**Endometrial Cancer Detection Using Convolutional Neural Networks**

**DESIGN PROJECT – 3 REPORT**

***Submitted by***

**P. SAI MANOJ (19113101)**

**R. NAVEEN (19113102)**

**K. SUMANTH KUMAR REDDY (19113105)**

**Under the guidance of**

**Dr. SATHYA PRIYA S**

**Associate Professor**

***in partial fulfillment for the award of the degree of***

**BACHELOR OF TECHNOLOGY**

**in**

**COMPUTER SCIENCE AND ENGINEERING**

A drawing of a face

Description automatically generated

**HINDUSTAN INSTITUTE OF TECHNOLOGY AND SCIENCE**

**CHENNAI - 603 103**

NOVEMBER 2022

A drawing of a face

Description automatically generated

**BONAFIDE CERTIFICATE**

Certified that this project report **“ENDOMETRIAL CANCER DETECTION USING CONVOLUTIONAL NEURAL NETWORKS”** is the bonafide work of **“P. SAI MANOJ (19113101), R. Naveen (19113102) and, K. SUMANTH KUMAR REDDY (19113105)”** who carried out the project work under my supervision during the academic year **2022-2023**.

**Dr. SATHYA PRIYA S**,

**SUPERVISOR**

ASSOCIATE PROFESSOR

Department of CSE.

**INTERNAL EXAMINER EXTERNAL EXAMINER**

Name: Designation:

Name:

Designation:

Project Viva - voce conducted on

**CHAPTER**

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **NO.** |  | |
|  | **Acknowledgement** | **vi** |
|  | **Abstract** | **vii** |
|  | **List of Tables**  **List of Figures** | **viii**  **ix** |
|  | **List of Abbreviations** | **x** |
| **1** | **INTRODUCTION** | **1-5** |
|  | **1.1** Overview | **1** |
|  | **1.2** Motivation for the project | **2** |
|  | **1.3** Problem Definition and Scenarios | **2** |
|  | **1.4** Organization of the report | **4** |
|  | **1.5** Summary | **5** |
| **2** | **LITERATURE REVIEW** | **6-10** |
|  | **2.1** Machine Learning for Endometrial Cancer Prediction and Prognostication | **6** |
|  | **2.2** A New Classification of Benign, Premalignant, and Malignant Endometrial Tissues Using Machine Learning Applied to 1413 Candidate Variables | **7** |
|  | **2.3** Computer-Aided Diagnosis in Histopathological Images of the Endometrium Using a Convolutional Neural Network and Attention Mechanisms | **7** |
|  | **2.4** Predicting endometrial cancer subtypes and molecular features from histopathology images using multi-resolution deep learning models | **8** |
|  | **2.5** The efficacy of deep learning models in the diagnosis of endometrial cancer using MRI: a comparison with radiologists | **9** |
|  |  |  |

**TITLE PAGE NO.**

1. **PROJECT DESCRIPTION 11-14**
   1. Objective of the Project work **11**
   2. Existing System **11**

* 1. Shortcomings of Existing System **12**
  2. Proposed System **13**
  3. Benefits of Proposed System **14**

1. SYSTEM DESIGN 15-18
   1. Architecture Diagram **15**
2. PROJECT REQUIREMENTS 19
   1. Hardware and Software Specification **19**
   2. Technologies Used **19**
3. MODULE DESCRIPTION 20-23
   1. Modules **20**
   2. Collection of samples **20**
   3. Image processing for model training **20**
   4. Evaluation Methods **21**
   5. VGG16 model **21**
   6. Evaluation Metrics **22**
4. IMPLEMENTATION 24-27
   1. [User Interface](#_TOC_250006) **24**
5. RESULT ANALYSIS 28-30
   1. Results obtained **28**
6. CONCLUSION 31
   1. [Conclusion](#_TOC_250003) **31**
   2. [Future Work](#_TOC_250002) **31**

**32-33**

1. INDIVIDUAL TEAM MEMBER’s

REPORT

* 1. [Individual Objective](#_TOC_250001) **32**
  2. Role of the Team Members **32**
  3. Contribution of Team Members **33**

[REFERENCES](#_TOC_250000)

**APPENDIX A: SAMPLE CODE**

**APPENDIX B: PLAGIARISM REPORT**

**APPENDIX C: TEAM DETAILS**

**ACKNOWLEDGEMENT**

First and foremost we would like to thank **ALMIGHTY** who has provided us the strength to do justice to our work and contribute our best to it.

We wish to express our deep sense of gratitude from the bottom of our heart to our guide **Dr Sathya Priya S, Associate Professor**, **Computer Science and Engineering,** for her motivating discussions, overwhelming suggestions, ingenious encouragement, invaluable supervision, and exemplary guidance throughout this project work.

We would like to extend our heartfelt gratitude to **Dr. J.Thangakumar, Ph.D., Head, Department of Computer Science and Engineering** for her valuable suggestions and support in successfully completing the project.

We thank the management of **HINDUSTAN INSTITUTE OF TECHNOLOGY AND SCIENCE** for providing us the necessary facilities and support required for the successful completion of the project.

As a final word, we would like to thank each and every individual who have been a source of support and encouragement and helped us to achieve our goal and complete our project work successfully.

**ABSTRACT**

One type of cancer that starts in the uterus is endometrial cancer. The pear-shaped, hollow pelvic organ known as the uterus is where foetal development takes place. The layer of cells that makes up the uterine lining (endometrium) is where endometrial cancer first manifests itself. Uterine cancer is another name for endometrial cancer.

Symptoms of endometrial cancer include over 90% of women with endometrial cancer experience abnormal vaginal bleeding. Bleeding or discharge not connected to your periods (menstruation). Bleeding after menopause urinating is difficult or painful. Pain during sexual activity. After menopause, endometrial cancer most frequently develops. Obesity. Obesity raises your chance of developing endometrial cancer. This might happen as a result of your body's hormone balance being altered by extra body fat. The most accurate method for identifying endometrial cancer is histological image analysis, which can have a major impact on a woman's reproductive system.

Current Computer-Aided Diagnosis (CADx) approaches utilizing conventional machine learning algorithms frequently failed to produce satisfactory results because of the limited capacity to model the complex links between histopathological pictures and their interpretations. In general, the treatment outcome will be good if endometrial lesions can be identified early utilizing widely used clinical screening and detection procedures, such as transvaginal ultrasonography, hysteroscopy, and hysterosalpingography. In this paper, created the Visual Geometry Group-16 (VGG-16) technique, which is based on an attention mechanism and a Convolutional Neural Network (CNN). VGG-16 is also called OxfordNet.

|  |  |  |
| --- | --- | --- |
|  | **LIST OF TABLES** |  |
| **FIGURE NO.** | **TITLE** | **PAGE NO.** |
| 1.1 | FIGO Cancer Staging System | 3 |
|  |  |  |

|  |  |  |
| --- | --- | --- |
|  | **LIST OF FIGURES** |  |
| **FIGURE NO.** | **TITLE** | **PAGE NO.** |
| 4.1 | System Architecture | 15 |
| 4.2 | Endometrioid Adenocarcinoma | 16 |
| 4.3 | Endometrial Hyperplasia-Complex | 16 |
| 4.4 | Endometrial Hyperplasia-Simplex | 16 |
| 4.5 | Endometrial Polyp | 16 |
| 4.6 | Follicular Phase | 17 |
| 4.7 | Luteal Phase | 17 |
| 4.8 | Menstrual Phase | 17 |
| 7.1 | ipynb notebook | 25 |
| 7.2 | Cross Validation for Epochs | 26 |
| 7.3 | Iteration no | 27 |
| 8.1 | Evaluation Metrics | 2 |
| 8.2 | Confusion Matrix | 2 |
| 8.3 | ROC Curve | 2 |
|  |  |  |

# LIST OF ABBREVIATIONS

|  |  |
| --- | --- |
| ADC | Apparent Diffusion Coefficient |
| AI | Artificial Intelligence |
| CAD | Computer-Aided Diagnosis |
| CET1-WI | Common Equity Tier 1 capital - Weighted Imaging |
| CNN | Convolutional Neural Network |
| EA | Endometrioid Adenocarcinoma |
| EH | Endometrial Hyperplasia |
| EIN | Endometrial Intraepithelial Neoplasia |
| EP | Endometrial Polyp |
| FIGO | International Federation of Gynecology and Obstetrics |
| GPU | Graphics Processing Unit |
| H&E | Hematoxylin and Eosin |
| IP | Image Processing |
| ML | Machine Learning |
| MRI | Magnetic Resonance Imaging |
| NE | Normal Endometrium |
| ROC | Receiver Operating Characteristic |
| TF | Tensor Flow |
| TPU | Tensor Processing Unit |
| VGG | Visual Geometry Group |
| WSI | Whole Slide Imaging |

**CHAPTER 1 INTRODUCTION**

## 1.1 Overview

One type of cancer that starts in the uterus is endometrial cancer. The pear-shaped, hollow pelvic organ known as the uterus is where fetal development takes place. The layer of cells that makes up the uterine lining (endometrium) is where endometrial cancer first manifests itself. Uterine cancer is another name for endometrial cancer. Uterine sarcoma is one of the other cancers that can develop in the uterus, however it is considerably less prevalent than endometrial cancer.

With the improvement of computer-based intelligence, profound learning with profound convolutional brain organizations has been demonstrated to be a strong calculation for progressing biomedical picture investigation. CNNs were effectively used to analyze breast cancer histological pictures, performing as well as a team of 11 histopathologists. CNNs and other deep networks have enabled unprecedented breakthroughs in a variety of computer vision tasks, from image classification to object detection, semantic segmentation, image captioning, and more recently. Numerous more reports of deep learning-based endoscopic CAD systems have shown good outcomes in cystoscopy, gastroscopy, enteroscopy, and colposcopy. Digital histopathological picture recognition is a completely appropriate problem for machine gaining knowledge of since the photos themselves comprise data enough for diagnosis. Deep learning has gathered much intrigued within the medical field since deep-learning methods are especially reasonable for image analysis.

**1.2 Motivation**

With an estimated 319,600 new cases reported in 2012, uterine cancer, also known as corpus uteri cancer (as opposed to cervical cancer), was the sixth most common malignancy among women worldwide. The most prevalent type of uterine cancer is endometrial cancer, which develops from the endometrium of the uterus. As a result, endometrial cancer is occasionally referred to as uterine cancer in a broad sense. Female endometrial cancer mortality and incidence rates are greater in industrialized nations, and the incidence rate is rising. Endometrial cancer makes up 7% of all new cancer diagnoses in women, making it the fourth most frequent cancer overall in the United States, according to research published by the American Cancer Society in 2018. Additionally, after receiving appropriate therapy, the five-year survival rate for endometrial cancer is over 80%.

**1.3** **Problem Definition and Scenarios**

Endometrial malignancies are histologically categorized into three groups, from those with well-differentiated cells (grade I) to those with extremely poorly-differentiated cells (grade III) (grade III). While grade III tumors are the most aggressive and most likely to recur, grade I tumors are the least aggressive and have the best prognosis. In terms of cell differentiation and disease aggressiveness, grade II tumors fall between grades I and III. Endometrial malignancies have a very varied histology. A well-differentiated endometrioid adenocarcinoma, which is made up of numerous, tiny, packed glands with varied degrees of nuclear atypia, mitotic activity, and stratification, is the most typical finding. On a background of endometrial hyperplasia, this often occurs. The International Federation of Gynecology and Obstetrics (FIGO) cancer staging system is used during surgery to stage endometrial carcinoma. Following is the 2009 FIGO stage system:

**Table 1.1** FIGO Cancer Staging System

|  |  |
| --- | --- |
| **Stage** | **Description** |
| IA | Tumor is confined to the uterus with less than half myometrial invasion |
| IB | Tumor is confined to the uterus with more than half myometrial invasion |
| II | Tumor involves the uterus and the cervical [stroma](https://en.wikipedia.org/wiki/Stroma_(animal_tissue)) |
| IIIA | Tumor invades [serosa](https://en.wikipedia.org/wiki/Serosa) or [adnexa](https://en.wikipedia.org/wiki/Adnexa_of_uterus) |
| IIIB | Vaginal and/or [parametrial](https://en.wikipedia.org/wiki/Parametrium) involvement |
| IIIC1 | Pelvic lymph node involvement |
| IIIC2 | Para-aortic lymph node involvement, with or without pelvic node involvement |
| IVA | Tumor invades bladder mucosa and/or bowel mucosa |
| IVB | Distant metastases including abdominal metastases and/or [inguinal lymph nodes](https://en.wikipedia.org/wiki/Inguinal_lymph_node) |

**1.4** **Organization of Report**

The overall report revolves around the objective of Endometrial Cancer Detection Using Convolutional Neural Networks.

First chapter deals with introduction of Endometrial Cancer Detection Using Convolutional Neural Networks. In that we have included overview, motivation, objectives, and scope.

Second chapter deals with literature review. In that we include details of every literature survey of we collected.

Third chapter deals with project description. In this we can see the objective of the project work existing system with disadvantages, proposed system, and its advantages.

Fourth chapter deals with system design. This includes system architecture.

Fifth module deals with project requirements. In this chapter we will see the Hardware and Software Specification and technologies which are used.

Sixth module deals with module description. In this chapter we will see in detail about each module and its functions.

Seventh chapter deals with implementation. In this chapter we will implement the project working process.

Eighth chapter deals with the results obtained during the implementation process of the project.

Nineth chapter deals with conclusion and future work of the project.

**1.5 Summary**

Under the summary it talks about the introduction of the project that includes Overview of the project, Motivation, Problem Definition and Scenarios, Organization of Report.

# CHAPTER 2

# LITERATURE REVIEW

**2.1 Machine Learning for Endometrial Cancer Prediction and Prognostication**

Michael J. Downing, David J. Papke Jr et.al suggested Pathologists have struggled for years to diagnose precancerous lesions of the endometrium (endometrial intraepithelial neoplasia, or "EIN") and distinguish them from a larger group of endometrial hyperplasia’s, which are primarily benign hormonally caused processes. An expert subspecialty pathologist anonymously reviewed glass Hematoxylin and Eosin (H&E) slides to confirm diagnoses centrally. A standard H&E slide was used to validate the identification of some fragments, and a green reporter immunostained cytokeratin in a close vicinity. Vector overlays at the tissue and cellular levels were produced as a result of segmenting microanatomic structures using image analysis, and these overlays served as the basis for measurements and summary statistics that were used as input for a random forest classifier. This model's limited diagnostic scope, which only pertains to endometrial tissues in the normal-EIN-cancer spectrum, is a drawback. This omitted polyps and the wide range of benign and cancerous disorders that pathologists frequently find during an endometrial curettage or biopsy. As the sampling device travels through the endocervical canal on its approach to the uterine cavity, incidental endocervical tissue, for instance, is frequently obtained. When used on tissues that have not been predefined to fall within the scope of the training set, our algorithm, like most others, performs inconsistently.

**2.2 A New Classification of Benign, Premalignant, and Malignant Endometrial Tissues Using Machine Learning Applied to 1413 Candidate Variables**

Hao Sun, Xianxu Zeng, Tao Xu et.al suggested HIENet, to carry out various classification tasks for endometrial histopathology images. HIENet introduces two crucial blocks constructed with the visual attention mechanism and is designed using a VGG-16 backbone network as its foundation. The configuration for HIENet and VGG-16 is the same. The core of HIENet extracts the image features from each input image and creates a feature map that is symbolized by a three-dimensional matrix. Deep learning always needs a huge amount of annotated picture data to train a high-quality machine learning classifier for a given cancer classification task. due to the high time and labour costs associated with picture annotation tasks, as well as the need to preserve patient privacy. Comparatively more false-negatives were reported by HIENet than by the three pathologists. In other words, HIENet's capacity to recognize "Malignant" patches is far inferior to that of human experts.

**2.3 Computer-Aided Diagnosis in Histopathological Images of the Endometrium Using a Convolutional Neural Network and Attention Mechanisms**

Runyu Hong, Wenke Liu et.al suggested H&E slide images of endometrial cancers were downloaded from databases, divided into training, validation, and test sets for each patient, cut into 299 x 299-pixel tiles without the background, and qualified tiles packaged into Tensor Flow (TF) record files for each set. The training and validation sets were used to train convolutional neural networks, and the testing set was used to evaluate trained models. Activation maps of the test set were used to determine how well the trained models performed. To get ready for Panoptes, slides were divided into matched tile sets at equivalent resolutions of 2.5x, 5x, and 10x of the same region. Clinical characteristics branch and optional 1 x 1 convolutional layer in the Panoptes architecture. The performance of these models may be considerably improved by creating a segmentation model or by setting a threshold to eliminate certain irrelevant non-tumor tissues, such as myometrium. The datasets were employed may not fully reflect the pathogenic diversity and feature distribution of endometrial cancer, even if they cover a range of endometrial carcinoma samples. More varied training sets could strengthen the models' robustness and, ideally, increase their performance in making predictions.

**2.4 Predicting endometrial cancer subtypes and molecular features from histopathology images using multi-resolution deep learning models**

Aiko Urushibara, Tsukasa Saida, Kensaku Mori et. al suggested A 32-channel phased-array body coil and 3 T or 1.5 T imaging technology were used for the Magnetic Resonance Imaging (MRI) scan. To construct a dataset, the picture slices that make up the endometrium were extracted. According to the consensus of two radiologists, only the picture slices showing the tumor were visualized and extracted from the cancer group after considering the sequences and pathological findings. For each sequence, the identical cross-sectional pictures were taken. The CNNs outperformed the radiologists when it came to diagnosing all five single image sets, and they performed noticeably better on the single image set of the axial Apparent Diffusion Coefficient (ADC) map and axial Common Equity Tier 1 capital - Weighted Imaging (CET1 – WI). Even though there were no appreciable variations, incorporating additional image sets into the training data enhanced the diagnostic performance.

**2.5 The efficacy of deep learning models in the diagnosis of endometrial cancer using MRI: a comparison with radiologists**

Vipul Bhardwaj, Arundhiti Sharma et.al suggested Machine Learning (ML) is now one of the main approaches for academics to address a variety of biological issues because it has become so widely used. The availability of increasing computational power, significantly improved pattern recognition algorithms, and improved Image Processing (IP) software operating at extremely fast acceleration has led to the emergence of computer-aided systems that have been programmed to carry out complex tasks in medical imaging, bioinformatics, and medical robotics. New and promising perspectives for the diagnosis and prediction of numerous malignancies, including breast, colorectal, and prostate cancer, are provided by advances in machine learning techniques. Recently, ML has significantly influenced the creation of prospective computational tools for classifying, scoring, and predicting cancer patients in order to increase patient survival. Limitations of ML approaches in endometrial cancer include difficulty comparing different algorithms, dataset change, algorithmic bias, and algorithmic skewedness. generalisation to new demographics and places poses difficulties, installing Artificial Intelligence (AI) systems presents logistical challenges, Human limitations to the use of AI in healthcare.

# CHAPTER 3 PROJECT DESCRIPTION

**3.1 Objective of the project**

The purpose of this study is to develop a CAD approach based on deep learning to assist pathologists in efficiently evaluating histological images from endometrial tissue samples stained with hematoxylin and eosin (H&E). In addition to accurate image classification, this attempt was to provide diagnostic inter- portability (i.e., histopathological correlation of H&E image features extracted by our CAD approach to different types of endometrial tissues, such as endometrial gland, stroma, and thick-walled vessel) for pathologists to interpret and analyse H&E images effectively. Therefore, it expects the work may help improve the efficiency and productivity of pathologists in diagnosing endometrial diseases.

**3.2 Existing System**

Existing Work proposed a capably planned network structure, entitled HIENet, to perform various kinds of order errands for histopathological pictures of the endometrium. As a broadly utilized class of profound learning calculations, convolutional brain organizations (CNNs) have demonstrated to find lasting success in the field of biomedical imaging. HIENet is planned in view of a spine organization of VGG-16 and presents two fundamental blocks worked with the visual consideration component. HIENet has a similar setup setting as VGG-16. For each info picture, the foundation of HIENet separates its picture elements and results a component map addressed by a three-layered (three dimensional) matrix. Two blocks of HIENet, to be specific Position Consideration and Channel Consideration, take the component map as an info and afterward produce two new component guides of equivalent size. In the wake of connecting the new element maps and the first one into a greater component map, worldwide normal pooling (Hole) and straightening tasks are applied to it independently to create two new component vectors. Then, three successive completely associated layers process the link of the two component vectors. Finally, the SoftMax capability of HIENet computes a likelihood dissemination north of four anticipated yield classes.

To look at the expectation execution of different profound learning calculations on the trial dataset, prepared unique classifiers utilizing four generally utilized CNNs, AlexNet, Inception-V3, and ResNet. Note that the four CNN classifiers were prepared start to finish with various hyper-boundaries.

**3.3 Shortcomings of Existing System**

The existing system used HIENet convolutional neural network for classifying the histopathological images. HIENet had a notably higher false-negative rate than the 3 pathologists. That is to say, the cap potential of HIENet to identify “Malignant” patches is notably weaker. Some EA patches were misclassified through HIENet into EH.

**3.4 Proposed System**

Gather the endometrial tissue tests from the dataset for each computerized picture got those are extricated to advanced histopathological picture patches (640 \* 480 pixels) of an injury or a sound tissue from the relating Whole Slide Imaging (WSI) utilizing Olympus ImageView. In the first place, resize each information fix to 224 \* 224 pixels (299\*299 pixels in Inception-V3) to meet the size necessities of the CNN. The info pictures were then standardized utilizing the standard score (otherwise called z-score) a usually utilized information standardization strategy.

The order consequences of CNN-based classifiers were assessed utilizing a 10-overlap cross-approval strategy. The exploratory dataset was haphazardly parted into 10 subsamples of equivalent size, one of which was utilized for testing and the excess 9 subsamples were utilized for preparing. The cross-approval process was rehashed multiple times utilizing every one of the 10 subsamples just a single time as test information. We then arrived at the midpoint of the 10 outcomes to ascertain a general rating. Capability of VGG16 to perform different kinds of grouping undertakings on histopathological pictures of the endometrium malignant growth. Now measure the assessment results utilizing three usually utilized measurements: precision, awareness, and particularity. Additionally envisioned proposed VGG16 and the analysts' characterization execution utilizing disarray lattices (otherwise known as mistake networks) and beneficiary working trademark (ROC) bends.

**3.5** **Benefits of Proposed System**

1. VGG16 is object detection and classification algorithm which can classify 1000 images of 1000 different categories with 92.7% accuracy.
2. VGG16 is that instead of having many hyper-parameters they focused on having convolution layers of 3x3 filter with a stride 1 and always used same padding and max-pool layer of 2x2 filter of stride 2.
3. It is one of the popular algorithms for image classification and is easy to use with transfer learning.

# CHAPTER 4

# SYSTEM DESIGN

**4.1 Architecture Diagram**

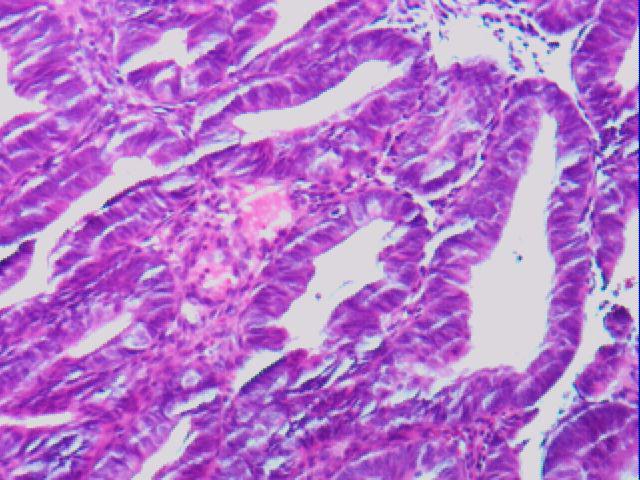
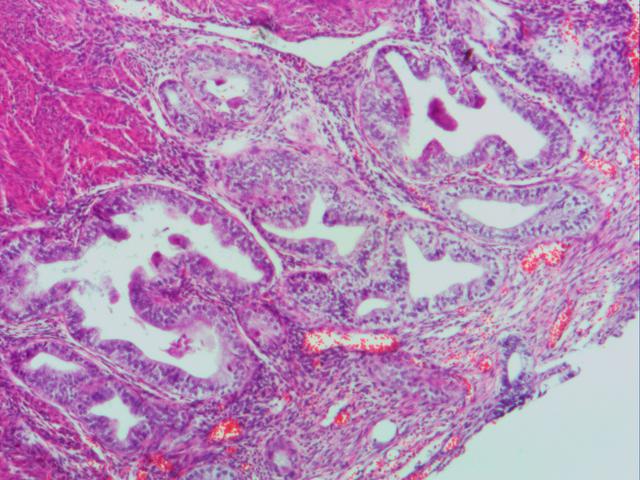
The below figure shows the architecture diagram for the VGG – 16 model.

Diagram

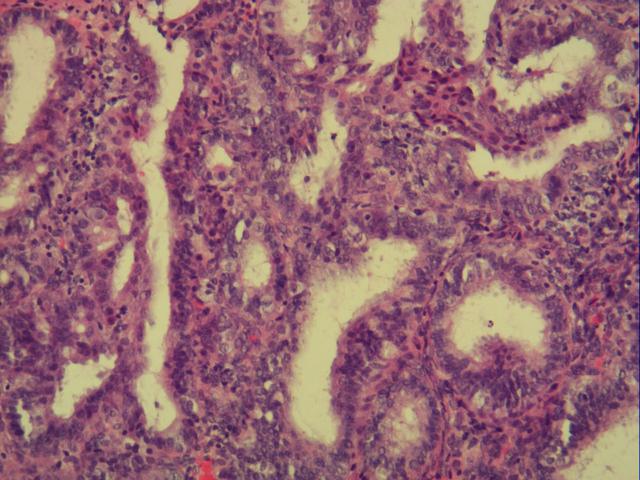
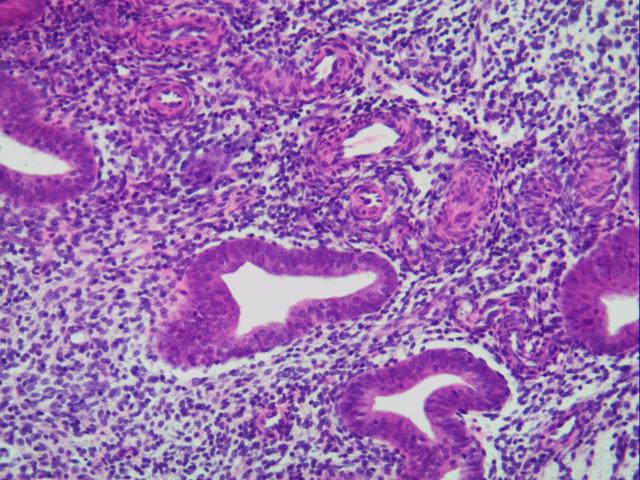
Description automatically generated

**Fig 4.1 System Architecture**

1. The collected data samples contain four general types of endometrial tissue, namely the Normal Endometrium (NE) within a regular menstrual cycle, Endometrial Polyp (EP), Endometrial Hyperplasia (EH), and Endometrioid Adenocarcinoma (EA), in this dataset.

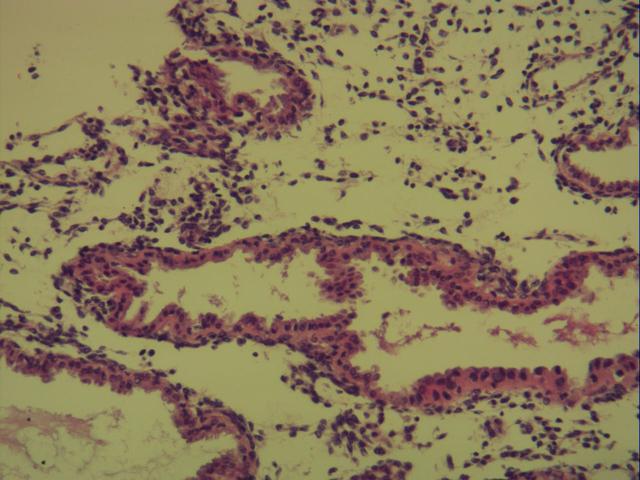
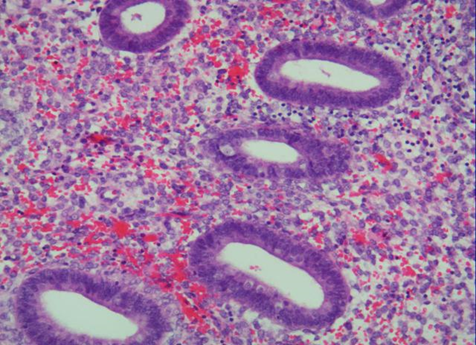


**Fig 4.2** Endometrioid Adenocarcinoma **Fig 4.3** Endometrial Hyperplasia- Complex

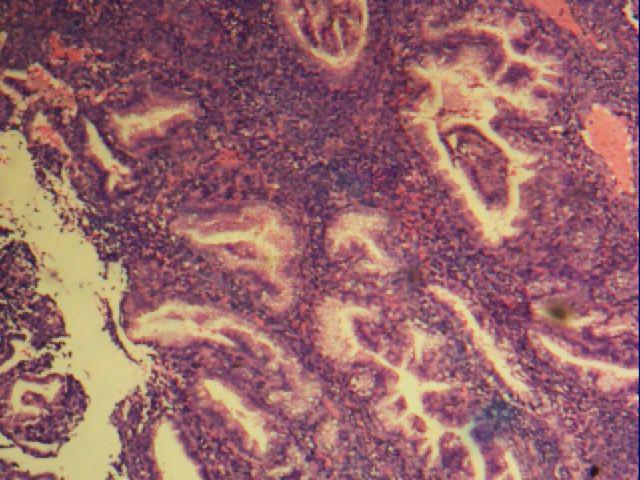


**Fig 4.4** Endometrial Hyperplasia-Simplex **Fig 4.5** Endometrial Polyp

1. The general class of NE has three subtypes defined by the phases of the menstrual cycle, namely the luteal phase, menstrual phase, and follicular phase.



**Fig 4.6** Follicular Phase **Fig 4.7** Luteal Phase



**Fig 4.8** Menstrual Phase

1. These collected digital image obtained from dataset are extracted to digital histopathological image patches (640 \* 480 pixels) of a lesion or a healthy tissue from the corresponding WSI using Olympus ImageView.
2. Then resize each input patch to 224 \* 224 pixels (299\*299 pixels in Inception-V3) to meet the size requirements of the CNN.
3. The input images were then normalized using the standard score (also known as z-score) a commonly used data normalization method.
4. The classification results of CNN-based classifiers were evaluated using a 10-fold cross-validation method. The experimental dataset was randomly split into 10 subsamples of equal size, one of which was used for testing and the remaining 9 subsamples were used for training.
5. The cross-validation process was repeated 10 times using each of the 10 subsamples only once as test data. Then average the 10 results to calculate an overall rating.
6. Now, VGG16 is used to perform various types of classification tasks on histopathological images of the endometrium cancer.
7. Quantifying the evaluation results using three commonly used metrics: accuracy, sensitivity, and specificity.
8. Visualizing the proposed VGG16 and the researchers' classification performance using confusion matrices (aka error matrices) and Receiver Operating Characteristic (ROC) curves.

**CHAPTER 5**

**PROJECT REQUIREMENTS**

* 1. **Hardware and Software Specification**

**Hardware Specification:**

RAM: 512MB

HDD/SSD: 50GB

Graphic Card: 512MB

**Software Specification**

Operating System: Windows, Mac OS, Linux

Development IDE: Anaconda navigator (Spyder, Jupiter, Pycharm)/ VS code, Colab

* 1. **Technologies Used**
* Python
* Google Colab
* VGG - 16

**CHAPTER 6**

**MODULE DESCRIPTION**

**6.1 Modules**

1. Collection of samples
2. Image processing for model training
3. Evaluation methods
4. VGG16 model
5. Evaluation metrics

**6.2 Collection of samples**

* From various sources histopathological images of endometrial cancer are collected.
* Each digital image obtained those are extracted to digital histopathological image patches (640 \* 480 pixels) of a lesion or a healthy tissue from the corresponding WSI using Olympus ImageView.

**6.3 Collection of samples**

* Resizing each information fix to 224 \* 224 pixels (299\*299 pixels in Beginning V3) to meet the size prerequisites of the CNN.
* These resized input images were then normalized using the standard score (also known as z-score) which is a commonly used data normalization method.

**6.4 Evaluation methods**

* The classification results of CNN-based classifiers were evaluated using a 10-fold cross-validation method.
* By training multiple ML models on subsets of the available input data and assessing them on the complementary subset of the data, a technique known as cross-validation is used to evaluate ML models.
* The experimental dataset was randomly split into 10 subsamples of equal size, one of which was used for testing and the remaining 9 subsamples were used for training.
* The cross-validation process was repeated 10 times using each of the 10 subsamples only once as test data. Then average the 10 results to calculate an overall rating.

**6.5 VGG – 16 Model**

* Qualification of VGG16 to perform various types of classification tasks on histopathological images of the endometrium cancer.
* VGG16 is an object recognition and classification method that has a 92.7% accuracy rate when classifying 1000 photos into 1000 different categories.
* It is a well-known technique for classifying images and is simple to deploy with transfer learning.
* There are 3 fully connected layers and 13 convolutional layers. Without altering the receptive fields, VGG uses 1x1 convolutional layers to make the decision function less linear.
* There are 138 million parameters in VGG16 altogether. Here, it's crucial to keep in mind that all of the conv kernels are 3x3 in size, while the maxpool kernels are 2x2 with a 2-stride.

**6.6 Evaluation Metrics**

* Evaluating the results using three commonly used metrics: accuracy, sensitivity, and specificity.
* Visualizing proposed VGG16 and the researchers' classification performance using confusion matrices (aka error matrices) and receiver operating characteristic (ROC) curves.
* A confusion matrix has rows for each actual class and columns for each expected class. The confusion matrix provides a wealth of data, but occasionally you would prefer a shorter metric.
* Precision is also known as positive predictive value. The proportion of correct positive predictions to total predicted positives.
* Recall is also known as Sensitivity, Probability of Detection, and True Positive Rate. The proportion of correct predictions made to total positive examples.
* Only when precision and recall are both strong can the F1 score rise. A more useful metric than accuracy is the F1 score, which is the harmonic mean of recall and precision.
* Support appears to suggest that the argument is the frequency of each specific type in the actual responses. It can be calculated by adding the confusion matrix's rows.
* The true positive rate and false positive rate for a single classifier at various thresholds are calculated and plotted to create the ROC curve.

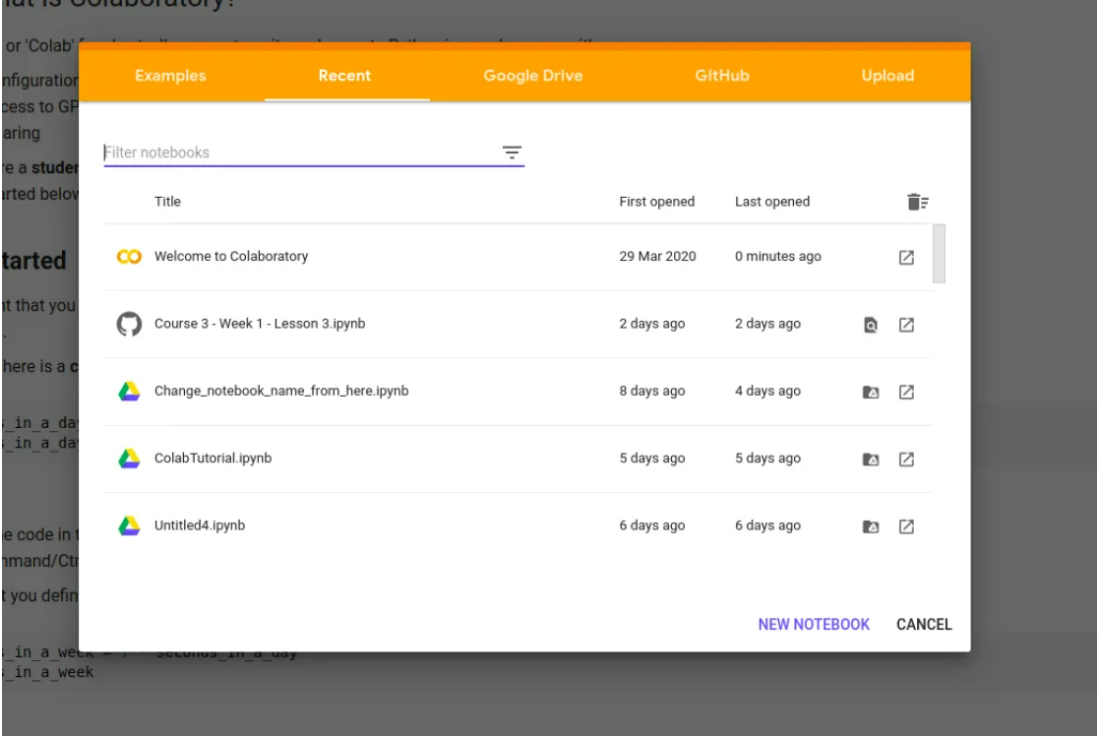
**CHAPTER 7**

**IMPLEMENTATION**

This project can be implemented in Google Colab. The Basics. Colaboratory, sometimes known as "Colab," is a Google Research product. Colab is particularly well suited to machine learning, data analysis, and education. It enables anyone to create and execute arbitrary Python code through the browser. This is essential because it enables the training of massive ML and Deep Learing models without access to strong hardware or high-speed internet. Due to the processing limits of local workstations, Google Colab supports both Graphics Processing Unit (GPU) and Tensor Processing Unit (TPU) instances, making it the ideal tool for deep learning and data analytics enthusiasts. A Colab notebook is suitable for business use as well because it can be accessed remotely from any computer using a browser.

**Creating. ipynb notebook in colab**

Go to colab.research.google.com in your preferred browser and sign in using your Google account. To create a new runtime instance, click on a new notebook.

****

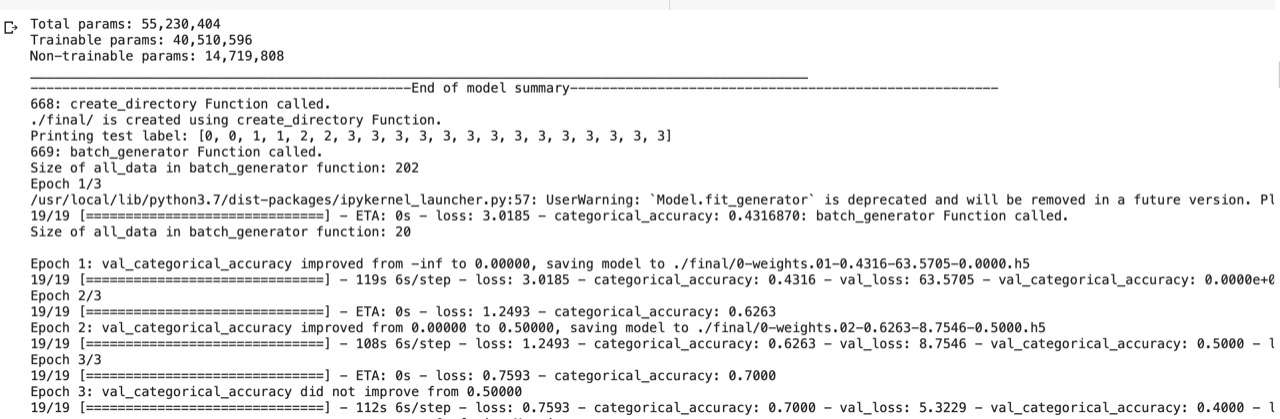
**Fig 7.1 ipynb notebook**

By clicking on the "Untitled.ipynb" in the upper-left corner, you can rename the notebook to anything you like. Where you write your code is in the cell execution block. Enter while holding down shift to execute the cell. The variable stated in one cell can be used as a global variable in other cells. If a variable is explicitly specified in the last line of code, the environment will print its value there.

To set up a selected model of TensorFlow use this command:

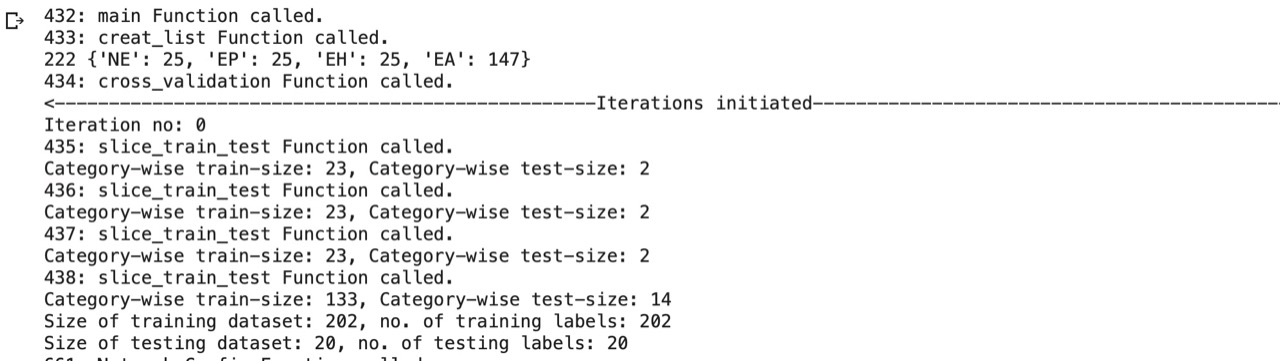
pip3 install tensorflow==1.5.0

The number histopathological images contain in the file contains 226 images which also includes the folders of the four different types of endometrial cancer. So, the actual number of images in the file are 222 images.



**Fig 7.2 Cross Validation for Epochs**

Fig 7.3 shows the 10-fold cross validation for three epochs and each epoch contains 19 testing labels. These three epochs are trained and tested for 10 times.

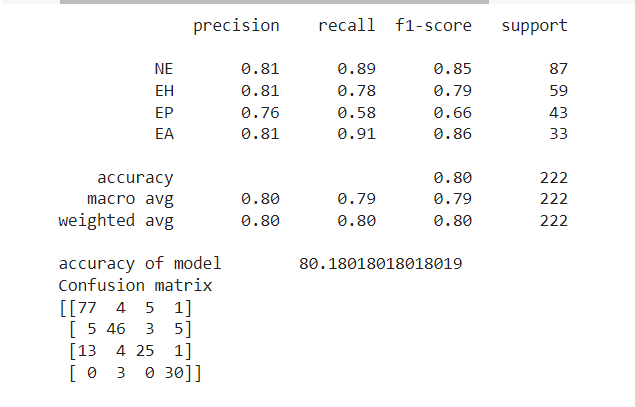


**Fig 7.3 Iteration no**

Fig 7.4 is the Iteration number 0 and it continues up to the Iteration number 9 which includes training and testing in each iteration.

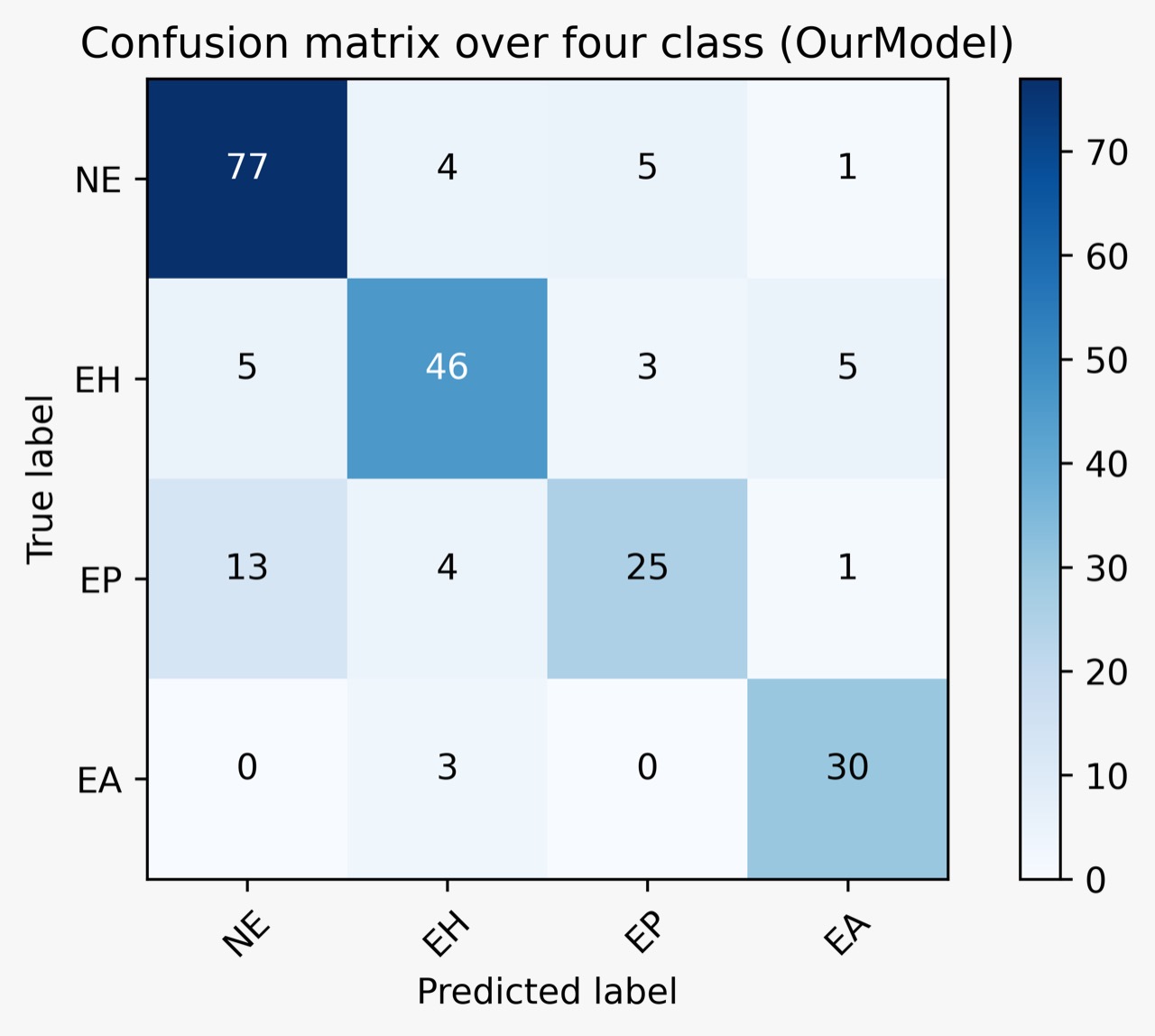
**CHAPTER 8**

**RESULT ANALYSIS**

****

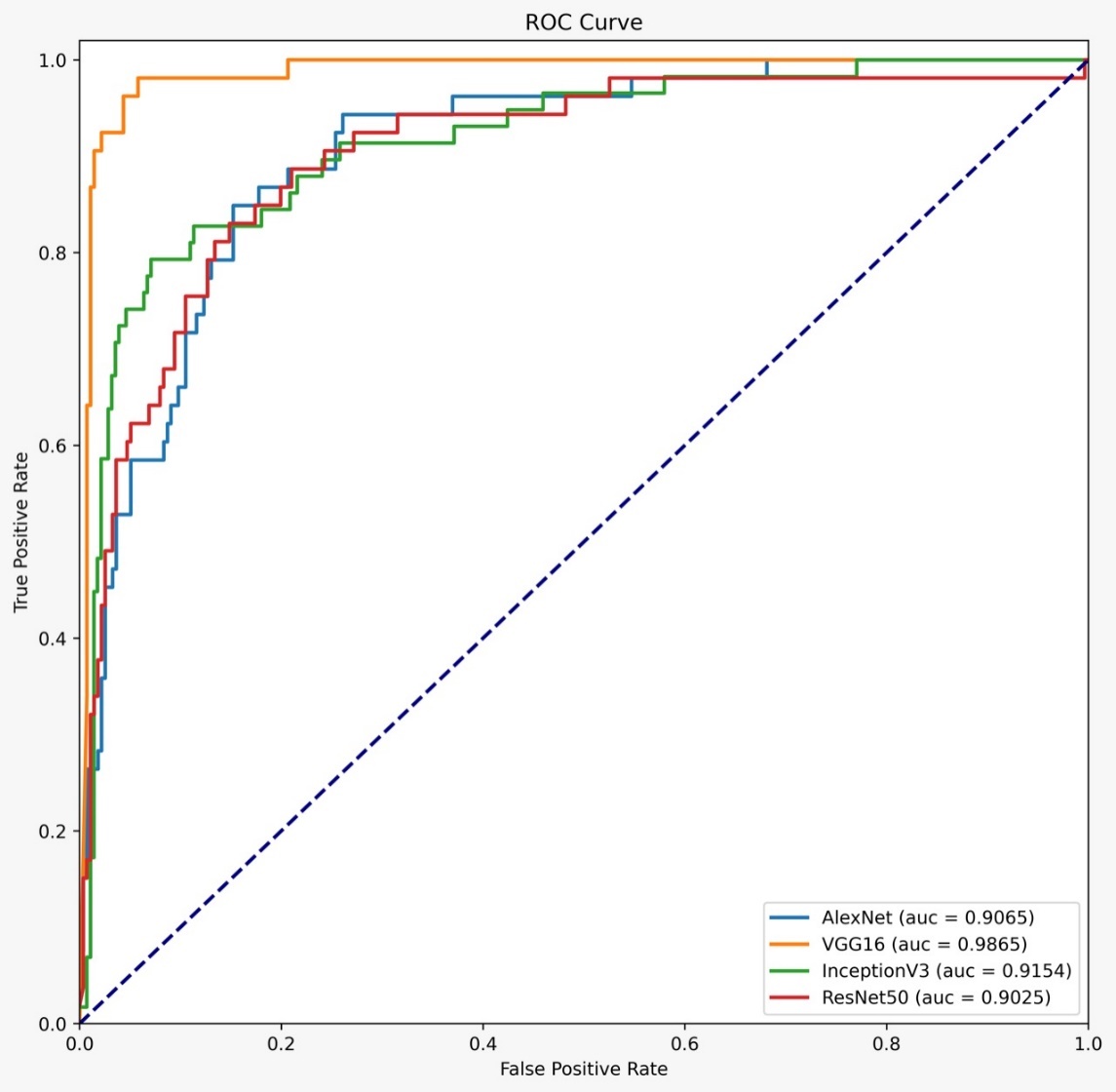
**Fig 8.1 Evaluation Metrics**

The above figure shows the result of evaluation metrics. This have quantified the evaluation results using three commonly used metrics: accuracy, sensitivity, and specificity for different of types of endometrial cancer histopathological images i.e for four types of endometrial cancer (NE, EH, EP, EA) and got the accuracy of 80.18 for three epochs.

****

**Fig 8.2 Confusion Matrix**

The given figure shows the confusion matrix over four classes of endometrial cancer histopathological images. X-axis contains the Predicted label and Y-axis contains the True label of four different types of endometrial cancer histopathological images. For NE - 77, EH – 46, EP - 25, EA – 30 images it predicted correctly.

****

**Fig 8.3 ROC Curve**

The above graph shows the ROC Curve as X - axis contains the False Positive Rate and Y – axis contains the True Positive Rate. The receiver operating characteristic curve (ROC curve) is a graph that displays how well a classification model performs across all categorization levels. Two parameters are plotted on this curve: % True Positive. Rate of False Positives. The accuracy of AlexNet, VGG16, InceptionV3, ResNet50 are as follows 90.65, 98.65, 91.54, 90.25 respectively.

**CHAPTER 9**

**CONCLUSION**

**9.1 Conclusion**

* This project developed an algorithm for training and testing the histopathological images. This paper suggests the concentration on classifying the types of endometrial cancer using histopathological images and improving accuracy with the help of convolutional neural networks. Moreover, evaluation metrics results are successful and then construct the confusion matrix using training and predicting values and calculated the confidence interval, ROC Curve and accuracy.

**9.2 Future Work**

* Determining the percentage of each form of endometrial cancer, or the percentage of endometrial cancer overall.
* Improving the accuracy for a greater number of histopathological image samples and reducing the time complexity.
* Finding the stage of the patient’s cancer and showing the details about it like medicines, consulting the doctor to that cancer.

**Chapter 10**

**INDIVIDUAL TEAM MEMBER’s REPORT**

* 1. **Individual Objective**
* Naveen objective was to maintain an updated data set samples for getting more accuracy and testing.
* Manoj and Sumanth objective were to design a model architecture and to develop the algorithm.
  1. **Role of the Team Members**

1. **P. Sai Manoj**

* Architecture diagram
* Modules and Modular descriptions
* Implementation of project code
* Literature review
* References
* Paper Writing completion

1. **R. Naveen**

* Abstract
* Objectives
* Literature review
* Existing work and proposed work
* Implementation of project code

1. **K. Sumanth Kumar Reddy**

* Project Idea
* Scope and motivation
* Proposed work
* Implementation of project code
* References
* Report Completion
  1. **Contribution of Team Members**
* Finding ways to implement the task.
* Splitting up of the document drafts to pitch in each member’s work.
* Contributions in implementing the desired final output.

**REFERENCES**

[1] Achim Hekler, Jochen Sven Utikal, et al. “Pathologist-level classification of histopathological melanoma images with deep neural networks”, European Journal of Cancer, Vol 115, Pages 79-83 (2019).

[2] Ramprasaath R. Selvaraju, et al. “Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization”, in proceedings of IEEE International Conference on Computer Vision (ICCV), pp. 618-626, doi: 10.1109/ICCV.2017.74. (2017).

[3] YunZheng Zhang, ZiHao Wang, et al. “Deep learning model for classifying endometrial lesions”, Journal of Translational Medicine 19, 10, <https://doi.org/10.1186/s12967-020-02660-x> (2021).

[4] Dong Wook Kim, Hye Young Jang, Kyung Won Kim, et al. “Design Characteristics of Studies Reporting the Performance of Artificial Intelligence Algorithms for Diagnostic Analysis of Medical Images: Results from Recently Published Papers”, Korean Journal of Radiology, Vol 20(3), Pages 405-410 (2019).

[5] Jakob Nikolas Kather, Alexander T. Pearson, et al. “Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer”, Nature Medicine, Vol-25, Pages 1054-1056 (2019 Jul).

[6] Daisuke Komura, Shumpei Ishikawa. “Machine Learning Methods for Histopathological Image Analysis”, Computational and Structural Biotechnology Journal, Vol 16, Issue 4, Pages 34-42 (2018).

[7] Yuqing Hou, Xiaoyang Xie, et al. “Bag-of-features-based radiomics for differentiation of ocular adnexal lymphoma and idiopathic orbital inflammation from contrast-enhanced MRI”, European Radiology, Vol 31, Pages 24–33 (2021).

[8] Yu Takahashi, Kenbun SoneID, et al. “Automated system for diagnosing endometrial cancer by adopting deep-learning technology in hysteroscopy”, PLoS ONE 16(3): e0248526. <https://doi.org/10.1371/journal.pone.0248526> (2021)”.

[9] BenTaieb and Ghassan Hamarneh. “Predicting cancer with a recurrent visual attention model for histopathology images”, In: Frangi, A., Schnabel, J., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds) Medical Image Computing and Computer Assisted Intervention – MICCAI 2018. Lecture Notes in Computer Science (), Vol 11071. Springer, Cham. <https://doi.org/10.1007/978-3-030-00934-2_15>, (2018).

[10] Asifullah Khan, Anabia Sohail1, et al. “A survey of the recent architectures of deep convolutional neural networks”, Artificial Intelligence Review, Vol 53, Pages 5455–5516 (2020).

[11] Arkadiusz Gertych, Zaneta Swiderska-Chadaj, et al. “Convolutional neural networks can accurately distinguish four histologic growth patterns of lung adenocarcinoma in digital slides 1483”, <https://doi.org/10.1038/s41598-018-37638-9> (2019).

[12] Jan P. Baak, George L. Mutter,Stanley Robboy, et al. “The Molecular Genetics and Morphometry-Based Endometrial Intraepithelial Neoplasia Classification System Predicts Disease Progression in Endometrial Hyperplasia More Accurately than the 1994 World Health Organization Classification System, Cancer;103(11):2304-12. doi: 10.1002/cncr.21058. PMID: 15856484; PMCID: PMC2600877. (2005).

[13] Georgios-Marios Makris, Abraham Pouliakis, et al. “Image Analysis and Multi-Layer Perceptron Artificial Neural Networks for the Discrimination Between Benign and Malignant Endometrial Lesions”, Diagn Cytopathol;45(3):202-211. doi: 10.1002/dc.23649. Epub 2017 Feb 3. PMID: 28160459 (2017).

[14] Baris Gecera, Selim Aksoya, et al. “Detection and classification of cancer in whole slide breast histopathology images using deep convolutional networks”, Pattern Recognition, Volume 84, Issue C, pp 345–356, <https://doi.org/10.1016/j.patcog.2018.07.022>, (2018).

[15] Zizhao Zhang, Pingjun Chen, et al. “Pathologist-level interpretable whole-slide cancer diagnosis with deep learning”, Nat Mach Intell, Vol 1, Pages 236–245, <https://doi.org/10.1038/s42256-019-0052-1>, (2019).

**APPENDIX A: SAMPLE CODE**

dir = "/content/drive/MyDrive/histopathological"

doc, k, l, lastLine, classnum = open("all\_files.txt",'w'), 0, 0, "", -1

for root, dir, files in os.walk(dir):

for file in files:

k += 1

if file.split('.')[-1] in ['jpg', 'JPG']:

l += 1

print(os.path.join(root,file), file = doc)

print(k, l)

doc.close()

list\_file = open("file\_list.txt","w")

with open("all\_files.txt") as f:

line = f.readline()

while line:

thisline = line.split("/")[-3]+line.split("/")[-2]

if lastLine != thisline:

classnum, lastLine = classnum + 1, thisline

list\_file.write(line.split("\n")[0]+"\t"+str(classnum)+"\n")

line = f.readline()

f.close()

list\_file.close()

def cross\_validation(categorised\_data, K, no\_of\_epochs, no\_of\_classes, batch\_size): # called inside main Function

global Step\_no

print(str(Step\_no) + ": cross\_validation Function called.")

Step\_no += 1

no\_of\_categories = len(categorised\_data) # categorised\_data = array of size 4, each index has an array of NE, EP, EH, EA

for each\_category in range(no\_of\_categories):

random.shuffle(categorised\_data[each\_category])

print("<--------------------------------------------------Iterations initiated-------------------------------------------------->")

for iteration in range(K):

print("Iteration no: %d" %iteration)

train\_data\_for\_present\_iteration, test\_data\_for\_present\_iteration, train\_data\_paths, train\_data\_labels, test\_data\_paths, test\_data\_labels = [], [], [], [], [], []

for category\_no in range(no\_of\_categories):

train\_data\_paths\_of\_each\_cat, test\_data\_paths\_of\_each\_cat = slice\_train\_test(categorised\_data[category\_no], iteration, K)

print("Category-wise train-size: " + str(len(train\_data\_paths\_of\_each\_cat))+ ", Category-wise test-size: " + str(len(test\_data\_paths\_of\_each\_cat)))

for each\_path in range(len(train\_data\_paths\_of\_each\_cat)):

train\_data\_paths.append(train\_data\_paths\_of\_each\_cat[each\_path])

train\_data\_labels.append(category\_no)

for each\_path in range(len(test\_data\_paths\_of\_each\_cat)):

test\_data\_paths.append(test\_data\_paths\_of\_each\_cat[each\_path])

test\_data\_labels.append(category\_no)

print("Size of training dataset: " + str(len(train\_data\_paths)) + ", no. of training labels: " + str(len(train\_data\_labels)))

print("Size of testing dataset: " + str(len(test\_data\_paths)) + ", no. of testing labels: " + str(len(test\_data\_labels)))

record = open("record\_of\_iterations.txt", 'a+')

record.write("Iteration no: " + str(iteration))

record.write(str(train\_data\_paths)+'\n')

record.write(str(test\_data\_paths)+'\n')

record.close()

for each\_path in train\_data\_paths:

train\_data\_for\_present\_iteration.append(read\_image\_from\_path\_as\_array(each\_path, 224, 224, True))

for each\_path in test\_data\_paths:

test\_data\_for\_present\_iteration.append(read\_image\_from\_path\_as\_array(each\_path, 224, 224, True))

Network\_Config(Train\_data = train\_data\_for\_present\_iteration, No\_of\_categories = no\_of\_categories, No\_of\_epochs = no\_of\_epochs, Train\_labels = train\_data\_labels,

Test\_data = test\_data\_for\_present\_iteration, Test\_labels = test\_data\_labels, Iteration\_no = iteration, Initial\_epoch = 0, Batch\_size = batch\_size)

def main():

global Step\_no

print(str(Step\_no) + ": main Function called.")

Step\_no += 1

list = creat\_list("file\_list.txt")

cross\_validation(categorised\_data = list, K = 10, no\_of\_epochs = 3, no\_of\_classes = 4, batch\_size = 10 )

main()

import numpy as np

import scipy as sp

import scipy.stats

import numpy

def confidenceinterval(array):

confidence = 0.95

a = 1.0 \* np.array(array)

m = np.mean(a)

fc = scipy.stats.sem(a)

h = fc \* sp.stats.t.\_ppf((1 + confidence) / 2., n - 1) / ((n - 1) \*\* 0.5)

return m - h, m + h

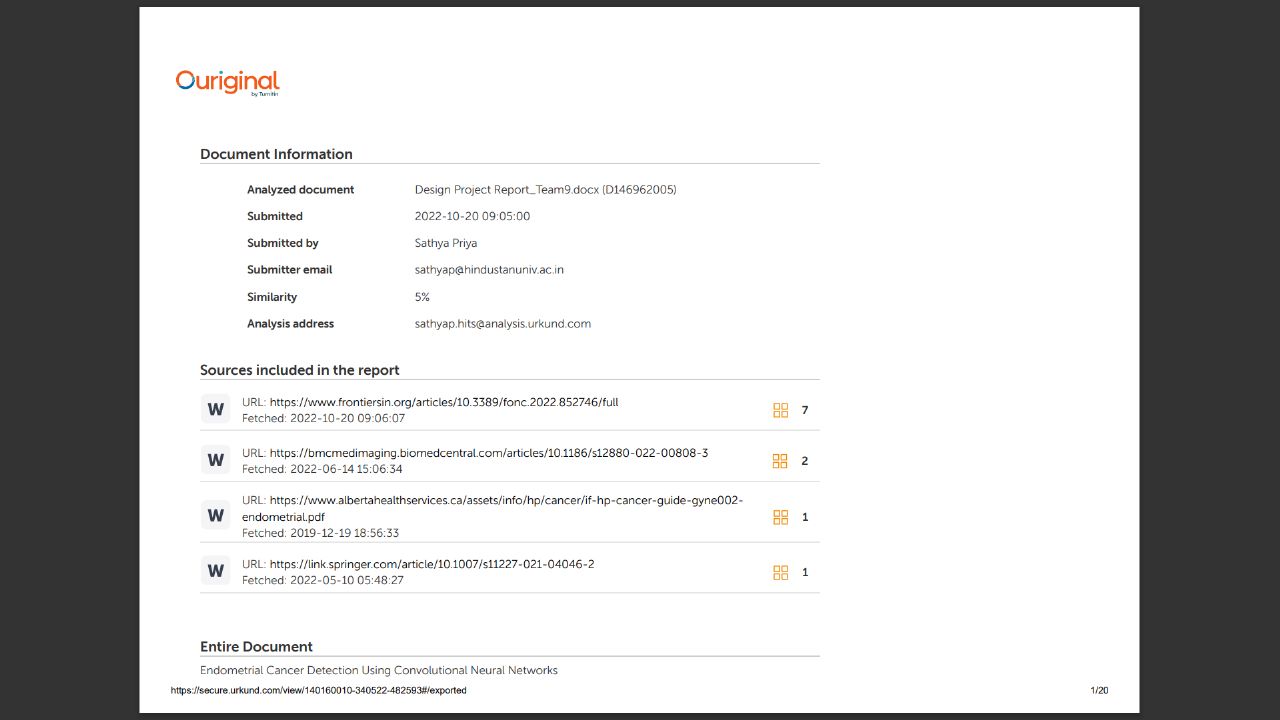
n = 10

a = [71,59,58]

array = numpy.asarray(a)

print(confidenceinterval(array))

**APPENDIX B: PLAGIARISM REPORT**

****

**APPENDIX C: Team Details**

1. **NAME:** P. SAI MANOJ

**CLASS:** CSE 7B

**ROLL NUMBER:** 19113101

**EMAIL:** [**19113101@student.hindustanuniv.ac.in**](mailto:19113101@student.hindustanuniv.ac.in)

1. **NAME:** R. NAVEEN

**CLASS:** CSE 7B

**ROLL NUMBER:** 19113102

**EMAIL:** [**19113102@student.hindustanuniv.ac.in**](mailto:19113102@student.hindustanuniv.ac.in)

1. **NAME:** K. SUMANTH KUMAR REDDY

**CLASS:** CSE 7B

**ROLL NUMBER:** 19113105

**EMAIL:** [**19113105@student.hindustanuniv.ac.in**](mailto:19113105@student.hindustanuniv.ac.in)