`

A SYSTEMATIC STUDY OF HISTOPATHOLOGICAL ORAL CANCER DATA CLASSIFICATION USING MACHINE LEARNING & DEEP LEARNING APPROACHES

A Project Report

Submitted by:

Ashutosh Mohapatra (2041013203)

Ayush Das Pattanaik (2041016183)

Swaraj Das (2041004013)

SK Abdul Rahman (2041018092)

in partial fulfillment for the award of the degree

of

BACHELOR OF TECHONOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Faculty of Engineering and Technology, Institute of Technical Education and Research

SIKSHA 'O' ANUSANDHAN (DEEMED TO BE) UNIVERSITY

Bhubaneswar, Odisha, India

(June 2024)

`

CERTIFICATE

This is to certify that the project report titled "A Systematic Study of Histopathological Oral Cancer Data Classification using Machine Learning & Deep Learning Approaches" being submitted Ashutosh Mohapatra, Ayush Das Pattnaik, Swaraj Das, SK Abdul Rahman of (Sec-'B') to the Institute of Technical Education and Research, Siksha 'O' Anusandhan (Deemed to be) University, Bhubaneswar for the partial fulfillment for the degree of Bachelor of Technology in Computer Science and Engineering is a record of original confide work carried out by them under my/our supervision and guidance. The project work, in my/our opinion, has reached the requisite standard fulfilling the requirements for the degree of Bachelor of Technology.

The results contained in this project work have not been submitted in part or full to any other University or Institute for the award of any degree or diploma.

(Name and signature of the Project Supervisor)

Department of Computer Science and Engineering

Faculty of Engineering and Technology;
Institute of Technical Education and Research;
Siksha 'O' Anusandhan (Deemed to be) University

`

ACKNOWLEDGEMENT

First and foremost, we would like to thank our project supervisor Dr. Rasmita Dash, project coordinator Dr. Mitrabinda Ray for their guidance, patience, and invaluable support. Their expertise and dedication have been instrumental in shaping the direction of this project. We are also grateful to our professors and faculty members who have imparted their knowledge and provided a strong foundation in computer science. Their passion for teaching and commitment to excellence have played a significant role in our academic journey. Furthermore, we would like to acknowledge the contributions of our teammates and fellow students who have been a source of inspiration and collaboration throughout this project. Their insights and discussions have enriched our understanding of the subject matter. To all those mentioned above we are truly grateful for all the contributions and support.

Place:

Institute of Technical Education and Research;

Siksha ‘O’ Anusandhan (Deemed to be) University

Signature of Students

Date: 17/06/2024

`

DECLARATION

We declare that this written submission represents our ideas in our own words and where other's ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea, fact and source in our submission. We understand that any violation of the above will cause for disciplinary action by the University and can also evoke penal action from the sources which have not been properly cited or from whom proper permission has not been taken when needed.

2041013203

2041016183

2041004013

2041018092

Signature of Students with Registration Numbers

Date: 17/06/2024

`

REPORT APPROVAL

This project report titled **“**A Systematic Study of Histopathological Oral Cancer Data Classification using Machine Learning & Deep Learning Approaches **“**submitted by Ashutosh Mohapatra, Ayush Das Pattanaik, Swaraj Das, SK Abdul Rahman is approved for the degree of *Bachelor of Technology in Computer Science and Engineering*.

Examiner(s)

_____

_____

_____

Supervisor

_____

Project Coordinator

_____

PREFACE

This study aims to conduct a systematic study on the classification of histopathological oral squamous cell carcinoma images using machine learning and deep learning approaches. Squamous cell carcinoma is the most common oral cancer which is a significant health concern globally, with early detection being crucial for successful treatment. This study addresses the problem of accurate classification of oral squamous cell carcinoma from biopsy slides, which plays an important role in diagnosing and treating the patient.

The data for this study were collected from two reputable healthcare institutions: Ayursundra Healthcare Pvt. Ltd and Dr. B. Borooah Cancer Institute. Biopsy slides from 230 patients recommended for oral biopsy tests were collected over a period from October 2016 to November 2017. Images were captured using a Leica DM 750 microscope connected to a high-configured computer and software, with 100x and 400x magnifications. The collected dataset contains a diverse range of histopathological images representing various stages and types of oral cancer.

The research methodology involves the development and evaluation of machine learning and deep learning models for the classification of histopathological oral cancer data. Various techniques such as image preprocessing, image denoising, feature extraction, and model training are used to enhance the accuracy of classification. The classification task involves distinguishing between normal epithelium and squamous cell carcinoma.

The benefits of this study include improved diagnostic accuracy and efficiency in identifying squamous cell cancer from histopathological images. This study contributes to the field of healthcare by offering a systematic approach to oral cancer classification using advanced models. The outcomes of this study have the potential to enhance clinical decision-making and positively impact patient care in the diagnosis and management of oral cancer

`

INDIVIDUAL CONTRIBUTIONS

Ashutosh Mohapatra	- Machine Learning Model Building (Random Forest, SVM, Logistic regression) - Custom CNN model Building - Model Evaluation
Ayush Das Pattanaik	- Wavelet Transformation - Feature vector conversion - Building CNN Architecture
Swaraj Das	- Data Augmentation - Pixel Normalization - Image Resizing - Color Normalization
SK Abdul Rahman	- LAB Conversion - Edge Detection - Color Preservation

`

TABLE OF CONTENTS

`

LIST OF FIGURES

`

LIST OF TABLES

1. INTRODUCTION

1.1 Introduction:

Oral cancer, specifically squamous cell carcinoma (SCC), is a big concern for global health. SCC is the second most common type of skin cancer and poses a significant risk due to its aggressive nature and high mortality rate if not caught early. In most cases, squamous cell carcinoma diagnosis necessitates an exhaustive study of cells obtained from tissue samples using a microscope which may also be limited in terms of manpower, accuracy, uniformity as well as time. Because of this, there is an urgent necessity for mechanized approaches designed to enable pathology specialists to make precise and prompt decisions concerning SCC. Image sorts are commonly worked by use Support Vector Machines, SVMs, Random Forests, k-NN and so on while deep learning techniques in particular Convolutional Neural Networks (CNNs) have proved to be very useful when extracting complicated attributes from medical images. In the unification of traditional machine learning techniques with recent deep learning advances such as CNNs, we aim to combine mutual strengths to improve the performance of our models so that they get very high percentages trustworthy in SCC detection. During our examination of existing scholarship, we discovered a number of frameworks and approaches that can be employed when categorizing histological images. Machine learning has formed a core part of many such studies, though its results have not always been satisfactory. By contrast, the advent of deep learning has greatly enhanced the speed and accuracy of this process. Many currently existing techniques do not fully incorporate every possible example category; so, it is clear that there should be full-block solutions able to handle various types of images. The project was developed to solve the problem of early diagnostics and treatment of squamous cell carcinoma because it is urgent. Detection of SCC in its early stages can save patients' lives and reduce death rates. Nowadays, it is difficult for people to manually detect this disease due to the complexity of the histopathological images and an accurate interpretation requires detailed knowledge. This is the main reason why machine learning (ML) and deep learning (DL) techniques are used. We are basically using those capabilities to develop computer-aided systems. Our main task is analyzing and striving to remove standard systems' shortcomings. This will require inspecting numerous machine learning and deep learning algorithms with the intention of unearthing the most appropriate ones for classifying histopathological images. This includes some activities such as; pre-processing to assess the model's performance, parameter optimization as well as image quality enhancement among others. Beginning with creating a schematic or prototype diagram of the system design proposed, the process involves several significant stages. We are going to use different methods, tools, and algorithms including Convolutional Neural Networks (CNNs) for histopathological image processing and segmentation. Testing involves training and testing the model using securely protected data

followed by analyzing results and fine-tuning the model for best performance. In order to make our findings clearer and more reproducible, we will describe our experiments' procedural details, data collection specifics, as well as the assumptions that were made about them. In tables and graphs, we will show experimental results that underline important discoveries and prove the efficacy of the proposed method. Our study focuses on helping to find squamous cell carcinoma early by constructing the robust automatic categorization model that uses training tactics both from machine learning and deep learning practices.

1.2 Project Overview:

Histopathological examination, which involves the microscopic evaluation of tissue, is considered the gold standard for diagnosing various types of cancer, including OSCC. It provides pathologists with a detailed view of cell structure, enabling them to identify abnormalities that indicate potential danger. In recent years, the integration of machine learning (ML) and deep learning (DL) techniques in medical imaging has shown great promise in improving the accuracy and efficiency of cancer diagnosis. These advanced computer techniques can assist in classifying histopathology images, potentially reducing the workload of pathologists and enhancing diagnostic accuracy. For this study, a dataset of histopathological images of nasal tissues at 100x magnification was utilized. These high-resolution images are essential for capturing intricate cellular details. Pre- processing steps like normalization, enhancement, and segmentation were performed to improve image quality and suitability for analysis. Data enhancement techniques, such as rotation, flipping, and scaling, were also employed to increase the diversity of the dataset and enhance ML robustness with DL models. To classify histopathologic images, traditional machine learning algorithms like random forests (RF), k-NN were employed. These algorithms involve extracting relevant features from images, such as patterns, shapes, and energy. Common extraction methods include local binary patterns (LBP), histogram of oriented gradients (HOG), and gray level co-occurrence matrix (GLCM). After feature extraction, ML algorithms were trained on labeled datasets to differentiate between healthy and cancerous interstitial tissue. Deep learning, particularly convolutional neural networks (CNNs), has revolutionized image classification tasks. CNNs can learn feature sequences directly from raw pixel data, eliminating the need for manual feature extraction. In this study, various CNN algorithms such as LeNet, AlexNet were investigated to determine their efficiency in classifying histopathological images. Transfer learning, where pre-trained models developed on large datasets like ImageNet are fine-tuned to match the histopathological dataset, was employed to leverage existing knowledge and improve performance. The performance of ML and DL models was evaluated using accuracy, precision,

recall, F1-score, and the area under the receiver operating characteristic curve (AUC-ROC). Cross-validation methods were employed to ensure the generalizability and robustness of the models. Confusion matrices provided a detailed breakdown of the models' performance, highlighting their strengths and areas that require improvement. The combination of machine learning and deep learning techniques in histopathology image analysis holds immense potential for improving the early detection of OSCC. By automating the classification process, these techniques can assist pathologists in making accurate and timely diagnoses. Future work involves exploring more advanced deep learning algorithms, expanding the dataset, and developing interpretable AI models to enhance result interpretability. The ultimate goal is to develop reliable and effective diagnostic tools that seamlessly integrate into clinical workflows, thereby contributing to improved patient care and outcomes. This comprehensive study highlights the transformative potential of ML and DL in medical research, enabling breakthroughs in cancer diagnosis and treatment.

1.3 Report Layout

Introduction

Introduction involves explaining what we have done in this research project. In this part we have explained statement, motivation and overview of the project. This part contains the problem statement. The introduction part tells how to make the research to get better results

Literature Survey

In literature survey we have discussed some of the previous studies that have been done in this field. Some of the authors have done machine learning part some have used deep learning part to get into a final conclusion. What preprocessing techniques they have used and what are the feature extraction used. They have used different techniques to find out how to get better results in their field.

Material and Methods used

Material and methods used contains the everything that is used for this research work. It includes tools used, the tools that we have used for writing the codes, the packages that we have imported for faster building of models, the algorithm used for different parts like image preprocessing, feature extraction, transformation and model building.

Results and Outputs

In results and output we have given detailed overview of all evaluation metrics for classification problem. We have shown different graphs like Accuracy curve, loss curve and ROC curve. Other mathematical metrics in a table to compare other models

Conclusion

In conclusion we are giving a main findings and results of our research, i.e. model that shows better result for both machine learning and deep learning. We explain why this is better model by using different references.

2 Literature Survey

2.1 Existing Systems

An overview of the various research papers that we have referenced and compared with our models. Our focus lies on the machine learning techniques used, along with the accuracy percentages achieved.

Rosado P, Lequerica-Fernández P, Villallaín L, Peña I. Sanchez Lasheras F, De Vicente JC. They have used SVM. They got results of 98 % accuracy. Chang SW, Abdul-Kareem S, Merican AF, Zain RB. They have used SVM. The result they got an accuracy of 75%. Das DK, Bose S, Maiti AK, Mitra B, Mukherjee G, Dutta PK. The technique they used are CNN, Gabor filter and Random Forest and got a accuracy of 96.88%. Krishnan MMR, Venkatraghavan V, Acharya UR,et al. The technique they have used is GMM and KNN. The Accuracy they got is 78.3% and 71.7% accuracy.

Table 1: This table gives previous studies that have been done on oral cancer data.

Sl No	Author	Title	Technique Used	Results
1	Rosado P, Lequerica-Fernández P, Villallaín L, Peña I. SanchezLasheras F, De Vicente JC.	Survival model in oral squamous cell carcinoma based on clinicopathological parameters, molecular markers and support vector machines.	SVM	Accuracy = 98%
2	Chang SW, Abdul-Kareem S, Merican AF, Zain RB.	Oral cancer prognosis based on clinicopathologic and genomic markers using a hybrid of feature selection and machine learning methods.	SVM	Accuracy = 75%
3	Das DK, Bose S, Maiti AK, Mitra B, Mukherjee G, Dutta PK.	Automatic identification of clinically relevant regions from oral tissue histological images for oral squamous cell carcinoma diagnosis.	CNN, Gabor Filter, Random Forests	Accuracy = 96.88%
4	Krishnan MMR, Venkatraghavan V, Acharya UR, et al.	Automated oral cancer identification using histopathological images: a hybrid feature extraction paradigm	GMM KNN	Accuracy = 78.3% Accuracy = 71.7%

These research papers provided a detailed overview of machine learning techniques used, highlighting their accuracy in results. Our research builds upon these foundations, further exploring and enhancing our machine learning models.

2.2 Problem Identification

We identified many challenges that impact the performance of our machine learning models. These are very crucial to address as to improve the accuracy of our results, and reliability of our model. The primary issues identified are as follows:

Data Imbalancing

The OSCC (Oral Squamous Cell Carcinoma) folder has almost five as many images as the Normal folder. If the model is trained on greater number of OSCC images than normal images, it can become biased, and give out biased results. Biased results mean it will predict the OSCC cells with good accuracy but fail with normal images. This can hamper the reliability of our model in real-world applications.

Data Variance

The images are stained using different H&E (Hematoxylin and Eosin) staining methods, that leads to variance in color modes. Feature extraction will become tough as there are unique color modes, and our model will not be capable of generalize the features. The model's precision and accuracy are affected making it extra challenging to extract features. This is a massive disadvantage as in medical picture analysis, precise and steady characteristic extraction is crucial.

Huge Data Size

Each picture in our dataset is 2048x1536 pixels. This big size of photos increases the computational load, as the system desires heavy computing sources. The big size ends in increased processing time and computational useful resource requirements, making it tough to train our model efficiently. This can bring about longer waiting instances, higher memory utilization, and the want for greater effective hardware, which won't always be possible.

By solving these problems, we make our machine learning models work better. Using methods like data augmentation to fix data imbalance. We can lower data variance by doing pixel normalization. Also, resizing images can help with big data sizes. By fixing these problems, we want to make our models more accurate and reliable, and help with our research project.

2.3 Problem Statement

This research seeks to identify and implement appropriate machine learning and deep learning approaches for the classification of oral cancer data. Given the complexities and challenges associated with this task, such as data imbalance, variance in staining methods, and large image sizes, our goal is to develop robust models that can accurately differentiate between OSCC (Oral Squamous Cell Carcinoma) and normal tissues.

Through our research, we aim to advance the field of medical image analysis, providing valuable tools for early detection and accurate diagnosis of oral cancer, ultimately contributing to improved patient outcomes.

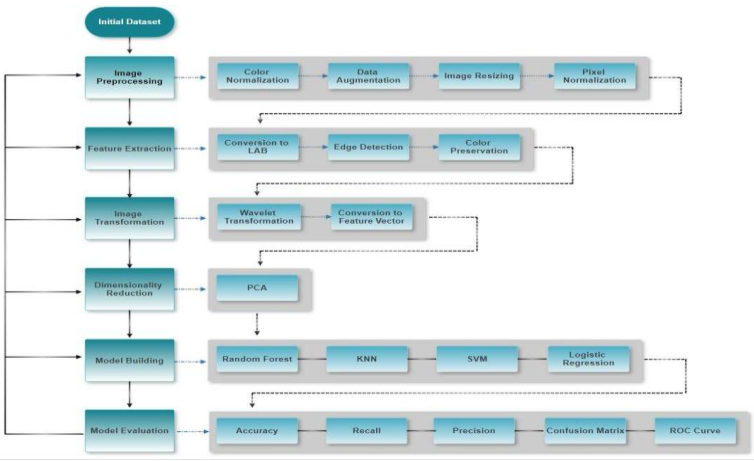
3. Materials and Methods

3.1 Dataset Description

We have Histopathological images of Oral cavity. Histopathological images are considered to be gold standard for detecting cancer from tissue. The Histopathological images are stained with Hematoxylin and Eosin(H&E). 100x magnified images of Histopathological images which has been taken from Leica DM Microscope. The Histopathological images are collected from 2 Cancer hospitals: Ayursundra Healthcare Pvt. Ltd, Guwahati, Assam, India and Dr B. Borooah Cancer Institute, Guwahati, Assam, India. The data are collected from 230 patients recommended for oral biopsy tests from the 2 hospitals. The image of size 2048*1536 pixels. It contains two folders. The Normal Epithelium contains 89 images and the OSCC folder contains 439 images. **3.2 Schematic layout**

Machine learning Layout

Fig 1: Machine Learning Flow Diagram



In Fig 1, the diagram we have shown all the steps that we have done before building the machine learning model. As this is a classification problem where we want to train a model that can predict whether a tissue image is infected by Oral squamous cell carcinoma or not. We have built various machine learning classifier to see the results.

Deep learning Layout

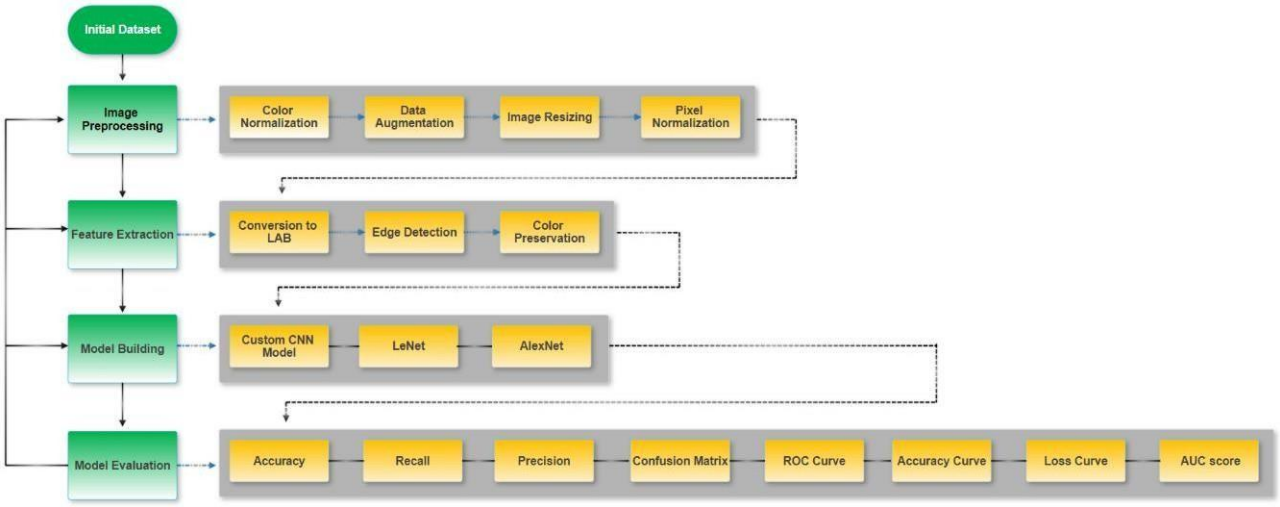


Fig 2: Deep Learning Flow Diagram

For Deep learning we have implemented a CNN Model. For the CNN evaluation we have used accuracy, precision, recall, F1 score, accuracy curve, loss Curve, ROC curve and AUC score. We have compared our model with other CNN models, like LeNet, AlexNet.

In Fig 2, the diagram we have shown all the steps that we have done before building the deep learning CNN model. As this is classification problem, where we want to train a model that can predict whether a tissue image is infected by Oral squamous cell carcinoma or not. We have built various CNN classifier to see the results.

3.2 Methods used

IMAGE PREPROCESSING

Color Normalization: Due to different level of H&E staining we have different color histopathological images.

Color Normalization solves the problem for Color variance.

Data Augmentation: In the initial dataset, we have 89 images of Normal Epithelium images and 439

images of OSCC infected images. If we train the model using this image then model will become bias on OSCC images as we have more images in OSCC folder. Now adding more images can be expensive. So, we generate more images from data augmentation. Machine Learning: Normal Epithelium contains 700 images and OSCC contains 700 images. Deep Learning: Normal Epithelium contains 2475 images and OSCC contains 2475 images.

Data Augmentation solves the problem of Biasing.

Image Resizing: The initial image size is 2048*1536 pixels. Now if we give this size of images to the model, then it will take longer time to train. We need to find a size of image that will make the image smaller and doesn't hamper the features in the images. The image size is reduced from 2048*1536 pixels to 500*500 pixels.

Image Resizing solves the problem of huge data size.

Pixel Normalization

We have RGB images. Each image has three channels, red channel, blue channel and green channel. Each of the channel is reduced to 500*500 pixels. But each pixel contains a value from 0-255. The range of values is high. For faster computation and to train the model faster we need to normalize the pixel values from 0-255 to 0-1.

Pixel Normalization solves the problem of slow training of model.

For feature extraction

Converting to LAB images

LAB images are device independent images. Converting our RGB images to LAB images can help us to detect important features from it.

Canny edge detection

Canny edge detection in histopathological images helps us to detect tissue and cells in the image by drawing edges. This can be useful in detecting the objects, the number of objects, the shape of image etc.

For image transformation

Wavelet transformation

Wavelet transformation of images is an image transformation technique that decomposes the images into different class of wavelet coefficients and can help us to capture important features from data (in our case its image).

Feature vector conversion

After doing wavelet transformation, the next task is to convert the image to NumPy nd array which

is feature vector and then converting the nd array to feature vector which can be useful for faster training and predicting.

Machine learning model building

Random Forest

Random Forest is a very popular supervised machine learning model. It is an ensemble technique. Random Forest builds over Decision Trees. In this research we are going to use Random Forest classifier for classifying the images. We have used Grid search cv to find out the best hyperparameters for classifying the problem. For hyperparameters we have used n_estimators, and criterion. For n_estimators we have used values like 10,50,100,200 and 300. For criterion we have used gini and entropy.

SVM

SVM is a popular machine learning model that is used for both classification as well as regression problems. SVM stands for Scalar Vector Machine. As we are doing a classification problem so we are using Scalar Vector Classifier (SVC). Scalar Vector classifier tries to different hyperplanes to classify the data. We have used Grid Search CV for different hyperparameters like kernel and regularization parameters. For kernel we have used kernel types like linear, poly, rbf, sigmoid. For regularization parameter we have used different values like 1,10,50 and 100.

Logistic Regression

Logistic regression using a mathematical function called as logistic function, also known as sigmoid function. It calculates the probability between different classes and tries to map them. For different Hyperparameters we have used different values like

KNN

KNN stands for K Nearest Neighbors. It tries to classify images using different hyperparameters like n_neighbours and p. For n_neighbours we taken 3,5 and 7. p is called Power parameter, for p=1 we are going to use Manhattan distance and for p=2 we are going to use Euclidean distance. For other p values, Minkowski distance is used.

Deep learning model building

Custom CNN architecture

CNN is a deep learning model which stands for Convolutional Neural Network. Convolutional Neural Network is highly used for image data. In CNN the initial layers are convolutional layers which helps in extracting high- and low-level features from the images. The next phase is to flatten

the image. After that we add some dense layer that will help to calculate the weights and biases of the dense layer that helps in calculating the probabilities of each class and after that it is given to the output layer.

LeNet 5

LeNet-5 is popular CNN model. LeNet-5 is a simple CNN model that has been built for digit classification problem. We have used LeNet-5 for our image data. The results we got are evaluated with various classification metrics. LeNet-5 contains 5 layers which includes input layer that takes input image of size 32*32 pixels. Both contains 2 convolutional layers of each 5*5 kernel size and stride size of 1. For each coevolution layer we have MaxPooling layer which is of 2*2 kernel and of stride size 1.

The Dense layer contains three fully connected layers. The first layer contains 120 neurons the second layer contains 84 neurons and the last layer which is the output layer contains 10 neurons. As our classification is binary classification problem so we used 2 neurons at the output layer.

AlexNet

AlexNet was designed for image classification for ILSVRC competition. During that time AlexNet was showing very impressive results for the input image given. In AlexNet it contains input layer of size 224*224 size. AlexNet contains five convolutional layers. The first convolutional layer contains 11*11 kernel size stride size of 4. The second convolutional layers contains kernel size of 585 and a stride size of 1. The rest convolutional layers contains kernel size of 3*3 kernel size and of stride size of 1. Only the first, second and fifth convolutional layers contains pooling layer. The MaxPooling layers layer size of 3*3 and stride of 2. Then the image is flattened. After that we give the image to fully connected layers. The Fully Connected layers contains three layers. The first two contains 4096 neurons and the last layer which is the output layer contains 1000 neurons. As our problem is binary classification so we have taken 2 neurons at the output layers.

3.3 Tools used:

NumPy:

NumPy is used for array operations. NumPy process multi-dimensional operation very fast. Whenever working with images, we need to convert it into multi-dimensional array. It can be 2D array, 3D arrays and 4D arrays. 2D array is used to represent the height and width of image. As we are using RGB images, it contains red channel, blue channel and green channel. So, in order to represent this, we use 3D NumPy array. While doing deep learning we need to train the images using batches. When batches are added then we are dealing with 4D arrays.

Pandas:

While dealing with images we need to see it in 2D arrays. So, save that image using 2D image we use Dataframes for it for different experimentation in images.

Matplotlib:

Matplotlib is used for printing the images using pyplot and to plot different graphs like

accuracy curve, loss curve and ROC curve.

Scikit learn

Scikit-learn is used for sklearn contains all model algorithm like Random Forest, KNN, SVC, Logistic regression. sklearn contains different metrics like accuracy, precision, recall, f1 score and confusion matrix. sklearn contains algorithm for dimensionality reduction i.e. Principal Component Analysis (PCA).

OpenCV:

OpenCV is used for Reading images, Conversion of images into different color spaces, Merging different channels of images, Resizing the images, Printing the images.

Keras (TensorFlow)

Keras is a TensorFlow library and is used for ImageDataGenerator for data augmentation, Fetching images from directory by batches, For Resizing and Rescaling, Contains different layers Conv2D, Maxpooling2D, Dense, Dropout etc.

PyWavelet

PyWavelet is used for wavelet transformation like discrete wavelet decomposition and reconstruction and inverse discrete decomposition and transformation.

Platform Used

Python:

Python is a programming language that is used for developing the code. It contains libraries for faster development of code.

Jupyter notebook:

Jupyter notebook is an open-source web application of running python notebooks.

Google Colab:

Google Colab is cloud-based platform is used for running python notebooks. It gives integrated GPUs and TPUs for faster processing of data. Many of the libraries are pre- installed as it is running on the cloud.

Anaconda:

Anaconda is the single platform which contains all the tools for Data analysis, Data science, Machine learning and AI. It contains platform like Jupyter notebook, Conda PowerShell, Conda prompt, Jupyterlab, spyder etc.

Conda prompt:

Conda prompt is a command line interface which is used for creating the python environment, activating conda environment, deactivating the environment, installing packages etc.

3.4 Evaluation Metrics

Accuracy

Accuracy is the widely used evaluation metrics. Accuracy can be measured as sum of true positives and true negatives divided by all measures (True Positives, True Negatives, False Positives and False Negatives). When choosing accuracy as the evaluation metric we should be very careful that our data is not imbalanced. If our data is imbalanced then accuracy should not be the only option for evaluation, we should take other metrics into consideration.

Precision

Precision is used for calculating the type and quality of positive predictions done by our model. It is calculated by correctly predicted positive instances out of all measures of positives predicted by our model. Higher precision tells that our model correctly predicts the true labels. However, while increasing the precision value sometimes recall value decreases. So, if both precision and recall show high score then we have a very good model.

Recall

Recall is also called as True Positive Rate (TPR) or Sensitivity. It measures the correctly predicted positive values by all actual positive measures in our dataset. The higher value of recall tells that our model can identify the positive cases effectively. Sometimes, while increasing the recall value, precision value decreases. So, if both recall and precision shows high score then we have a very good model.

Confusion matrix

Confusion is an important classification metrics. It gives tabular data about the Actual values and

Predicted values. The data the confusion matrix has are True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN). By seeing confusion matrix, we can predict our model's strengths and weaknesses. Various other evaluation metrics uses confusion matrix for finding their values like Accuracy, Precision, Recall.

ROC curve

The ROC curve is a graph plotted between True Positive Rate (TPR) and False Positive Rate (FPR). It shows the graph whether our model is able to classify normal and OSCC images or not. We calculate the area under the curve score. If the area under the curve is close 1 then our model is performing very well i.e. the model can easily classify the normal and OSCC images.

AUC score

AUC stands for Area Under the Curve. The Area Under the Curve is calculated from the ROC curve plotted for our model. It gives the measure of our model classifying different classes. From the AUC score we can tell our model is performing good or not.

Accuracy curve

Accuracy curve is metrics that shows how our model is learning the different features epoch wise. It plots a curve between the accuracy obtained and the Epochs. After seeing the accuracy curve, we can identify whether our models overfits or learning very slowly etc.

Loss curve

While building our deep learning model, we specify the loss function. The Loss Curve is plotted with Loss obtained in every epoch to Number of Epochs. Generally, the loss should decrease with epochs.

4. RESULTS AND OUTPUT

4.1 System Specification

Table 2: System Specifications

NAME	SPECIFICATION
CPU	AMD Ryzen 7 5700U
OS	Windows 11 Home 23H2
GPU	AMD Radeon Graphics 2GB
Memory	16.0 GB (13.9 GB usable)
Disk	512 GB SSD
Python	3.11

TensorFlow	2.10

The system specification of the system where we have run all the code and executed it. The training and model building depends on the specifications of the code. The specification of the table is shown in Table 2.

4.2 Results and Outcomes

Model Evaluation:

Classifiers we have built are Random Forest, KNN, Logistic regression and SVC. In order to evaluate our model, we have calculated different metrics like accuracy, precision, recall, F1 score, confusion matrix, ROC curve and AUC score. The Table shown below describes all the metrics and their corresponding values.

In Table 3 we have shown for both Training and test metrics of all the model that we have built. Now doing this helps us to check whether our model overfits or whether our model is able to learn from the images we are providing.

Other Evaluation metrics like ROC curve and confusion matrix are shown after Table 3. It shows how our model predicts both Normal and OSCC images for both training and testing data.

After going through Table 3, we can tell that the Random Forest shows good results for both Training and Testing data.

Table 3: Comparison between different model

	Random Forest		KNN		SVC		Logistics Regression	
	Training	**Test**	**Training**	**Test**	**Training**	**Test**	**Training**	**Test**
Accuracy	99.92 %	98.57 %	91.82 %	92.14 %	80.63 %	81.42 %	78.80 %	78.57 %
Precision	99.84 %	98.66 %	89.78 %	89.02 %	82.45 %	83.56 %	79.73 %	80.82 %
Recall	100 %	99.84 %	94.24 %	97.33 %	77.44 %	81.33 %	76.80 %	78.66 %
F1	99.92 %	98.66 %	91.95 %	92.99 %	79.86 %	82.43 %	78.23 %	79.72 %
AUC score	0.99	0.99	0.98	0.98	0.86	0.88	0.86	0.87

Random Forest

The Random Forest confusion matrix for both training and test data. In the train data, 1 image got misclassified out of 1260 and 2 images from test data are misclassified out of 140.

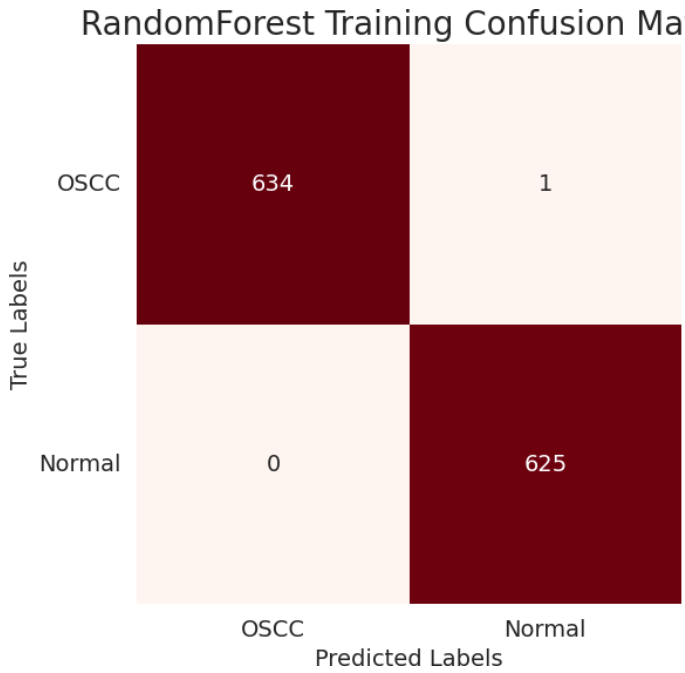


Fig 3.1: Random Forest Training Confusion Matrix

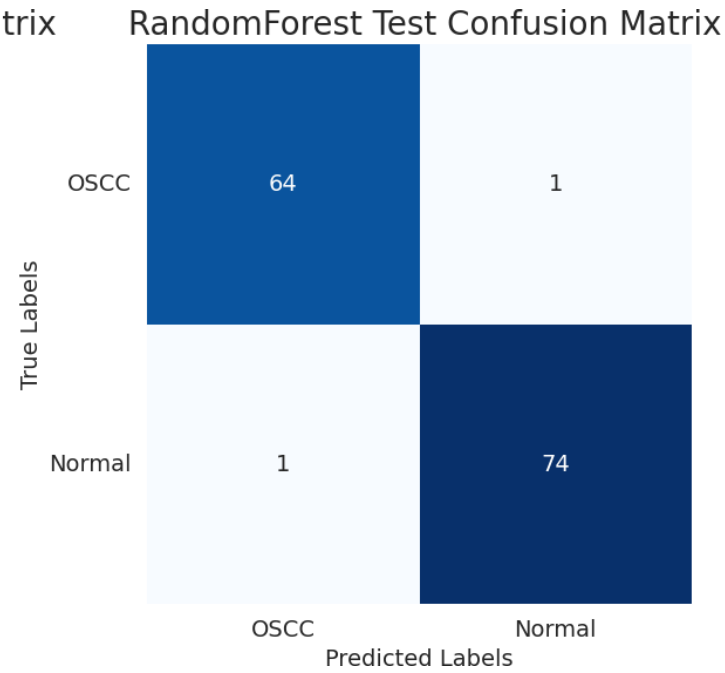


Fig 3.2: Random Forest Test Confusion Matrix

Fig 3: Random Forest Confusion Matrix

The Random Forest classifier is also very good at classifying both normal and OSCC images. This can be clearly be seen from the ROC curve plotted down. The AUC score for both train and test is close to 1.

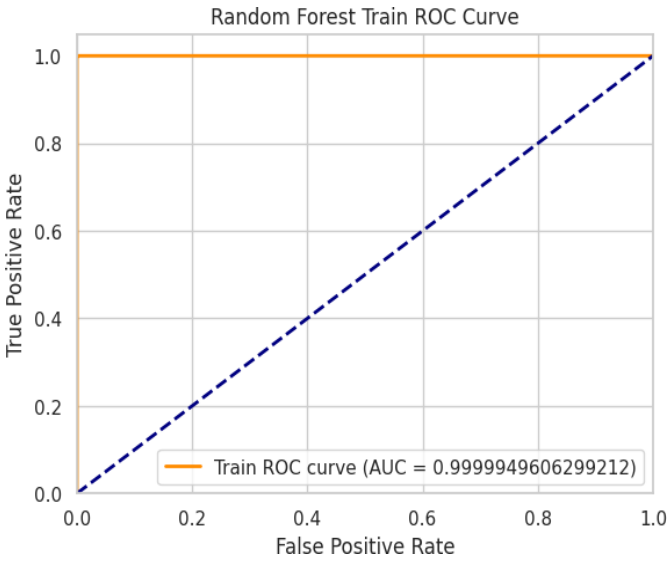


Fig 4.1: Random Forest Training ROC Curve

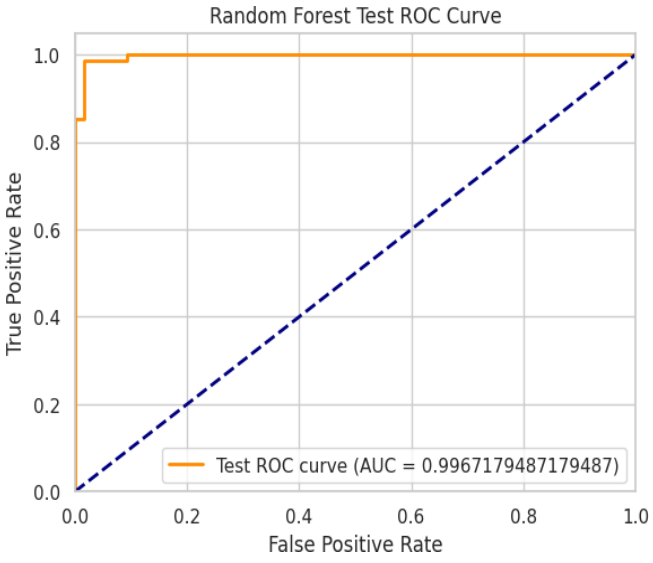


Fig 4.2: Random Forest Test ROC Curve

Fig 4: Random Forest ROC curve

KNN

The KNN confusion matrix for both training and test data. In the train data, 103 image got misclassified out of 1260 and 11 images from test data are misclassified out of 140.

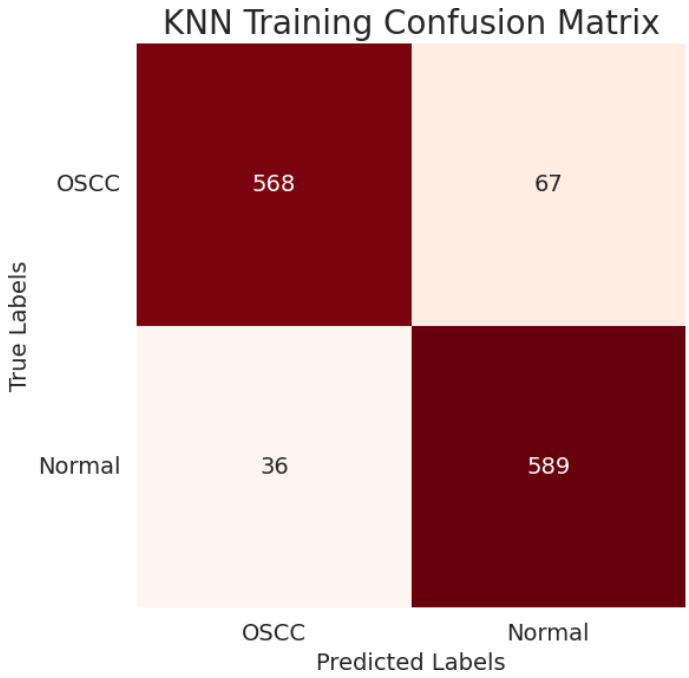


Fig 5.1: KNN Training Confusion Matrix

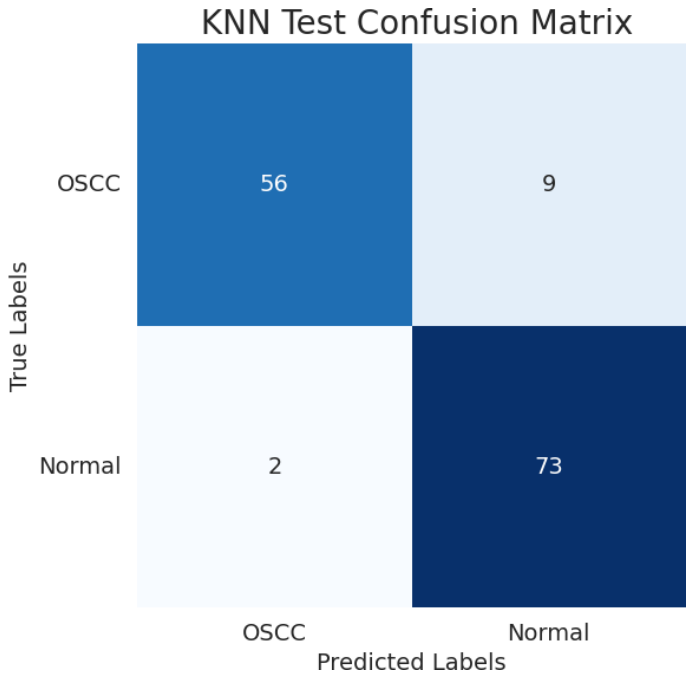


Fig 5.2: KNN Test Confusion Matrix

Fig 5: KNN Confusion Matrix

The KNN confusion matrix shows that KNN has done misclassification, but the ROC curves shows that it tries to classify the classes better. For training data, the AUC score is 0.98 and for test data the AUC score is 0.98.

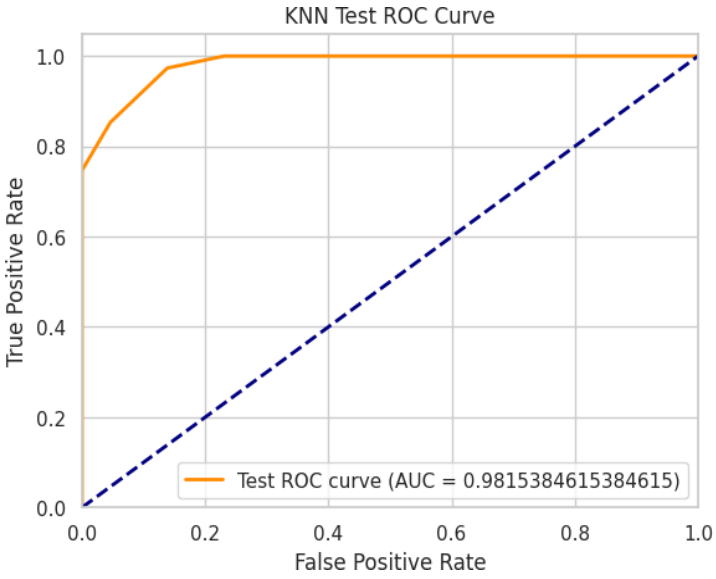


Fig 6.1: KNN Training ROC Curve

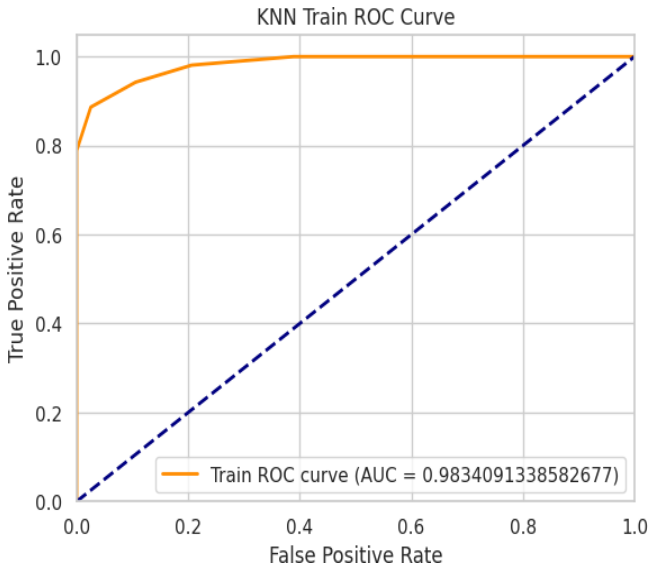


Fig 6.2: KNN Test ROC Curve

Fig 6: KNN ROC Curve

SVC

The SVC confusion matrix for both training and test data. In the train data, 244 image got misclassified out of 1260 and 26 images from test data are misclassified out of 140.

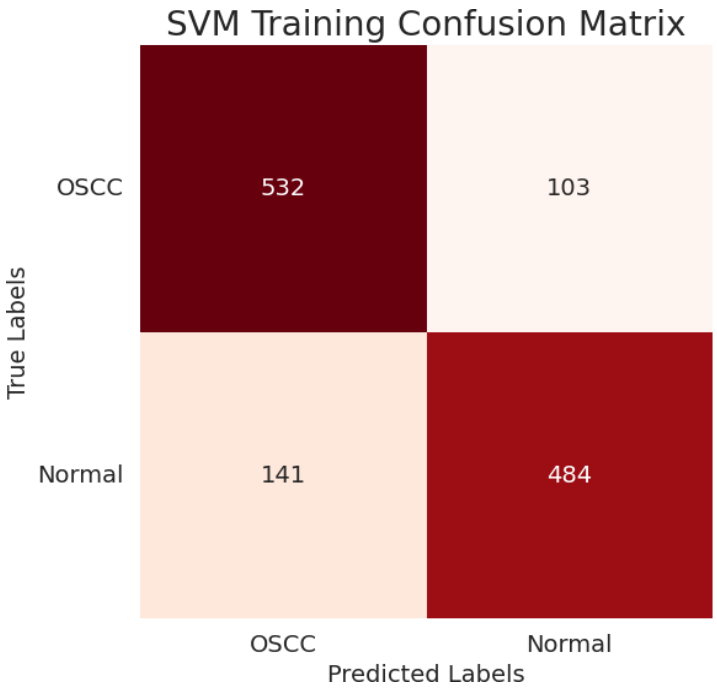


Fig 7.1: SVC Training Confusion Matrix

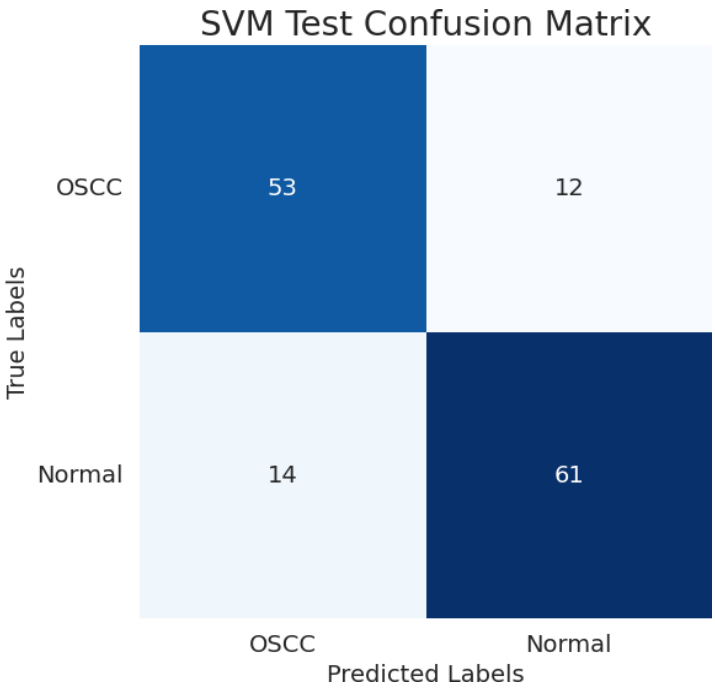


Fig 7.2: SVC Test Confusion Matrix

Fig 7: SVC Confusion Matrix

The SVC ROC curves shows that it tries to classify the classes better. For training data, the AUC score is 0.86 and for test data the AUC score is 0.88.

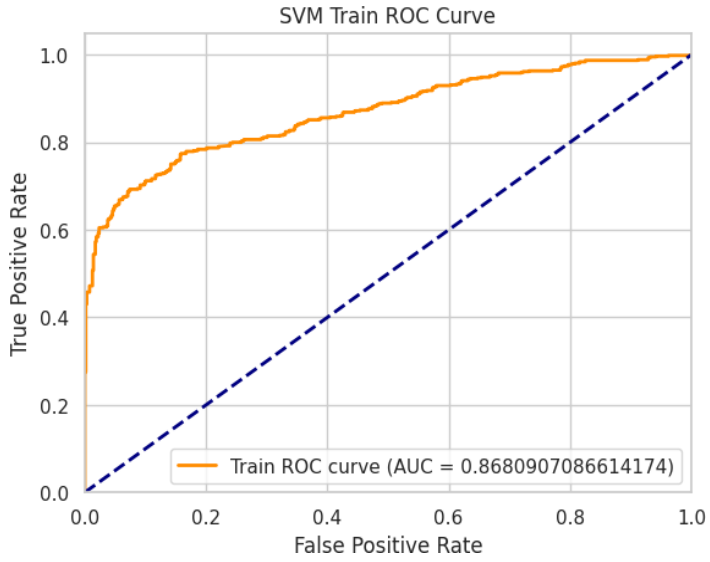


Fig 8.1: SVC Training ROC Curve

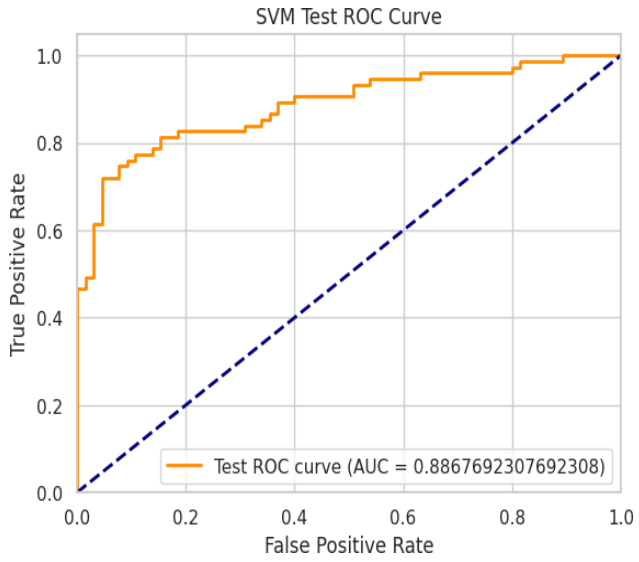


Fig 8.2: SVC Test ROC Curve

Fig 8: SVC ROC Curves

Logistic Regression

The Logistic regression confusion matrix for both training and test data. In the train data, 267 image got misclassified out of 1260 and 26 images from test data are misclassified out of 140.

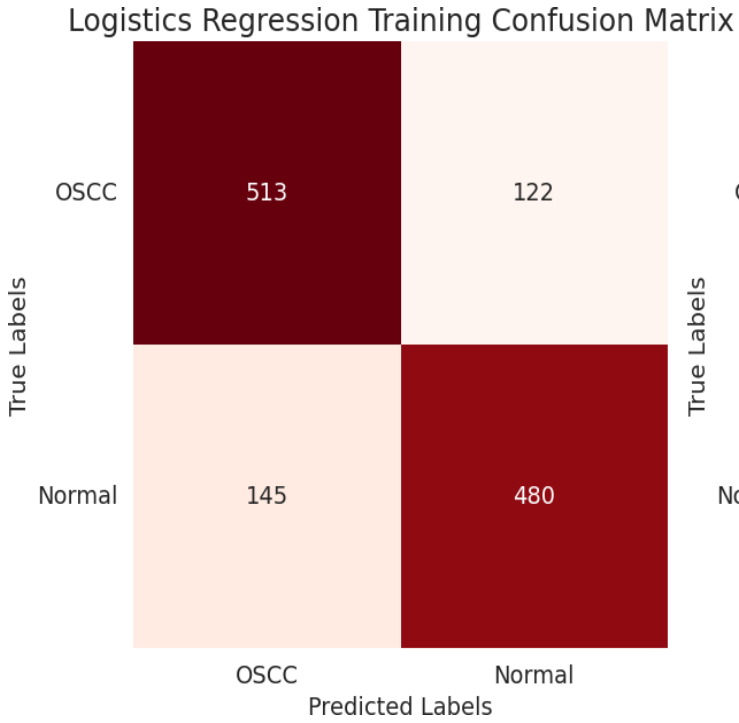


Fig 9.1: Logistic Regression Training Confusion Matrix

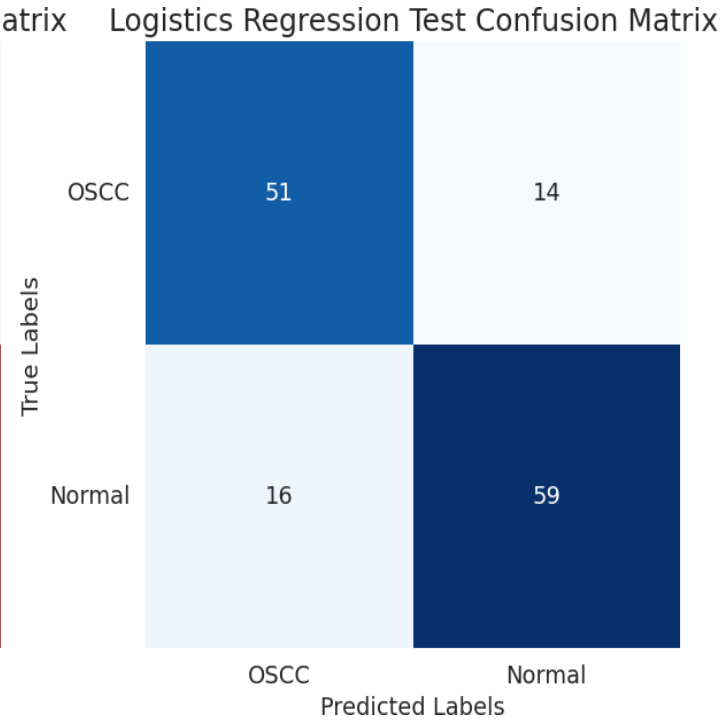


Fig 9.2: Logistic Regression Test Confusion Matrix

Fig 9: Logistic Regression Confusion Matrix

The Logistic regression ROC curves shows that it tries to classify the classes better. For training data, the AUC score is 0.86 and for test data the AUC score is 0.87.

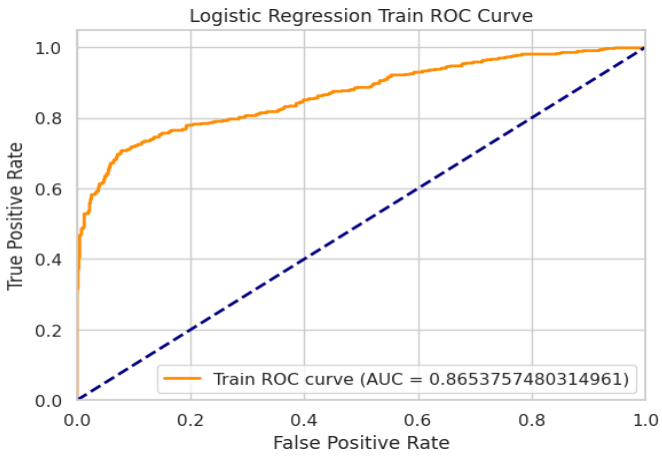


Fig 10.1: Logistic Regression Training ROC Curve

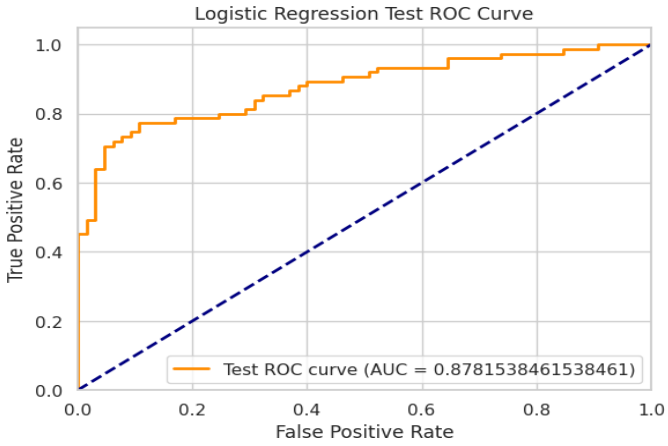


Fig 10.2: Logistic Regression Test ROC Curve

Fig 10: Logistic Regression ROC Curve

Deep learning

For Deep learning we have implemented a CNN Model. For the CNN evaluation we have used accuracy, precision, recall, F1 score, accuracy curve, loss Curve, ROC curve and AUC score. We have compared our model with other CNN models, like LeNet, AlexNet. In Fig 10, the diagram we have shown all the steps that we have done before building the deep learning CNN model. As this is classification problem, where we want to train a model that can predict whether a tissue image is infected by Oral squamous cell carcinoma or not. We have built various CNN classifier to see the results.

Table 4: Comparison between different models

		Accuracy	Precision	Recall	F1	AUC
CNN Model	Training	98.27 %	97.07 %	98.29 %	99.79 %	1.0
	Validation	98.88 %	99.10 %	98.66 %	98.88 %	1.0
	Test	98.80 %	97.65 %	100 %	98.81 %	1.0
Lenet	Training	50 %	50 %	100 %	66.66 %	0.5
	Validation	50 %	50 %	100 %	66.66 %	0.5

Alexnet	Test	50 %	50 %	100 %	66.66 %	0.5
	Training	50 %	0 %	0 %	0 %	0.5
	Validation	50 %	0 %	0 %	0 %	0.5
	Test	50 %	0 %	0 %	0 %	0.5

CNN classifiers we have built are our custom CNN model, LeNet, AlexNet. In order to evaluate our model, we have calculated different metrics like accuracy, precision, recall, F1 score, confusion matrix, ROC curve and AUC score, accuracy curve and loss curve. The Table shown below describes all the metrics and their corresponding values.

Table 4 down below shows for Training, validation and test metrics of all the model that we have built. Now doing this helps us to check whether our model overfits or whether our model is able to learn from the images we are providing.

Other Evaluation metrics like Accuracy curve, Loss curve, ROC curve and confusion matrix are shown in Table 4. It shows how our model predicts both Normal and OSCC images for both training and testing data.

CNN Model

Accuracy and Loss Curves

Accuracy curve and loss curve here shows how the model gets trained with epoch wise. With increase in epoch our accuracy should increase and loss should decrease. Number of epochs are 30. The curves are shown below.

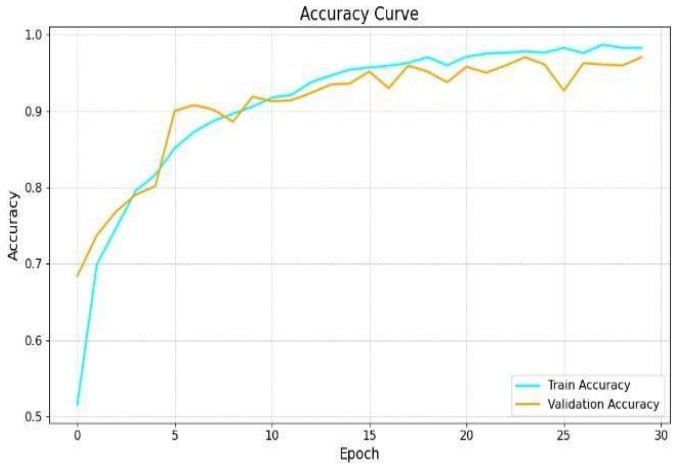


Fig 12: CNN Accuracy Curve

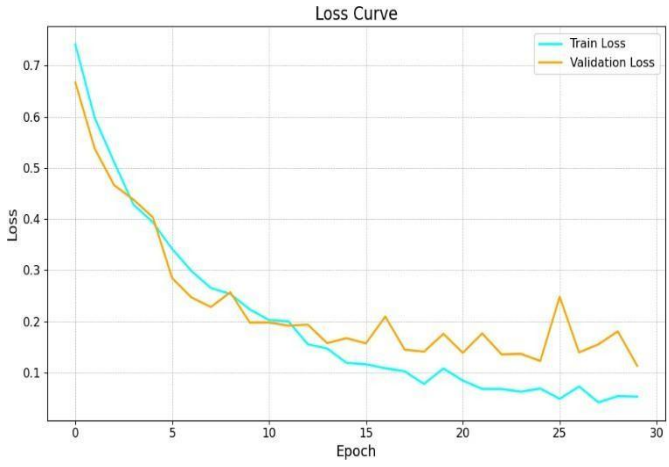
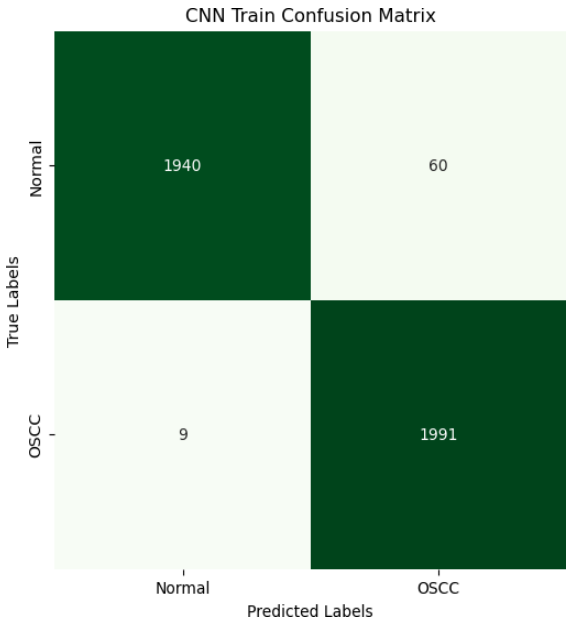


Fig 13: CNN Loss Curve

Train Dataset

The number of misclassifications shown in the confusion matrix are 69 out of 4000 images.

The ROC curve is shown with AUC score of 1.



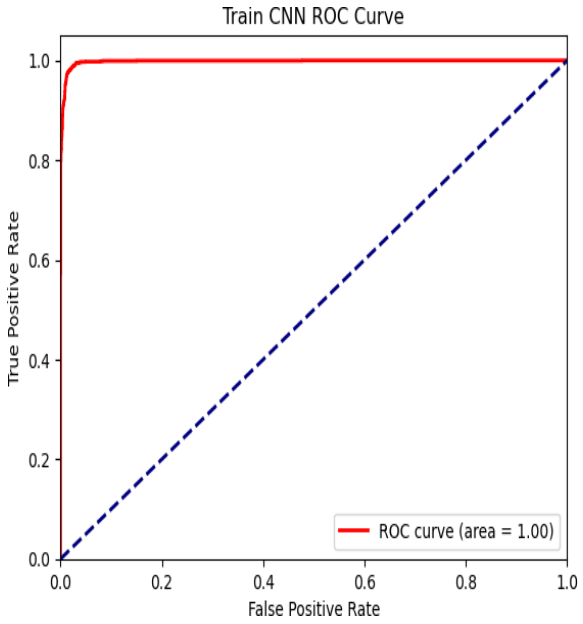


Fig 14.1: CNN Train Confusion Matrix

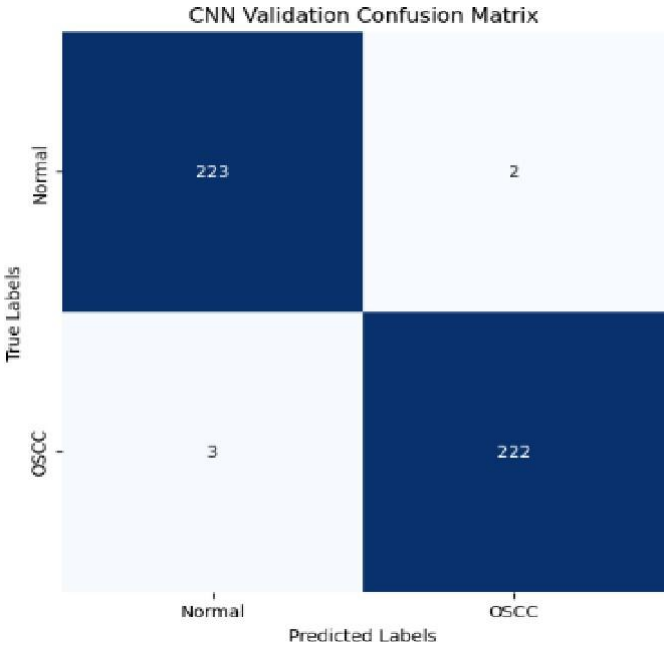
Fig 14.2: CNN Train ROC Curve

Fig 14: CNN Train Confusion Matrix and ROC Curve

Validation Dataset

The number of misclassifications shown in the confusion matrix are 5 out of 450 images.

The ROC curve is shown with AUC score of 1.



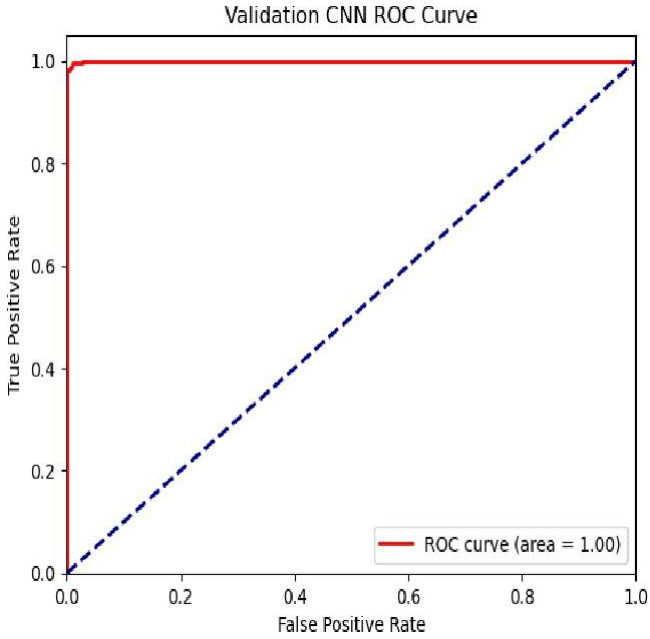


Fig 15.1: CNN Test Confusion Matrix

Fig 15.2: CNN Test ROC Curve

Fig 15: CNN Test Confusion Matrix and ROC Curve

Test Dataset

The number of misclassifications shown in the confusion matrix are 6 out of 500 images. The ROC curve is shown with AUC score of 1.

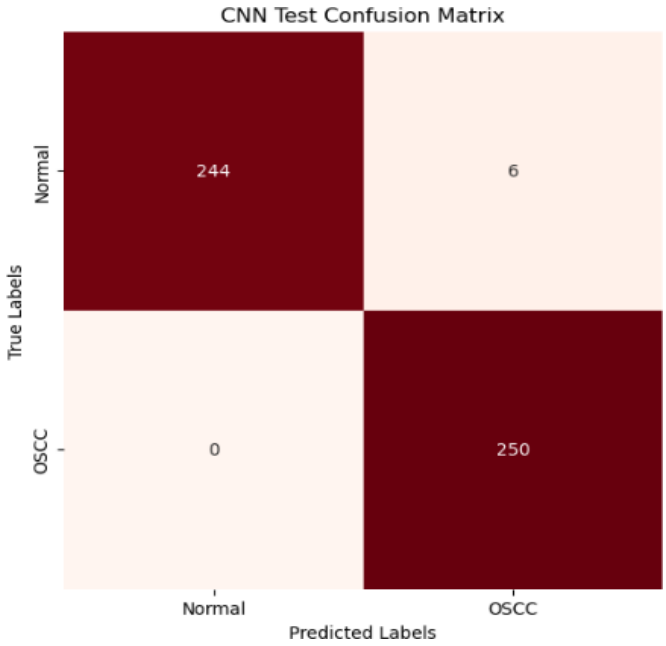


Fig 16.1: CNN Test Confusion Matrix

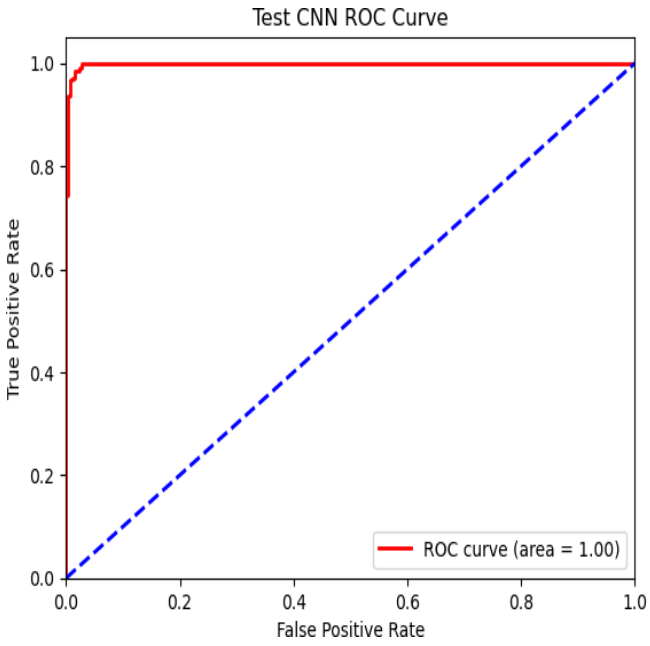


Fig 16.2: CNN Test ROC Curve

Fig 16: CNN Test Confusion Matrix and ROC Curve

LeNet

Accuracy and Loss Curve

Accuracy curve and loss curve of LeNet here shows how the model gets trained with epoch wise. With increase in epoch our accuracy should increase and loss should decrease in LeNet. Number of epochs are 30. The curves are shown below. The LeNet model is not able to learn properly.

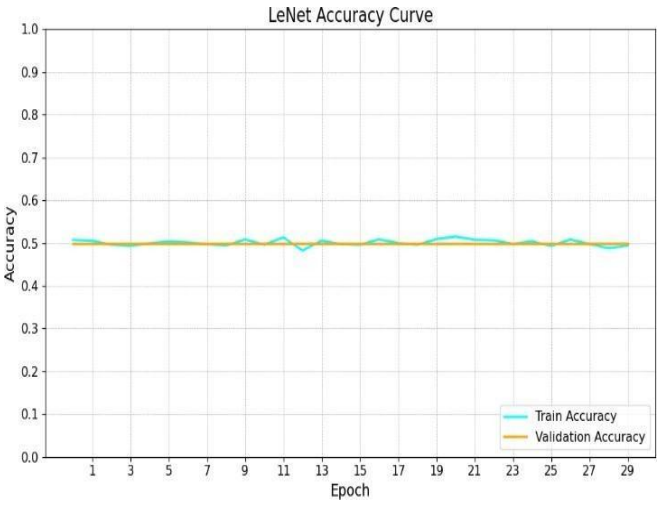


Fig 17.1: LeNet Accuracy Curve

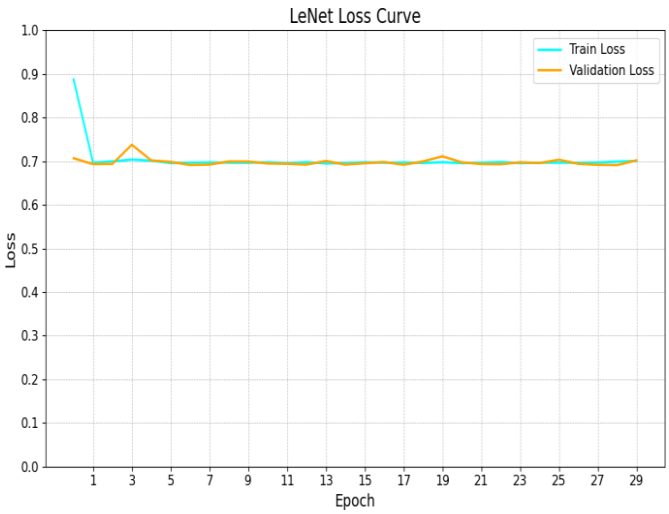


Fig 17.2: LeNet Loss Curve

Fig 17: LeNet Accuracy and Loss Curve

Train Dataset

The number of misclassifications shown in the confusion matrix are 2000 out of 4000 images. The ROC curve is shown with AUC score of 0.5.

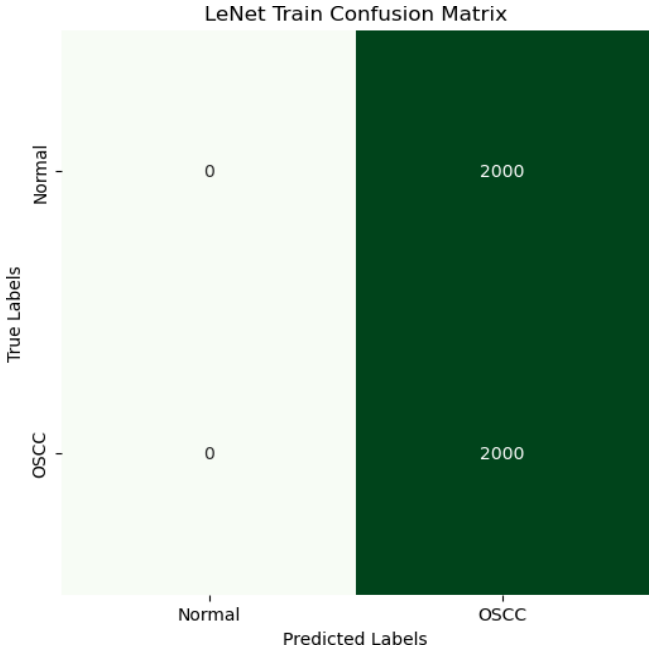


Fig 18.1: LeNet Train Confusion Matrix

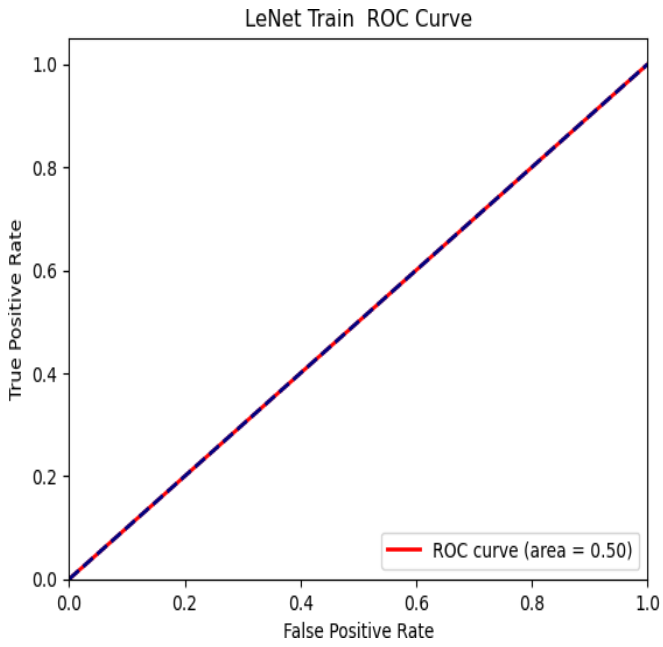


Fig 18.2: LeNet Train ROC Curve

Fig 18: LeNet Confusion Matrix and ROC Curve

Validation Dataset

The number of misclassifications shown in the confusion matrix are 225 out of 450 images. The ROC curve is shown with AUC score of 0.5.

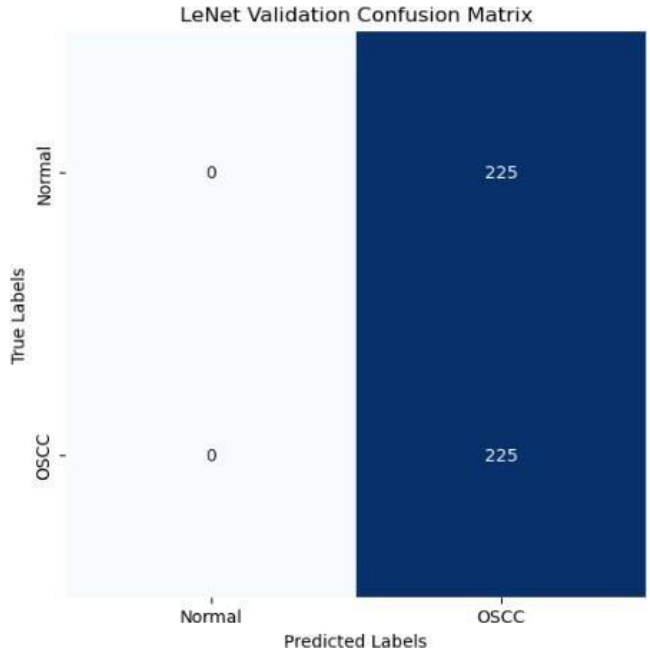


Fig 19.1: LeNet Validation Confusion Matrix

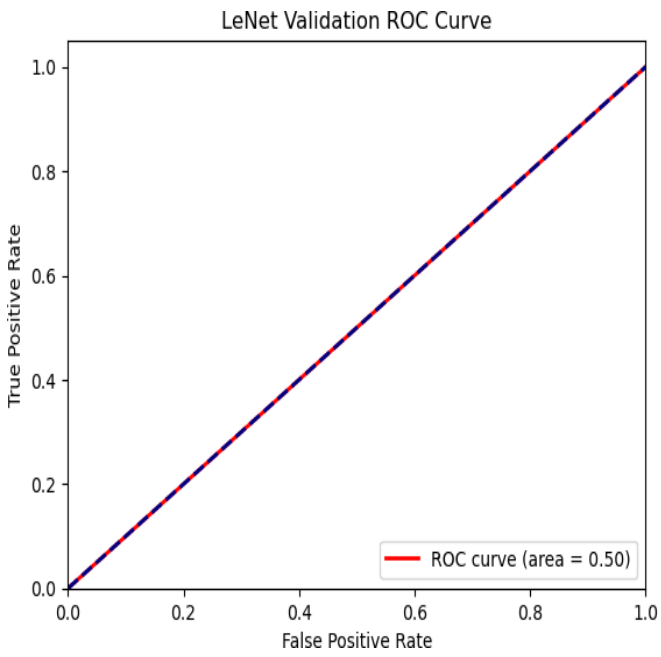


Fig 19.2: LeNet Validation Confusion Matrix

Fig 19: LeNet Validation Confusion Matrix and ROC Curve

Test Dataset

The number of misclassifications shown in the confusion matrix are 250 out of 500 images. The ROC curve is shown with AUC score of 0.5.

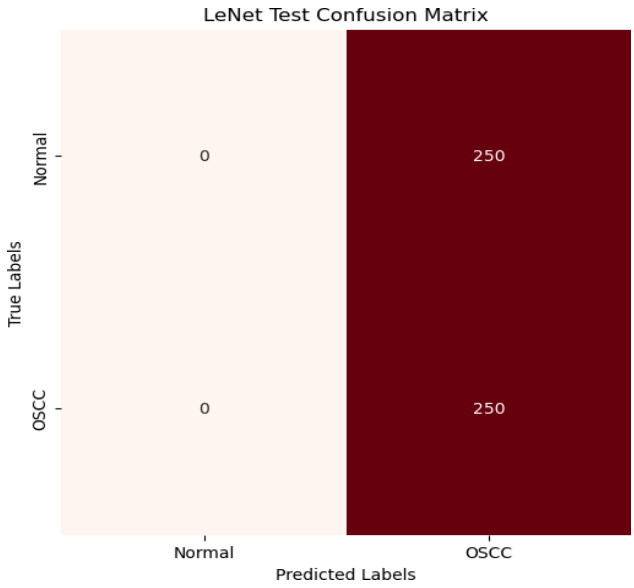


Fig 20.1: LeNet Test Confusion Matrix

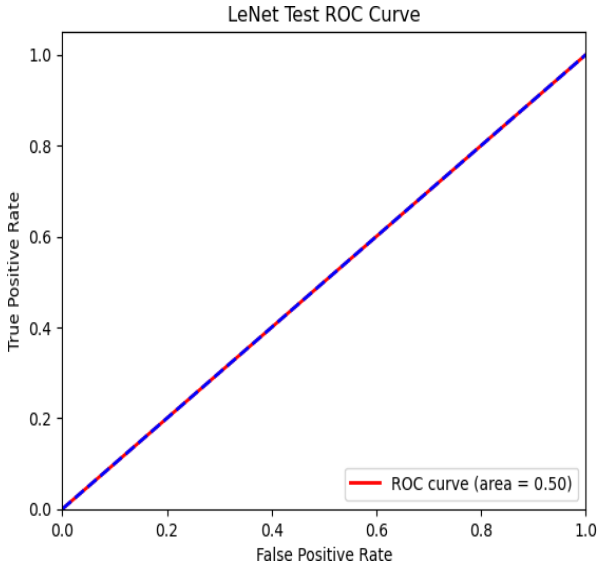


Fig 20.2: LeNet Test ROC Curve

Fig 20: LeNet Test Confusion Matrix and ROC Curve

AlexNet

Accuracy and Loss Curve

Accuracy curve and loss curve of AlexNet here shows how the model gets trained with epoch wise. With increase in epoch our accuracy should increase and loss should decrease. Number of epochs are 30. The curves are shown below. The AlexNet model is not able to learn properly.

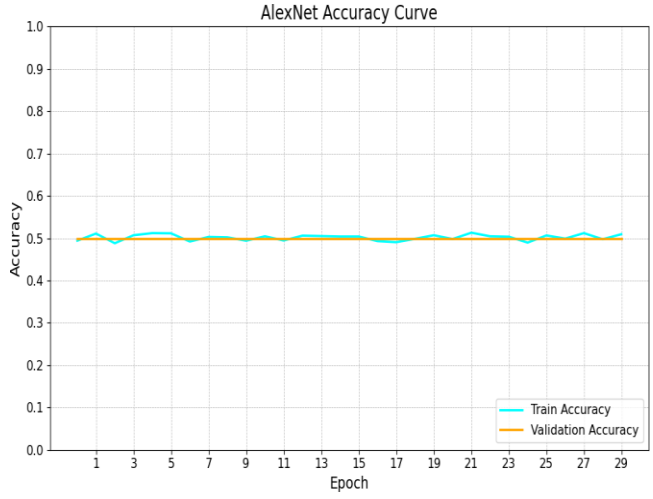


Fig 21: AlexNet Accuracy Curve

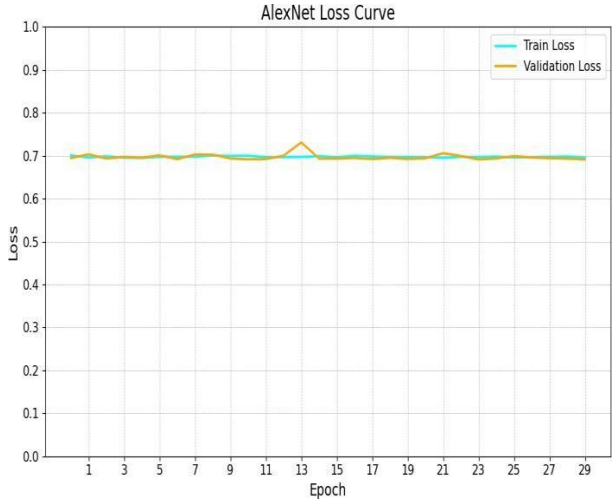


Fig 22: AlexNet Loss Curve

Train Dataset

The number of misclassifications shown in the confusion matrix are 2000 out of 4000 images. The ROC curve is shown with AUC score of 0.5

Fig 23.1 AlexNet Train Confusion Matrix

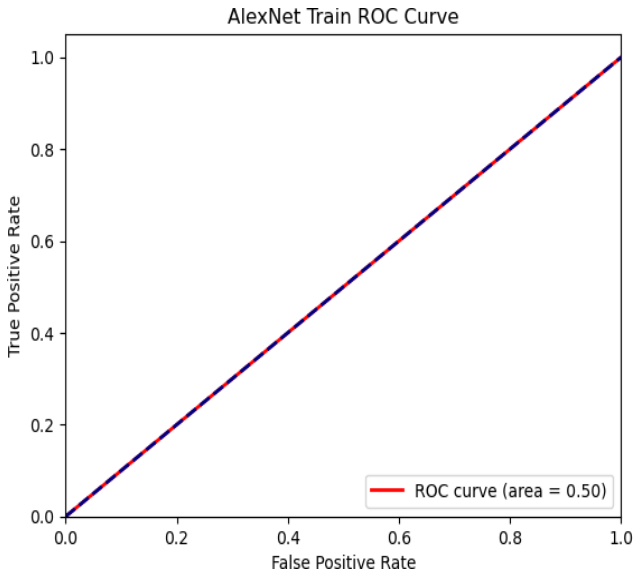


Fig 23.2 AlexNet Train ROC Curve

Fig 23 AlexNet Test Confusion Matrix and ROC Curve

Validation Dataset

The number of misclassifications shown in the confusion matrix are 225 out of 450 images. The ROC curve is shown with AUC score of 0.5

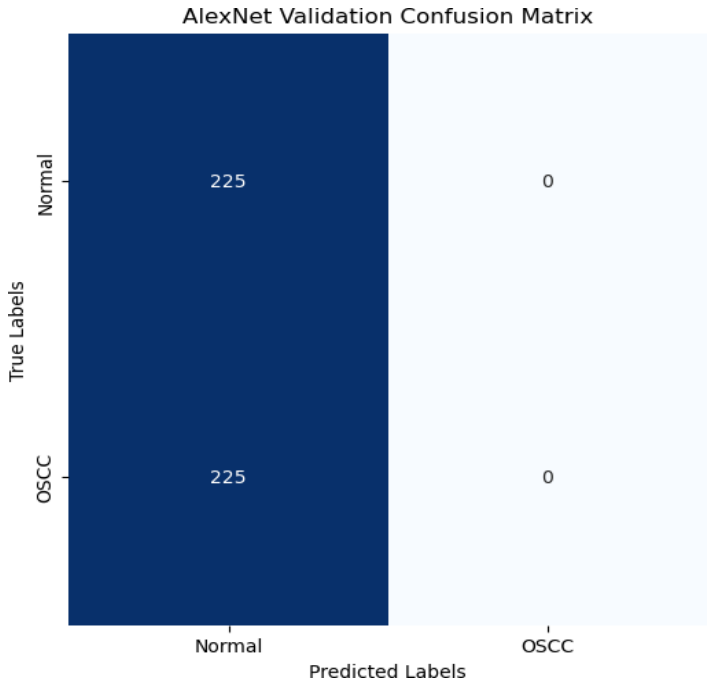


Fig 24.1 AlexNet Validation Confusion Matrix

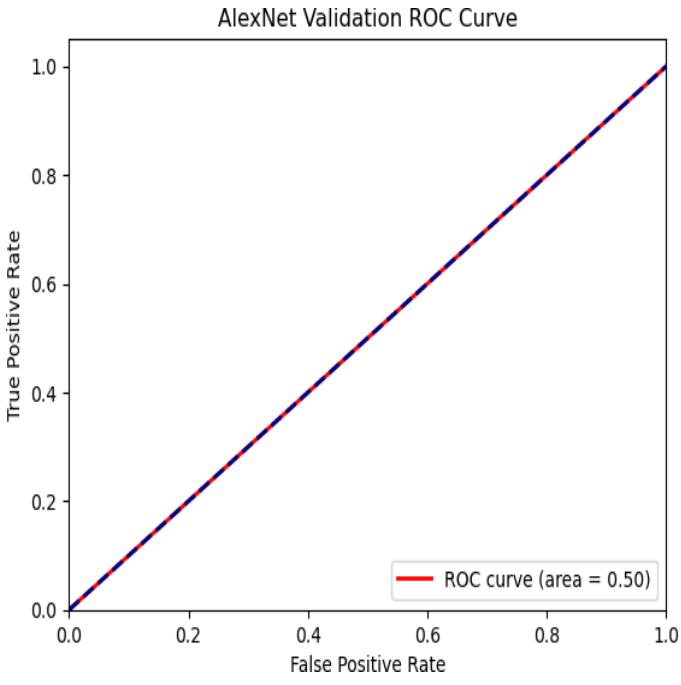


Fig 24.2 AlexNet Validation ROC Curve

Fig 24 AlexNet Validation Confusion Matrix and ROC Curve

Test Dataset

The number of misclassifications shown in the confusion matrix are 250 out of 500 images. The ROC curve is shown with AUC score of 0.5

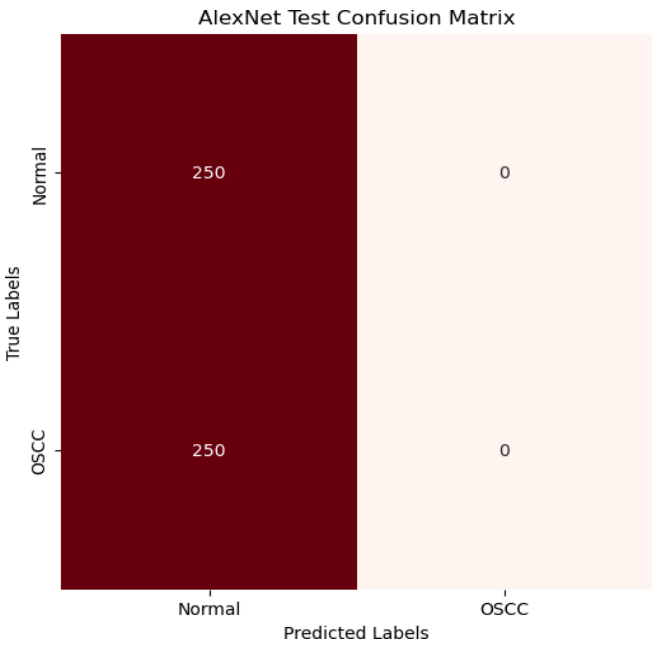


Fig 25.1 AlexNet Test Confusion Matrix

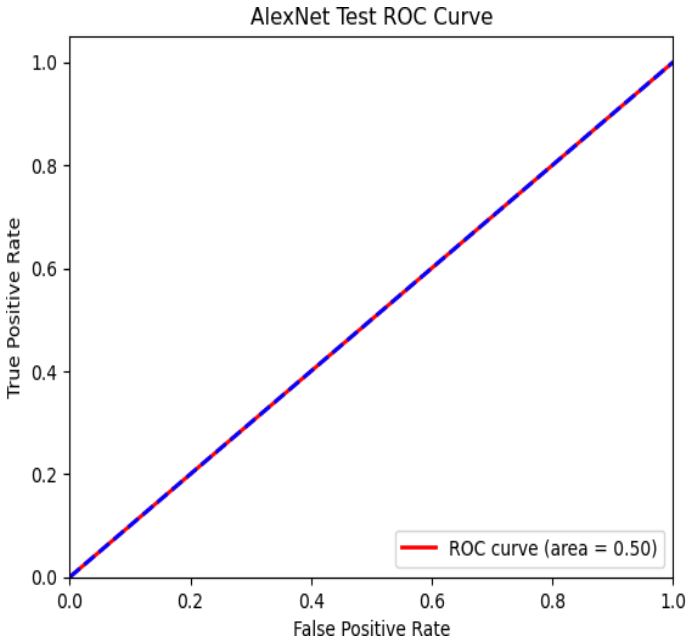


Fig 25.2 AlexNet Test ROC Curve

Fig 25 AlexNet Test Confusion Matrix and ROC Curve

5. Conclusions

Our research gives a detailed overview of analysis of histopathological oral cancer dataset. The histopathological image we have is of 100x magnification, and is fetched from 2 cancer hospitals. This study gives detailed overview of preprocessing histopathological image data, extracting features from it, transforming the image and do apply machine learning and CNN model to it. The initial data we have has 89 images of normal epithelium and 439 images of cancer infected cell. The dataset is imbalanced, and to add more histopathological is expensive. So, we do data augmentation to generate new data. In this data is balanced. Balancing of data eliminated the chances of model biasing. The data we have different colors because of different level of H&E staining, due to this our model can get confuse. So, in order to eliminate color variance problem, we implemented color normalization technique. Each pixel has value has 0-255. Training image with pixel value can take huge time and the original image is of size 2048*1536 pixels. We have to resize the image also. In order to remove the model overfitting problem and slow training we used Image resizing and Pixel normalization. Feature extraction is an important step for extracting features from the data. Extracting important features from the images can help building better prediction from model. Converted the preprocessed image we converted to LAB. LAB is a device independent color space. Edge Detection can help us detect objects in images. Now preserving the color help to preserve the important features in the images. Wavelet Transformation is the process of transforming images into series of wavelets. This can help in detecting more features from it. After that converting it into feature vectors. Machine learning model like Random Forest, KNN, SVM, Logistic regression is built. Random Forest is our best model which is showing 99.57% accuracy, 98.66% precision, 99.84% recall and F1 score of 98.66% F1 score. After that we have compared other models with it. As we are dealing with images so we use CNN for deep learning. We have built our own CNN model to predict the normal or cancer infected cell or not. CNN model is showing 98.80% accuracy, 97.65% precision, 100% recall, 98.81% F1 score. This experimental research can be a step forward for fast detecting of cancer, that can help in curing and defeating the cancer disease.

6. References

[1] Bakare, Y. B. (2021). Histopathological image analysis for oral cancer classification by support vector machine. *International journal of advances in signal and image sciences, 7*(2), 1-10.

[2] Brinker, T. J., Hekler, A., Utikal, J. S., Grabe, N., Schadendorf, D., Klode, J., ... & Von Kalle, C. (2018). Skin cancer classification using convolutional neural networks: systematic review. *Journal of medical Internet research, 20*(10), e11936.

[3] Fati, S. M., Senan, E. M., & Javed, Y. (2022). Early diagnosis of oral squamous cell carcinoma based on histopathological images using deep and hybrid learning approaches. *Diagnostics, 12*(8), 1899.

[4] Jiang, X., Hu, Z., Wang, S., & Zhang, Y. (2023). Deep learning for medical image-based cancer diagnosis. *Cancers, 15*(14), 3608.

[5] Jeyaraj, P. R., & Samuel Nadar, E. R. (2019). Computer-assisted medical image classification for early diagnosis of oral cancer employing deep learning algorithm. *Journal of cancer research and clinical oncology, 145,* 829-837.

[6] Kaur, C., & Garg, U. (2023). Artificial intelligence techniques for cancer detection in medical image processing: A review. *Materials Today: Proceedings, 81,* 806-809.

[7] Panigrahi, S., Nanda, B. S., Bhuyan, R., Kumar, K., Ghosh, S., & Swarnkar, T. (2023). Classifying histopathological images of oral squamous cell carcinoma using deep transfer learning. *Heliyon, 9*(3).

[8] Parida, P., & Bhoi, N. (2017). Wavelet based transition region extraction for image segmentation. *Future Computing and Informatics Journal, 2*(2), 65-78.

[9] Rajaguru, H., & Prabhakar, S. K. (2017). Performance comparison of oral cancer classification with Gaussian mixture measures and multi-layer Perceptron. In *The 16th International Conference on Biomedical Engineering: ICBME 2016, 7th to 10th December 2016, Singapore* (pp. 123-129). Springer Singapore.

[10] Rahman, T. Y., Mahanta, L. B., Das, A. K., & Sarma, J. D. (2020). Automateds oral squamous cell carcinoma identification using shape, texture and color features of whole image strips. *Tissue and Cell, 63,* 101322.

7.APPENDICES

APPENDIX A: Image Preprocessing

In this section we have described the image preprocessing done int the image dataset

I. **Data Augmentation:**
   We increase the number of images from both the folders in order to remove class imbalancing.

II. **Color normalization:**
   Color normalization is used to solve the color variance problem of our data.

III. **Image resizing:**
   The image we have is of huge size. Resizing the image will help us to train our model better and faster.

IV. **Pixel value normalization:**
   The model can take more time to train. So, pixel normalization is done in order to bring them from 0-255 pixels value to 0-1.

APPENDIX B: Feature Extraction

I. **Converting to LAB images**
   Converting our RGB images to LAB images can help us to detect important features from it.

II. **Canny edge detection**
   This can be useful in detecting the objects, the number of objects, the shape of image etc.

APPENDIX C: Image Transformation

I. **Wavelet transformation**
   Wavelet transformation of images is a data transformation technique that decomposes the data into different series of wavelet coefficients and can help in capturing important features from data.

II. **Feature vector conversion**

APPENDIX D: Model Building

Machine learning model building

I. **Random Forest**

It shows an accuracy of 98.57% accuracy, 98.66% precision, 99.84% recall and 98.66% F1 score

II. **SVM**

It shows an accuracy of 92.14% accuracy, 89.02% precision, 97.33% recall and 92.99% F1 score

III. **Logistic Regression**

It shows an accuracy of 78.57% accuracy, 80.82% precision, 78.66% recall and 79.72 % F1 score

IV. **KNN**

It shows an accuracy of 81.42% accuracy, 83.56% precision, 81.33% recall and 82.43% F1 score

Deep learning model building

V. **Custom CNN architecture**

It shows an accuracy of 98.80% accuracy, 97.65% precision, 100% recall and 98.61 % F1 score

VI. **LeNet 5**

It shows an accuracy of 50% accuracy, 50% precision, 100% recall and 66.66 % F1 score

VII. **AlexNet**

It shows an accuracy of 50% accuracy, 0% precision, 0% recall and 0% F1 score

8. Reflection of the team Members on the project

In this section we are going to discuss team works, their experiences, learning new thing and challenges;

8.1 Reflection of Ashutosh Mohapatra

Throughout the research I contributed on model building models for both Machine learning and Deep learning. I have built machine learning model like Random Forest, SVM, logistic regression and KNN. I have this model using different hyperparameters to find out the best hyperparameter. I have used Grid Search CV to find those hyperparameters. I have built architecture of CNN by

changing the number of neurons, number of layers, activation function and loss function. Add different regularizer, Dropout layers, different learning rates and optimizers. During this project I have faced many challenges like choosing the right value. How to train the model with different size of images and different number of images etc.

8.2 Reflection of Ayush Das Pattanaik

Throughout this research I have given the responsibility of Image transformation. Mainly the Wavelet transformation. I faced many challenges like finding the correct wavelet family and its sublets. How much layer of discrete wavelet transformation are going to use. Extracting different layers of image and doing wavelet transformation to extract valuable features. After doing wavelet transformation I have converted it to feature vector and send it for model building. During building of CNN architecture, I have also contributed in dealing with different CNN architecture. I have learned about wavelet transformation, how to use different wavelet transformation and building CNN architecture.

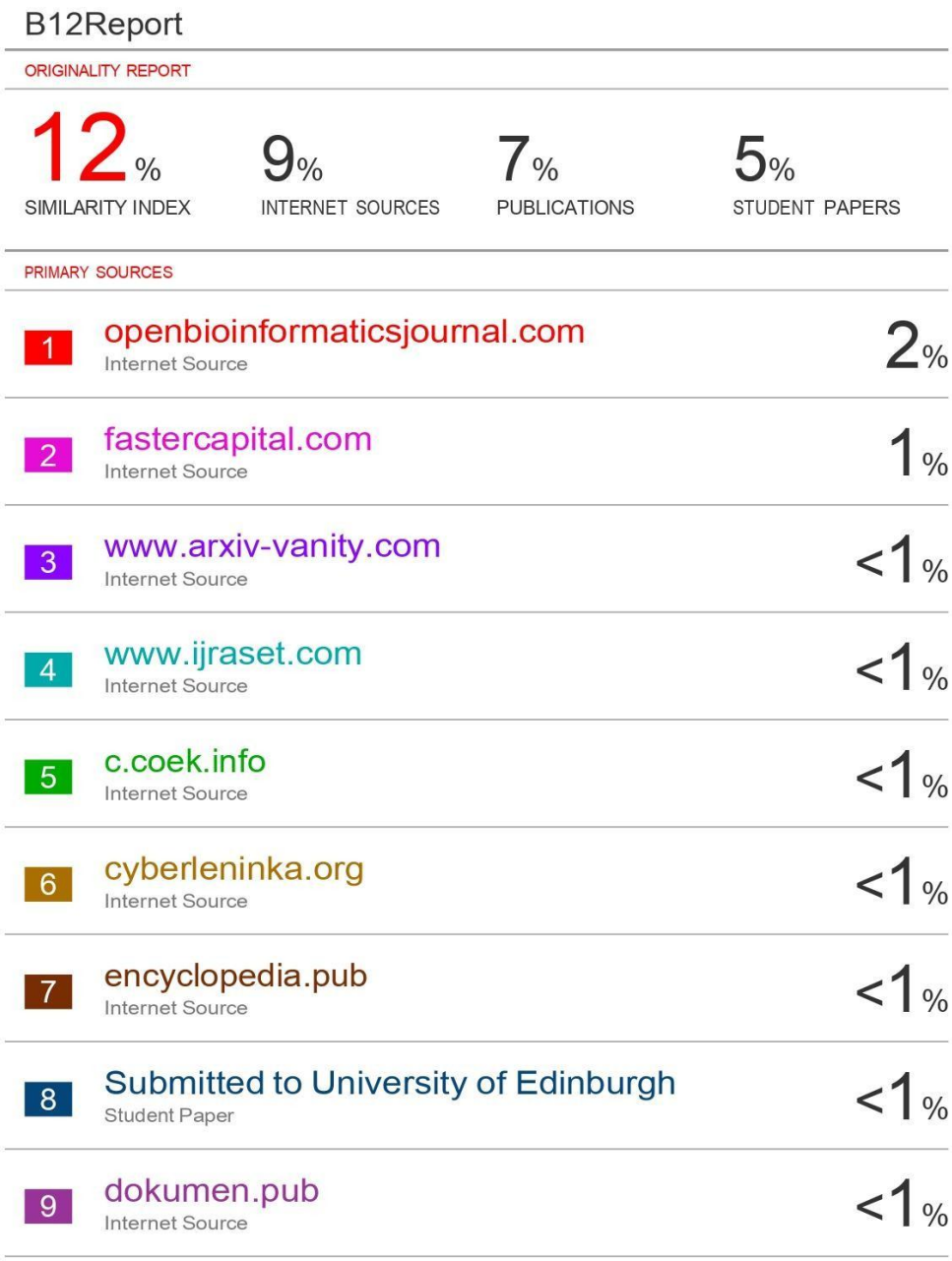
8.3 Reflection of Swaraj Das

I my part I have done Image preprocessing part. This was a repetitive process where I have to gone through many image preprocessing parts like image resizing, color normalization, pixel normalization, data augmentation, image denoising, contrast brightness, saturation adjustment etc. Finding the right preprocessing part is very crucial and was a very challenging task. Extracting features from histopathological images was a very challenging task and I have learned many preprocessing tasks.

8.4 Reflection of SK Abdul Rahman

I have done the feature extraction task for both machine learning and deep learning part. As input, I got the preprocessed images and my task was to extract features from it. Feature extraction working for machine learning and deep learning part it becomes a very crucial part. If we extract important features from, he images then our model can learn better. I task I have done is converting the preprocessed images into LAB images. Then detecting edges from the images can helped us detecting the tissue and cells there. Now preserving the color with edge detection can help in better feature extraction. I have done this many times to get valuable features for both machine learning and deep learning part.

9. SIMILARITY REPORT



B12Report

ORIGINALITY REPORT

12%	9%	7%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	openbioinformaticsjournal.com Internet Source	2%
2	fastercapital.com Internet Source	1%
3	www.arxiv-vanity.com Internet Source	<1%
4	www.ijraset.com Internet Source	<1%
5	c.coek.info Internet Source	<1%
6	cyberleninka.org Internet Source	<1%
7	encyclopedia.pub Internet Source	<1%
8	Submitted to University of Edinburgh Student Paper	<1%
9	dokumen.pub Internet Source	<1%

10	M. Harivirat, D. Manisha, N. Shesha Sarathi, V. Kakulapati, Shaik Subhani. "Chapter 40 Comparative Analysis of Lung Sac Inflation", Springer Science and Business Media LLC, 2024 Publication	<1%
11	preview-biomarkerres.biomedcentral.com Internet Source	<1%
12	Submitted to The Robert Gordon University Student Paper	<1%
13	Nikhil Bhade, Ms Kishan, Anand Sankar, Sv Shruthi. "Latin Square Image Cipher for Medical Images", 2021 IEEE International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS), 2021 Publication	<1%
14	Swathi Prabhu, Keerthana Prasad, Thuong Hoang, Xuequan Lu, Sandhya I.. "Multi-organ squamous cell carcinoma classification using feature interpretation technique for explainability", Biocybernetics and Biomedical Engineering, 2024 Publication	<1%
15	deepai.org Internet Source	<1%
	ecampus.ius.edu.ba	

16	Internet Source	<1%
17	ojs.acad-pub.com Internet Source	<1%
18	www.iieta.org Internet Source	<1%
19	www.itssi-journal.com Internet Source	<1%
20	www.dtic.mil Internet Source	<1%
21	www.frontiersin.org Internet Source	<1%
22	www.researchgate.net Internet Source	<1%
23	www2.mdpi.com Internet Source	<1%
24	"Advances in Intelligent Control Systems and Computer Science", Springer Science and Business Media LLC, 2013 Publication	<1%
25	dspace.univ-ouargla.dz Internet Source	<1%
26	flore.unifi.it Internet Source	<1%

27	mdpi-res.com Internet Source	<1%
28	www.coursehero.com Internet Source	<1%
29	"Deep Learning for Targeted Treatments", Wiley, 2022 Publication	<1%
30	Fan, Qi, and Daqi Gao. "A fast BP networks with dynamic sample selection for handwritten recognition", Pattern Analysis and Applications, 2016. Publication	<1%
31	aranne5.bgu.ac.il Internet Source	<1%
32	www.springerprofessional.de Internet Source	<1%
33	"Hybrid Artificial Intelligent Systems", Springer Science and Business Media LLC, 2019 Publication	<1%
34	Chala Sembeta, Amansisa Embabo, Swapna Gangone, G. J. Bharat Kumar. "Chapter 66 Aspect Based Sentiment Analysis for Hotel Services in Afaan Oromo Text Using Deep Learning", Springer Science and Business Media LLC, 2024 Publication	<1%

35	Submitted to Heriot-Watt University Student Paper	<1%
36	Khaled Mohammed Elgamily, M. A. Mohamed, Ahmed Mohamed Abou-Taleb, Mohamed Maher Ata. "A Novel Pyramidal CNN Deep Structure for Multiple Objects Detection in Remote Sensing Images", Journal of the Indian Society of Remote Sensing, 2023 Publication	<1%
37	Submitted to Sunway Education Group Student Paper	<1%
38	Zhen Zuo, Bing Shuai, Gang Wang, Xiao Liu, Xingxing Wang, Bing Wang, Yushi Chen. "Learning Contextual Dependence With Convolutional Hierarchical Recurrent Neural Networks", IEEE Transactions on Image Processing, 2016 Publication	<1%
39	discovery.researcher.life Internet Source	<1%
40	effiloop.com Internet Source	<1%
41	journalofbigdata.springeropen.com Internet Source	<1%
42	www.analyticsvidhya.com Internet Source	<1%

43	www.irjmets.com Internet Source	<1%
44	www.mdpi.com Internet Source	<1%
45	www.wseas.us Internet Source	<1%
46	Godfrey Perfectson Oise, Susan Konyeha. "Harnessing Deep Learning for Sustainable E-Waste Management and Environmental Health Protection", Research Square Platform LLC, 2024 Publication	<1%
47	Santisudha Panigrahi, Bhabani Sankar Nanda, Ruchi Bhuyan, Kundan Kumar, Susmita Ghosh, Tripti Swarnkar. "Classifying Histopathological Images of Oral Squamous Cell Carcinoma using Deep Transfer Learning", Heliyon, 2023 Publication	<1%
48	Barun Barua, Kangkana Bora, Anup Kr.Das, Gazi N. Ahmed, Tashnin Rahman. "Stain color translation of multi-domain OSCC histopathology images using attention gated cGAN", Computerized Medical Imaging and Graphics, 2023 Publication	<1%

49	Madhusmita Das, Rasmita Dash, Sambit Kumar Mishra. "Automatic Detection of Oral Squamous Cell Carcinoma from Histopathological Images of Oral Mucosa Using Deep Convolutional Neural Network", International Journal of Environmental Research and Public Health, 2023 Publication	<1%
50	Santisudha Panigrahi, Tripti Swarnkar. "Machine Learning Techniques used for the Histopathological Image Analysis of Oral Cancer-A Review", The Open Bioinformatics Journal, 2020 Publication	<1%
51	Tabassum Yesmin Rahman, Lipi B. Mahanta, Anup K. Das, Jagannath D. Sarma. "Automated oral squamous cell carcinoma identification using shape, texture and color features of whole image strips", Tissue and Cell, 2020 Publication	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		