HISTOPATHOLOGY ORAL CANCER DETECTION

INT 625 Internship 3

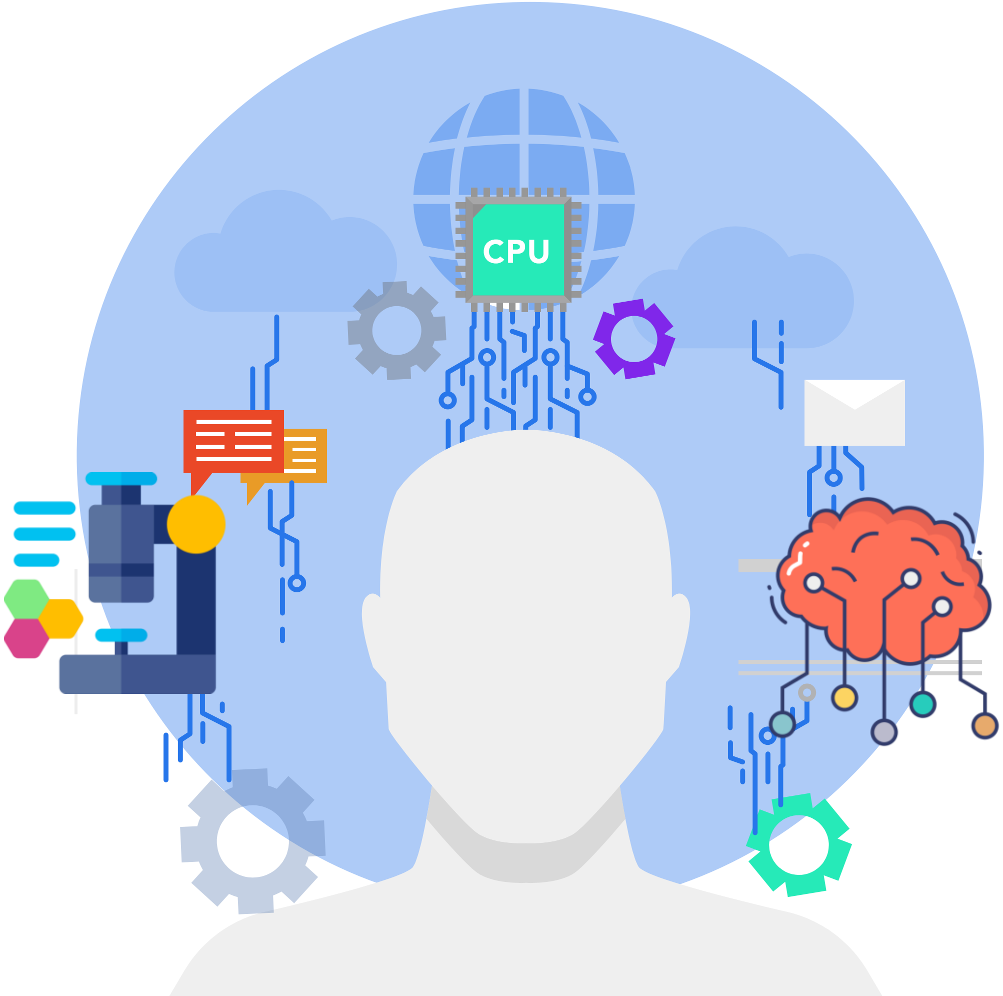
Final Review

Team Members

1. Sree lakshmi A - E7321005
2. Prem Kumar K S - E7321013
3. Divya D - E7321010
4. Pavithran V - E7321014

Project Guide
Dr. S. Pitchumani Angayarkanni

SRI RAMACHANDRA
INSTITUTE OF HIGHER EDUCATION AND RESEARCH
(Category - I Deemed to be University) Porur, Chennai
SRI RAMACHANDRA FACULTY OF ENGINEERING AND TECHNOLOGY



HISTOPATHOLOGY ORAL CANCER DETECTION

Introduction

Histopathology is a field of medicine that involves the microscopic examination of tissue samples to diagnose diseases. Oral cancer is a type of cancer that affects the tissues of the mouth and throat, and histopathology is a critical tool in its detection and diagnosis. By analyzing tissue samples, histopathologists can identify the presence of cancerous cells, determine the stage of the disease, and guide treatment decisions.

However, histopathology can help identify the presence of abnormal or cancerous cells in tissue samples taken from the mouth.

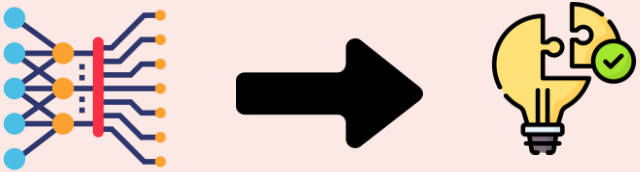
By using deep learning algorithms, histopathologists can automate the analysis of large, complex tissue samples and accurately identify

Problem statement/Objectives

How Deep Learning can help Oral cancer Detection?

Oral cancer is a significant health problem worldwide, and it is the sixth most common cancer. Early detection of oral cancer can significantly improve the chances of successful treatment and patient survival.

The problem statement for histopathology oral cancer detection using deep learning is to develop a deep learning-based model that can accurately and reliably detect and classify oral cancer in histopathology images. The model should be able to differentiate oral cancer or Not with a large dataset of histopathology images.



Review of Literature

SNO	PUBLICATIONS	DATASET TYPE	METHODS	ACCURACY	FINDINGS
[1]	Shams w.k.et al.[2016]	Gene Expression	DNN	86%	Slide 6
[2]	H.Alkhadar [2020]	Clinical and histopathological	Decision Tree	76%	Slide 6
[3]	V.Shavlokhova [2021]	Histopathological	CNN	77.89%	Slide 6
[4]	H.Aberville [2017]	Laserendomicroscopy	Deep Learning	80.01%	Slide 6

1. Investigated using machine learning techniques with gene expression profiling to predict the possibility of oral cancer development in OPL patients.

1. Machine learning analyses of cancer outcomes for oral cancer remain sparse compared to other types of cancer like breast or lung. The purpose of the present study was to compare the performance of machine learning algorithms in the prediction of global, recurrence-free five-year survival in oral cancer patients based on clinical and histopathological data.

1. In this preliminary work, we trained a CNN model on a limited number of ex vivo FCM images and obtained promising results in the automated classification of cancerous tissue. Further studies using large sample, sizes are warranted to introduce this technology into clinics.

1. Evaluated a novel automatic approach for OSCC diagnosis using deep learning technologies on CLE images. The method is compared against textural feature-based machine learning approaches that represent the current state of the art.

REFERENCES:

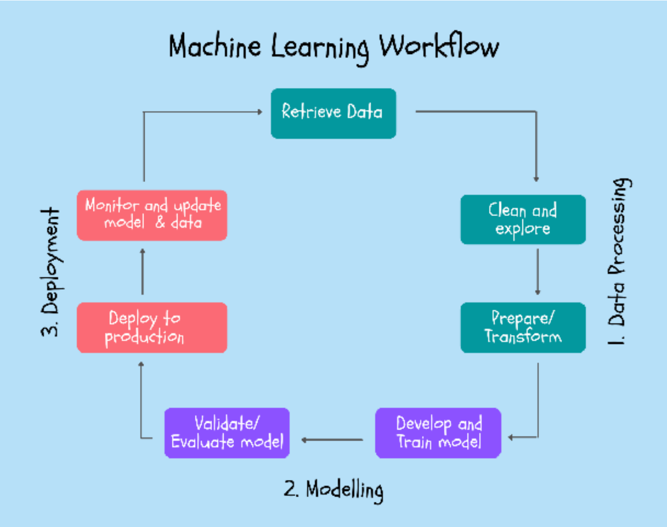
1. Shams WK, Htike ZZ. Oral cancer prediction using gene expression profiling and machine learning. Int J Appl Eng Res 2017; 12: 4893-8.

1. Alkhadar H., Macluskey M., White S., Ellis I., Gardner A. Comparison of machine learning algorithms for the prediction of five-year survival in oral squamous cell carcinoma. J. Oral Pathol. Med. 2021;50:378–384. doi: 10.1111/jop.13135. [PubMed] [CrossRef] [Google Scholar]

1. Shavlokhova V., Sandhu S., Flechtenmacher C., Koveshazi I., Neumeier F., Padrón-Laso V., Jonke Ž., Saravi B., Vollmer M., Vollmer A., et al. Deep Learning on Oral Squamous Cell Carcinoma Ex Vivo Fluorescent Confocal Microscopy Data: A Feasibility Study. J. Clin. Med. 2021;10:5326. doi: 10.3390/jcm10225326. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

1. Aubreville M., Knipfer C., Oetter N., Jaremenko C., Rodner E., Denzler J., Bohr C., Neumann H., Stelzle F., Maier A. Automatic Classification of Cancerous Tissue in Laserendomicroscopy Images of the Oral Cavity using Deep Learning. Sci. Rep. 2017;7:11979. doi: 10.1038/s41598-017-12320-8. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

Work Flow



1 Data preparation

Collection of dataset for the model

2 Data pre-processing

Preparing the image data

- tain ,test,val
- cropping
- augmenting

3 Model selection

deep learning architecture

- Vgg 16
- Renet 50

4 Train the model

model learns to recognize patterns in the data

5 Hyperparameter tuning

such as the learning rate or number of layers, to improve its performance

6 Deploy the model

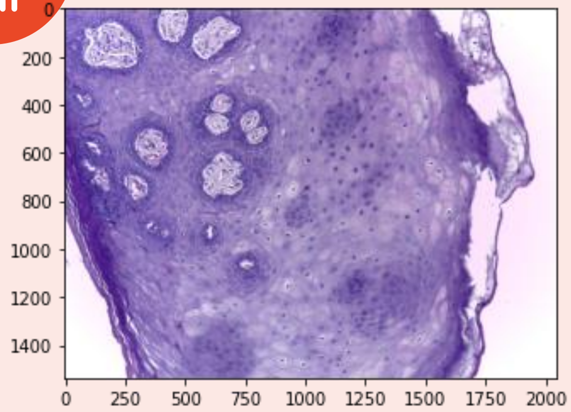
After testing the model deploying it for the Real-time access

Data preparation

- Download and extract the dataset files from Kaggle.
- Perform data cleaning to remove any corrupted or invalid data.
- Normalize the image pixel values to a common scale to improve model training and reduce bias.
- Augment the dataset by applying transformations to the images such as rotation, flipping, or zooming to increase the dataset's size and improve model robustness.
- Annotate the images to identify regions of interest or label the images according to the presence or absence of cancer. This annotation process can be done manually or through automated tools.

observed

- Dataset Size - 2.97 GB
- Contains 3 main folders train test val
- Normal and OSCC
- Dataset contains 5192 JPG files.



OSCC - Oral squamous cell carcinoma

Data pre processing

- Split the dataset into training, validation, and testing subsets to evaluate the performance of the machine learning models accurately.

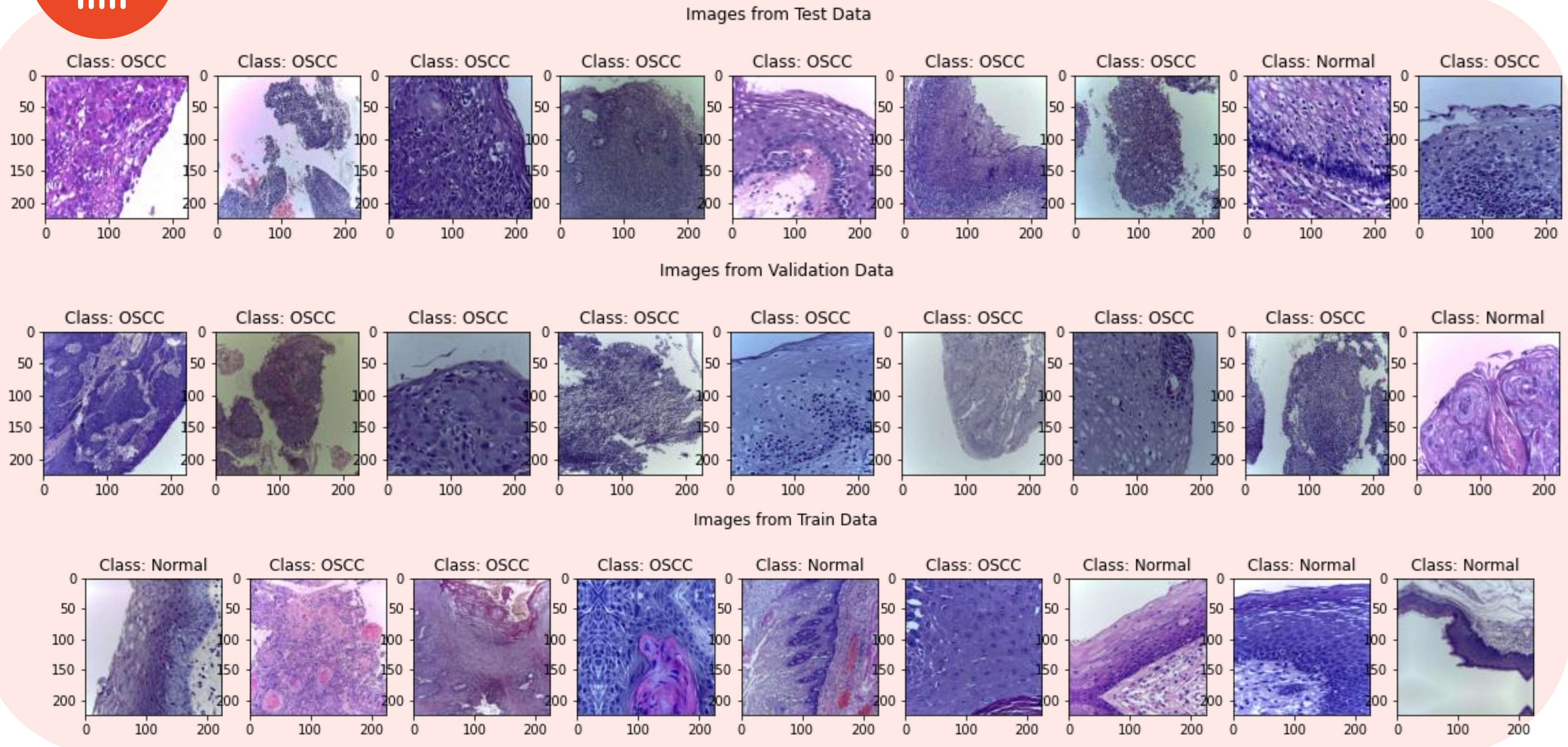
- Verify that the dataset is balanced and representative of the target population to reduce potential bias and improve model generalization.

  - Image normalization
  - Image augmentation

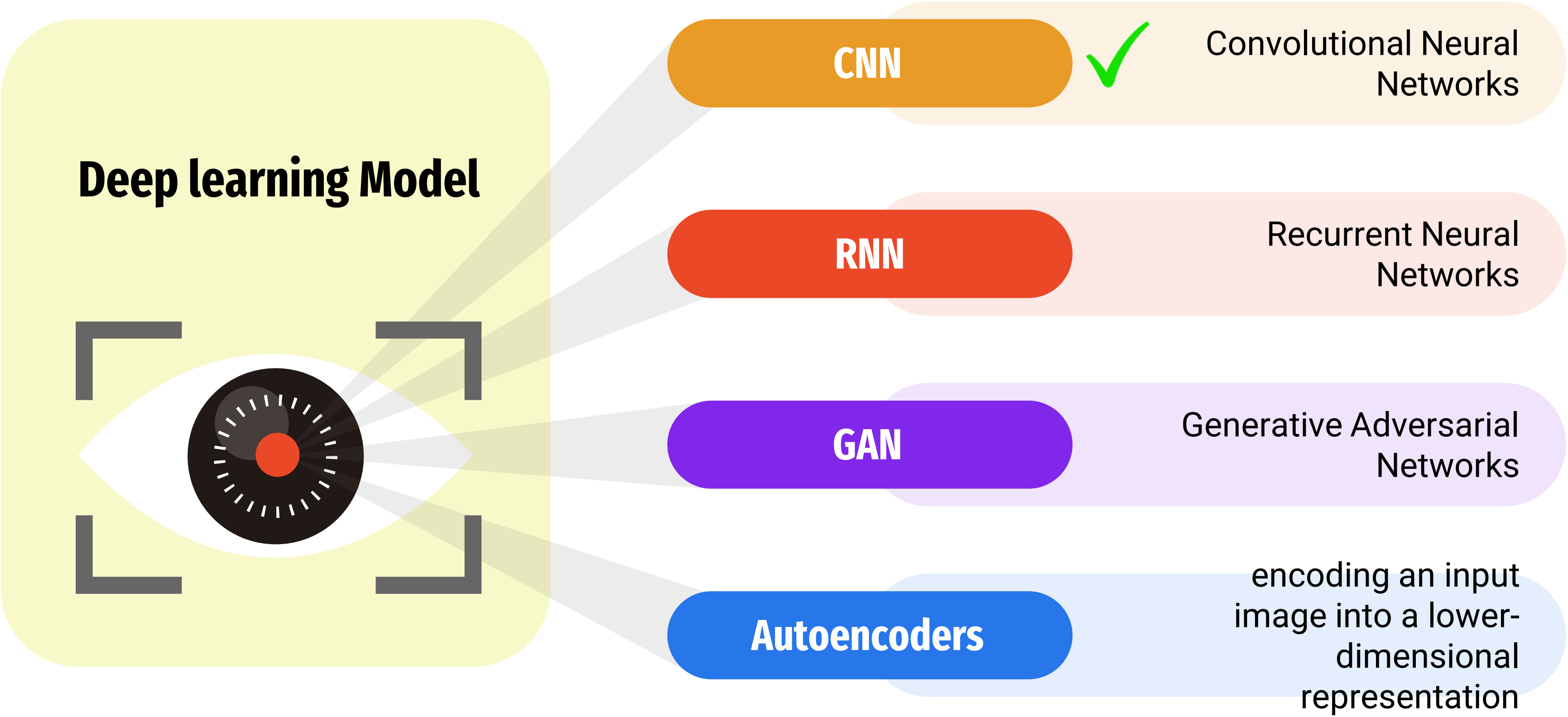
```
Found 4946 images belonging to 2 classes.
Found 120 images belonging to 2 classes.
Found 126 images belonging to 2 classes.
```

Image datagenerator

Images in Datagenerator



Model Selection



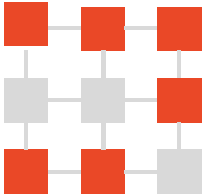
CNN Model

01

Traditional CNN model

Trained from the Scratch

02



Pre-trained CNN model ✓

already been trained on a large dataset

Pre-trained CNN model

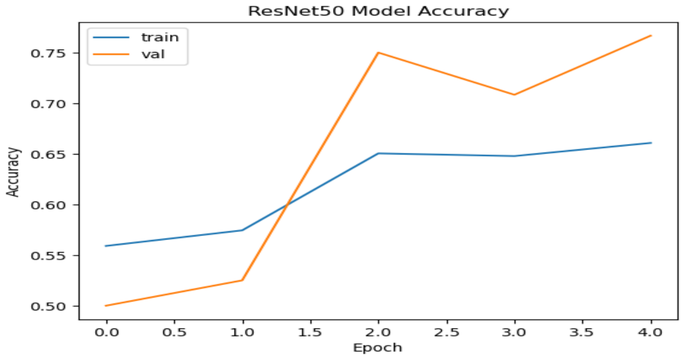
VGG16 and VGG19: These models were developed by the Visual Geometry Group at the University of Oxford and are named after the number of layers they have (16 and 19, respectively). They consist of a series of convolutional and pooling layers followed by a few fully connected layers.

ResNet: ResNet stands for "Residual Network" and was developed by Microsoft Research. It uses a residual architecture that allows for very deep networks without causing the vanishing gradient problem.

InceptionNet/GoogLeNet: InceptionNet, also known as GoogLeNet, was developed by Google and won the 2014 ImageNet competition. It uses "inception modules" that allow for parallel processing of the input data.

MobileNet: MobileNet was developed by Google and is designed for mobile and embedded devices. It uses depth-wise separable convolutions to reduce the number of parameters and computations required.

ResNet 50

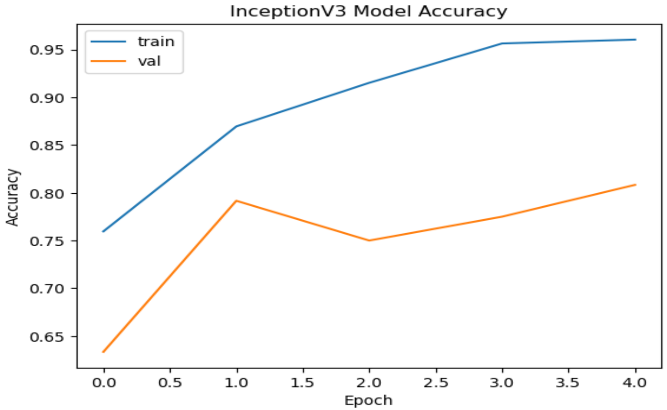


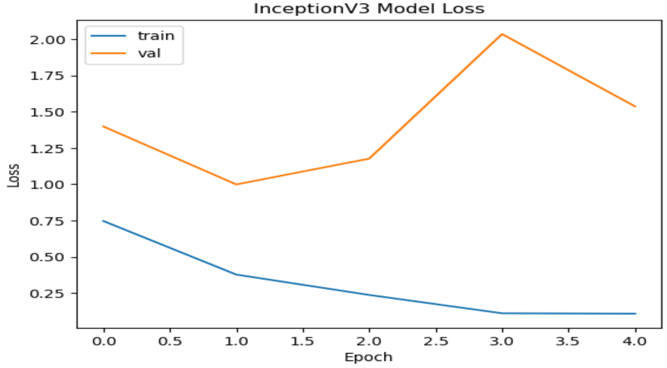
ResNet50 Test Loss: 0.6760708689689636

ResNet50 Test Accuracy: 0.7460317611694336

Here we can observe that , for ResNet50 we got accuracy as 74% and 67% test loss

INCEPTION V3



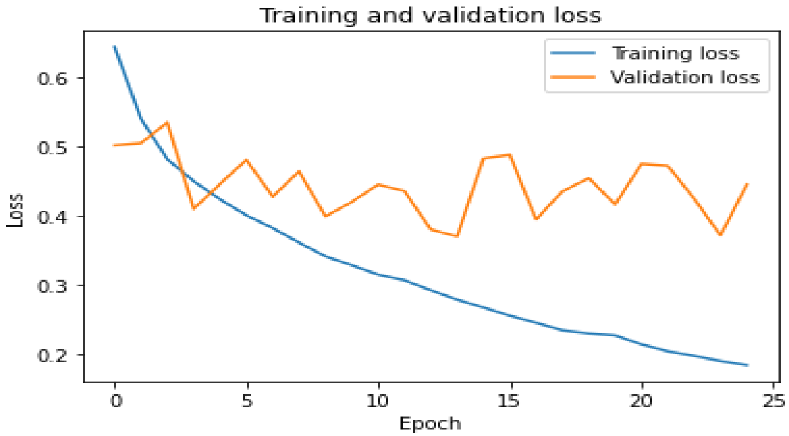


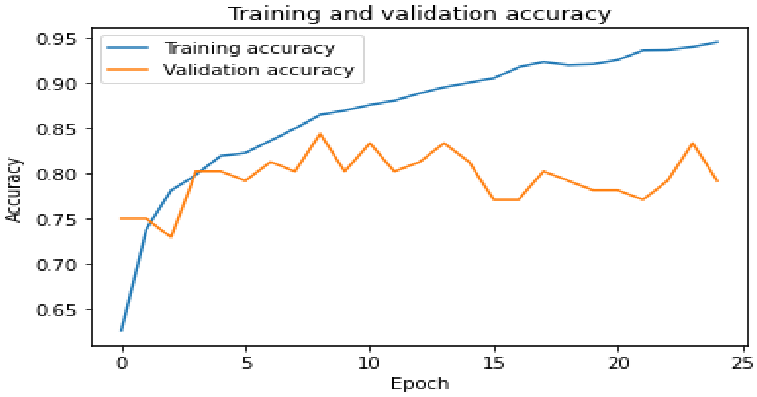
InceptionV3 Test Loss: 1.0103096961975098

InceptionV3 Test Accuracy: 0.8095238208770752

Here we can observe that , for InceptionV3 we got accuracy as 80%.

VGG 16

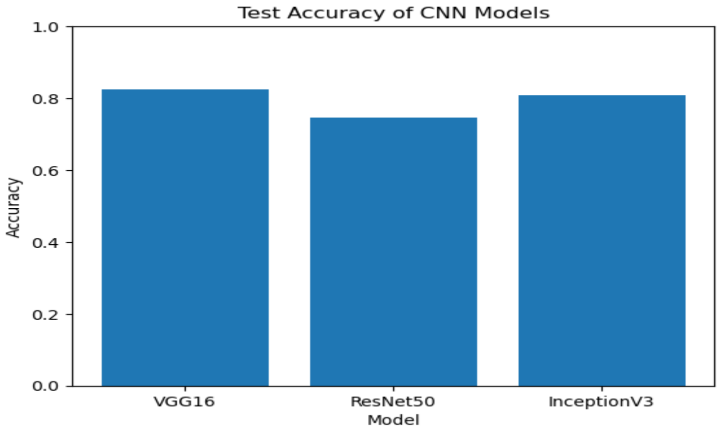


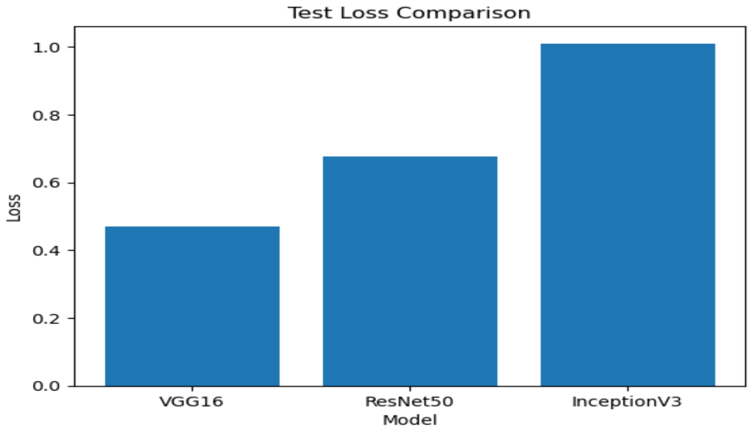


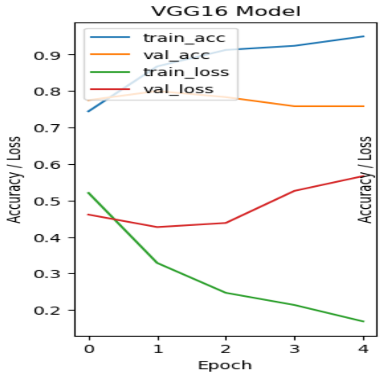

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 vgg16 (Functional)          (None, 7, 7, 512)         14714688

 flatten (Flatten)           (None, 25088)             0

 dense (Dense)               (None, 256)               6422784

 dropout (Dropout)           (None, 256)               0

 dense_1 (Dense)             (None, 1)                 257

=================================================================
Total params: 21,137,729
Trainable params: 6,423,041
Non-trainable params: 14,714,688
_________________________________________________________________
```

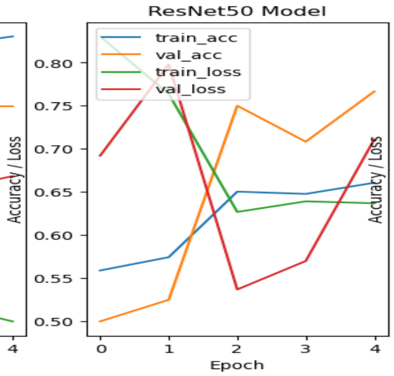
```
VGG16 Test Loss: 0.4692530035972595
VGG16 Test Accuracy: 0.8253968358039856
```

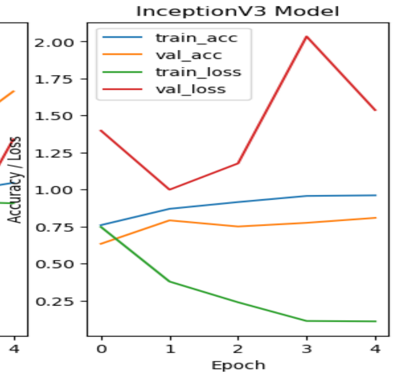
MODEL COMPARISON





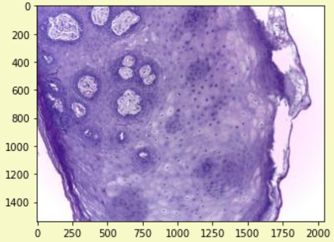


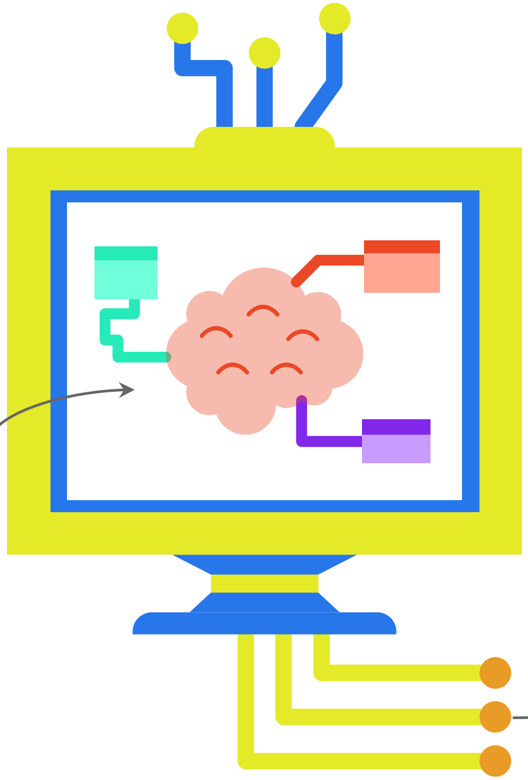




Machine Learning Infographics

Inputs





Outputs

OSCC

OSCC

CONCLUSION

Deep learning-based histopathology analysis has shown promising results in the detection of oral cancer. By leveraging the power of deep neural networks, researchers have been able to develop robust and accurate models for analysing histopathological images and identifying cancerous cells or tissues.

While comparing VGG16,ResNet50 and InceptionV3 it is observed that VGG16 pre-trained model performed well with 83% accuracy. The ultimate objective in Oral cancer detection is to develop a high accuracy model that is effective and affordable.

FUTURE ENHANCEMENT

In the future work, we will continue to propose more efficient and rapid methods for Oral cancer recognition. The target is to realize multi-class recognition of oral cancer based on the research of benign and malignant tumour recognition. In addition to improving the recognition accuracy, we also hope to extract more effective information about cancer, which can help doctors find the lesion faster and reduce the workload on doctors.

WORKLOG

Date	Duration in Hrs.	Task Planned	Task Completed	Description of the Task	Scope of the task	Skills Acquired
01.03.2023 - 02.03.2023	12 hours	Abstract	Completed	Referring to the concept about the given dataset	Abstract done	Understanding the dataset
04.03.2023 - 09.03.2023	30 hours	Exploring the dataset	Completed	Exploring the dataset in different platforms.	Dataset Found	Anayzing the dataset
11.03.2023 - 15.03.2023	36 hours	Learning about oral cancer histopathology	Completed	Studying about histopathology, different types of level and more	Got some knowledge about histopathology	Learning About the histopathology
27.03.2023	6 hours	PPT - 1	Completed	Creating a ppt based on the completed task Literature review	Finalizing a better ppt	Power point tools
28.03.2022 - 30.03.2023	30 hours	Data Preparation	Completed	Collection of dataset for the model	Properly oragnised the Dataset	using image datagenerataor
04.04.2023	6 hours	Model Selection	Completed	selecting model for image dataset	selecting the CNN model	deep learning python
06.04.2023 - 08.04.2023	18 hours	Model training	Completed	Training the selected model	To Analyze the Model	Model training python deep learning architecture
10.04.2023 - 13.04.2023	14 hours	Hyperparameter tuning	Completed	optimizing the hyperparameters of the model	To improve Model performance.	Optimization techniques

THANK YOU