

# Configuration Manual

MSc Research Project  
Masters in Data Analytics

Minakshi Tikone  
Student ID: x19235283

School of Computing  
National College of Ireland

Supervisor: Hicham Rifai

**National College of Ireland**  
**MSc Project Submission Sheet**  
**School of Computing**



**Student Name:** Minakshi Tikone  
**Student ID:** X19235283  
**Programme:** Msc Data Analytics **Year:** 2021-2022  
**Module:** Msc Research Project  
**Lecturer:** Hicham Rifai  
**Submission Due Date:** 31-01-2022  
**Project Title:** Ovarian cancer classification using histopathological image dataset.  
**Word Count:** 4075 **Page Count: 21**

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

**Signature:** Minakshi Tikone

**Date:** 31-01-2022

**PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST**

Attach a completed copy of this sheet to each project (including multiple copies)	<input type="checkbox"/>
<b>Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).</b>	<input type="checkbox"/>
<b>You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.</b>	<input type="checkbox"/>

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

<b>Office Use Only</b>	
Signature:	
Date:	
Penalty Applied (if applicable):	

# Classification Of Ovarian Tumor Using Histopathological Images

MSc in Data Analytics

Thesis Project

Minakshi Goraksh Tikone

x19235283

Lecturer Name: Prof. Hicham Rifai

## 1 Abstract

Histopathology and allied sciences such as cytology have gained new perspectives as a result of the rise of digital pathology. Computer algorithms, specifically artificial-intelligence algorithms, can work on digitized slides to aid pathologists with theranostic and diagnostic duties. In recent years, machine learning has come a long way. Machine learning approaches such as segmentation and classifications have apparent advantage of doing pathological image analysis. In this research, I provide the findings of a search of the world's biggest public archive ([TCGA]The Cancer Genome Atlas program run by the National Cancer Institute in the United States) of entire slide pictures of over 50 individuals with ovarian tumor cells. I successfully indexed and searched around high-resolution digital slides totaling 8 giga bytes of data composed of 8000 x 8500 pixels picture patches for the first time. This study demonstrate a pretrained DCNN architecture -VGG16 which consist of 16 layers. The model was trained with 46 images, achieved an accuracy score of 44% due to the limitation issue. The proposed study is to use VGG16 model to classify the ovarian tumor using histopathological images.

Keywords: Histopathology, machine learning, DCNN, pre-trained model, VGG16

## 2 Introduction

Cancer is a deadly illness that is the main cause of mortality among women in medium and high-income nations. The female population in the world accounts for around 49.5 percent of the total. Ovarian cancer is, in fact, one of the most common and fatal gynecologic malignancies. The yearly frequency is predicted to be 225.000 cases, with 140.000 fatalities globally.

The ovary is a female reproductive organ that produces oocytes, which are egg cells. After that, the oocytes are moved to the fallopian tube, where they can be fertilized by sperm. After fertilization, the ovum is moved into the uterus and develops into a baby. The female reproductive system also generates hormones including progesterone and estrogen

Ovarian cancer occurs when cells in ovaries proliferate uncontrollably, resulting in the creation of tumor, which is a mass of tissue. Ovarian cancer is caused due to many factors such as hormonal changes after menopause, family history, smoking and so on. Women over the age of 50 are most likely to acquire the cancer. Abdominal discomfort, swollen stomach, pelvic pain, bloating and so on are the indications of ovarian cancer. Ovarian cancer is known as a lethal cancer since the signs are hard to detect in the preliminary stages and often go untreated until the cancer cells have spread to the abdomen and pelvic. As a result, it is exceedingly difficult to treat cancer if it is not detected at an early stage. Appropriate detection of stage and type of tumor is very necessary for further diagnosis. Ovarian tumor consist of different types of tumor cells such as stromal cells, necrosis cells, epithelial cells and so on. In this study, I classify the ovarian tumor using histopathological images based on grade levels as well as the percentage of normal cells, stromal cells, and tumor location are visualised.

Histology is the examination of a biopsy or tissue material under a microscope. Understanding microscopic structures and their activities at cellular, sub cellular, tissue and organ levels is necessary in order to understand the illness progression and prognosis. Pathologists also discover the morphological traits of tissue that signal the presence of illness, such as cancer, when studying and analyzing histological images under a microscope. If the disease is detected, a grading procedure is carried out, which deals with the growth of diseased cells throughout the tissue. The prognosis and subsequent therapy for each patient are then determined based on the grade of disease Belsare & Mushrif (2012). Manual detection is a time-consuming and can be prone to human mistake since most portions of the cell are typically part of irregular, unpredictable, and arbitrary visual angles. (Cancer diagnosis in histopathological image: CNN based approach ). Thus,

computer assisted disease diagnosis (CAD) plays a critical role and has evolved into a major research topic in histopathological imaging and diagnostics, where various image processing algorithms may be employed to interpret these pictures for disease diagnosis and treatment. Histology, as a result, serves as a scientific support for clinical research, teaching, and practice.

The purpose of this research is to study at robust and reliable image analysis techniques for digital histopathology picture interpretation. Various image processing methods will be used for image tissue classification, tissue masking, different cell counts, identification of cells type, or classification in order to derive quantitative measurements of disease characteristics from histological images and automatically evaluate whether a disease is grade level 2 or grade level 3.

In this research paper, several image segmentation techniques, feature extraction and classification techniques used by various researchers are reviewed in section 3. The section 4 describes about the Methodology used during this research paper. Dataset with the flow of basic histological image processing and the technique for preparing histopathology slides for tissue examination, including the image zooming/resolution required for studies, are described. Finally the classification method for histopathological images is introduced to get the appropriate output, followed by the conclusion on this study.

## **2.1 Research Question**

How Deep convolutional neural network model can help in classifying ovarian cancer based on grade level using histopathological images?

## **3 Literature Review**

### **3.1 Machine learning used in field of Cancer**

An study of the risk variables for ovarian cancer utilizing several machine learning techniques such as SVM, random forest, logistic regression, KNN etc on hospital data. According to the study by (Ahmed et al. 2021), logistic regression had the highest accuracy of 0.933, and variables such as,menopausal issues pregnancy, abortion, and so on were revealed to be important risk factors.

Gene expression can help is addressing the problems related to the cancer diagnosis. DNA micro array technique helps in monitoring the gene expressions.As the number of the gene expression are huge, the author (Lu & Han 2003) presented various types of classification methods which are based on the computation time, accuracy score and ability to predict gene information.

Cancer has been described as a complex illness with several subgroups.Hence classifying the cancer at an early stage has become important in cancer research. In this study the author (Kourou et al. 2015) has discussed the importance of the machine learning methods. Also illustrated about the different machine learning models such as ANN's, SVM,BN for resulting in accurate results in cancer research field.

Cervical pap cytology abnormal cell detection is critical for early cervical cancer screening and therapy. The effectiveness of an automation-based cervical screening and diagnostic system is dependent on precise cell type identification. The authorK & C (2022) in this study proposed an efficient framework which is automated for cell classification.Various machine learning classifiers were used and out of which CNN model performed the best with AUC 0.99 and accuracy of 99.7

### **3.2 Tissue Analysis**

By analyzing the texture of the tissue, different kinds of cancer tissues may be recognized.The forms of cancer tissue were examined in this study by (Linder et al. 2012) utilizing histopathology images of epithelial and stromal.SVM model was used to train the data with a local binary patter which gave the accuracy score of 0.995. To identify the cancer kinds, tissue analyzation will be carried out in this research.

In recent years, a number of powerful pattern recognition algorithms have emerged, notably in the field of texture categorization. Many of these strategies,

in general, need that textures be rendered uniformly and in the same way. To analyze tumor tissue, slices of tumor tissue are cut in various planes and placed at various angles on slides. As a result, it is impossible to achieve constant orientation, and tissue texture examination should be independent of orientation. Furthermore, the system should be able to withstand fluctuations in picture contrast induced by tissue processing and image capturing factors. The author (Ahonen et al. 2006) suggested utilizing the LBP method to analyze tissue texture.

For determining the risk and seriousness of ovarian cancer, histological study of tissue samples is required. For a better evaluation of histological pictures of ovarian cancer, the Hematoxylin and Eosin (HE) staining method is used, however the procedures are sophisticated and require more criteria. To solve this difficulty, the author (Meng et al. 2021) suggested employing a Generative Adversarial Network as a virtual histological staining. Simultaneously, they demonstrated an effective autofluorescence picture production approach in the absence of real data, which might save time and effort in many cases.

### **3.3 Techniques used for classification for ovarian cancer**

Deep learning techniques can be used in the health care field because computer-based equipment such as CT scans and MRIs are expensive and not always available to people. Using tissue samples, the researchers (Kasture et al. 2021) introduced a Deep convolutional neural network model which is a type of deep learning algorithm, to predict ovarian cancer and its subtypes, such as whether the tumor cells are non-cancerous, serous, clear cell, mucinous or endometrioid, with 83% of an accuracy score. Deep learning techniques has been crucial in prior research, hence this approach will be used in my study.

Ovarian cancer comes in a variety of medical stages and histopathologies, each with its own grading system. According to the research by (Kawakami et al. 2019), epithelial ovarian cancer was found in patients by identifying multiple clinical attributes from the patient's blood report and age using supervise machine learning techniques such as support vector, gradient boosting, neural network, random forest and others, with random forest performing the best in separating epithelial ovarian cancer from benign tumor with accuracy of 92.4 percent and ROC curve of 0.968. Through patient classification prior to treatment, the use of machine learning techniques may allow for personalized treatment options.

Machine learning plays a vital role in image classification. Machine learning has enormous promise for health screening, and it poses no threat to technical

businesses or health-care organizations. Artificial Intelligence (AI) has recently been utilized in health screening to detect sickness. The author of this paper, (Booma et al. 2020), offered the Enhanced Max Pooling approach for recognizing and categorizing an ovarian cancer tumor utilizing deep learning techniques with a 45 percent accuracy score. For improved outcomes, the author suggested that other machine learning methods be used.

The cancerous cell if not identified at early stage spread rapidly, making the condition difficult to cure. Various machine learning algorithms were used by the author (Bhattacharjee et al. 2017) to classify if the cells were benign or cancerous, out of which multilayer perceptron appeared to be better option by considering the performance outcomes such as accuracy, sensitivity, specificity and error whereas in other study the author (Wadhwa 2021) classified the cells using CNN algorithm with an accuracy score of 94.73 percent by using the histopathological images.

Epithelial ovarian cancer consist of histology subtypes which can be categorise into clinical, genetic and morphological features. All of this details can be determined by examining the whole slide images. The author (Farnell et al. 2020) did the slide classification of ovarian cancer by utilizing the two stage transfer learning technique which was based on CNN and gained the accuracy score of 87.54% which performed better than the normal CNN. The author claimed the results will be promising for future validation for the large group of patients.

In this paper, the author (Komura & Ishikawa 2018) specifies many challenges such as image size issue,improperly labeled images etc and their solutions during pathological image analysis which is more pathology-focused viewpoint. The author suggested many problems can be solved with the given solution with the large annotated WSI files.

Precise prediction and characterization of Ovarian cancer subtypes are critical in the varied diagnosis. The author (Kasture 2021) demonstrated the prediction and subtype classification of ovarian cancer with histopathological images using the pre-trained model DCNN -VGG-16 which received the accuracy score of 84.6% after augmenting the images whereas the original image without augmentation gave the score of 50%. The author proposed of augmenting the images before training the model.

Manual examination of tissue, tissue cell distribution,shape regularities, and tumor image categorization by pathologists takes time and can result in human error. To overcome this issue, the author Belsare & Mushrif (2012) discuss various procedure for analyzing and segmenting the histopathological images. The



image processing techniques such as image segmentation, feature extraction, classification and so on are mentioned in this study. The author proposed of using low magnification(10x) for tissue classification and high magnification(40x) for cell segmentation and analysis.

Due to the small size of data and due to high dimension data, the performance of the machine model decreases, whereas the deep learning model can process effectively with high dimension but not with small amount of size data. To overcome this issue the author Wei et al. (2022) proposed to use GAN (generative adversarial model) model. The GAN model it creates its own target data if not present for training purpose. The author executed the GAN model which received the accuracy score of 92.6%.

As per the previous studies review, it is concluded that Densenet model, AlexNet model, GAN model, MLP performs better with classification of histopathological images. In my research, I will be using VGG16 model with CNN for classifying the Ovarian cancer tumor. In previous study, the researcher Kasture (2021) used VGG16 model with DCNN to predict ovarian cancer and its subtypes, using histopathological images, with an accuracy score of 84%. My focus would be to classify the ovarian tumor using histopathological images with more accuracy score.

## **4 Methodology**

The aim of the model built in this research paper is to detect and classify the ovarian slide images into tumor and normal slides by using the whole slide image. Also to showcase the ovarian tumor grade level and location of the tumor cell if present in the tissue. A methodology is a structured way of acquiring knowledge. KDD (Knowledge Discovery in Database) approach will be used in my research. KDD is a pioneering data mining approach. It is a one-to-one approach with several phases to guarantee that the work is accomplished correctly and that the results are properly fed into the following step. KDD consists of many procedures that provide relevant knowledge data as a result, which includes selection of data, preprocessing data, data transformation, and the last phase, interpretation/evaluation. The required research paper, conference proceedings, journals which are related to Ovarian cancer detection and classification with histopathological images were searched.

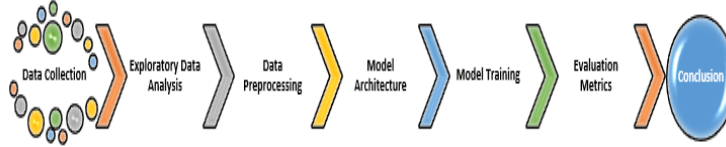


Figure 1: KDD Flow Diagram

## 4.1 DataSet and Preparation

This study comprised 46 patients with serous ovarian cancer who participated in The Cancer Genome Atlas (TCGA). Histopathological whole slide images were acquired from Genomic Data Common (GDC). TCGA's main mission is to enhance our capacity to detect, cure, and prevent cancer. The National Human Genome and the National Cancer Institute adopted a phase in method to launch TCGA in order to accomplish this goal in a scientifically manner. These slides are in format of '.svs' which contains high resolutions, colored and 2-D images. SVS files are digital slide image file which contains various specimen images that are saved at different resolutions. Total of 46 slide images in SVS format were downloaded and each slide images size is around 389 MB and the whole dataset size is of approximately of 8GB . The dataset contains the ovarian tumor tissue images. This dataset contains different data categories such as RNA-seq, miRNA-seq, Clinical, biospecimen information out of which I am using Biospecimen information which contains the patient user ID, stage level, grade level of the cancer, percent of stromal cells, necrosis cell and tumor cells present in the tissue along with its '.svs' file. The stage level defines the level of the cancer whether the stage is I or IV, where I is the 1st early stage and IV is the last stage level. The necrosis cell percent determines if the cancer tumor cell is aggressive or not and The stromal cell percent shows if the tumor contains stromal cells in the tissue. Biospecimen file is gathered in the form of '.csv' format. This data is publicly available from GDC portal<sup>1</sup> and have been utilized widely. Using these dataset, tumor and non tumor cells of ovarian cancer will be classified along with its stage level by showing the tumor cell location. The dataset will be split into two sets which are train and validation sets for training, testing and validating the model to classify slides based on grade of ovarian cancer.

<sup>1</sup><https://gdc.cancer.gov>

## 4.2 Data Preprocessing

For my study, I utilized jupyter notebook to perform my research code. To initialize the code, several types of packages that were necessary for study were installed. For example, to open TCGA files in notebook that are in '.svs' format, the openslide package was installed. The '.svs' file cannot be viewed until the openslide package is imported. OpenSlide is a C library for processing whole slide images, often known as virtual slides, which are large resolution images used in computer - aided diagnosis. I had to supply the path as well as the platform system to import the openslide package. In order to import the openslide package, the openslide directory must also be called. The TCGA ovarian slide images were called when the open slide package was successfully deployed. For the biospecimen csv file, pandas package was used for reading the file. After reading the file into the notebook, the missing values were checked by using isnull() function. As my dataset is limited, with only 17 columns and 46 rows, any missing data causes significant information to be lost. Hence missing data leads to unbalanced observations, skewed estimates, and, in the worst-case scenario, erroneous conclusions. I discovered that there were no missing values present in my dataset . After finding missing value step, various columns were eliminated from the csv file as they were not needed for further research by using drop function.

### 4.2.1 Reading the images by applying the pixel size

By using the openslide library, the SVS images were zoomed to closest resolution to view the tumor cells in small pixel form. The svs images were zoomed into patches with the pixel size of 8000x8500 for better visualising the tissue cells.

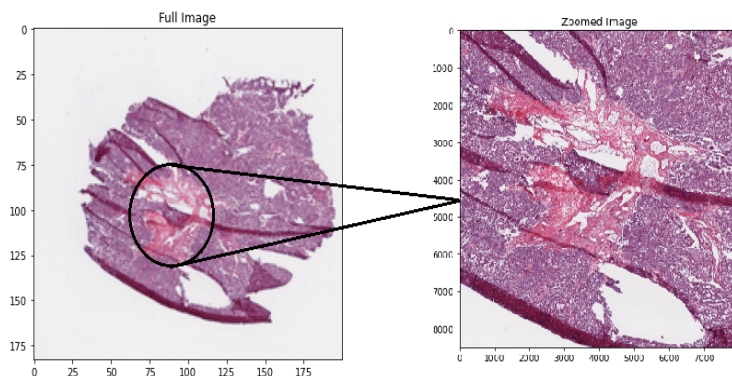


Figure 2: Patch Zooming

### 4.2.2 Converting SVS to TIFF

The availability of appropriate data is a critical component in assessing the performance of deep convolutional neural networks. A whole-slide picture is a computerized representation of a microscopic slide at a very higher resolution, such as 20x or 40x. Whole slide images are often relatively huge in size as a result of the high resolution. In our training data set, the largest file size for a single whole-slide image is 436 MB, with an average of more than 8 GB of overall file size. The width of my 46 image training dataset ranged from 6475 to 44002 pixels, with an average of 25238 pixels. The height ranged between 8099 and 32398 pixels, with an average of 20248 pixels. The total pixel sizes of the images ranged from 50002x27585 pixels. The 46 training images occupy 8 GB of storage space. Set of filters needed to apply to the entire set of WSI slides. The two issue which I faced was regarding the size and format of the data. As previously stated, the average .svs file size for my training dataset is above 1 GB, and have a total of 55 slides. Furthermore, the svs format is a quite unique format that is often not displayed visually by default by standard apps and operating systems. Hence, using OpenCV and Python to transform the WSI data set to smaller images in a common format, thereby reformulating a huge data problem as a smaller data problem. The .svs images were converted to .tiff image which is also a graphic file format supported by OpenSlide. The conversion from .svs to .tiff leads to information loss of slide images. This resulted in problems opening .tiff images using openslide, therefore using tifftools<sup>2</sup> the metadata of the images were restored for further processing, training and validation. Converting 46 images from .svs to .tiff took less 60 seconds to convert in windows 11. These converted files are then stored in a directory folder.

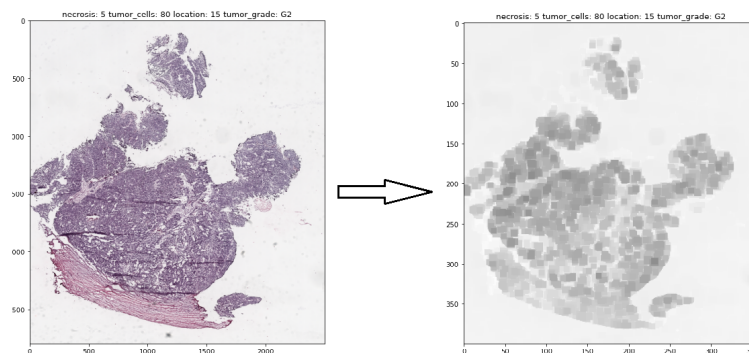


Figure 3: Image Conversion from .svs format to .tiff format

<sup>2</sup><http://www.libtiff.org/tools.html>

## 5 Design Specification

### 5.1 DCNN Architecture - VGG16

Convolution is the process of applying a filter on an image or a signal in order to change it. Pooling is the process of discretizing samples in order to minimize the dimensions of the input. Convolutional neural networks (CNN) are made up of hidden layers such as convolutional, pooling, fully connected, and normalizing. CNN architecture is a stack of these layers combined with a differentiable function. Arranging these layers in a certain order has led in the development of numerous CNN designs during the last decade.

In 2012, the world learned about AlexNet, a novel DCNN architecture with 5 convolutional, 3 max-pooling, 2 normalization, 2 completely connected, and one softmax layer. There are 60 million parameters in this model (Masood et al. 2018). VGG-16 is the successor to the AlexNet architecture developed by the University of Oxford's Visual Geometry Group (VGG). VGG's Karen Simonyan and Andrew Zisserman created this CNN to win the ImageNet Competition 2014's identification and classification track. VGG-16 is an improvisation on the AlexNet design that is significantly deeper, resulting in more layers and greater processing costs.

As the name implies, VGG-16 has sixteen layers: 13 convolutional layers, 3 fully connected (FC) layers, 5 max-pool layers, and a single softmax layer. The pretrained VGG-16 architecture is fed an input picture of size 256x256x3 as seen in figure.4 and derived from the dataset.

### 5.2 DCNN Architecture - VGG19

As the name implies, VGG-19 model contains 19 layers as shown in figure.5 in which there are 16 convolutional layers, three fully connected layers, 5 maxpool layers and 1 softmax layer. The input image size is 224. The batch size used in the model is of 32. VGG19 consists of various arguments as follows:

- include\_top: This argument is used whether to add three fully connected layers at top of network
- weights: weights can be set as 'none' or 'imagenet'
- input\_tensor: to input the image for the model
- input\_shape: It should have 3 input such as width, height and channels.
- pooling: This argument is carried out in order to provide an output with a smaller dimension than the source images. pooling consist of 3 types such as None, average and max.

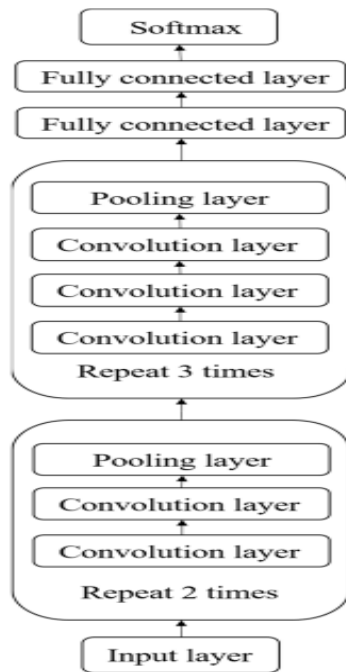


Figure 4: VGG16 Architecture

classifier\_activation: This argument is used to load pre-trained weights and can be set to 'none' or 'softmax'

## 6 Implementation

This section delves further into the classification of ovarian tumor by Data Mining. It begins with the hardware / software characteristics that will be required for proper execution, and then moves on to the models.

### 6.1 Environment Configuration and Requirements

This section contains details on the software Figure.8 and hardware 9ure.6 requirements for implementing this proposal, as well as a decent network connection. To avoid taking up a lot of local space, storage can alternatively be done via cloud.



Figure 5: VGG19 Architecture

RAM	Minimum 8gb ram
Processor	I5 and above
Speed	1.99 GHz

Figure 6: Hardware Requirements

Backup Storage	OneDrive, Google Drive
Language and Tools	Google Collab, Python, Web browser
Libraries	PyTorch, NumPy, pandas

Figure 7: Software Requirements

## 6.2 Implementation Requirements

The implementation of this study was executed on jupyter notebook. Python is used as the operational language to translate the design specification to executable code. This is due to Python's excellent support for machine learning and deep learning model development. Some of the libraries utilized in this implementation are:

1. Pandas (Data Handling)
2. Openslide (Image handling)
3. OpenCV (Image Processing)
4. Numpy (Array processing)
5. Matplotlib (Visualizations)
6. Keras (Model development and testing)

## 6.3 Classification Model

The processed data is then split into train set and validation set using 80/20 split with a random state of 101. Before applying the pre-processed data to the model, the split data is stored into different directories depending upon the Grade level of the ovarian cancer using the shutil module. Python's shutil module is used to manipulate files or their collections, which are directories. This module is comparable to Python's OS module (Meng et al. 2021).



Figure 8: training and validation dataset



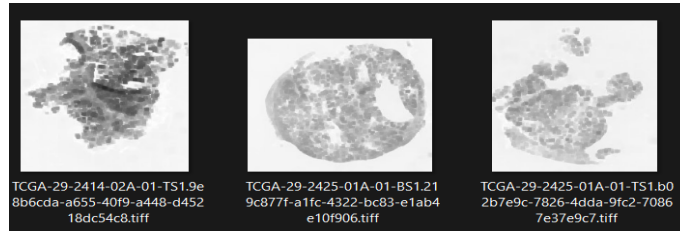


Figure 9: Grade 2 sample images



Figure 10: Grade 3 sample images

## 6.4 Image Augmentation

In my work, I utilized Keras ImageDataGenerator, which allows to enhance the model throughout the training process. It provides several Techniques such as flips, rotation, normalization, and so forth. At each epoch, the ImageDataGenerator class ensures that The model is updated with new picture variants. However, it just returns the converted photos and does not include them. It also consumes less RAM than the original corpus The ImageDataGenerator class consists of horizontal flip and vertical flip. Flip is a tool for flipping pictures along the vertical and horizontal axes. I implemented both horizontal and vertical flip parameter in my research.

## 6.5 VGG16 Model

The flattened weighted feature map created from the last pooling layer is utilized as input to the fully connected network, which determines the loss and alters the weights of the hidden nodes accordingly during the classification process. Using the Vgg16 model the process of classification is carried out. The vgg16 model can be summarized as follows:

The Vgg16 network was constructed with 512 inputs layer and 1 dense layer of 1024 units with 'relu' activation function. The output layer consist of sigmoid activation function . This model uses a batch size of 32. The optimizer parameter used in the model is 'adam' and 'binary\_crossentropy' parameter was used for

calculating the loss function.

```
Model: "sequential"
Layer (type)                Output Shape                Param #
-----
vgg16 (Functional)          (None, 8, 8, 512)          14714688
Flatten (Flatten)           (None, 32768)               0
dense (Dense)                (None, 1024)                33555456
dropout (Dropout)           (None, 1024)                0
dense_1 (Dense)             (None, 2)                   2050
-----
Total params: 48,272,194
Trainable params: 48,272,194
Non-trainable params: 0
```

Figure 11: VGG16 Model Summary

## 6.6 VGG19 Model

After execution of the VGG16 model, the same preprocessed data is applied to the VGG19 model. The VGG19 model can be summarized as follows:

The Vgg19 network was built with 512 inputs and 1 dense layer of 1024 units with the 'relu' activation function. The sigmoid activation function is used in the output layer. The batch size for this model is 32. The optimizer parameter 'adam' was utilized in the model, and the loss function was calculated using the 'binary\_crossentropy' parameter.

```
Layer (type)                Output Shape                Param #
-----
vgg19 (Functional)          (None, 512)                 20024384
flatten_3 (Flatten)         (None, 512)                 0
dense_8 (Dense)              (None, 1024)                525312
dropout_3 (Dropout)         (None, 1024)                0
dense_9 (Dense)             (None, 2)                   2050
-----
Total params: 20,551,746
Trainable params: 20,551,746
Non-trainable params: 0
```

Figure 12: VGG19 Model Summary

## 7 Evaluation and Results

The aim of this part is to offer a complete overview of the study's results and major findings. As this research paper consist of two implementations, two different evaluation reports are shown below.

### 7.1 Evaluation and result based on VGG16 model

The VGG16 model is implemented using 15 epochs and batch size of 32. The shuffle mode is set as false. Relu function is employed in this model, and sigmoid

activation function is utilized in the final output layer. The rate of learning is 0.0001. The learning rate parameter has a significant impact on the performance of the VGG16 model. The iteration will begin with a low beginning value, and the learning rate will grow with each iteration. Each iteration will display the loss gain, and if the loss value increases rapidly, the process will be terminated.

The below figure(12) shows the accuracy and validation graph of the VGG16 model. As per the graph it can be seen that the Training loss at the start of the iteration was 0.74 and decreased to 0.56 whereas the validation loss at the start of iteration was 0.93 and was decreased to 0.72. As per the Training and Validation accuracy, the validation accuracy was constant during the whole iteration time at 0.44 and training accuracy was 0.38 at the start of the iteration and it decreased to 0.72.

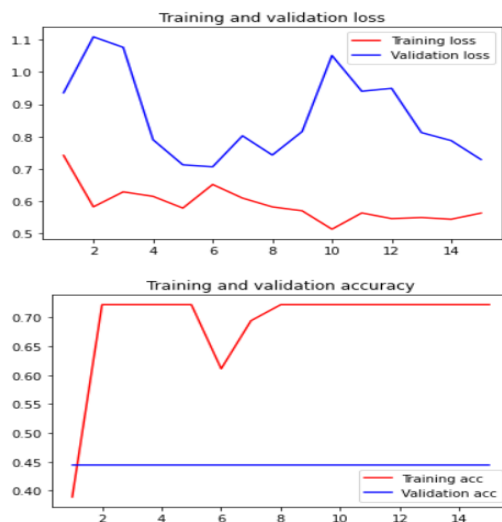


Figure 13: VGG16 training and validation loss versus epochs

## 7.2 Evaluation and result based on VGG19 model

The VGG19 model is implemented using 10 epochs and batch size of 32. The shuffle mode is set as false. Relu function is used in this model, and sigmoid activation function is utilized in the final output layer. The rate of learning is 0.001. The learning rate parameter has a significant impact on the performance of the VGG19 model. The iteration will begin with a low beginning value, and the learning rate will grow with each iteration. Each iteration will display the loss

gain, and if the loss value increases rapidly, the process will be terminated.

The below figure(13) shows the accuracy and validation graph of the VGG19 model. As per the graph it can be seen that the Training loss at the start of the iteration was 0.79 and decreased to 0.61 whereas the validation loss at the start of iteration was 0.76 and was climbed to 1.11. As per the Training and Validation accuracy, the validation accuracy was constant during the whole iteration time at 0.44 and training accuracy was 0.38 at the start of the iteration and it decreased to 0.72.

The below figure(12) shows the accuracy and validation graph of the VGG19 model.

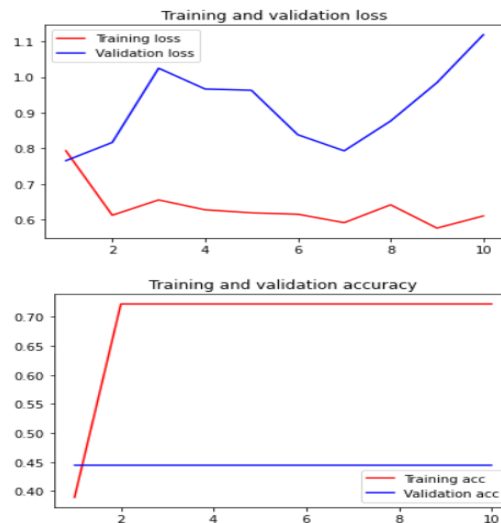


Figure 14: VGG19 training and validation loss versus epochs

## 8 Conclusion and Discussion

In this research paper, Deep learning architectures such as VGG16 and VGG19 were used. Various models were studied and based on that the classification models were selected. Both the models were trained on 46 images. As per the result, both the models did not performed good while classifying the ovarian tumor grade level. Both the model gave the validation accuracy score of 0.44%. The models needs to have more improvements. Model testing was done using the TCGA image dataset. In addition, several iterations of testing were performed to acquire more accurate findings by dividing the data into different sizes of test and train

and randomizing the data. This research paper consist of 3 limitations. The first limitation is the size of the data. As mentioned before, the data size of 46 images was 8 giga bytes with high resolution, hence only 46 images were used. The second limitation is regarding the bias between grade level 2 and grade level 3 images. The grade level 3 images are more than the grade level 2 images. The third limitation is regarding the computing power. With more computaional power, the model can be trained well.

Future work for this project will be to deal with all the above mentioned limitations. Model can also include other medical data such as genetic data for more analysis. Also using the same dataset, different pre-trained model such as Google-Net, Mobile-Net can be used with larger data for analysing and classifying the accuracy futher.

## 9 Acknowledgement

I'd like to express my heartfelt gratitude to Prof. Hicham Rifai for his valuable and constant guidance, direction, inspiration, and ongoing assistance throughout the research, without which this study would not have been possible.

## References

- Ahmed, M. R., Rehana, H. & Asaduzzaman, S. (2021), 'Ovarian cancer substantial risk factor analysis by machine learning: A low incoming country perspective', *Biointerface Research in Applied Chemistry* **11**(1), 8457–8466.
- Ahonen, T., Hadid, A. & Pietikäinen, M. (2006), 'Face description with local binary patterns: Application to face recognition', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28**(12), 2037–2041.
- Belsare, A. D. & Mushrif, M. M. (2012), 'H ISTOPATHOLOGICAL I MAGE A NALYSIS U SING I MAGE P ROCESSING T ECHNIQUES : A N O V E R V I E W', **3**(4), 23–36.
- Bhattacharjee, S., Singh, Y. J. & Ray, D. (2017), 'Comparative performance analysis of machine learning classifiers on ovarian cancer dataset', *Proceedings - 2017 3rd IEEE International Conference on Research in Computational Intelligence and Communication Networks, ICRCICN 2017* **2017-Decem**, 213–218.
- Booma, P. M., Thiruchelvam, V., Ting, J. & Ho, S. (2020), 'MAX POOLING TECHNIQUE TO DETECT AND CLASSIFY MEDICAL IMAGE FOR Max

- Pooling Technique to Detect and Classify Medical Image for Ovarian Cancer Diagnosis’, (April).
- Farnell, D., Farahani, H., Nursey, M., Tessier-cloutier, B., Jones, S. J. M., Huntsman, D. G., Gilks, C. B. & Bashashati, A. (2020), ‘Classification of Epithelial Ovarian Carcinoma Whole-Slide Pathology Images Using Deep Transfer Learning’, pp. 1–5.
- K, S. & C, G. (2022), ‘A hybrid model for efficient cervical cell classification’, *Biomedical Signal Processing and Control* **72**(PA), 103288.  
**URL:** <https://doi.org/10.1016/j.bspc.2021.103288>
- Kasture, K. R. (2021), ‘Multi-class Classification of Ovarian Cancer from Histopathological Images using Deep Learning -’, **16**, 1–6.
- Kasture, K. R., Shah, D. D. & Matte, P. N. (2021), ‘Research Article A New Deep Learning method for Automatic Ovarian Cancer Prediction Subtype classification Turkish Journal of Computer and Mathematics Education Research Article’, **12**(12), 1233–1242.
- Kawakami, E., Tabata, J., Yanaihara, N., Ishikawa, T., Koseki, K., Iida, Y., Saito, M., Komazaki, H., Shapiro, J. S., Goto, C., Akiyama, Y., Saito, R., Saito, M., Takano, H., Yamada, K. & Okamoto, A. (2019), ‘Application of artificial intelligence for preoperative diagnostic and prognostic prediction in epithelial ovarian cancer based on blood biomarkers’, *Clinical Cancer Research* **25**(10), 3006–3015.
- Komura, D. & Ishikawa, S. (2018), ‘Machine Learning Methods for Histopathological Image Analysis’, **16**, 34–42.
- Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V. & Fotiadis, D. I. (2015), ‘Machine learning applications in cancer prognosis and prediction’, *Computational and Structural Biotechnology Journal* **13**, 8–17.  
**URL:** <http://dx.doi.org/10.1016/j.csbj.2014.11.005>
- Linder, N., Konsti, J., Turkki, R., Rahtu, E., Lundin, M., Nordling, S., Haglund, C., Ahonen, T., Pietikäinen, M. & Lundin, J. (2012), ‘Identification of tumor epithelium and stroma in tissue microarrays using texture analysis’, *Diagnostic Pathology* **7**(1), 1–11.
- Lu, Y. & Han, J. (2003), ‘Cancer classification using gene expression data’, *Information Systems* **28**(4), 243–268.

- Masood, A., Sheng, B., Li, P., Hou, X., Wei, X., Qin, J. & Feng, D. (2018), 'Computer-Assisted Decision Support System in Pulmonary Cancer detection and stage classification on CT images', *Journal of Biomedical Informatics* **79**(June 2017), 117–128.
- Meng, X., Li, X. & Wang, X. (2021), 'A Computationally Virtual Histological Staining Method to Ovarian Cancer Tissue by Deep Generative Adversarial Networks', *Computational and Mathematical Methods in Medicine* **2021**.
- Wadhwa, G. (2021), 'A Deep Convolutional Neural Network Approach for Detecting Malignancy of Ovarian Cancer Using Densenet Model', **25**(2), 4449–4458.
- Wei, K., Li, T., Huang, F., Chen, J. & He, Z. (2022), 'Cancer classification with data augmentation based on generative adversarial networks', *Frontiers of Computer Science* **16**(2).