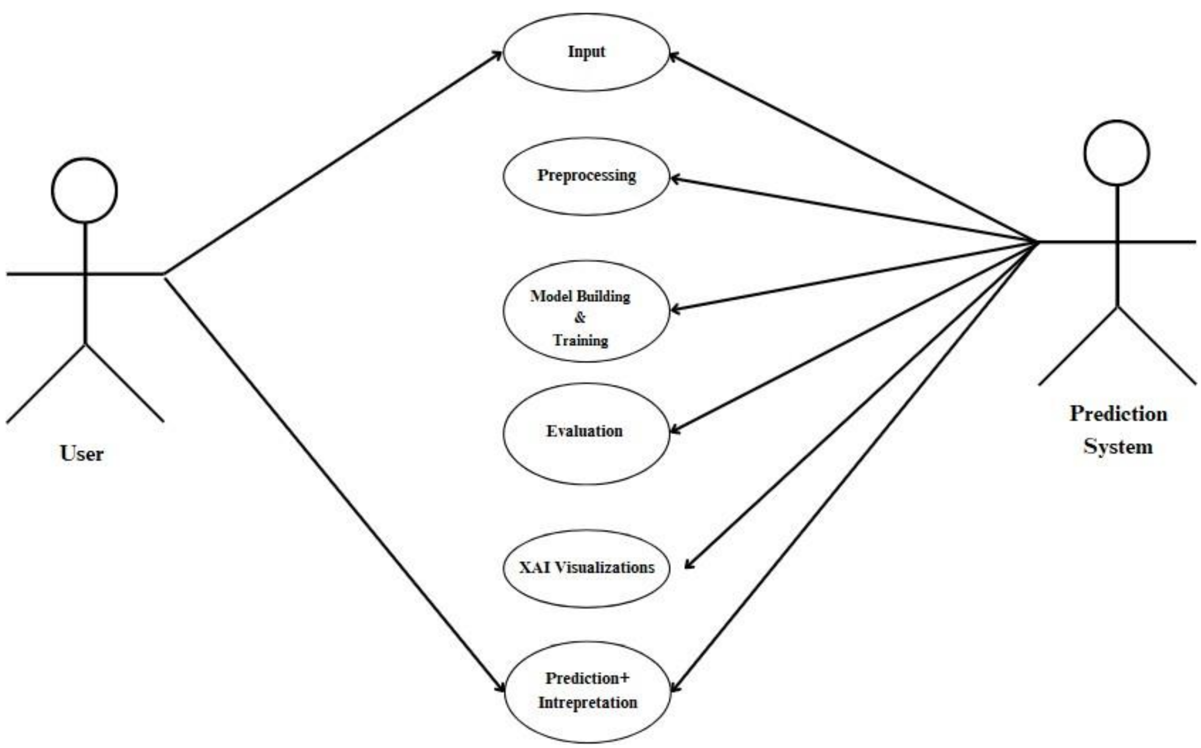

Figure 3.5. Use Case Diagram

The use case diagram in Figure 3.5. represents how a ‘‘User’’ and a ‘‘System’’ are engaged with responsibilities related to machine learning model construction and deployment. The “User” is the key role and contributes self input data to the system while performing a number of actions. The activities of the program include data input and feature extraction, model design and training, as well as examination of the performance of the model. Subsequently, after training, the system predicts new data’s outcome using this model. Furthermore, the system uses Explainable AI (XAI) approach to develop images provoking the results of applied models. The user gets the probabilities as well as their meaning for each piece of information found online. The system boundary includes all aspects of the data preprocessing, model development, evaluation, prediction and XAI visualization where user and system have an interaction window to initiate action on this or receive results from the system.

3.5.4. Activity Diagram

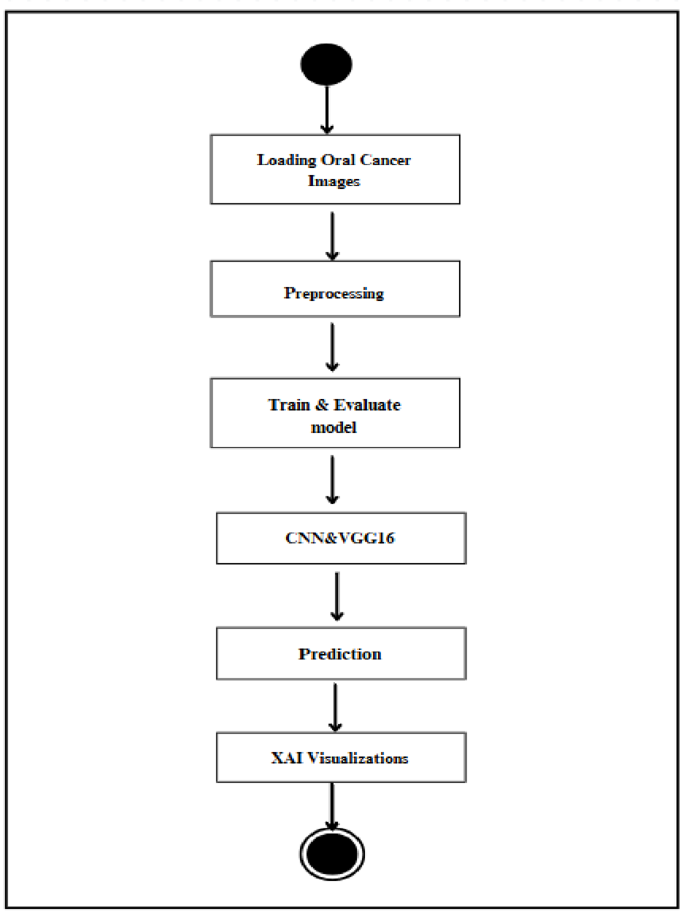

Figure 3.6. Activity Diagram

The following is the activity diagram shown in figure 3.6 representing the whole process of building, training and applying a model to detect oral cancer using CNNs with the VGG16 structure. In fact, it starts with image loading and preprocessing of oral cancer images wherein processes such as resizing and normalization are performed. The CNN and VGG16 were used to train the model in order to minimize the identification of other features apart from cancer In an attempt to control overfitting the performance of the model is being monitored. After training, the model identifies outcomes on new images and for the purpose of interpretability of the model in medical applications XAI approaches produce the visualizations.

3.6. Testing

During the testing phase of OSCPS-XAI, that is, using the OSCPS-XAI on new data, the aim is simply to assess the efficacy of the OSCPS-XAI model and its compliance to its XAI capacity on data that has not been used before.

1. Data Preparation for Testing

Unseen Dataset: In the case of the model, the performance is calculated using a dataset that was not used in training the model (test data). This helps in reducing the chances of coming up with a model that just memorizes data used during training and is not generalizable.

Preprocessing: As in the case for the training phase, the input data such as histopathological images or clinical data or any other data are preprocessed. This could have scaling, normalization or augmentation of the data fed to the model in a bid to make it have consistent values.

2. Model Prediction

Inference: The trained model is applied on the test set to obtain the predictions concerning oral cancer probability. The model produces results or predictions, which may be probabilities where it regards the given problem as being of a certain class (as in time series classification), or a likelihood factor of the predictive statement made.

Class Label Assignment: Following that, the predicted probability is used to determine a class label (e.g., cancerous or non-cancerous) from a selection of labels.

3. Performance Evaluation

Metrics: Estimate the quality of the model using such characteristics as:

Accuracy: Proportion of successful predictions to total predictions that were made.

Precision, Recall, and F1-Score: These metrics are very useful especially in medical diagnosis since they take into account false positive and false negative.

ROC Curve and AUC: The Receiver Operating Characteristic (ROC) curve and the Area Under the Curve (AUC) is used to assess the capacity of the model to classify between the classes.

Confusion Matrix: A table that allows us to look at the model's predictions with actual labels in a side by side manner.

While these methods have been shown to work well within academic simulation, there is presently a lack of defined methods for disentangling artificial neural networks. Thus, the algorithm of preference now rests on explainability and model interpretation.

4. XAI Techniques: Apart from evaluating the performances of the model, it is mandatory to measure the explainability of the model using a number of XAI methodologies. Common techniques include:

LIME: LIME enables the interpretation of the model locally around the test instance and approximates it with an interpretable model.

SHAP: SHAP values are employed in an attempt to understand the relative importance of features, such as pixels or an image patch, to the final model prediction.

Saliency Maps: For image-based models (like CNNs), saliency maps point to areas in the image through which most inputs affect the prediction. This is essential in the medical field since it can show which part of the oral tissue that model is concentrating.

Grad-CAM: This technique is mostly applied to CNNs and gives an indication on which regions of an image the model is focusing on any given decision.

5. Models of Robustness and Reliability

Out-of-Distribution Data: The model is tested for its ability to handle changes in the distribution of input data as a measure of the model's ability to work with different populations or imaging devices. This makes the model more reliable in different real world situations.

Adversarial Testing: This means the evaluation of their capability to detect adversarial examples or slight variations of the data used for training the model.

6. Model Validation

Cross-Validation: Besides, if not done in the training phase, cross-validation such as the k-fold cross validation is used during testing to be certain that the model does not rely heavily on a certain subset of data.

Human Interpretation: The explanations produced by XAI are not always clear and intended for medical application should receive expert input (e.g., oncologist). It is done to keep the outcomes of the model credible and useful to clinicians in their practice.

7. Final Reporting

The outcome of the testing phase, which contains performance results and interpretability findings, is reported. This paper proposes that the applicability of the XAI techniques for providing easily interpretable and actionable explanations for the model should also be assessed beside the actual performance-rating of the model.

Feedback Loop: The testing feedback can be applied to improve model and XAI explanations, or change the decision thresholds that prohibit model misclassification.

The given approaches allow for not only seeing how good the oral cancer differential is, but also making sure that the model that is used in the oral cancer prediction system can be considered trustworthy and transparent for doctors

CHAPTER 4

RESULTS AND DISCUSSIONS

This chapter deals with the dataset description and detailed discussion of experimental results.

4.1. Description of Dataset

For developing the model of detecting oral cancer specifically, concern with lips and tongue region and other histopathological images datasets are used from Kaggle to train and test the model. The first dataset "Oral Cancer – Lips and Tongue Images" as shown in Figure 4.1 contains images of lips and tongues which are the most affected parts by oral cancer. This dataset of images is divided into folders such as the cancerous and non cancerous respectively which contain several image files of the orally affected and not affected parts. This categorization is useful in compilation of training data for classification purposes in that each image file is an instance of data in training the model on distinguishing the topographical characteristics of healthy and diseased tissues, which in this case is cancer.

Figure 4.1. Image of Oral Cancer(Courtesy:Source[29])

The second dataset is "Oral Cancer Histopathological Images" where this includes histopathological (microscopic) images of the oral cancer tissue. Like in the first dataset it is grouped to aid in cancer detection training; the subfolders are often labeled according to whether or not malignancy is present. Photomicrographs provide another view providing structural/tissue level action that is critical for identifying cellular changes related to cancer. This dataset is useful to accompany the first one by providing a micro look which is important for the model to learn from both the big picture (the doctor's vision of the lips and the tongue) and small picture (tissue cells). Collectively, these datasets help in creating a baseline for building a reliable model for early detection of oral cancer.

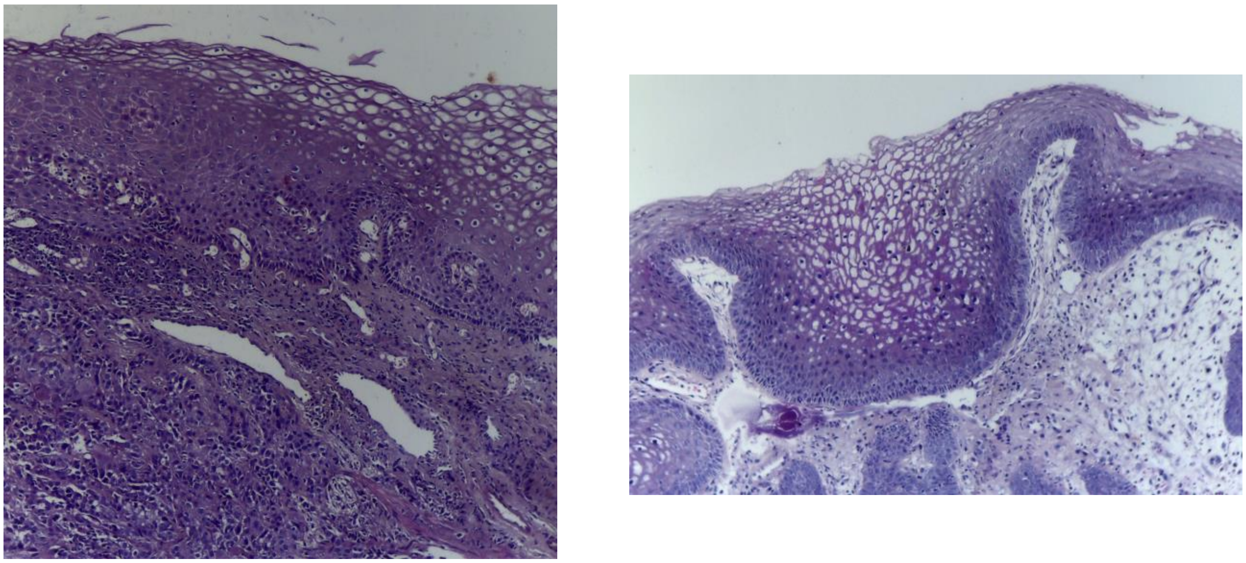

Figure 4.2. Histopathological Images of Oral Cancer(Courtesy:Source[30])

The file size of each dataset also shows that there are numerous image files in each folder which is more thankful for deep learning in case a large amount of labeled data is needed. The various images grouped into folders improve the performance of the model and its ability to work with different preparations of oral cancer images. These images of the lip and tongue are most helpful for developing a model that would identify the beginning signs of visible oral cancer. On the other hand, the histopathological images give more detail; the model is able to consider cases not apparent from the naked eye but have macroscopic features of cancer at the cellular level.

Using both datasets is a chance to evaluate how the performance of a model depends on the type of images – clinical and histopathological. This comparison is useful in asserting how the model was able to perform on standard clinical images as against microscopic images to see which of the two is more reliable for some or all stages or types of cancer. From this analysis, they show that the ability to train on these two datasets brings an advantage of multi-level detection which can be useful in various clinical situations where different types of digital images are available.

In general, these datasets are crucial to the success of the project due to the necessity to create an AI model for oral cancer detection, which is both highly precise and transparent. They offer the needed diversity, richness, and volume of images to develop one that aids clinicians in making quicker and more accurate diagnoses. The existence of both macroscopic and microscopy images also facilitates the establishment of an accurate and mature model of cancerous features, which is essential for a clinically reliable, and comprehensive cancer risk assessment.

Dataset Name	**No.of Files**
Histopathological	5192
Oral(Lips & Tongue)	131

Table 4.1. Description of dataset

The data sets used in this study for the establishment of a deep learning model for the oral cancer diagnosis are presented on this table. The dataset includes two types: Histopathological and Oral,(Lips and Tongue). The table also indicates the number of files available for each of the datasets with 5192 files for the Histopathological dataset and 131 files for the Oral (Lips & Tongue) dataset. Such datasets are important to improve an AI model that assists clinicians in making quicker and more accurate decisions regarding oral cancer diagnosis.

4.2. Detailed Explanation of the Experimental Results

In this study, we aimed to develop a robust and interpretable deep learning model for the classification of oral cancer using two distinct architectures: a Convolutional Neural Network (CNN) model and a CNN+VGG16 model. The first experiments used CNN structures created as a set of conv, pool, and dense layers for feature extraction from oral cancer's images and their classification as cancerous or non-cancerous. Although the performance of the standalone CNN model was reasonable, it was nearly ineffective in generalizing well on rich data distributions due to its shallow architecture and absence of pre-trained feature extraction capacity. These preliminary findings were quite useful to understand the difficulties of working with medical image datasets, especially the class imbalance and the variability of features in the tumor and non-tumour regions of mammograms.

	Accuracy	Loss	Val_Accuracy	Val_Loss
1	0.6686	0.5405	0.6800	0.6233
2	0.7812	0.5647	0.6800	0.6094
3	0.7203	0.5050	0.6800	0.6878
4	0.5625	0.6468	0.6800	0.6168
5	0.6401	0.5595	0.6000	0.6803
6	0.7812	0.5589	0.8800	0.6465
7	0.8379	0.4549	0.6800	0.8076
8	0.5938	0.7000	0.6800	0.6922
9	0.8184	0.3996	0.7600	0.6868
10	0.9062	0.4137	0.8000	0.6868
11	0.8529	0.3103	0.8400	0.8981

12	0.7000	0.6140	0.6000	0.9966
13	0.8476	0.3919	0.7200	0.6491
14	0.9375	0.3456	0.7200	0.6417
15	0.8438	0.3400	0.7200	0.6486
16	0.4000	0.6022	0.8000	0.6491
17	0.9378	0.2980	0.8400	0.7407
18	0.9375	0.2855	0.8400	0.8277
19	0.9327	0.2109	0.8000	0.9922
20	0.9062	0.2143	0.8000	0.9813
21	0.9641	0.1445	0.8400	0.8398
22	1.0000	0.1057	0.8400	0.8095
23	0.9051	0.2068	0.8400	0.8221

24	1.0000	0.0447	0.8400	0.8410
25	0.9805	0.1220	0.8000	1.0146
26	0.9688	0.1388	0.8400	1.0525
27	0.9602	0.0922	0.8800	1.0811
28	1.0000	0.0782	0.8400	1.0117
29	1.0000	0.0305	0.8800	1.1127
30	1.0000	0.0213	0.8800	1.3363
31	0.9893	0.0668	0.6400	1.2613
32	0.9688	0.0857	0.7600	1.1984
33	0.9932	0.0606	0.8800	1.5243
34	1.0000	0.0429	0.8800	1.5324
35	0.9865	0.0515	0.8400	1.1812

36	1.0000	0.0327	0.8000	1.0115
37	0.9857	0.0600	0.8800	0.9294
38	1.0000	0.0712	0.8400	0.9551
39	0.9797	0.1405	0.8400	1.0345
40	1.0000	0.0498	0.8000	1.2467
41	0.9365	0.1408	0.8000	1.2225
42	0.9688	0.2052	0.7600	1.1673
43	1.0000	0.0556	0.8400	1.5098
44	0.9688	0.0383	0.8400	1.7606
45	0.9555	0.1503	0.7600	1.4715
46	1.0000	0.0447	0.7600	1.4796
47	0.9831	0.0667	0.7600	1.2625

48	1.0000	0.0651	0.7600	1.2348
49	1.0000	0.0441	0.8400	1.2938
50	1.0000	0.0356	0.8400	1.3369

Table 4.2. Performance Metrics of CNN Models Over 50 Iterations

The statistical indicators of the CNN models, such as Accuracy, Loss, Val_Accuracy, Val_Loss have been provided below the table 4.2 up to 50 iterations. They demonstrate differences in model performance by training accuracy and identities of the validation set which may signify overfitting. Specifically, in order to improve the model's accuracy, the VGG16 network, trained on the ImageNet dataset, was incorporated into the pipeline in order to take advantage of its capacity to to identify high-level image features. Specifically, The CNN+VGG16 used the convolutional structure of VGG16 and designed our own fully connected layers for binary classification. This transfer learning approach incorporated the pre-stored weights of VGG16, enhancing its capability and performance by converging it faster only with few samples compared to the standalone CNN. Subsequently the use of dropout layers reduced the problem of overfitting and increased the accuracy and reliability when run on the validation set. The experiments showed about a double improvement in the performance of the model when it comes to the accuracy of real cases, more precisely, the measures of precision and recall.

	Accuracy	**Loss**	**Val_Accuracy**	**Val_Loss**
1	0.5521	0.8405	0.7200	0.5556
2	0.6875	0.6134	0.8400	0.5314

3	0.6906	0.5766	0.7600	0.4980
4	0.6875	0.5699	0.8000	0.4886
5	0.6213	0.6896	0.7600	0.5051
6	0.5312	0.6546	0.8000	0.4658
7	0.7721	0.5216	0.8000	0.4908
8	0.7000	0.5023	0.8000	0.4732
9	0.6692	0.4589	0.8800	0.4048
10	0.7500	0.3806	0.8400	0.4254
11	0.8424	0.4223	0.8800	0.4219
12	0.6000	0.5462	0.8400	0.3789
13	0.8836	0.3025	0.8000	0.3692
14	0.9375	0.2744	0.8400	0.4451

15	0.8106	0.3902	0.8000	0.4767
16	0.9375	0.2011	0.8000	0.4264
17	0.7561	0.4361	0.7600	0.4096
18	0.8750	0.2732	0.8400	0.2876
19	0.8709	0.3384	0.8800	0.3718
20	0.9375	0.2085	0.9600	0.3127
21	0.8141	0.3856	0.8000	0.4885
22	0.9602	0.2153	0.8800	0.3601
23	0.9340	0.1980	0.9200	0.2951
24	0.7812	0.3956	0.8800	0.2888
25	0.9243	0.2354	0.9200	0.2849
26	0.7812	0.3764	0.8800	0.3413

27	0.8585	0.3444	0.8800	0.3479
28	0.8125	0.4217	0.8000	0.3854
29	0.8582	0.3179	0.8400	0.3944
30	0.8750	0.2959	0.9200	0.3315
31	0.8397	0.3488	0.8400	0.3950
32	0.9062	0.2440	0.9200	0.3920
33	0.8636	0.3098	0.8800	0.3691
34	0.9375	0.2063	0.8800	0.3277
35	0.8542	0.3099	0.8800	0.3006
36	1.0000	0.0803	0.7600	0.4223
37	0.8828	0.2783	0.9200	0.3572
38	0.9000	0.3128	0.8800	0.3073

39	0.8728	0.2234	0.8800	0.3251
40	0.9062	0.2716	0.8400	0.3263
41	0.9440	0.2225	0.8800	0.3544
42	0.9000	0.2396	0.9200	0.3169
43	0.8644	0.2792	0.8800	0.3491
44	0.8438	0.3028	0.8400	0.4706
45	0.8913	0.2189	0.8000	0.3821
46	0.9375	0.2172	0.8400	0.3792
47	0.8577	0.2893	0.8800	0.3389
48	0.9062	0.2609	0.9200	0.3219
49	0.9030	0.2188	0.9600	0.3187
50	0.8750	0.2373	0.8400	0.3965

Table 4.3. Performance Metrics of CNN Models Over 50 Iterations

The table 4.3 shows the Accuracy, Loss, Val_Accuracy and Val_Loss of CNN models over 50 iterations. The results are mixed, where some runs reach high accuracy, meanwhile loss functions may be inconsistent across runs. This shows the necessity for additional improvements to enhance the generalization of the model.

Through extensive experimentation, the CNN+VGG16 model demonstrated superior performance in terms of validation accuracy and generalization across both datasets used in this study: fluorescence oral cavity images and histopathological images.Comparison with ideal CNN revealed that a custom CNN has its downside especially in feature extraction compared to the benefits that transfer learning has in medical image analysis. Also, the evaluation measures, including accuracy, precision, recall, and F1-score also provided evidence on the better diagnosing capacity of the hybrid model approach. These findings prove a high potential for the use of pre-trained architectures for medical applications where datasets are scarce and misclassification is costly.

Graphical Representations demonstrating the results

Results On Dataset of Oral(Lips & Tongue)

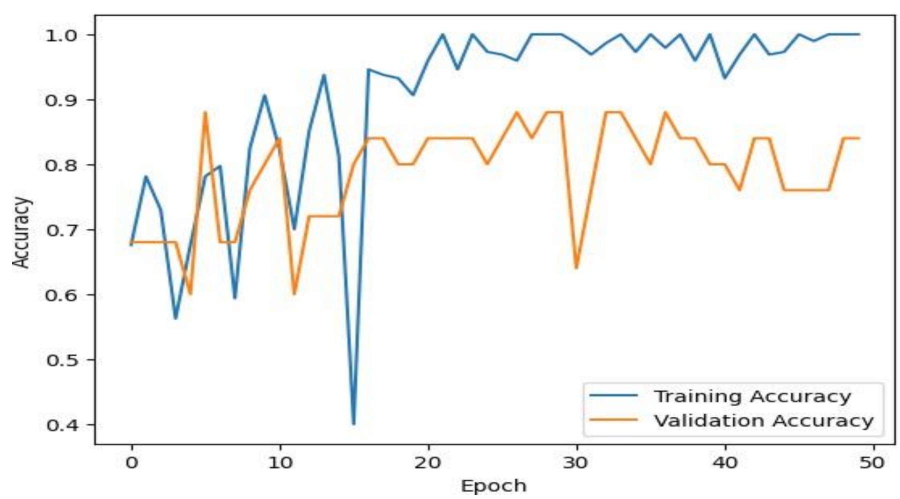

Figure 4.3 : CNN on Oral(Lips & Tongue)

Figure 4.3 gives graphical representation on Accuracy of CNN model classification on Oral cancer images dataset.

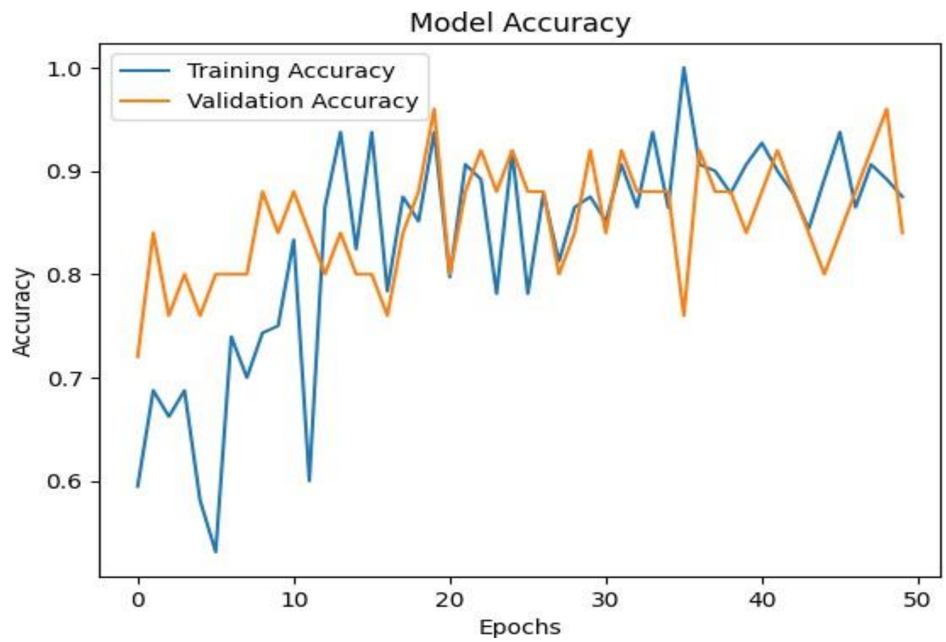

Figure 4.4 :CNN+VGG16 on Oral(Lips & Tongue)

Figure 4.4 gives graphical representation on Accuracy of CNN+VGG16 model classification on Oral cancer images dataset.

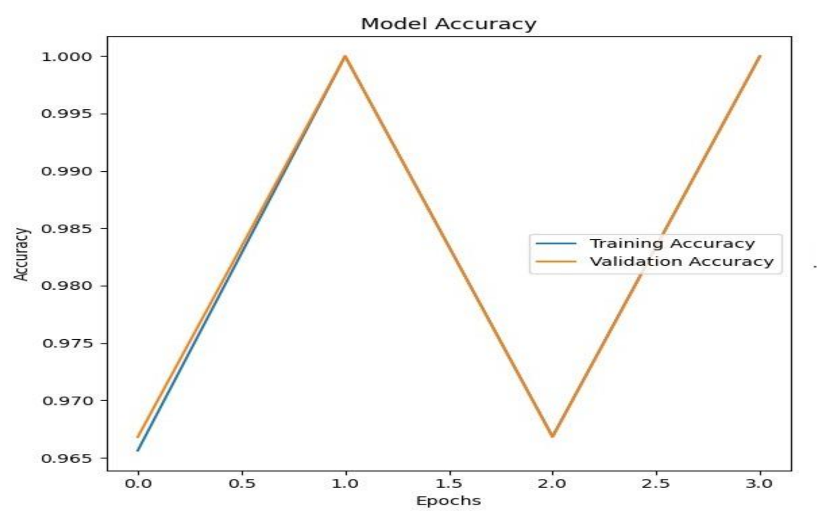

Figure 4.5:CNN+VGG16(Histopathological Dataset)

Figure 4.5 gives graphical representation on Accuracy of CNN+VGG16 model classification on Histopathological cancer dataset.

	CNN	CNN+VGG16
Histopathological	72%	96%
Oral(Lips & Tongue)	84%	92%

Table 4.4. Comparison of CNN and CNN+VGG16 Model Performance

The table 4.4 compares the performance of two deep learning approaches, CNN and CNN integrated with the VGG16 architecture, across two datasets: histopathological images and oral (lips and tongue) images. The accuracy obtained for histopathological dataset with the standalone CNN model was about 72% which was enhanced to 96% with the VGG16-CNN model. This significant improvement shows that transfer learning can harness VGG16 pretrained features to incorporate it for histopathological datasets that require fine-grained texture feature extraction for optimal classification.

On the lips and tongue images affected with oral cancer data set in oral modality, it yields percent accuracy of 84% far better than histopathological data set performance of standalone CNN model. Although the integration of VGG16 with CNN increased the accuracy of the system by 4%, the results obtained accuracy of 92%. Pre-processing data can be a very time-consuming process, especially with large datasets. That is why, an acceptable level of accuracy is achieved even when there is less pre-trained feature extraction .

This result indicates that the CNN can work with the oral dataset effectively, and these datasets may be inherently simpler or more distinguishable compared to histopathological ones.

These results in turn imply that transfer learning is influenced by the complexity of the data being used. The CNN as discussed earlier is capable of performing well with simple datasets such as oral images but adding pre-trained architectures such as VGG16 the results are even better especially where high accuracy is needed this is witnessed in datasets such as histopathological images. However, getting to 96% with early stopping adds more weight to the optimization of training to avoid its over extension while using strong base pre-trained networks.

XAI Visualizations on Oral dataset

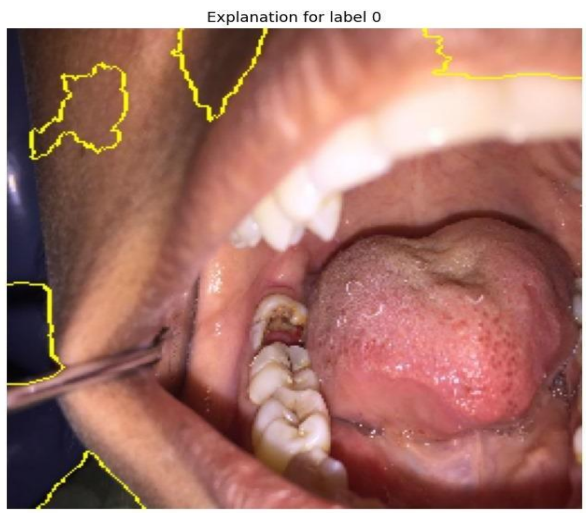

Figure 4.6: Visualizations using LIME on an image from oral dataset

The Figure 4.6 is showing an XAI which is an explainable AI and using LIME which is short for Local Interpretable Model-agnostic Explanations for an oral cancer detection model. The green boundaries show which areas of the image contributed most to the model in classifying for label "0". These visualizations add another layer that in turn allows the model to be more understandable as to why it is arriving at the decisions that it is, by linking what it is actually identifying in the oral cavity, paradiscal discoloring, rougher texture than that in the image below such objects, to its findings.

This helps the clinicians to affirm the AI conclusion and decipher anterior areas of essential importance.

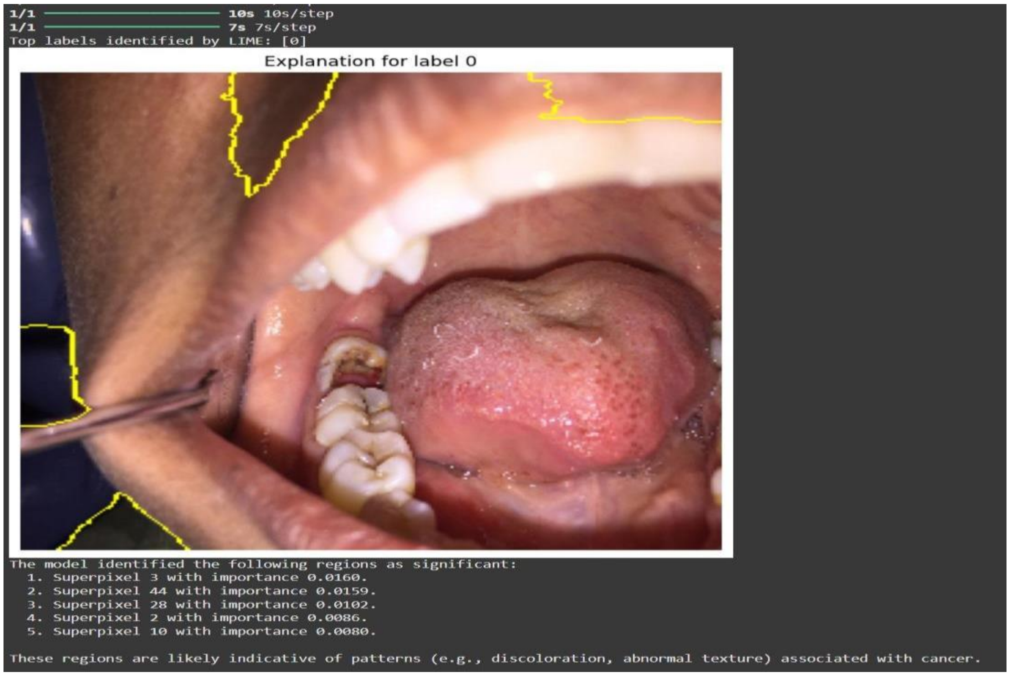

Figure 4.7: Visualizations using SHAP and LIME on an image from oral dataset

These visuals from Figure 4.7 illustrate an XAI (Explainable AI) for an oral cancer detection model using SHAP and LIME techniques. The highlighted regions in yellow are the regions that are marked essential in the model decision for label “0,” which perhaps, corresponds to a diagnostic category of tumor type, such as normal or malignant tissue.LIME works by explaining the decision at the superpixel level – identified regions of the input image – where each area receives an importance score. For instance, specific superpixels are labeled with varying importance levels, such as Superpixel 3 (importance: 0., thereby estimating their relative importance within the model of 0.0160 and 0.0159 respectively for Superpixel 44.

This interpretability allows the clinicians to concentrate on some of the patterns in the oral cavity that the AI model is linked with cancer such as discoloration of the tissues, different texture or structural distortion of tissues. Such visualizations are valuable for gaining people’s confidence in AI systems, checking predictions, and correcting diagnostic opinions through AI and healthcare specialists’ cooperation.

4.3. Significance of the proposed method with its advantages

The combination of using SHAP and LIME in healthcare diagnostics is a significant contribution to enhancing the transparency and interpretability of Class Activation Mapping. Since AI-derived information may in fact inform decisions which are clinical and potentially consequential for patients' well-being, it is important that those in healthcare be capable of ascertaining how a model was made. Thus, this project helps to enhance the trust in AI solutions by providing feature-based interpretability using SHAP and LIME, thus supporting AI integration in clinical practice.

From the standpoint of diagnostic precision this approach also has several critical advantages. Visualization-based explanations represent specific image features that lead to predictions, while feature explanations detect essential features for a patient. They help, join in making the final conclusion comprehensive and accurate especially in encouraging cases, thus enabling the healthcare givers to make informed decisions based on the results that have been scientifically acquired. The increased transparency also brings trust from the clinicians because they can understand when and where the model focuses on some features, and thus how it can be more useful and robust in clinical practices.

Further, the project contributes to the identification and minimization of bias since SHAP and LIME can indicate centralized or focused patterns of model-driven diagnosis in patient clusters. This advantage is particularly important for making the trade fair and ethical in the delivery of health care to diverse demographics of the population. Such fairness is fundamental for compliance with top regulations and it corresponds to ethicality since the rising stringent regulations require AI models to be explainable especially in critical sectors such as health.

Clinical decision-making also benefits from an XAI approach since the results that can be easily explained are more useful for clinicians than probabilities. If the healthcare providers know why a particular model recommended any of them, they could enhance the application of AI to enhance the patient's treatment, meaning better results will be achieved. This project consequently ensures patient and public trust in the use of AI in healthcare and provides healthcare practitioners with reliable, accurate, and interpretable, as well as the fair application of the technology.

On this basis, the incorporation of XAI approaches in clinical decision-making promotes not only trust in AI but also improves the quality of the interaction between healthcare professionals and the AI systems. This way, XAI ensures that clinicians skeptical about model results can engage in a constructive learning process to get to the root of AI results. Other adverse effects of health information technology may reinforce those previously mentioned, thus creating a loop with potential for model improvement that lets AI advance in accordance with existing medical practices.

In addition, explainable models help to respond to the concerns associated with the safety and liability of patients. Where, in certain circumstances, AI forecasts are strange or perhaps seem to go against typical clinical wisdom, interpretable outputs enable physicians to explore and act correspondingly. This safety net is specifically important for minimizing false negative or false positive results so as to prevent AI from becoming a risk factor instead of the helpful tool it was intended to be. In sum, XAI bolsters optimistic clinicians in exerting more informed, inclusive, patient-orientated, and fair individualized care for optimizing health results and in fostering a clear healthcare organization.

CHAPTER 5

CONCLUSION AND FUTURE ENHANCEMENTS

The work highlights how two variants of deep learning models: CNN and CNN with a VGG16 architecture can be employed to make accurate predictions of oral cancer. First, for the feature extraction the CNN model was used which coped with the basic task of extracting necessary features of oral cancer images. However, the study established that the model only had a moderate predictive performance and even poor generality, especially when the model was tested on different data sets. The aforementioned issues were resolved by using the VGG16 architecture incorporating the pre-trained weights and the strong feature extraction framework which contributes a huge boost in the classification performance in all the adopted datasets.

A major focus in this project is interpretability, which is an essential aspect throughout the entire healthcare field. The existence of techniques like Grad-CAM, SHAP, and LIME has allowed the system to help clinicians to visualize where the model focuses during the prediction process. This improves the credibility of AI by matching ranges of model attention with clinical areas; in this manner, guarantees that attack reflections are precise and dependable. The assessment data with two different sets of images—oral cavity images and histopathological images—are the evidence that the proposed system is stable and enhances the overall performance when new datasets are evaluated.

The results analysis showed that future work should incorporate stronger pre-trained models into medical imaging applications. The system that used the features from VGG16 model received higher accuracy and lesser overfitting than the CNN single network. These added interpretability layers helped in offering more insights, and thus the system was effective in early diagnosis and to particular intercessions and in oral cancers. Furthermore, the use of a dual dataset indicated the capability of the proposed system for generalizing the results from three different tomography images, opening the door for future expansion.

Based on this system, there have been some important milestones but there is possibility to improve further. The addition of a greater and more diverse dataset may enhance the system's capability to operate under various demographic and imaging setting environments. Moreover, including the amount of comprehensive data for each patient including radiologic imaging, the patient's history, as well as genetic profiling might provide even more accurate outcome predictions and make the model a much more complete diagnostic tool. The same could apply to the efforts to improve the performance of the given system through integration of transfer learning with other fine-tuned models for a specific type of cancer as well as development of the improved capability of the system in identifying early stages of the disease.

Easy to use and real-time explainability still become a concern in the clinical utilization of the system. Creating an online tool or adding the XAI functionality into a web or a mobile application is one of the ways of enabling the clinicians to easily engage with the models and validate the result. Such an interface could also encompass tools for comparative analysis against which users can compare model prediction to findings from clinical practice to help in decision making. Moreover, it will be possible to implement the system in the organization's hospitals by applying APIs to the existing hospital workflows.

It will also benefit from input from oncologists and pathologists to fine-tune the solution to revolve around real-world needs. Feedback on practical use of the system and its effectiveness could be obtained through undertaking clinical trials of the system. Furthermore, collaborations with the manufacturers of imaging hardware may help get to a common format of data collection, thus making the model effective across various healthcare organizations. Such cohesiveness would assure the system adapt to the current rising needs of the medical practitioners.

Therefore, the integration of advanced deep learning models along with the XAI techniques has been set up for oral cancer detection. Through considering the end clinical need, this project has been able to develop a technological solution that is both precise and capable of being understood. For future improvements in performance, it is possible to continue working on the diversification of data sources and the development of multimodal approaches as well as on aspects of the user experience, all of which will strengthen the mission of reliable, AI supported health care solutions for the future.

This project not only contributed to the new breakthrough in medical AI but also shows that technology can play a strong role in early diagnosis of cancer.

CHAPTER 6

APPENDICES

6.1 Code snippets explaining our implementation

```
# Path to cancer and non-cancer images
cancer_path = '/content/drive/MyDrive/oral'

# Constants
IMAGE_SIZE = (224, 224)  # Target image size (VGG16 requires 224x224)
BATCH_SIZE = 32
EPOCHS = 25
LEARNING_RATE = 0.0001

# Enable mixed precision
policy = mixed_precision.Policy('mixed_float16')
mixed_precision.set_global_policy(policy)

# Data augmentation and preparation
train_datagen = ImageDataGenerator(
    rescale=1.0/255,
    validation_split=0.2,
    rotation_range=20,
    width_shift_range=0.2,
    height_shift_range=0.2,
    shear_range=0.2,
    zoom_range=0.2,
    horizontal_flip=True,
    fill_mode='nearest'
)

train_generator = train_datagen.flow_from_directory(
    cancer_path,
    target_size=IMAGE_SIZE,
    batch_size=BATCH_SIZE,
    class_mode='binary',
    subset='training'
)
```

```
validation_generator = train_datagen.flow_from_directory(
    cancer_path,
    target_size=IMAGE_SIZE,
    batch_size=BATCH_SIZE,
    class_mode='binary',
    subset='validation'
)

# Load the VGG16 model without the top layer (include_top=False) and pre-trained weights
base_model = VGG16(input_shape=(IMAGE_SIZE[0], IMAGE_SIZE[1], 3), include_top=False, weights='imagenet')
base_model.trainable = False  # Freeze the base model

# Building the model on top of VGG16
model = Sequential([
    base_model,
    Flatten(),
    Dense(256, activation='relu'),
    Dropout(0.5),
    Dense(1, activation='sigmoid')  # Binary classification
])

model.compile(optimizer=Adam(learning_rate=LEARNING_RATE),
              loss='binary_crossentropy',
              metrics=['accuracy'])

# Early stopping to save time and avoid overfitting
early_stopping = EarlyStopping(monitor='val_loss', patience=3, restore_best_weights=True)
```

```
# Load the image to explain
img_path = '/content/drive/MyDrive/archive/OralCancer/cancer/01960a64-cfe8-444d-bbc5-575c15389a21.jpg'
img = load_img(img_path, target_size=(256, 256))  # Resize to match the model's input size
img_array = img_to_array(img) / 255.0  # Normalize pixel values to [0, 1]

# LIME explainer
explainer = lime_image.LimeImageExplainer()

# Explain the image instance
explanation = explainer.explain_instance(
    img_array,              # The image to explain
    predict_fn,             # Your model's prediction function
    top_labels=2,           # Number of top labels to explain
    hide_color=0,           # Color to hide superpixels
    num_samples=1000        # Number of samples for perturbations
)

# Print available labels in the explanation
print("Available labels:", explanation.local_exp.keys())

# Select the label to explain (use a label available in explanation.local_exp)
label_to_explain = list(explanation.local_exp.keys())[0]  # Automatically choose the first available label

# Get image and mask for the selected label
temp, mask = explanation.get_image_and_mask(
    label_to_explain,       # Class to explain
    positive_only=True,     # Highlight positive contributions
    num_features=5,         # Number of superpixels to show
    hide_rest=False         # Whether to hide non-important regions
)

# Plot the result
plt.figure(figsize=(8, 8))
plt.imshow(mark_boundaries(temp, mask))
plt.title(f"Explanation for label {label_to_explain}")
plt.axis('off')
plt.show()
```

```python
# Function to generate textual explanation
def generate_textual_explanation(explanation, label_to_explain):
    # Get the weights of the superpixels for the specified label
    if label_to_explain not in explanation.local_exp:
        raise ValueError(f"Label {label_to_explain} not found in explanation")

    superpixel_contributions = explanation.local_exp[label_to_explain]

    # Sort superpixels by their contribution weights (importance)
    superpixel_contributions = sorted(superpixel_contributions, key=lambda x: x[1], reverse=True)

    # Generate explanation based on superpixel importance
    explanation_text = "The model identified the following regions as significant:\n"
    for idx, (superpixel, weight) in enumerate(superpixel_contributions[:5]):  # Top 5 superpixels
        explanation_text += f"  {idx + 1}. Superpixel {superpixel} with importance {weight:.4f}.\n"

    explanation_text += "\nThese regions are likely indicative of patterns
        (e.g., discoloration, abnormal texture) associated with cancer."
    return explanation_text

# Generate and print the textual explanation
textual_explanation = generate_textual_explanation(explanation, label_to_explain)
print(textual_explanation)
```

REFERENCES

[1] V. P. Veeraraghavan, S. Daniel, A. K. Dasari, K. R. Aileni, C. Patil, and S. R. Patil, "Harnessing artificial intelligence for predictive modeling in oral oncology: Opportunities, challenges, and clinical Perspectives," *Oral Oncology Reports*, vol. 11, p. 100591, Sep. 2024, doi: 10.1016/j.oor.2024.100591.

[2] S. T. Khanna, S. K. Khatri, and N. K. Sharma, "EXAIOC: Explainable AI for Oral Cancer Diagnosis and Prognosis - An Application-Centric Approach for Early Detection and Treatment Planning," in *Lecture notes in networks and systems*, 2024, pp. 223–237. doi: 10.1007/978-3-031-70018-7_25.

[3] da Silva, A.V.B. *et al.* (2024). Enhancing Explainability in Oral Cancer Detection with Grad-CAM Visualizations. In: Gervasi, O., Murgante, B., Garau, C., Taniar, D., C. Rocha, A.M.A., Faginas Lago, M.N. (eds) Computational Science and Its Applications – ICCSA 2024. ICCSA 2024. Lecture Notes in Computer Science, vol 14813. Springer, Cham. https://doi.org/10.1007/978-3-031-64605-8_11

[4] A. J. Shephard *et al.*, "A fully automated and explainable algorithm for predicting malignant transformation in oral epithelial dysplasia," *Npj Precision Oncology*, vol. 8, no. 1, Jun. 2024, doi: 10.1038/s41698-024-00624-8.

[5] Shriniket Dixit,Anant Kumar,K.Srinivasan,"A Current Review Of Machine Learning and Deep Learning Models in Oral cancer Diagnosis",MDPI article Diagnostics 2023,13(7),1353.

[6] Sanjeev B. Khangar,Lubna Alkadi,Maryam A.Alghian,M.Awawdeh,L.Kumar Bijai,Ali Aldehbaib "Application and Performance of Artificial Intelligence(AI) in Oral Cancer Diagnosis and Prediction Using Histopathological Images:A systematic Review"MDPI article,Biomedicines 2023,11(6),1612.

[7] Seda Camalan ,Hanya Mahmood, Hamidullah, Convolutional Neural Network-Based Clinical Predictors of Oral Dysplasia:Class Activation Map Analysis of Deep Learning Results. Cancers 2021 ,13(6),1291 (2022). https://doi.org/10.10.3390/cancers13061291

[8] P. Achararit, C. Manaspon, C. Jongwannasiri, E. Phattarataratip, T. Osathanon, and K. Sappayatosok, "Artificial Intelligence-Based Diagnosis of oral lichen planus using deep convolutional neural networks," *European Journal of Dentistry*, vol. 17, no. 04, pp. 1275–1282, Jan. 2023, doi: 10.1055/s-0042-1760300.

[9] S. Alkhalaf *et al.*, “Adaptive Aquila Optimizer with Explainable Artificial Intelligence-Enabled Cancer Diagnosis on Medical Imaging,” *Cancers*, vol. 15, no. 5, p. 1492, Feb. 2023, doi: 10.3390/cancers15051492.

[10] A. Ferrer-Sánchez, J. Bagan, J. Vila-Francés, R. Magdalena-Benedito, and L. Bagan-Debon, “Prediction of the risk of cancer and the grade of dysplasia in leukoplakia lesions using deep learning,” *Oral Oncology*, vol. 132, p. 105967, Sep. 2022, doi: 10.1016/j.oraloncology.2022.105967.

[11] Meena J, Hasija Y. Application of explainable artificial intelligence in the identification of Squamous Cell Carcinoma biomarkers. Comput Biol Med. 2022 Jul;146:105505. doi: 10.1016/j.compbiomed.2022.105505. Epub 2022 Apr 17. PMID: 35477047.

[12] Ducret M, Mörch CM, Karteva T, Fisher J, Schwendicke F. Artificial intelligence for sustainable oral healthcare. J Dent. 2022 Dec;127:104344. doi: 10.1016/j.jdent.2022.104344. Epub 2022 Oct 21. PMID: 36273625

[13] Figueroa KC, Song B, Sunny S, Li S, Gurushanth K, Mendonca P, Mukhia N, Patrick S, Gurudath S, Raghavan S, Imchen T, Leivon ST, Kolur T, Shetty V, Bushan V, Ramesh R, Pillai V, Wilder-Smith P, Sigamani A, Suresh A, Kuriakose MA, Birur P, Liang R. Interpretable deep learning approach for oral cancer classification using guided attention inference network. J Biomed Opt. 2022 Jan;27(1):015001. doi: 10.1117/1.JBO.27.1.015001. PMID: 35023333; PMCID: PMC8754153.

[14] 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

[15] de Chauveron J, Unger M, Lescaille G, Wendling L, Kurtz C, Rochefort J. Artificial intelligence for oral squamous cell carcinoma detection based on oral photographs: A comprehensive literature review. Cancer Med. 2024 Jan;13(1):e6822. doi: 10.1002/cam4.6822. Epub 2024 Jan 2. PMID: 38164652; PMCID: PMC10807632.

[16] A. H. A. Teo and C. P. Goh, "Oral Disease Image Detection System Using Transfer Learning," 2024 3rd International Conference on Digital Transformation and Applications (ICDXA), Kuala Lumpur, Malaysia, 2024, pp. 194-198, doi: 10.1109/ICDXA61007.2024.10470514.

[17] J. Vijaya, K. Rishabh and P. Parashar, "CanScan: Non-Invasive Techniques for Oral Cancer Detection," 2023 International Conference on Electrical, Electronics, Communication and Computers (ELEXCOM), Roorkee, India, 2023, pp. 1-5, doi: 10.1109/ELEXCOM58812.2023.10370254.

[18] Taaresh, Sakshi. , Kumar, Sunil. , Kumar, Neeraj. Neutrosophic ANFIS Machine Learning Model and Explainable AI Interpretation in Identification of Oral Cancer from Clinical Images. *International Journal of Neutrosophic Science*, vol. , no. , 2024, pp. 198-221. DOI: https://doi.org/10.54216/IJNS.240218

[19] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017, pp. 618-626, doi: 10.1109/ICCV.2017.74.

[20] Nagi R, Bibra A, Rakesh N, Patil DJ, Vyas T. Artificial intelligence-integrated optical coherence tomography for screening and early detection of oral cancer. Gen Dent. 2024 Jan-Feb;72(1):46-52. PMID: 38117641.

[21] Nguyen, Hung & Cao, Hung & Nguyen, Vo & Pham, Dinh. (2021). Evaluation of Explainable Artificial Intelligence: SHAP, LIME, and CAM

[22] Taaresh, Sakshi. Kumar, Sunil. , Kumar, Neeraj. Neutrosophic Meta SHAP and Neutrosophic Meta LIME: An Efficient Framework for Explainable AI in Oral Cancer Detection. *International Journal of Neutrosophic Science*, vol. , no. , 2024, pp. 373-245. DOI: https://doi.org/10.54216/IJNS.230328

[23] Amparore E, Perotti A, Bajardi P. To trust or not to trust an explanation: using LEAF to evaluate local linear XAI methods. PeerJ Comput Sci. 2021 Apr 16;7:e479. doi: 10.7717/peerj-cs.479. PMID: 33977131; PMCID: PMC8056245.

[24] K. Davagdorj, J. -W. Bae, V. -H. Pham, N. Theera-Umpon and K. H. Ryu, "Explainable Artificial Intelligence Based Framework for Non-Communicable Diseases Prediction," in *IEEE Access*, vol. 9, pp. 123672-123688, 2021, doi: 10.1109/ACCESS.2021.3110336.

[25] L. H. Gilpin, D. Bau, B. Z. Yuan, A. Bajwa, M. Specter and L. Kagal, "Explaining Explanations: An Overview of Interpretability of Machine Learning," *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*, Turin, Italy, 2018, pp. 80-89, doi: 10.1109/DSAA.2018.00018.

[26] https://www.mdanderson.org/content/mda/en/cancerwise/2022/12/what-does-mouth-cancer-look-like/jcr:content/blog/par/adaptiveimage.resize.648.0.medium.jpg/1671032269715.jpg

[27] https://media.licdn.com/dms/image/v2/C4D12AQG0eOeYZpAIRQ/article-cover_image-shrink_600_2000/article-cover_image-shrink_600_2000/0/1520176370451?e=2147483647&v=beta&t=icSQfzdayvLZP_3KgK_gc6Pn-od0rjFZG3GXq1FM_ds

[28] https://getyoursmileback.co.uk/wp-content/uploads/2017/08/image-result-for-oral-cancer-incidence-300x257.png

[29] https://www.kaggle.com/datasets/shivam17299/oral-cancer-lips-and-tongue-images

[30] https://www.kaggle.com/datasets/ashenafifasilkebede/dataset