



KLE Technological University
Creating Value
Leveraging Knowledge

School
of
Computer Science and Engineering

Mini Project Report
on
Histopathology for Oral Cancer Diagnosis

By:

- | | |
|---------------------------|--------------|
| 1. Shirisha K M | 01FE21BCS167 |
| 2. Vijayalaxmi Aralikatti | 01FE21BCS181 |
| 3. Srivaths Acaharya | 01FE21BCS083 |
| 4. Saisamarth Udikeri | 01FE21BCS003 |

Semester: V, 2023-2024

Under the Guidance of

**Ujwala Patil,
Ramesh Ashok Tabib.**

**K.L.E SOCIETY'S
KLE Technological University,
HUBBALLI-580031**
2023-2024



SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that project entitled “**Histopathology for Oral Cancer Diagnosis**” is a bonafide work carried out by the student team of “**Shirisha K M (01FE21BCS167), Vijayalaxmi A Aralikatti (01FE21BCS181), Srivaths Acharya (01FE21BCS083), Saisamarth Udikeri (01FE21BCS003)**”. The project report has been approved as it satisfies the requirements with respect to the mini project work prescribed by the university curriculum for BE (V Semester) in the School of Computer Science and Engineering of KLE Technological University for the academic year 2023-2024

Ujwala Patil
Ramesh Ashok Tabib
Guide

Vijayalakshmi M
Head of School

B.S.Anami
Registrar

External Viva:

Name of Examiners

Signature with date

1.

2.

ACKNOWLEDGMENT

The project Histopathology for Oral Cancer Diagnosis became successful with the kind support and help of many individuals. We the team under the Centre of Excellence in Visual Intelligence(CEVI) extend our sincere thanks to each one respectfully. The team expresses its deepest gratitude to all the people who have assisted us in the process of completion of this project. All their contributions are deeply appreciated and acknowledged. The team would like to place on record our deep sense of respect for Prof. Suneetha Budihal, Head of the Department of School of Electronics and Communication Engineering, Prof. Vijayalakshmi M. Head of the Department of School of Computer Science and Engineering for allowing us to extend our skills in the direction of this project. The team expresses heartfelt gratitude to our guide Prof. Uma Mudenagudi, Prof. Ujwala Patil, and Ramesh Ashok Tabib sir whose valuable insights proved to be vital in contributing to the success of this project.

By:

Project Team

ABSTRACT

Oral cancer, affecting areas like the mouth, throat, and oral tissues, poses a significant health threat globally. Early detection is crucial for effective treatment. The study focuses on segmenting and classifying histopathology images to enable early oral cancer identification. Our approach employs machine learning algorithms to annotate features that indicate oral cancer in input images and create accurate masks for U-Net segmentation. After segmenting the images, the Inception v3 model is used to classify them into four severity levels: normal, mild, moderate, and severe. The dataset comprises annotated images for segmentation and labeled images for classification, providing a comprehensive training and testing set. This innovative methodology aims to enhance oral cancer diagnosis through advanced image analysis, offering a valuable contribution to medical research and patient care.

Contents

1	Introduction	9
1.1	Motivation	10
1.2	Objectives	10
1.3	Literature survey	10
1.4	Problem statement	14
1.5	Application in Digital Histopathology	14
2	System design	15
2.1	Design alternatives	15
2.2	Final design	18
3	Implementation details	19
3.1	Specifications and final system architecture	19
3.2	Flowchart	28
4	Results and discussions	29
4.1	Datasets	29
4.2	Evaluation metrics	30
4.3	Experimental Setup	30
4.4	Experimental Results	31
5	Conclusions and future scope	32
5.1	Conclusion	32
5.2	Future scope	32

List of Figures

1.1	Difference between traditional and digital pathology	9
1.2	Overall architecture of DCGANOCIS	11
1.3	Block diagram representation of proposed methodology	12
1.4	Structure of the histopathological image diagnostics methodology for early diagnosis of OSCC	12
1.5	Proposed methodology of OSCC	13
2.1	Deeplab V3 Architecture.	15
2.2	GoogleNet Architecture.	16
2.3	Deeplab V3	16
2.4	Final design of Overall Architecture.Input images are annotated and segmented with U-Net before classification using Inception v3, categorizing them into four oral cancer severity levels.	18
3.1	Block diagram of overall architecture.Sequential process from feature annotation to segmentation and classification using U-Net and Inception v3 models for comprehensive analysis and categorization of oral cancer images into severity levels.	19
3.2	Basilar Hyperplasia as marked in green.	20
3.3	In the right image the loss of stratification is visible while no such thing is visible in the left image.	20
3.4	In the right image there is more darker region on the epithelium region than that in the left image.	20
3.5	In the left image there is loss of cohesion while it is not present in the right image.	20
3.6	In the left image there are cells which are not oval shaped and have unusual shape.	21
3.7	The image has a visible pink shaped structure.	21
3.8	The left image has a visible pink shaped structure which is absent in the right image.	21
3.9	Features identified to annotate the images. Tissue images are color-coded in Photoshop to aid feature visualization.	22
3.10	U-net Architectural Model.The U-Net architecture comprises an encoder-decoder structure with skip connections, enabling precise segmentation by integrating high-resolution features with context information.	22
3.11	Segmentation Module.The U-Net architecture integrates image and mask-generated inputs through an encoder-decoder structure with skip connections to achieve precise semantic segmentation by merging high-resolution features with contextual information.	23
3.12	The basic module of the inception V1 module.	24
3.13	Module 1 of inception module by reducing the parameter.	25
3.14	Module 2 Structure of Asymmetric Convolutions.	26
3.15	Module 3 after applying the optimization techniques.	26
3.16	Module 4 The above image shows how the grid size is reduced efficiently while expanding the filter banks.	27

3.17	Final Module The above image shows how the grid size is reduced efficiently while expanding the filter banks.	27
3.18	The detailed components of the Inception V3 model.	28
3.19	Complete flowchart of the overall architectural design	28
4.1	Classified images of Oral Cancer Images consists of normal, mild, moderate, severe classes.	29
4.2	Segmentation results using Unet	31

Chapter 1

Introduction

Histopathology, a crucial digital imaging technique, leverages RGB channels to offer an insightful analysis of tissue samples, particularly instrumental in the diagnosis of cancer. This method provides intricate details about cellular structures and abnormalities within the tissues, enabling pathologists to discern subtle nuances that are essential for accurate diagnosis.

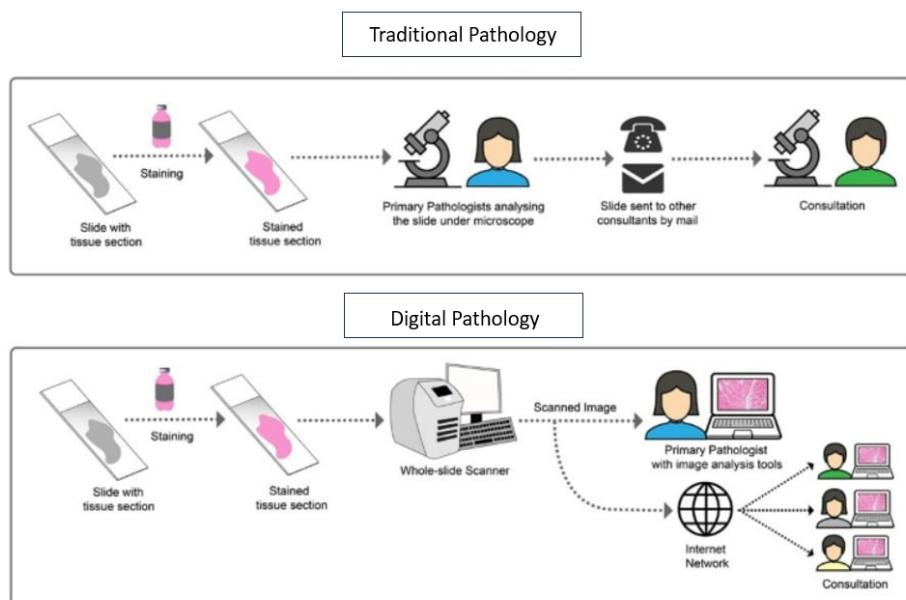


Figure 1.1: Difference between traditional and digital pathology

Conventional pathology methods have been crucial in the diagnosis, disease classification, and semi-quantitative or qualitative evaluation of protein expression. The development of digital pathology-based approaches for quantitative pathologic assessments, such as whole slide imaging and artificial intelligence (AI)-based solutions, has recently been made possible by technological advancements and the growing emphasis on precision medicine. These approaches enable us to explore and extract information beyond human visual perception. In the realm of cancer immunotherapy,

Deciphering complex pathophysiology and discovering new biomarkers and therapeutic targets have become immensely easier thanks to the deployment of such approaches in translational research and drug development. The problem for practitioners in choosing the best course of action for each patient is growing as more and more treatment choices become accessible for

any given ailment. The increasing use of AI-based methods contributes significantly to our understanding of the tumor microenvironment. Digital methods of patient selection and stratification for diagnostic assays enable the determination of the best course of treatment based on individual patient profiles. This paper addresses how developments in digital pathology have affected biomarker discovery and patient selection, and it gives an outline of the potential and constraints associated with using AI-based techniques.

Artificial intelligence (AI) and its associated subcategories, machine learning and deep learning, is of interest to pathologists because it can increase diagnosis accuracy and impartiality while reducing effort and time spent, all of which can impact the decision-making process' accuracy. Unfortunately, there are now several challenges associated with the implementation of artificial intelligence, including the suitability and verification of computational technologies and algorithms. Convolutional neural networks is one of the many machine learning algorithms that have been developed to automatically segment pathological images. Among these, fully convolutional networks and other segmentation deep learning methods are particularly notable for their accuracy, computational efficiency, and generalizability.

1.1 Motivation

Histopathology enables histopathologists to understand and detect diseases. and The need to incorporate histopathology image analysis techniques to improve the diagnostic procedure for oral cancer. The worldwide prevalence of oral cancer presents formidable obstacles, and the traditional techniques for diagnosing the disease are laborious and dependent on the subjective interpretation of humans. Our suggested approach aims to speed up the identification of malignant spots in histopathology pictures, enabling a quicker and more precise diagnosis, by automating segmentation and classification operations. The integration of machine learning algorithms endeavors to mitigate the constraints of conventional diagnostic methodologies while simultaneously alleviating the burden on pathologists. This novel strategy not only promises to enhance patient outcomes by expediting treatment decisions but also advances the more general objectives of telepathology and precision medicine, encouraging a change in the diagnosis of oral cancer.

1.2 Objectives

1. Apply machine learning algorithms to classify the given histopathological slide images based on the severity of the cancer
2. Apply machine learning algorithms to detect and localize cancerous regions in histopathological slides
3. Comparison with state-of-the-art methods on benchmark datasets

1.3 Literature survey

1. **DCGANOCIS: Convolutional Generative Adversarial Networks Based on Oral Cancer Identification System (IJISAE, July 2023) [3]**

This research offers a unique feature extraction model that combines convolutional neural networks (CNN) and modified deep convolutional generative adversarial networks (MDC-GAN) for precise oral cancer detection. The main goal is to categorize input photos of oral cavity squamous cell carcinoma (OCSCC) as either healthy or diseased. The suggested method makes use of images

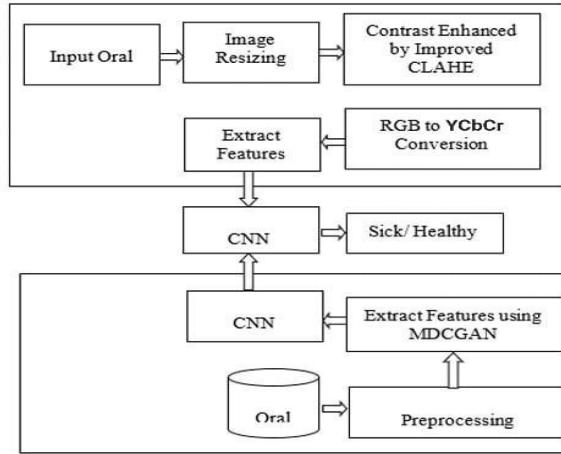


Figure 1.2: Overall architecture of DCGANOCIS

enhancement, in which the Improved CLAHE method is used to downsize, contrast enhance, and transform the input image from RGB to YCbCr color space. The primary innovation in this work is the feature extraction model, MDCGAN, which is based on deep learning and uses a different approach than conventional GANs. The Generator (G) component of the suggested MDCGAN model is used to increase the number of samples of each image in the dataset, therefore growing characteristics and raising forecast accuracy. Unlike traditional GANs, a Modified Convolutional Neural Network (MCNN) is used in place of the Discriminator (D). The results show that the suggested method performs better than previous methods, obtaining impressive testing phase results with 96.26 percentage classification accuracy, 98.96 percentage precision, 94.18 percentage recall, and 96.34 percentage f-measure. Compared to more conventional deep learning techniques, MDCGAN is a highly recommended model for image classification applications since the accuracy and quantity of features obtained from OCSCC images determine how well oral cancer will be predicted. In conclusion, the study presents a novel method for the identification of oral cancer by merging CNN for classification with MDCGAN for feature extraction. The approach outperforms current methods, highlighting the significance of the magnitude of the obtained features in attaining more precision. Since MCNN serves as the discriminator and GANs are creatively used for feature extraction, oral cancer prediction accuracy is increased, making MDCGAN a good option for these kinds of picture classification tasks.

2. An Enhanced Histopathology Analysis: An AI-Based System for Multiclass Grading of Oral Squamous Cell Carcinoma and Segmenting of Epithelial and Stromal Tissue(MDPI, April 2021) [4]

Although it is localized in an area that is visible and can be discovered fairly early, oral squamous cell carcinoma is the most common histological neoplasm of head and neck cancers. However, this usually does not occur. The conventional method for diagnosing oral cancer involves a histological examination. The primary issue with this type of procedure is tumor heterogeneity, since a subjective aspect of the test may have an immediate effect on the treatment plan for a given patient. To lessen inter- and intra-observer variability, artificial intelligence (AI) algorithms are frequently employed as a computational help in the diagnosis for tumor categorization and segmentation. This study used a two-step AI-based approach for segmenting the data and automating multiclass grading (the first

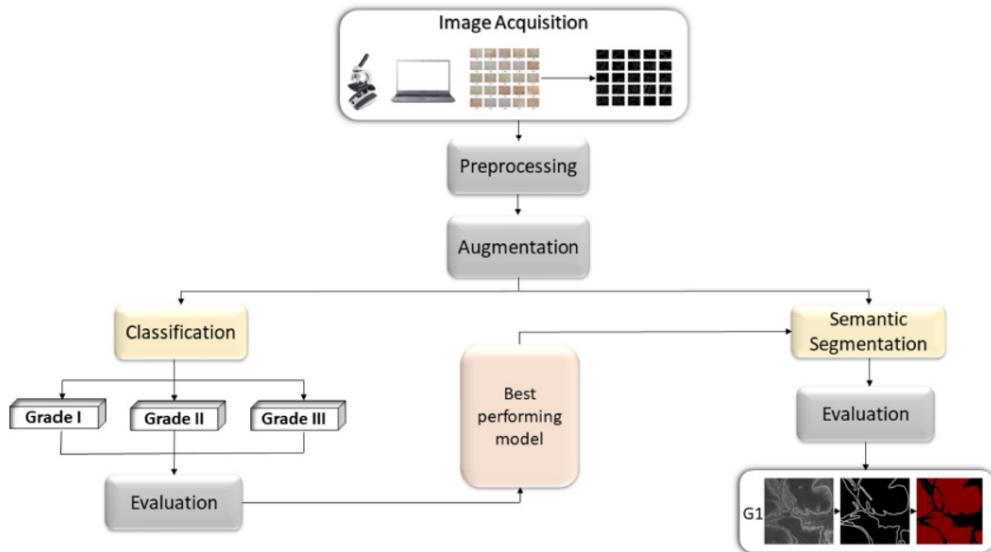


Figure 1.3: Block diagram representation of proposed methodology

stage). To help a physician diagnose oral squamous cell carcinoma, epithelial and stromal tissue (the second stage) from oral histopathology pictures is suggested. When using DeepLabv3+ with Xception65 as the backbone and data preprocessing, semantic segmentation prediction produced 0.878 ($\sigma = 0.027$) mIOU and 0.955 ($\sigma = 0.014$) F1 score.

The integration of Xception and SWT resulted in the highest classification value of 0.963 ($\sigma = 0.042$) AUCCmacro and 0.966 ($\sigma = 0.027$) AUCCmicro. The results obtained show that the suggested AI-based method has a lot of promise for OSCC diagnosis.

3. Early Diagnosis of Oral Squamous Cell Carcinoma Based on Histopathological Images Using Deep and Hybrid Learning Approaches (MDPI, August 2022)

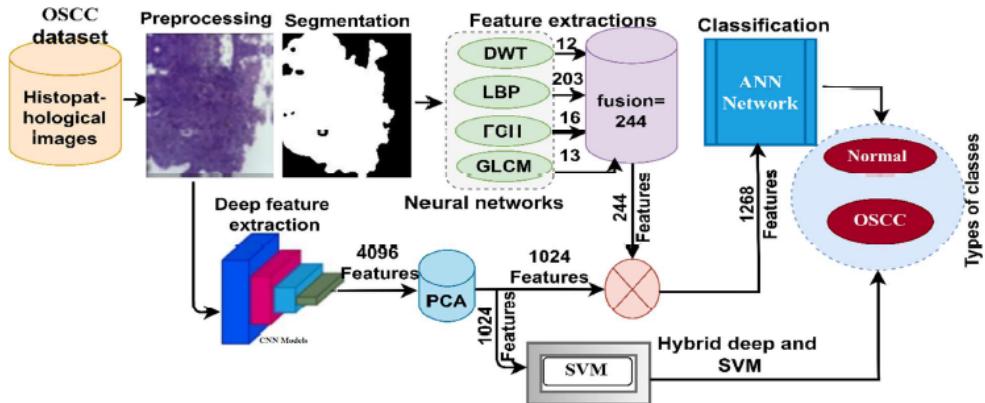


Figure 1.4: Structure of the histopathological image diagnostics methodology for early diagnosis of OSCC

Ranked as the seventh most frequent cancer, oral squamous cell carcinoma (OSCC) is one of the most prevalent kinds of head and neck cancer. Since OSCC is a histological tumor, the gold standard for diagnosis is based on histopathological scans. Tumor heterogeneity, however, makes such a diagnosis a laborious and highly efficient human experience. Thus, the use of artificial intelligence tools aids in the accurate diagnosis that professionals like doctors make. This study applied hybrid approaches based on fused features in an attempt to obtain good findings for the early diagnosis of OSCC. The first strategy is based on a hybrid approach that combines the support vector machine (SVM) algorithm with CNN models (AlexNet and ResNet-18). This approach performed better while diagnosing the OSCC data set. The second suggested approach combines color, texture, and shape features extracted via the fuzzy color histogram (FCH), discrete wavelet transform (DWT), local binary pattern (LBP), and gray-level co-occurrence matrix (GLCM) algorithms with the hybrid features extracted by CNN models (AlexNet and ResNet-18). Due to the large dimensionality of the data set features, the dimensionality was reduced using the principle component analysis (PCA) algorithm, which then sent the data to the artificial neural network (ANN) program for a potentially accurate diagnosis. In terms of histological image diagnosis of OSCC, all suggested methods performed better than expected. The ANN network based on hybrid features with AlexNet, DWT, LBP, FCH, and GLCM achieved an accuracy of 99.1 percentage, specificity of 99.61 percentage, sensitivity of 99.5 percentage, and precision of 99.71 percentage, and AUC of 99.52 percentage.

4. Multi-Method Analysis of Histopathological Image for Early Diagnosis of Oral Squamous Cell Carcinoma Using Deep Learning and Hybrid Techniques (MDPI, : 31 October 2023)[1]

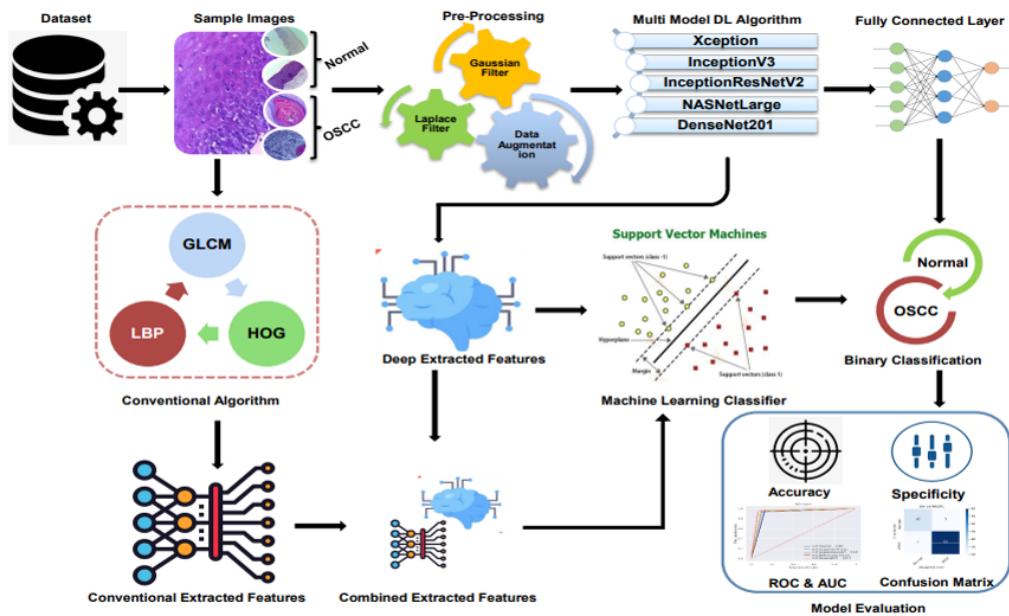


Figure 1.5: Proposed methodology of OSCC

The goal of this study was to create hybrid approaches based on fused features in order to improve outcomes for OSCC early diagnosis. Three distinct methodologies, each utilizing

five different models, were used in this investigation. Using the Xception, Inceptionv3, InceptionResNetV2, NASNetLarge, and DenseNet201 models, transfer learning is the initial tactic. The second tactic combines a Support Vector Machine (SVM) for classification with a pre-trained CNN for feature extraction. Specifically, pre-trained models such as Xception, Inceptionv3, InceptionResNetV2, NASNetLarge, and DenseNet201 were used to extract features, which were then fed into the SVM method to assess classification accuracy. The last approach uses a state-of-the-art hybrid feature fusion method to extract the deep features of the previously described models by using an art-of-CNN model. These profound features become more dimensional by principal component analysis (PCA) reduction. The form, color, and texture characteristics that were extracted using the gray-level co-occurrence matrix (GLCM), Histogram of Oriented Gradient (HOG), and Local Binary Pattern (LBP) techniques are then integrated with the low-dimensionality features. The SVM was modified to include hybrid feature fusion in order to improve classification performance. The suggested approach produced encouraging outcomes for the quick diagnosis of OSCC from histological pictures. Based on the hybrid feature fusion of DenseNet201 with GLCM, HOG, and LBP features, the support vector machine (SVM) algorithm's accuracy, precision, sensitivity, specificity, F-1 score, and area under the curve (AUC) were 97.00 percentage, 96.77 percentage, 90.90 percentage, 98.92 percentage, 93.74 percentage, and 96.80 percentage respectively.

1.4 Problem statement

Histopathology image analysis for Oral cancer detection using segmentation and classification for improved and faster diagnosis.

1.5 Application in Digital Histopathology

Digital histopathology, when integrated with advanced technologies such as machine learning, revolutionizes pathology by automating tasks and enhancing diagnostic capabilities.

1. It enables accurate tumor detection and grading, aiding pathologists in cancer diagnosis and prognosis. Furthermore, the technology facilitates the identification of predictive biomarkers, contributing to personalized medicine and targeted therapies.
2. ML algorithms enable tissue segmentation in histopathology images, facilitating the analysis of specific regions, while automation of cellular annotation streamlines the identification and annotation of cells, ensuring consistent and efficient analysis processes.
3. It provides tissue data by analyzing images of histology slides using image analysis and machine learning.
4. Digital histopathology assists in drug development by assessing toxicity and offers automated pathology reporting, streamlining workflows, and improving report accuracy.
5. Additionally, it supports education through annotated images, standardizes diagnostic criteria and contributes to research by exploring large datasets for novel patterns and biomarkers, thereby advancing our understanding of diseases.

Chapter 2

System design

2.1 Design alternatives

1. Alternative design using VGG Net and Deeplab V3+ [2]

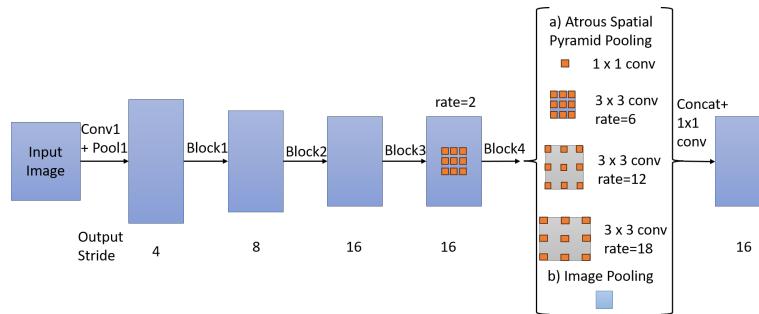


Figure 2.1: Deeplab V3 Architecture.

VGG Net, or VGG-18, is a deep convolutional neural network architecture noted for its simplicity and efficacy in image classification applications. VGG Net, with its 18 layers, has been widely used for numerous classification tasks, exhibiting its capacity to learn hierarchical features from input photos. Deeplab V3+, a semantic segmentation model, performs well on pixel-level segmentation tasks. Deeplab V3+ successfully captures contextual information with dilated convolutions and an atrous spatial pyramid pooling module, allowing for reliable object segmentation in images. While VGG Net is used for image classification, Deeplab V3+ is useful for jobs that need accurate segmentation, such as recognizing and distinguishing objects within an image at a finer level.

2. Alternative design using GoogleNet and MDCGAN [5]

GoogleNet (also known as Inception) is a deep convolutional neural network architecture created by Google. The design is distinguished by its distinctive "inception modules," which use filters of varying sizes inside the same layer to collect features at several scales. This method enables GoogleNet to effectively learn a wide range of features from input images, making it a popular choice for high-accuracy image classification jobs, particularly in later versions such as Inception-v3. Concurrently, MDCGAN (Multi-Discriminator Conditional Generative Adversarial Network) is a GAN-based segmentation model. MDCGAN, which is optimized for image segmentation, excels in producing detailed and realistic segmented images. Conditional generative adversarial techniques involve conditioning the

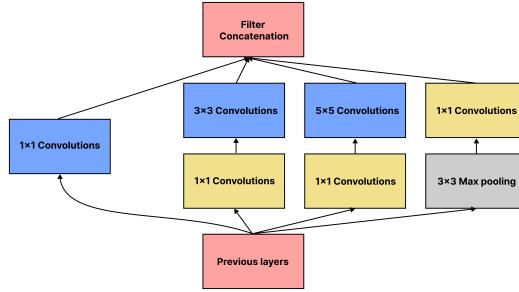


Figure 2.2: GoogleNet Architecture.

generator on both input images and segmentation masks, which allows for exact object delineation. GoogleNet is extensively used.

3. Alternative design using BN-Inception and HR-Net

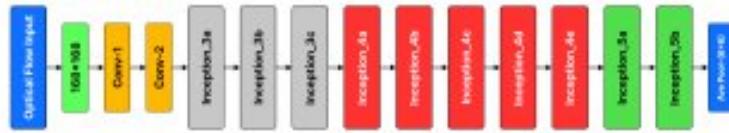


Figure 2.3: Deeplab V3

BN-Inception, a convolutional neural network design, has found popularity in image classification problems. BN-Inception, which incorporates batch normalization (BN) layers, enhances training efficiency and convergence by normalizing intermediate feature maps. It employs inception modules, combining filters of varying widths within the same layer to collect a wide range of information. HRNet (High-Resolution Network), on the other hand, specializes in picture segmentation and excels at jobs that need detailed spatial information. HRNet maintains high-resolution representations throughout the network, maintaining fine-grained characteristics that are critical for accurate segmentation. While BN-Inception is chosen for its classification efficacy, HRNet is useful in segmentation applications, demonstrating the complementary nature of both architectures in handling different elements of computer vision problems.

Final architectures

UNet for segmentation and Inception V3 for classification are considered to be superior in certain aspects to the approaches discussed above. UNet is well-known for segmentation jobs because of its unusual architecture, which includes a contracting path to collect context and a symmetric expansive path for precise localization. The skip connections help to preserve small features, making UNet particularly useful in medical image segmentation and other tasks that need accurate object delineation.

In classification, Inception V3 performs better in picture classification tasks than not just BN-Inception but also GoogleNet and VGG-18. The advanced inception module design of Inception

V3 is what makes it so good at capturing features at different scales. Its incorporation of factorized convolutions further sets it apart by decreasing computational complexity and parameter counts while increasing efficiency. Inception V3 has shown state-of-the-art performance across multiple picture classification benchmarks when compared to GoogleNet and VGG-18, demonstrating its adaptability in handling a variety of datasets. Inception V3 is a popular option for tasks demanding precise and effective image classification because of its creative architectural choices, which also add to its success.

Both UNet and Inception V3 are well-liked options for a variety of computer vision applications because of their creative architectures and advanced capabilities, which have been demonstrated to make them superior in their respective domains—segmentation and classification.

2.2 Final design

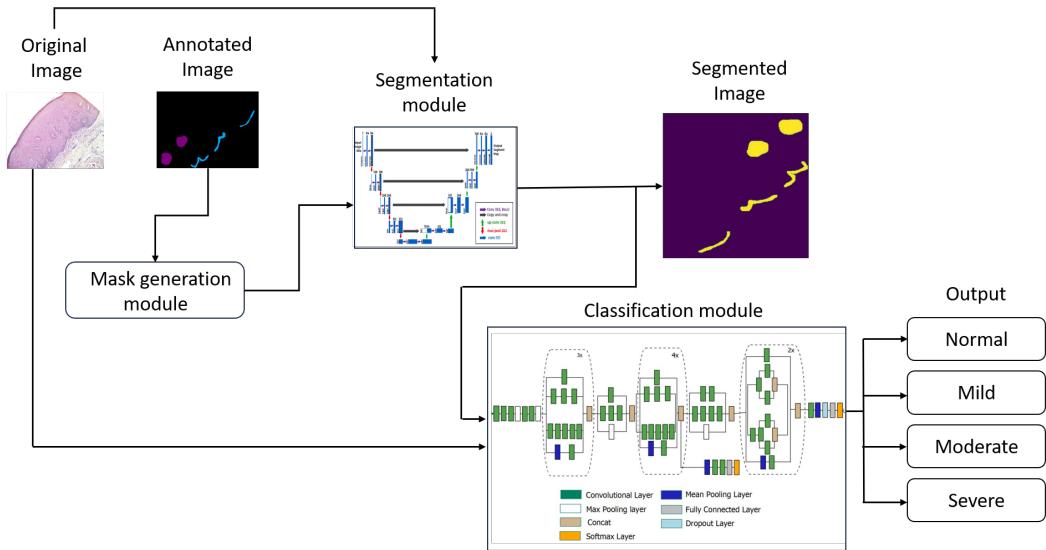


Figure 2.4: Final design of Overall Architecture. Input images are annotated and segmented with U-Net before classification using Inception v3, categorizing them into four oral cancer severity levels.

In the initial step, input images are annotated based on oral cancer features, creating mask-image files. Subsequently, both the original image files and their corresponding mask-image files undergo segmentation using the U-Net architecture. The segmentation module partitions the images into distinct regions. The segmented images then progress to the classification module, which employs the Inception v3 model. Through multi-class classification, the module assigns each segmented image to one of four categories: mild, moderate, normal, or severe. This sequential process ensures a comprehensive analysis, from feature annotation to segmentation and classification, ultimately resulting in the categorization of images according to their oral cancer severity levels.

Chapter 3

Implementation details

3.1 Specifications and final system architecture

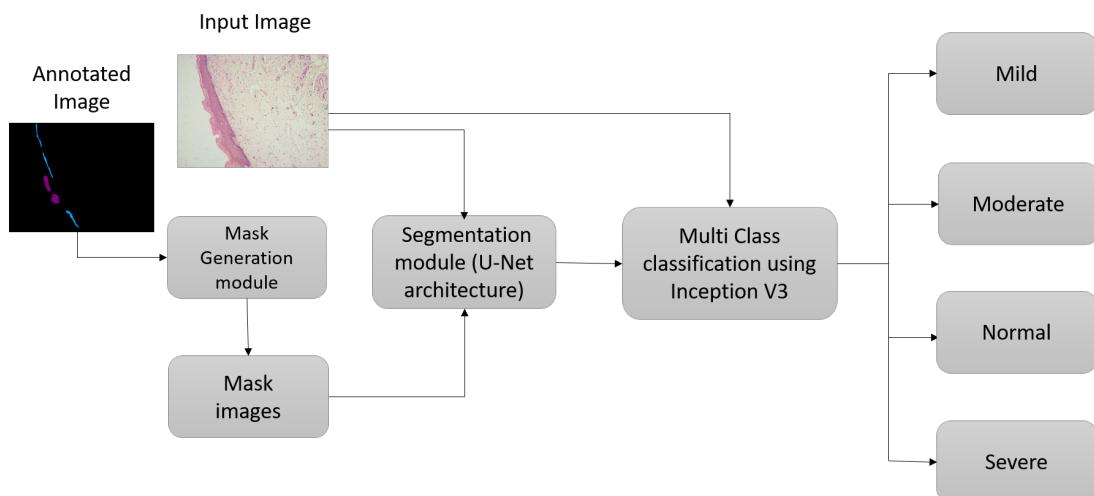


Figure 3.1: Block diagram of overall architecture. Sequential process from feature annotation to segmentation and classification using U-Net and Inception v3 models for comprehensive analysis and categorization of oral cancer images into severity levels.

1. Mask Generation Module

Oral cancer refers to a type of cancer that develops in the tissues of the mouth or throat. It can affect various areas, including the lips, tongue, gums, cheeks, and the roof or floor of the mouth.

Features to Identify Oral Cancer

These are the following features to identify cancerous tissue in a particular image

- Basilar Hyperplasia:** There will be an increase in the number of cells at the boundary of the epithelium and connective tissue. Commonly affected classes: Mild and Moderate.

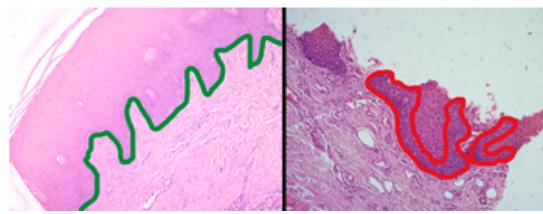


Figure 3.2: Basilar Hyperplasia as marked in green.

- (b) **Loss of Stratification:** Proper strata or layers will be absent of the arrangement of cells. Commonly affected classes: Mild, Moderate, and Severe.

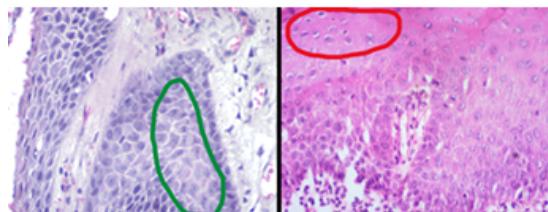


Figure 3.3: In the right image the loss of stratification is visible while no such thing is visible in the left image.

- (c) **Hyperchromatism:** There will be nuclei of certain cells staining more deeply than normal. Commonly affected classes: Mild, Moderate, and Severe.

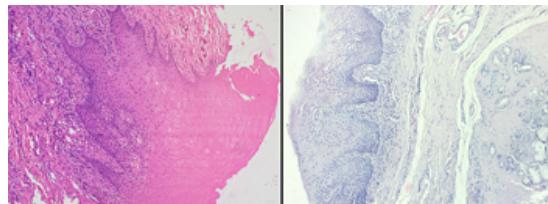


Figure 3.4: In the right image there is more darker region on the epithelium region than that in the left image.

- (d) **Loss of Cellular Cohesion:** There will be tissues having uneven and non-uniform spacing between them. Commonly affected classes: Moderate and Severe.

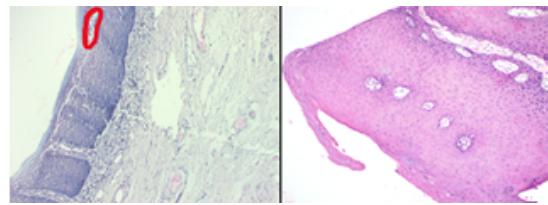


Figure 3.5: In the left image there is loss of cohesion while it is not present in the right image.

- (e) **Pleomorphism:** There will be irregular and variant forms of cells. Commonly affected classes: Moderate and Severe.

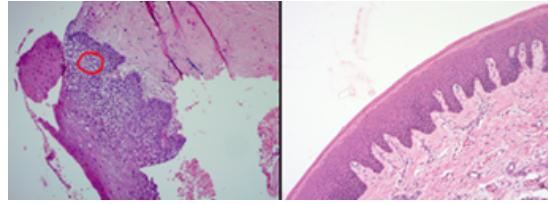


Figure 3.6: In the left image there are cells which are not oval shaped and have unusual shape.

- (f) **Keratin Pearl:** There will be a keratinized structure found in regions where abnormal squamous cells form concentric layers in the epithelium layer. Commonly affected classes: Severe.

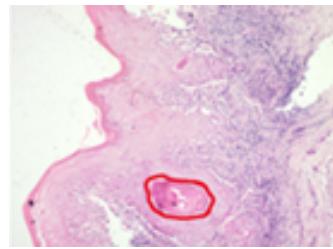


Figure 3.7: The image has a visible pink shaped structure.

- (g) **Drop-Shaped Structures:** There will be an abnormal shape of the epithelium-connective tissue boundary that resembles a drop. Commonly affected classes: Moderate and Severe.

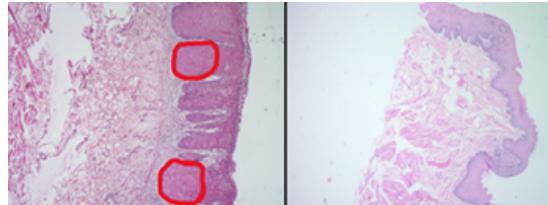


Figure 3.8: The left image has a visible pink shaped structure which is absent in the right image.

We use these features to annotate 10x tissue images with different RGB colors assigned to each feature. To do this, we used the Photoshop tool to annotate the original image. This involves placing a black mask over the image, along with masks highlighting specific features identified in that particular image. This approach helps us visually distinguish the features. These annotated images will be given as input for the mask generation. The color representation of the mask are converted from BGR to RGB format. Then we categorize pixel values in the mask into nine classes(include one negative class) based on specific RGB color values. For example, pixels with the RGB values (255, 0, 255) are assigned class 1, and similar categorizations are done for other color combinations. Then we create a new binary mask, where each pixel is assigned an integer value corresponding to its class. Finally, the new masks are created.

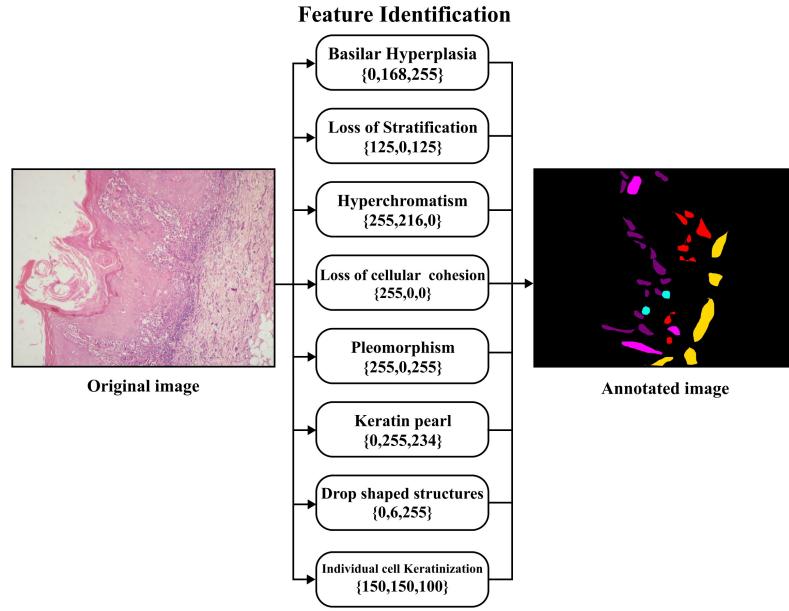


Figure 3.9: Features identified to annotate the images. Tissue images are color-coded in Photoshop to aid feature visualization.

2. Segmentation Module(U-net Architecture)

In the segmentation module, we are using U-Net architecture, a well-liked neural network structure for computer vision problems including semantic segmentation. The encoder-decoder architecture of the U-Net design is distinctive; skip links are included to enable precise image segmentation.

The U-Net architecture is made up of various essential parts. The DoubleConvolution class describes a fundamental building block that consists of two convolutional layers that are followed by rectified linear unit (ReLU) activation and batch normalization (BN) in order of sequence. The network's encoding and decoding parts both use this block.

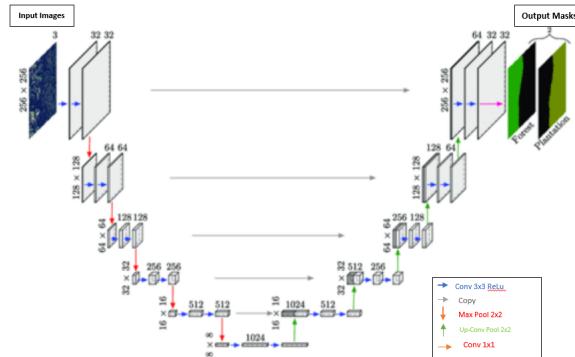


Figure 3.10: U-net Architectural Model. The U-Net architecture comprises an encoder-decoder structure with skip connections, enabling precise segmentation by integrating high-resolution features with context information.

The U-Net's encoding portion is represented by the Down class, which carries out downscaling procedures. To extract features and minimize the spatial dimensions of the input, it uses max pooling and a DoubleConvolution block. On the other side, the U-Net's decoding is handled by the Up class. Either bilinear upsampling or transposed convolutions along with DoubleConvolution blocks are used to accomplish upscaling. To help preserve spatial information, skip connections are integrated to mix upsampled data with matching features from the encoder. Using a 1x1 convolution to limit the number of channels to the required number of segmentation classes, the OutConvolution class defines the final output layer.

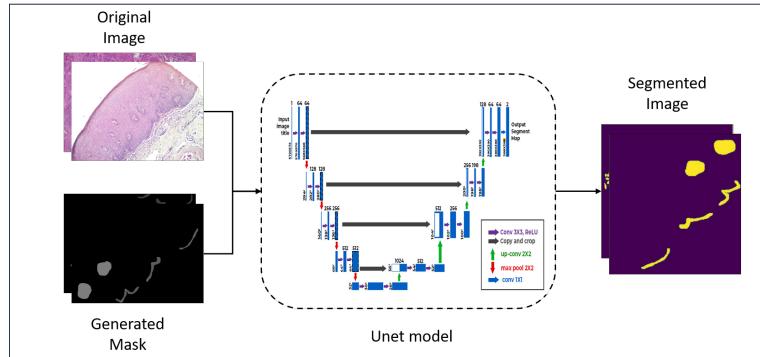


Figure 3.11: Segmentation Module. The U-Net architecture integrates image and mask-generated inputs through an encoder-decoder structure with skip connections to achieve precise semantic segmentation by merging high-resolution features with contextual information.

The entire U-Net architecture is produced by integrating these elements through the UNet class. It consists of a decoder with four Up blocks (Up1 to Up4) and an encoder with four Down blocks (Down1 to Down4). The OutConvolution layer is used to obtain the final output. Stochastic gradient descent (SGD) with a learning rate of 0.02 and momentum of 0.5 is used by the algorithm as the optimizer, while the Dice loss function is used for segmentation. For a predetermined number of epochs, the training loop iterates through the dataset, computing the model output, figuring out the Dice loss, and using backpropagation to update the model parameters.

In the testing part of the algorithm, input images, model predictions, and ground truth labels are visualized after processing photos from a test loader and obtaining the encoder's output. The matplotlib library is used to visualize the data, demonstrating how well the U-Net architecture performs in semantic segmentation tasks.

3. Classification Module (Inception V3)

The advanced deep learning model architecture known as InceptionV3 was mainly created for image classification applications. The "V3" stands for the third version that Google developed, and it shows improvements in terms of both performance and efficiency. The architecture makes use of Inception blocks, which combine various convolutional layers with different filter sizes and pooling strategies to simultaneously capture features at multiple scales. Notably, InceptionV3 effectively manages computational complexity by employing factorization techniques like 1x1 convolutions. Specifically, we use models. Inception v3(pretrained=True) in our code to obtain pre-trained weights, an important step that

makes use of information from large-scale datasets such as ImageNet to improve model performance. You adjust the fully connected layers for your custom classification task, making the network output fit the desired number of classes. Furthermore, to accommodate input images of different sizes, the adaptive average pooling layer dynamically modifies spatial dimensions to 1x1. Using the Adam optimizer and CrossEntropyLoss as the classification loss function, the model is trained over several epochs. Utility functions such as val loss and accuracy evaluate the model's performance on training and validation datasets during training. All things considered, InceptionV3 is a strong architecture that shows flexibility in how it can be applied to different tasks and effectively used in your scenario of multi-class classification.

Importance of Inception V3

Convolutional neural networks serve as the foundation for the Inception V3, a deep learning model for image classification. The Inception V3 unveiled as GoogLeNet in 2014, is an improved iteration of the foundational model Inception V1. Released in 2015, the Inception v3 model boasts 42 layers overall and a lower error rate than its predecessors. Let's examine the various optimizations that improve the Inception V3 model. The principal adjustments made to the Inception V3 model are

- (a) Factorization into Smaller Convolutions
 - (b) Spatial Factorization into Asymmetric Convolutions
 - (c) Utility of Auxiliary Classifiers
 - (d) Efficient Grid Size Reduction
- (a) Factorization into Smaller Convolutions

The generous dimension reduction of the Inception V1 model was one of its main advantages. The model's larger convolutions were factorized into smaller convolutions to further improve it.

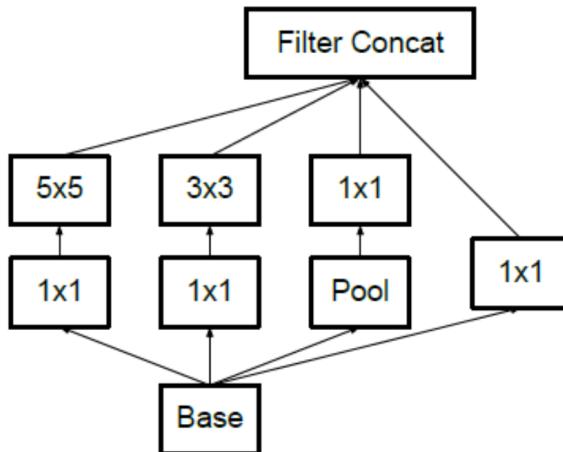


Figure 3.12: The basic module of the inception V1 module.

As previously mentioned, its 5x5 convolutional layer was computationally costly. Hence, as can be seen below, two 3x3 convolutional layers were used in place of the 5x5 convolutional layer to lower the computational cost.

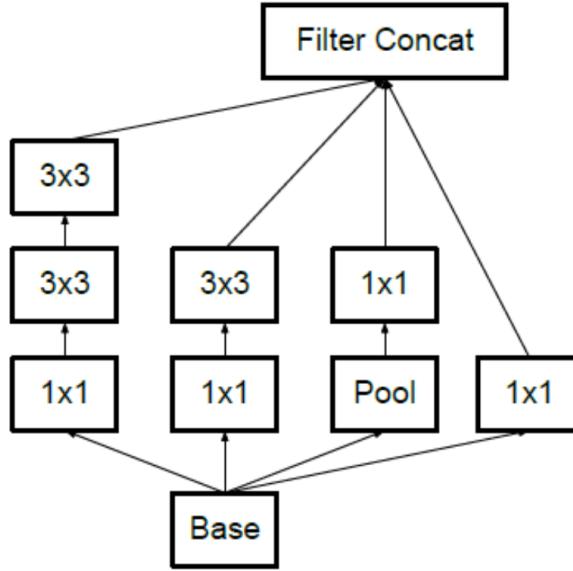


Figure 3.13: Module 1 of inception module by reducing the parameter.

See how the number of parameters is reduced by employing two 3×3 convolutions to gain a better understanding of it. The decrease in the number of parameters also leads to a decrease in computational costs. A 28 percent relative gain was obtained from factorizing larger convolutions into smaller convolutions.

(b) Spatial Factorization into Asymmetric Convolutions

Even though smaller convolutions are produced by factorizing the larger convolutions. You might be wondering what would happen if we could factorize even more, like to a 2×2 convolution. However, asymmetric convolutions were a superior substitute to increase the model's efficiency. The form of asymmetric convolutions is $n \times 1$. Consequently, they substituted the 1×3 and then the 3×1 convolutions for the 3×3 convolutions. Sliding a two-layer network with the same receptive field as in a 3×3 convolution is equivalent to doing this.

Assuming an equal number of input and output filters, the two-layer solution comes in at 33 percent less cost for the same number of filters. The inception module appears as follows following the application of the first two optimization techniques.

(c) Utility of Auxiliary classifiers

Using an auxiliary classifier aims to increase very deep neural networks' convergence. In very deep networks, the auxiliary classifier is primarily employed to counter the vanishing gradient issue. Early in the training process, there was no improvement from the auxiliary classifiers. However, in the end, the network with the auxiliary classifiers outperformed the network without them in terms of accuracy. The Inception V3 model architecture uses the auxiliary classifiers as a regularizer as a result.

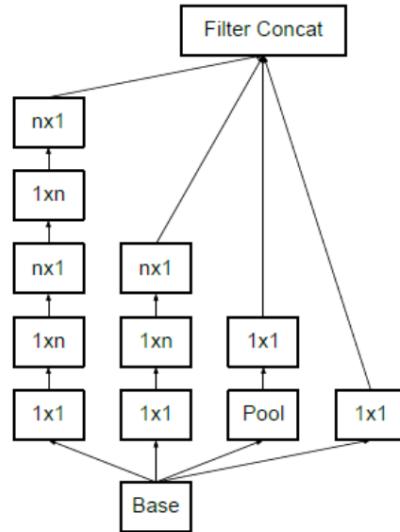


Figure 3.14: Module 2 Structure of Asymmetric Convolutions.

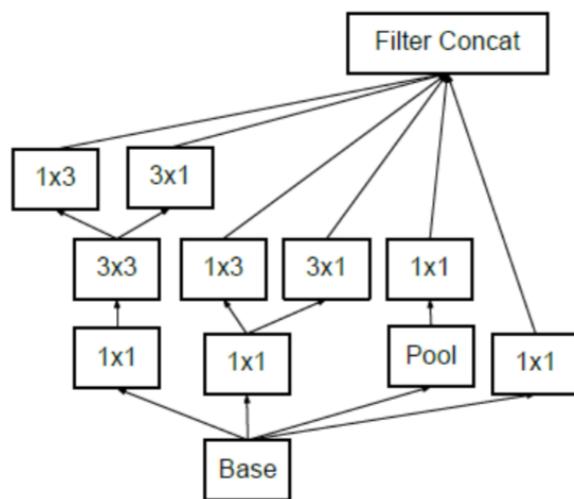


Figure 3.15: Module 3 after applying the optimization techniques.

(d) Efficient Grid Size Reduction

Traditionally, the feature maps' grid sizes were decreased by using max pooling and average pooling. The activation dimension of the network filters is expanded in the Inception V3 model to reduce the grid size efficiently. For instance, reduction yields a $d/2 \times d/2$ grid with $2k$ filters from a $d \times d$ grid with k filters. To accomplish this, two parallel blocks of pooling and convolution are later concatenated.

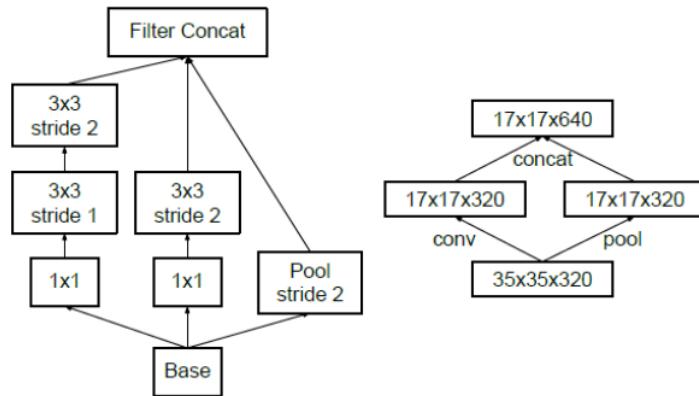


Figure 3.16: Module 4 The above image shows how the grid size is reduced efficiently while expanding the filter banks.

The final Inception V3 model After performing all the optimizations the final Inception V3 model looks like this

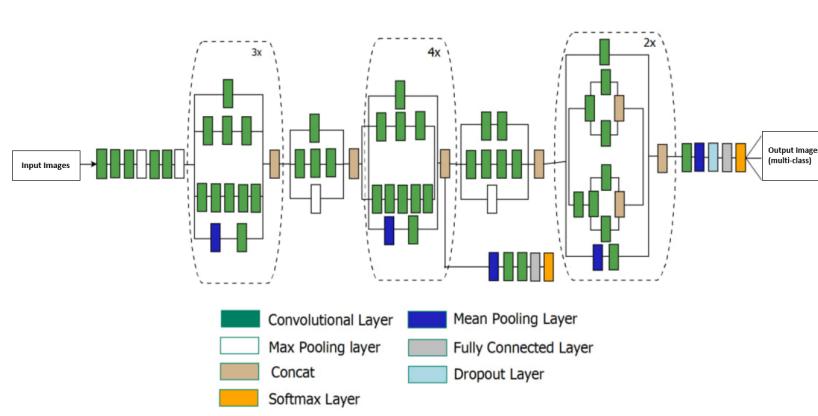


Figure 3.17: Final Module The above image shows how the grid size is reduced efficiently while expanding the filter banks.

The inception V3 model has 42 layers overall, which is a few more than the inception V1 and V2 models had. However, this model's efficiency is incredibly remarkable. We'll discuss it shortly, but for now, let's take a closer look at the parts that make up the Inception V3 model.

The above table describes the outline of the inception V3 model. Here, the output

TYPE	PATCH / STRIDE SIZE	INPUT SIZE
Conv	3 x 3/2	299 x 299 x 3
Conv	3 x 3/1	149 x 149 x 32
Conv padded	3 x 3/1	147 x 147 x 32
Pool	3 x 3/2	147 x 147 x 64
Conv	3 x 3/1	73 x 73 x 64
Conv	3 x 3/2	71 x 71 x 80
Conv	3 x 3/1	35 x 35 x 192
3 x Inception	Module 1	35 x 35 x 288
5 x Inception	Module 2	17 x 17 x 768
2 x Inception	Module 3	8 x 8 x 1280
Pool	8 x 8	8 x 8 x 2048
Linear	Logits	1 x 1 x 2048
SoftMax	Classifier	1 x 1 x 1000

Figure 3.18: The detailed components of the Inception V3 model.

size of each module is the input size of the next module.

3.2 Flowchart

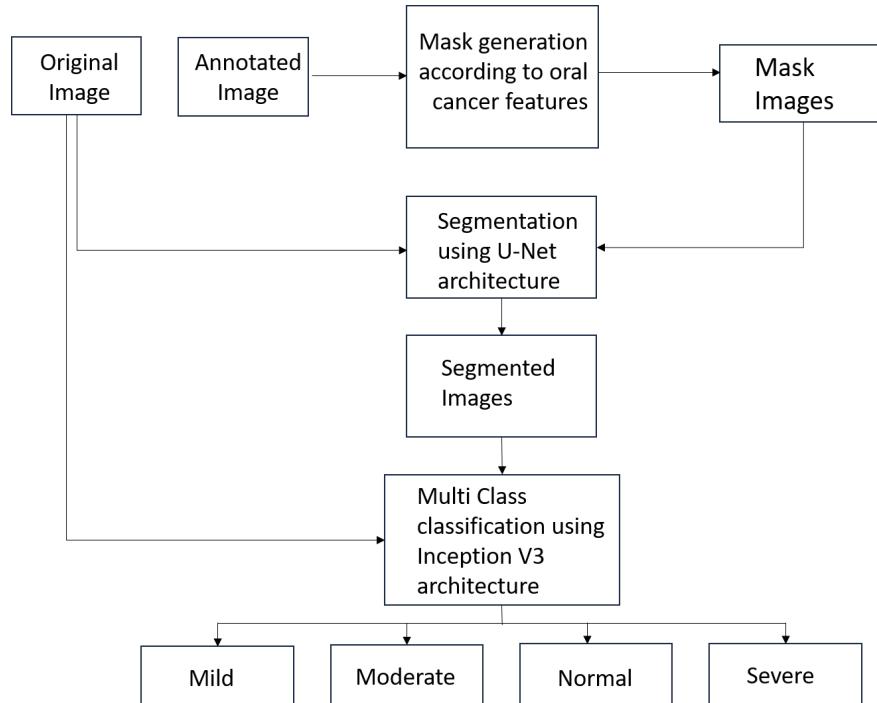


Figure 3.19: Complete flowchart of the overall architectural design .

Chapter 4

Results and discussions

4.1 Datasets

The dataset comprises 10x magnification images of oral cancer tissue. For the segmentation task, the dataset is annotated with pixel-level labels, distinguishing between different tissue classes such as Normal, Mild, Moderate, and Severe. For the classification task, the dataset is divided into training and testing sets, with corresponding counts for each class. In the classification dataset, each image is labeled with one of the classes (Normal, Mild, Moderate, or Severe), providing a comprehensive dataset for training and evaluating models.

Classes	Segmentation	Classification	
	Count	Train Count	Test Count
Normal	127	164	37
Mild	120	152	37
Moderate	123	107	39
Severe	48	166	37

Table 4.1: Table shows the Dataset (Number of images) per class

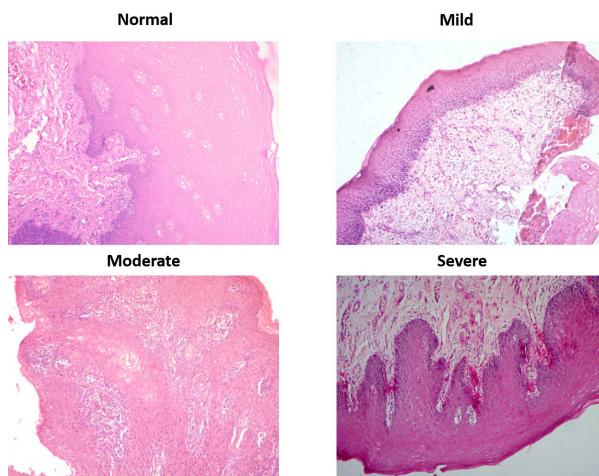


Figure 4.1: Classified images of Oral Cancer Images consists of normal, mild, moderate, severe classes.

4.2 Evaluation metrics

- Accuracy: We evaluate the results generated by the proposed architecture considering accuracy evaluation metric. The number of classifications a model successfully predicts divided by the total number of predictions is known as model accuracy. The easiest way to see how well a model performs on a particular dataset is to look at its accuracy. The better a model's ability to generalise to 'unknown' data, the better predictions and insights it can generate.

$$Accuracy = \frac{\text{Number of predicted images}}{\text{Total number of images}}$$

- Pixel Accuracy: is a semantic segmentation metric that measures the proportion of correctly identified pixels in the image. This measure determines the ratio of pixels in the image that are correctly identified to all of the pixels in the image

$$\frac{\sum_j^k = 1^n jj}{\sum_j^k = 1^t j}$$

- Dice Score: A reproducibility validation metric and an index of spatial overlap are both used by Dice Similarity Coefficient. A DSC's value ranges from 0, which denotes entire spatial overlap between two sets of data from binary segmentation, to 1, which denotes total spatial overlap. To validate the segmentation, the Dice Similarity Coefficient was used. It is used for pixel wise comparison between the segmented mask and the ground truth.

$$Dicescore = \frac{(2 \times \text{area of overlap})}{\text{Total area}}$$

4.3 Experimental Setup

1. Data Preprocessing and Mask Generation:

Annotated images are loaded, and masks are generated based on color-coded annotations. Masks are created for different classes (e.g., severe, normal, mild, moderate) by mapping specific RGB values to corresponding class labels.

2. UNet Model Training:

UNet is initialized and trained for 100 epochs. Training samples consist of pairs of original images and generated segmentation masks. The Adam optimizer is used to optimize the UNet model with a momentum of 0.5 and a learning rate of 0.02. Cross-Entropy Loss is the loss function that is applied. The loss function used is Cross-Entropy Loss.

3. InceptionV3 Model Training:

InceptionV3 is initialized with a modified final fully connected layer for classification. The model is trained for 50 epochs using a custom dataset created from the preprocessed original

images and segmentation masks. A learning rate of 1e-4 is applied when using the Adam optimizer. Cross-Entropy Loss is employed as the loss function for classification.

4.Experimental Configuration:

The experiments are conducted on a system with GPU assistance (CUDA-enabled) for quicker model training. Every epoch, the training process is observed and relevant metrics including accuracy, learning rate, training and validation loss are recorded. Data is split into training and testing sets, with a subset reserved for validation during training.

4.4 Experimental Results

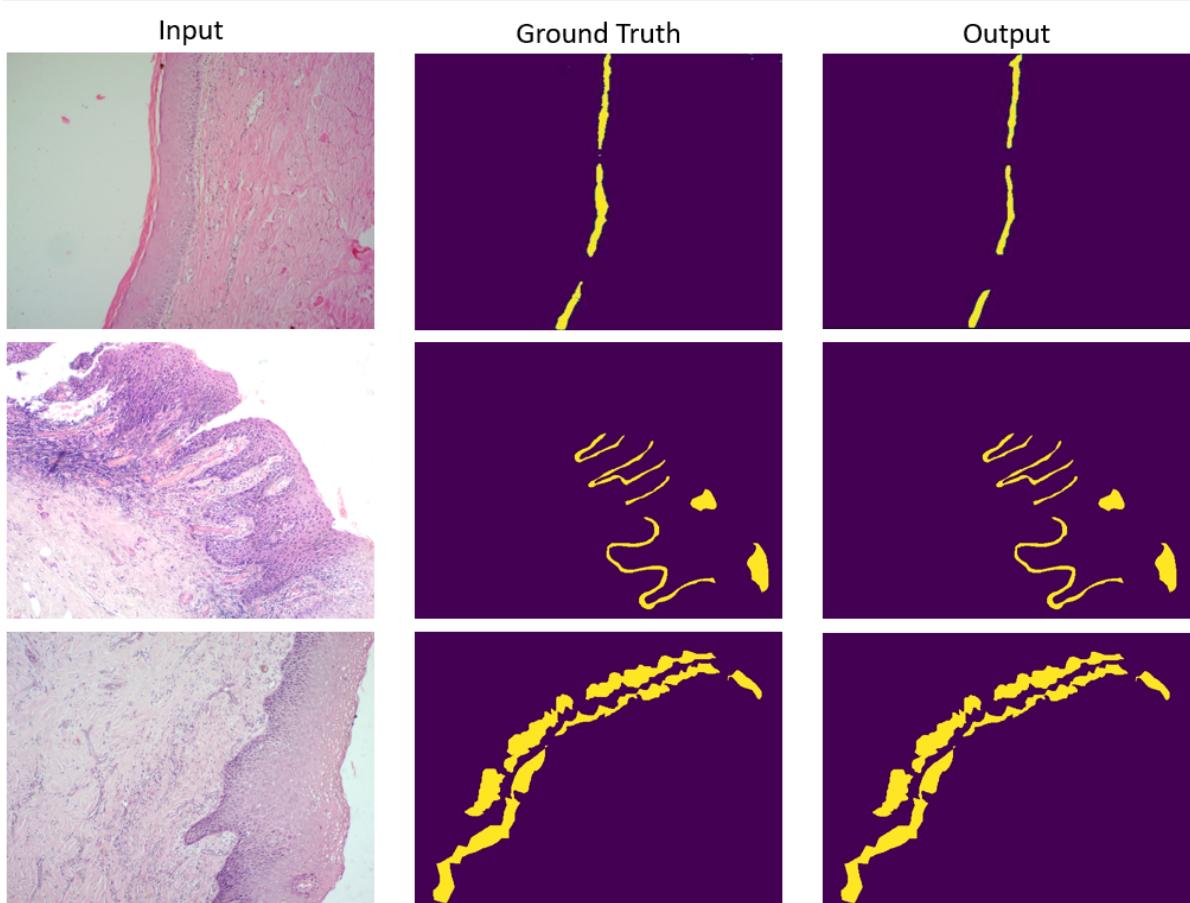


Figure 4.2: Segmentation results using Unet

Evaluation Metrics	Value
Validation Loss	1.25
Validation Accuracy	56.25%

Table 4.2: Classification results using Inception-v3

Chapter 5

Conclusions and future scope

5.1 Conclusion

After the images were annotated and subsequently validated by a medical expert, we implemented a U-Net model for segmentation. The annotated images were utilized to generate masks, and the original images, along with their corresponding masks, were fed into the U-Net model for training. The segmentation model yielded the outcomes shown in Figure 4.2 after 100 epochs. In the evaluation of our oral cancer image classification using the Inception V3 model, we obtained a validation loss of 1.25 after 100 epochs. Additionally, the validation accuracy reached 56.25%, indicating a moderate level of performance. It is crucial to note that these results were achieved without incorporating segmented features into the model. Further iterations of our research could explore the integration of segmented features to enhance the model's sensitivity and improve its performance in distinguishing between mild, normal, and moderate-severe oral cancer cases.

5.2 Future scope

There is a lot of promise for early detection and improved diagnostic accuracy with segmentation classification in the RGB space for oral cancer diagnosis in the future. Improved segmentation techniques will probably be seen in the further development of computer-aided diagnosis (CAD) expert systems, allowing for more accurate delineation of oral tissue regions in RGB images. Subsequent advancements could concentrate on utilizing deep learning algorithms for segmentation, which would enable the extraction of complex patterns and nuanced characteristics suggestive of early-stage oral cancer. Adding cutting-edge imaging technologies, like hyperspectral or multispectral imaging, could improve segmentation accuracy even more.

The validation of segmentation algorithms' efficacy will depend on the creation of strong performance evaluation metrics tailored to oral cancer. It may be possible for feature extraction techniques to advance in order to gather a wider variety of discriminative data, which would help in the characterization of various cancer stages. It is anticipated that the interaction of feature selection and image pre-processing methods will maximize the quality and relevance of features utilized in the classification procedure. Working together, pathologists and computer scientists will be essential to improving and confirming segmentation models for practical clinical use.

The use of segmentation classification tools in the medical field will require careful consideration of ethical issues, such as patient privacy and openness. All things considered, segmentation classification has bright futures ahead of it, establishing it as a crucial element in the development of computer-aided diagnosis systems for the early detection of oral cancer.

Bibliography

- [1] Mehran Ahmad, Muhammad Abeer Irfan, Umar Sadique, Ihtisham ul Haq, Atif Jan, Muhammad Irfan Khattak, Yazeed Yasin Ghadi, and Hanan Aljuaid. Multi-method analysis of histopathological image for early diagnosis of oral squamous cell carcinoma using deep learning and hybrid techniques. *Cancers*, 15(21):5247, 2023.
- [2] Xiaodong Chen, Qiongyu Duan, Rong Wu, and Zehui Yang. Segmentation of lung computed tomography images based on segnet in the diagnosis of lung cancer. *Journal of Radiation Research and Applied Sciences*, 14(1):396–403, 2021.
- [3] R Dharani, S Revathy, K Danesh, R Deeptha, and S Preethi Parameswari. Degancis: Convolutional generative adversarial networks based on oral cancer identification system. *International Journal of Intelligent Systems and Applications in Engineering*, 11(3):673–679, 2023.
- [4] Jelena Musulin, Daniel Štifanić, Ana Zulijani, Tomislav Čabov, Andrea Dekanić, and Zlatan Car. An enhanced histopathology analysis: An ai-based system for multiclass grading of oral squamous cell carcinoma and segmenting of epithelial and stromal tissue. *Cancers*, 13(8):1784, 2021.
- [5] R Pandian, V Vedanarayanan, DNS Ravi Kumar, and R Rajakumar. Detection and classification of lung cancer using cnn and google net. *Measurement: Sensors*, 24:100588, 2022.