# LEVERAGING DEEP LEARNING FOR OVARIAN CANCER CLASSIFICATION USING IMAGE DATA

**BY**

**Gazi Adnan**
**ID: 202-15-3796**

**AND**

**Md. Tanvir Mahmud**
**ID: 202-15-3811**

## FINAL YEAR DESIGN PROJECT REPORT

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

## Supervised By

**Dr. S. M. Aminul Haque**
Professor and Associate Head
Department of Computer Science and Engineering
Daffodil International University

## Co-Supervised By

**Mr. Md. Mubtasim Fuad**
Lecturer
Department of Computer Science and Engineering
Daffodil International University

## DAFFODIL INTERNATIONAL UNIVERSITY

**DHAKA, BANGLADESH**

**1 July 2024**

## APPROVAL

This Project titled **"Leveraging Deep Learning for Ovarian Cancer Classification Using Image Data"**, submitted by **Gazi Adnan, ID:202-15-3796** and **Md. Tanvir Mahmud, ID:202-15-3811** to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 15-07-2024.

### BOARD OF EXAMINERS

Hossain 15.07.2024

**Dr. Md. Fokhray Hossain** — **Board Chairman**
**Professor**
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Firoz Hasan 15.7.24

**Md. Firoz Hasan** — **Internal Examiner 1**
**Sr. Lecturer**
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Lamia Rukhsara 15.07.24

**Lamia Rukhsara (LR)** — **Internal Examiner 2**
**Sr. Lecturer**
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

15.7.24

**Dr. S. M. Hasan Mahmud (SMH)** — **External Examiner**
**Assistant Professor**
Department of Computer Science
American International University Bangladesh

*©Daffodil International University, Bangladesh* ii

## DECLARATION

We hereby declare that this project has been done by us under the supervision of **Dr. S. M. Aminul Haque, Professor and Associate Head, Department of Computer Science and Engineering, Daffodil International University**. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

**Supervised by:**

**Dr. S. M. Aminul Haque**
Professor and Associate Head
Department of CSE
Daffodil International University

**Co-Supervised by:**

15.9.24

**Mr. Md. Mubtasim Fuad**
Lecturer
Department of CSE
Daffodil International University

**Submitted by:**

Gazi Adnan

**Gazi Adnan**
ID: 202-15-3796
Department of CSE
Daffodil International University

Tanvir
15-9-24

**Md. Tanvir Mahmud**
ID: 202-15-3811
Department of CSE
Daffodil International University

©Daffodil International University, Bangladesh iii

# ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to almighty for His divine blessing making it possible for us to complete the final year thesis successfully.

We are grateful and wish our profound indebtedness to **Dr. S. M. Aminul Haque**, **Professor and Associate Head**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Data Mining*" to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting them at all stages have made it possible to complete this project.

We would like to express our heartiest gratitude to **Dr. Sheak Rashed Haider Noori**, Professor & Head, Department of CSE Daffodil International University, for his kind help in finishing our project and also to other faculty members and the staff of the Department of CSE, Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discussion while completing the course work.

Finally, we must acknowledge with due respect the constant support and patience of our parents.

# ABSTRACT

Ovarian cancer presents a significant challenge to global healthcare, demanding accurate diagnosis and specialized treatment plans. Leveraging advancements in deep learning, particularly in medical imaging, offers a promising approach to enhance subtype classification. This study introduces a robust deep learning framework for precise categorization of ovarian cancer subtypes from image data. Driven by the urgent need to improve diagnostic accuracy, our system integrates deep learning methods with convolutional neural networks (CNNs) to deliver reliable classification outcomes. Through preprocessing of image data and utilization of pre-trained CNN models for feature extraction, complex patterns corresponding to various ovarian cancer subtypes are identified. Classification is further enhanced through fine-tuning and ensemble learning techniques to maximize performance. The proposed framework addresses critical gaps in current diagnostic methods, providing clinicians with a rapid and accurate tool for subtype classification. Rigorous experimentation and validation aim to develop a system capable of generalizing across diverse patient populations, ultimately improving patient outcomes in ovarian cancer treatment. Experimental results demonstrate that our proposed model achieves a significant accuracy of 96%, along with 96% precision, recall, and F1-score, and a 97% area under the ROC curve (AUC-ROC) score. While our model's accuracy surpasses that of traditional transfer learning models such as VGG16, VGG19, Xception, and EfficientNetB4, further refinements are essential to maximize its potential in clinical applications. This study underscores the importance of advancing medical research through data-driven approaches and highlights the critical role of ongoing progress in healthcare.

**Keywords:** Ovarian Cancer Classification, Histopathological Images, Deep Learning, Convolutional Neural Networks (CNNs), Transfer Learning, EfficientNet, VGG16, VGG19, Xception, Medical Image Analysis, Cancer Subtypes.

## TABLE OF CONTENTS

## LIST OF FIGURES

## LIST OF TABLES

# CHAPTER 1
# INTRODUCTION

## 1.1 Overview

An artificial intelligence method for computer-aided diagnosis, such as the diagnosis of ovarian disease, is called deep learning (DL). While professional researchers, doctors, and physicians, classification are criticized for being costly and time-consuming, deep learning (DL) has been demonstrated to be successful and efficient in the classification and detection of illnesses using photos [1]. It is often known that the most lethal type of cancer affecting the female reproductive system is ovarian cancer. Ovarian cancer ranks sixth in women's cancer-related fatalities and has the highest mortality rate among gynecologic malignancies [2]. This one influences the death rates related to cancers of the female reproductive system. More women than any other cancer of the reproductive system pass away from ovarian cancer. For women, the lifetime risk of ovarian cancer is 1 in 78. They have a 1 in 108 lifetime risk of developing ovarian cancer. Older women are primarily affected by this cancer. Most patients with ovarian cancer are sixty-three years of age or older. Compared to black women, white women receive the diagnosis more often [3]. The awareness and application of subtype-specific therapy approaches have grown as a result of the pursuit of improved patient outcomes. Nevertheless, precisely defining subgroups is essential for such tailored therapy. At the moment, pathologists' skill is mostly responsible for the identification of ovarian cancer. However, this reliance poses considerable challenges, including disagreements between observers, doubts regarding the reproducibility of diagnoses, and the severely limited supply of pathologists with specialized knowledge, especially in underprivileged areas. Given that ovarian cancer has a 4.7% death rate and an incidence rate of 3.4%, it poses a serious danger to the general public's health. This condition has a significant yearly burden, impacting over 300,000 women and causing about 152,000 fatalities. These figures highlight the serious effects that ovarian cancer has on the general prognosis and health of women [4-5]. The fact that only 30% of those who are diagnosed with ovarian cancer go on to survive indicates that the prognosis is not good [6]. The current standard of care for ovarian cancer is platinum-based chemotherapy and cytoreductive surgery [7]. According to the 2020 World Health Organization, there are at least five main categories of ovarian carcinomas, which are identified by things like immunological profile, histology, and molecular analysis. The many forms of ovarian carcinoma may be divided into the following categories: High-grade serous Carcinoma (HGSC), which accounts for 70–80% of cases, is the most prevalent variety. A few less frequent subtypes are Mucinous Carcinoma (MC, 3-4%), Clear-Cell Carcinoma (CC, 6–10%), Low-Grade Serous Carcinoma (LGSC, 5%), and Endometrioid Carcinoma (EC, 10%) [8-10]. The current hierarchy-based subclassification scheme for ovarian carcinomas begins with the classification of the tumors into histotypes according to traditional histological characteristics. Based on a number of variables, such as the cell of origin, molecular alterations, clinical traits, and therapeutic modalities, histotypes are categorized as discrete. A variety of malignancies with unique precursor lesions, origin, metastatic patterns, response to treatment, and prognosis are characteristics of the heterogeneous disease known as ovarian carcinoma Prat et al. and Kobel et al. In order to make the transition from the broad talk about cancer subtypes to our particular technique seamless, we describe in detail how our research tackles the issues raised. The objective of this study is to examine the intricate variations observed in various types of ovarian cancer, which provide formidable obstacles to the resolution of cases using current diagnostic methods. Our thorough study of cancer subtype categorization and outlier identification concentrated on the intricate field of ovarian cancer. A variety of tissue microarray (TMA) and whole-slide pictures made up the dataset used in this investigation. There are 5000 samples in the collection altogether for our study. This brought up a number of problems with size variations, source origin, picture quality, and staining techniques. This emphasizes how crucial it is to put in place a robust error-handling system. Our goal is to accurately classify CC, EC, HGSC, LGSC, and MC. This work attempts to leverage the Attention Embedder model's capabilities to improve customized treatment approaches.

## 1.2 Motivation

The devastating impact of ovarian cancer on women's health necessitates a paradigm shift in diagnostic approaches. This research is driven by the critical need to address the limitations of traditional methods, particularly the subjectivity and inter-observer variability inherent in histology analysis for ovarian cancer subtype classification. These limitations can lead to misdiagnosis, delayed treatment initiation, and ultimately, poorer patient outcomes. Deep learning has revolutionized image analysis, offering exceptional capabilities for pattern recognition and feature extraction from complex medical images. This project delves into the exciting potential of deep learning, specifically leveraging convolutional neural networks (CNNs), to develop a robust and objective framework for ovarian cancer subtype classification based on image data. By harnessing the power of CNNs to identify subtle yet critical features within ovarian tissue images, we aim to surpass the limitations of human analysis and achieve a highly accurate, reliable, and efficient classification system. The ultimate motivation for this endeavor lies in its potential to transform the landscape of ovarian cancer diagnosis. A more precise and objective classification system, enabled by deep learning, can empower clinicians to make informed treatment decisions with greater confidence. This could lead to earlier intervention, improved treatment personalization, and ultimately, a significant improvement in patient survival rates. Furthermore, this research contributes to advancements in computational pathology, paving the way for a deeper understanding of the biological characteristics of different ovarian cancer subtypes. By unraveling these complexities, we can pave the way for the development of more targeted therapies in the future.

## 1.3 Rational of the Study

The necessity to solve pressing issues in the fields of respiratory illness diagnostics and medical imaging is what led to the creation of the "Leveraging Deep Learning for Ovarian Cancer Classification Using Image Data" project. Ovarian sickness is still a major global health concern, and the management and treatment of this condition depend heavily on a prompt and precise diagnosis. The model's capacity to recognize patterns of ovarian sickness in medical photos is improved by carefully incorporating a potent artificial intelligence technique called transfer learning, which leverages historical data from unrelated tasks. Because deep convolutional neural networks (CNNs) have been proved to be effective in image processing, using them boosts the potential of the study. This study analyzes segmentation and detection strategies in addition to basic detection. Understanding the subtleties of these techniques in depth is essential to understanding the geographic scope and complexity of ovarian cancer patients. By carefully analyzing both detection and segmentation, the research seeks to shed light on the advantages and disadvantages of each method, so making a substantial new contribution to the field of science. This comparative study is essential for improving the current diagnostic tools and encouraging the creation of more accurate and dependable systems.

### 1.4 Objective

The ultimate objective of the project is to provide a revolutionary increase in the precision and efficacy of detection and segmentation methods, paving the way for thereafter significant progress in the diagnosis of ovarian cancer. By improving our knowledge of the most effective ways to use deep CNNs and transfer learning in this context, we intend to enhance both the theoretical and applied disciplines of healthcare through this work. Better diagnostics directly and favorably affect medical outcomes as well as crop health in general.

### 1.5 Expected Output

The goal of this study is to increase knowledge in medical science and technology while also producing theoretical and practical results. The expectations that have been expressed include:

- Building a Sturdy Disease Detection System: The main outcome of this work is the development of a reliable and accurate method for ovarian disease detection. Modern deep learning techniques, including our suggested model and models like VGG19, VGG16, Xception, and EfficientNetB4, will be used in this system. The strategy seeks to diagnose different ovarian illnesses with a high degree of accuracy to support early intervention efforts.

- Assessment and Partial Comparison of Deep Learning Models: There will be a thorough evaluation and analysis of the selected deep learning models. The efficiency, accuracy rates, and performance indicators of each model will be covered in this study so that readers may decide for themselves how useful it is for ovarian illness detection.

- Integration of Predictive Strategies: One essential expected outcome is the integration of detection methods into the disease detection system. The technique aims to identify illnesses that are present as well as predict potential outbreaks in their early phases. This proactive approach, which aims to lower losses and increase the overall resilience of health management, is in line with medical best practices.

- Application of Image Processing Techniques: The goal of the project is to increase the efficacy and accuracy of disease diagnosis by utilizing sophisticated image processing techniques including picture segmentation and clustering. The system combines several techniques to provide comprehensive insights on the characteristics and geographic distribution of ovarian disorders.

- Validation of Findings by Extensive Testing: Strict procedures will be developed to verify the efficacy of the suggested illness detection method. Several datasets will be used in the validation process to verify the system's adaptability and resilience in a variety of situations and environmental changes that may affect the environments where ovarian illnesses are found.

- Contribution to Scholarly Conversation: In order for our proposed method to deliver the most correct answers to patients' problems in the real world, the study aims to further the scholarly conversation on the intersection of deep learning and medical expertise.

## 1.6 Research Questions

The project intends to address a number of significant questions in the area of "Deep Learning Models for Ovary Disease Detection". These inquiries function as guidance markers, directing the research toward a thorough comprehension of the difficulties associated with using cutting-edge artificial intelligence systems for improved medical diagnostics. The primary research questions are as follows:

- What are the expected milestones and project development course?
- In what ways can we evaluate how well our model detects objectionable content?
- What are the advantages of our proposed model?
- What are the limitations of the study?
- What is the future plan of the study?
- What are the techniques of data preprocessing of the study?

These research questions collectively give the thesis its conceptual framework and guide the investigation of the complexities of deep CNNs, segmentation methods, transfer learning, and ovarian illness diagnosis. The project intends to contribute to the ongoing discussion regarding the link between artificial intelligence and health care by providing meaningful information that will impact medical diagnostics going ahead through a methodical approach to these topics.

# CHAPTER 2
# LITERATURE REVIEW

## 2.1 Overview

Deep learning-based skin disease detection has gained substantial traction in the medical community, evident from the wide array of research and methodologies documented in the literature. Numerous scholars have explored automated diagnosis and classification of various skin disorders using deep learning, particularly convolutional neural networks (CNNs). Studies have utilized diverse datasets, ranging from publicly available repositories to meticulously curated collections of clinical photographs, to train and evaluate deep learning models for identifying skin diseases.

## 2.2 Related Works

Prior to our starting to use image databases for Ovary Disease classification, several significant research initiatives were conducted. The tremendous insights and methods that these ground-breaking studies have given to the field have allowed for improvements in therapeutic efficacy and diagnostic precision. Bedi et al., an innovative approach utilizing AResUNet for ovarian cancer detection demonstrated a remarkable 98% detection rate, showcasing the model's effectiveness [11]. Farahani et al. developed a model that achieved 81.38% diagnostic concordance on the training set and 80.97% on an independent test set from a different institution, suggesting its potential to augment diagnostic accuracy in conjunction with traditional histopathology [12]. Nobel SMN. et al. reported a model with notable accuracies: 96.42% in training, 95.10% in validation, and 91.37% accuracy in testing, indicating its reliability [13]. Vázquez MA et al. introduced two methods: a Bayesian hierarchical model and a recurrent neural network (RNN) for binary classification, both achieving an ROC curve (AUC) of 0.98%, demonstrating high predictive power [14]. Akazawa et al. explored the use of the XGBoost algorithm, achieving a top accuracy of 0.80, which signifies its potential for application in this field [15]. Ziyambe et al. implemented a CNN model that achieved a 94% accuracy rate, further affirming the utility of deep learning in ovarian cancer detection [16]. To demonstrate the predictive power of the Inception V3 deep learning model, HoD.J. et al. used it to forecast

how patients with high-grade serous ovarian cancer will respond to platinum-based chemotherapy [17]. Similarly, Liu et al. showed the viability of the strategy by using the InceptionV3 model to predict ovarian cancer responses to platinum-based chemotherapy based on pre-treatment histopathology scans. [18]. A substantial advancement in diagnostic techniques has been demonstrated by Kasture et al.'s development of a unique DCNN that considerably improves the detection of ovarian cancer through high classification accuracy [19]. Finally, Gajjela et al. achieved a classification accuracy of 0.98 for ovarian tissue subtype categorization by combining deep learning with mid-infrared spectroscopic imaging. This marks a significant advancement in quantitative, label-free cancer diagnostics [20]. Jopek et al. introduce an innovative multiclass cancer classification method using deep learning on liquid biopsy data, achieving notable accuracies and offering new insights into cancer detection through tumor-educated platelets, supported by explainable AI techniques for enhanced diagnostic trust [21]. Abuzinadah et al. present an advanced ovarian cancer prediction model using a stacked ensemble classifier integrated with SHAP for explainable AI, achieving an impressive accuracy of 96.87%, showcasing the model's effectiveness in enhancing diagnostic precision through comprehensive feature analysis [22]. Hema et al. propose an innovative ovarian cancer detection methodology utilizing a region-based Convolutional Neural Network (CNN), showcasing enhanced diagnostic precision through a novel segmentation and classification approach, achieving over 95% precision in detecting ovarian cancer from segmented histopathological images [23]. Guo et al. developed a novel deep learning framework for ovarian cancer subtype identification using multi-omics data and denoising autoencoders, achieving enhanced clustering performance and identifying significant molecular markers [24]. Wadhwa et al. develop an efficient framework using DenseNet-
201 for early ovarian cancer detection through histopathological image analysis, achieving a high accuracy of 94.73%. demonstrating the model's potential in improving early diagnosis and treatment strategies [25].

### 2.3 Comparative Analysis

This research on "Leveraging Deep Learning for Ovarian Cancer Classification Using Image Data" would not be possible without the comparative analysis, which highlights the subtle differences and complementing features of various methods. The study employs both conventional techniques and cutting-edge segmentation algorithms to thoroughly assess the effectiveness of deep Convolutional Neural Networks (CNNs) and transfer learning in the detection of ovarian illness. The research evaluates the advantages and disadvantages of each strategy through thorough testing and experimental result analysis, offering a thorough grasp of their distinct contributions to the diagnostic process. An in-depth investigation of the segmentation and detection algorithms' contributions to improving the precision and effectiveness of ovarian cancer diagnosis is made possible by their extensive examination. The study greatly increases the understanding of the subject and helps practitioners and researchers choose the best solutions for their particular situations by comparing the results of various approaches. The study does more than just point out distinctions; it also illustrates how segmentation and detection might cooperate to enhance our understanding of the characteristics and geographic distribution of hospitals affected by ovarian cancer.

Finally, the comparative analysis in this work offers an important lens through which to assess the efficiency of deep CNNs and transfer learning for ovarian cancer diagnosis. It highlights the need of a thorough approach in medical diagnosis and paints a complicated picture by fusing traditional and cutting-edge techniques. The results enhance our understanding of ovarian illness detection and offer helpful insights for enhancing existing diagnostic methods. In the next chapters, the comparative study provides a solid foundation for the discussion of the social effect, suggestions, and implications for more research.

TABLE 2.3.1: Comparative Analysis Table

| SL No | Author Name | Used Algorithm | Best Accuracy with Algorithm |
|---|---|---|---|
| 1. | Farahani et al. | Ensemble | 81.38% |
| 2. | Vázquez MA et al. | Bayesian hierarchical model, RNN | Bayesian hierarchical model, RNN both similar 98% Auc score. |

| | | | |
|---|---|---|---|
| 3. | Akazawa. et al. | XGBoost | 80% |
| 4. | Ziyambe et al. | CNN | CNN 94% |
| 5. | Kasture et al. | DCNN | 83.93% |
| 6. | Gajjela et al. | O-PTIR with deep learning | 98% |
| 7. | Jopek et al. | ResNet architecture | 66.51% |
| 8. | Hema et al. | Region-based CNN | 95% |
| 9. | Wadhwa et al. | DenseNet-201 | 94.73% |
| 10. | Our proposed work | 4 pre-trained models with one custom model | 96% |

## 2.4 Scope of the Problem

The "Leveraging Deep Learning for Ovarian Cancer Classification Using Image Data" aim of this work expands the problem's scope to encompass patient ailments in specific as well as medical diagnostics in general. This endeavor is motivated by the need to close the large gaps in current knowledge on the diagnosis and subtype classification of ovarian disease. The different kinds of ovarian illness cannot often be accurately and efficiently distinguished using conventional approaches, leading to treatment outcomes that fall short of ideal and diagnostic delays. This study attempts to develop a robust framework for categorizing Ovary Disease subtypes from visual data using convolutional neural networks (CNNs), one of the deep learning approaches. In the process, we want to revolutionize the diagnostic process and provide quick, trustworthy information to medical practitioners, enabling them to customize treatment regimens and ultimately improve patient outcomes in the battle against illness. The scope of the issue includes the difficulties in diagnosing ovarian sickness with conventional techniques, which may have limits on overall diagnostic sensitivity, specificity, and accuracy.

## 2.5 Challenges

The study explores the use of transfer learning, a rapidly developing artificial intelligence paradigm, to improve ovarian diagnostics. By using information from unrelated activities, transfer learning enables the model to possibly handle the complexity and unpredictability observed in medical imaging linked to ovarian sickness. Since deep convolutional neural networks (CNNs) are a good tool for extracting detailed patterns and features from complex visual input, integrating CNNs adds even more intricacy and sophistication to the system. The scope of the challenge also includes a comparative study of segmentation and detection techniques, including a thorough investigation of their advantages, disadvantages, and possible synergies. For a thorough diagnosis, an understanding of the architectural design and features of hospitals impacted by ovarian cancer is crucial, and the research attempts to provide insights that go beyond conventional screening techniques. The problem scope essentially includes the difficulties in detecting ovarian disease, the possible improvements provided by deep CNNs and transfer learning, and the comparative analysis of different methods. The initiative aims to improve the precision of presently existing diagnostic tools and lessen the impact of ovarian cancer on public health by tackling these issues.

# CHAPTER 3
# METHODOLOGY

## 3.1 Overview

The research project "Leveraging Deep Learning for Ovarian Cancer Classification Using Image Data" focuses on diagnosing ovarian diseases by utilizing state-of-the-art AI techniques in medical diagnostics. The primary emphasis of our study is ovarian disease, a common respiratory condition that may have dangerous consequences. The paper examines the limitations of current methods for diagnosing ovarian cancer and investigates how deep convolutional neural networks (CNNs) and transfer learning might be used to overcome these limitations.

Transfer learning is a unique notion in artificial intelligence that allows one to use knowledge from unrelated jobs to the challenging subject of diagnosing ovarian illness. When complex patterns and attributes are retrieved from medical photographs utilizing Deep CNNs—which are widely recognized for their efficacy in image analysis—the diagnostic process becomes more precise and detailed.

The comparative analysis portion of the study looks at the differences and synergies between various segmentation and detection methods. While detection determines whether the sickness is present in medical pictures, segmentation concentrates on characterizing the characteristics and spatial distribution of hospitals affected by ovarian disease. The goal of the study is to get a comprehensive understanding of the benefits and drawbacks of various techniques, which will aid in the development of more dependable and effective diagnostic tools. The intersection of medical imaging, respiratory health, and artificial intelligence is the primary emphasis of this work. In particular, it examines the application of deep CNNs for transfer learning to enhance ovarian illness detection. The findings aim to improve health performance and increase diagnostic accuracy, which will assist both the academic and practical facets of health care.

### 3.2 Instrumentation:

The testing for this study was conducted on Google CoLab using Kera's library. TensorFlow, one of the greatest deep learning frameworks, was used to handle Python machine learning techniques. After building each model and training it on a cloud-based Tesla GPU, Google made the models available through the Google Collaboratory platform. The Collaboratory framework provides up to 16GB of RAM (random-access memory) and around 360GB of GPU in the cloud for research purposes. The architectures for ovarian disease detection in this study employ the following models: EfficientNetB4, Vgg16, Vgg19, and our suggested model. From each category, one architectural design was selected. The study creates deep learning models and associated approaches using Python and TensorFlow, two computer languages and frameworks. Deep learning tasks are highly complex, and their effective completion requires computer infrastructure, including GPUs for quick processing.

Here's an overview of the materials used in Anaconda, Jupyter Notebook, NumPy, Pandas, Matplotlib, and Seaborn:

Anaconda: Anaconda is a distribution of the Python and R programming languages for scientific computing, aimed at simplifying package management and deployment.

- Anaconda Navigator: A desktop graphical user interface (GUI) that allows users to manage environments, packages, and channels easily.

- Conda: A package manager that installs, runs, and updates packages and their dependencies.

- Python Interpreter: The core Python programming language, including standard libraries and additional scientific computing packages.

- Jupyter Notebook: An interactive computing environment for creating and sharing documents that contain live code, equations, visualizations, and narrative text.

- Jupyter Notebook: Jupyter Notebook is an open-source web application that allows you to create and share documents containing live code, equations, visualizations, and narrative text.

- Markdown Cells: For text-based documentation and explanation.

- Code Cells: For writing and executing Python code interactively.

- Output Cells: Display the output of executed code, including text, images, plots, and interactive widgets.

- Kernel: The computational engine that executes the code contained in the notebook.

NumPy: NumPy is a fundamental package for scientific computing in Python. It provides support for arrays, mathematical functions, linear algebra operations, random number generation, and more.

- ndarray: An efficient multidimensional array object that supports mathematical operations.

- Mathematical functions: Functions for array manipulation, mathematical operations, linear algebra, Fourier transforms, and random number generation.

Pandas: Pandas is a powerful data manipulation and analysis library for Python. It provides data structures and functions for efficiently working with structured data.

- DataFrame: A two-dimensional labeled data structure with columns of potentially different data types, similar to a spreadsheet or SQL table.

- Series: A one-dimensional labeled array capable of holding any data type.

- Data manipulation functions: Functions for reading/writing data, cleaning, filtering, transforming, aggregating, and merging datasets.

Matplotlib: Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python. It provides a MATLAB-like interface for plotting.

- pyplot module: A collection of functions that provide a MATLAB-like interface for creating plots and visualizations.

- Figure and Axes objects: Components of the plot where data and visual elements are rendered.

- Various types of plots: Including line plots, scatter plots, bar plots, histogram plots, pie charts, etc.

### 3.3 Proposed Methodology

Our study outlines a comprehensive approach for ovarian cancer subtype classification using advanced deep learning techniques. We employed several Convolutional Neural

Network (CNN) architectures to automatically extract features from histopathological images, with a focus on evaluating and comparing their performance. The methodologies include the use of VGG16, VGG19, Xception, EfficientNetB4, and a newly proposed custom model. The trials are shown in Figure 3.3.1.

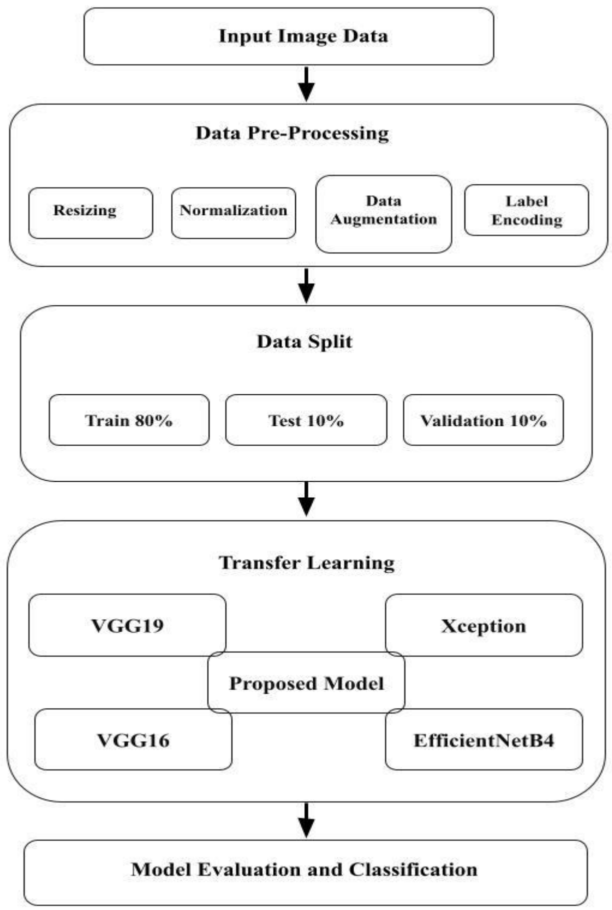


Figure 3.3.1: Methodology of the work

**3.3.1 Convolutional Neural Networks (CNNs):** Convolutional Neural Networks (CNNs) use the mathematical operation of convolution to blend data, forming feature maps through multiple layers, including input, convolutional, pooling, fully connected, and output layers.

Initially, CNNs detect simple patterns like lines and curves, and in later stages, they identify complex patterns such as objects and people, making them widely used for image recognition and classification. Transfer learning enhances CNN performance by utilizing pre-trained networks, adding new layers for specific tasks while retaining early layers' learned features. This approach is particularly effective with limited training data, improving efficiency and accuracy by fine-tuning the network for new tasks.

**3.2.2 VGG19:** One example of this is VGG19, a 19-layer variant of VGG that allows explicit and efficient spatial information transfer between neurons in the same CNN layer. This approach is very effective with things that have unusual forms. By utilizing tiny (3×3) convolution filters to evaluate networks with varying depth and detect left/right and up/down orientations, VGG19 achieves a major breakthrough. The input is linearly transformed using a Rectified Linear Unit (ReLU) following 1x1 convolution filters. The spatial resolution is preserved since the convolution stride is set at one pixel following convolution.

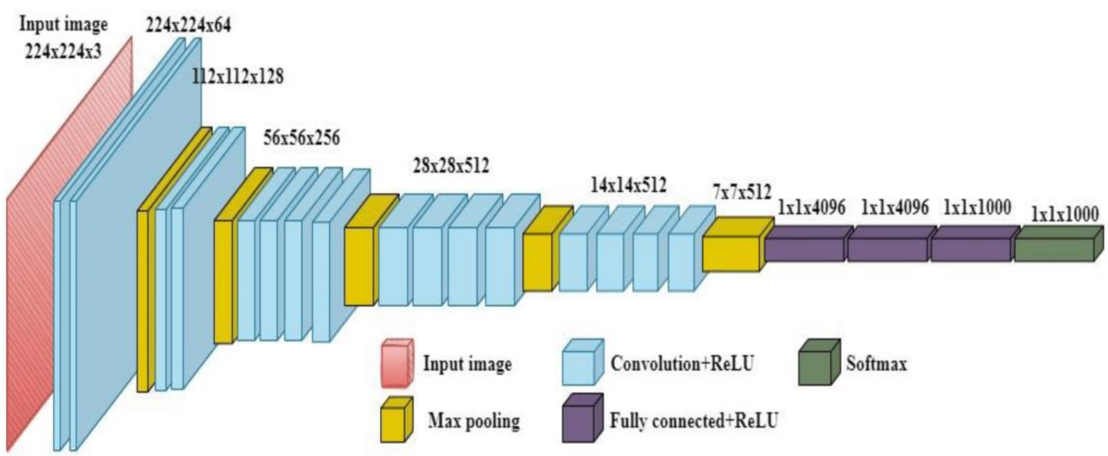


Figure 3.3.2: Architecture of VGG19 [26]

**3.3.3 VGG16:** The University of Oxford's Visual Geometry Group (VGG) designed the VGG16 architecture, a deep convolutional neural network. The 16-layer VGG16 architecture is noted by its uniformity and simplicity. It has three fully connected layers, thirteen convolutional layers, and a few max pooling layers scattered throughout, in addition to ReLU activation algorithms. Little 3x3 filters are used in the convolutional layers and layered to form a deep network that can recognize intricate characteristics. After a sequence of convolutional layers, max pooling layers minimize the spatial dimensions, which lowers the computational load and aids in the extraction of hierarchical features. To combine these features and provide final predictions, the fully linked layers at the

end of the network are helpful. Concerning image classification tasks, VGG16's consistent design and deep structure make it quite successful; it achieves notable performance on benchmark datasets such as ImageNet. Because of its many parameters, VGG16 has a lot of depth but is computationally costly; yet, because of its straightforward architecture, it is a fundamental model in the deep learning area.

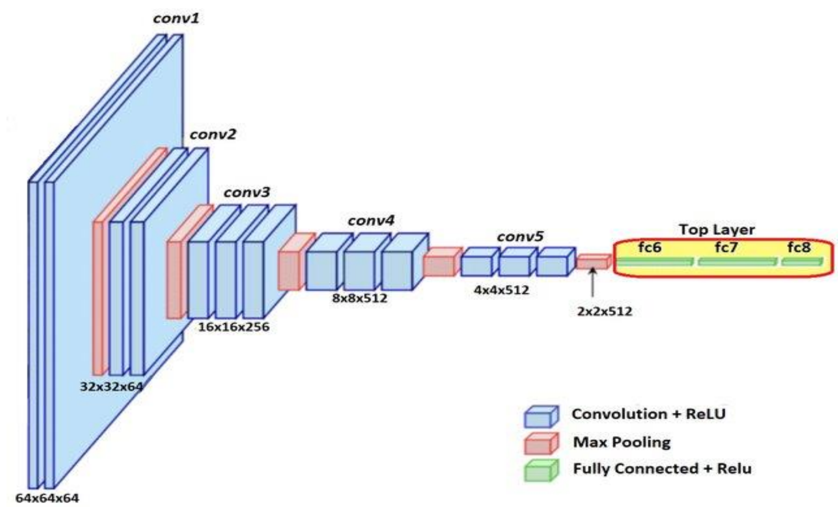


Figure 3.3.3: Architecture of VGG16 [27]

**3.3.4Xception:** François Chollet created the Xception architecture, a more sophisticated convolutional neural network (CNN) that expands on the Inception design by utilizing depthwise separable convolutions, a more effective type of convolution. Xception, which stands for "Extreme Inception," is a system that substitutes depthwise separable convolution layers for the normal Inception modules. These layers break down the convolution process into two more straightforward operations: a depthwise convolution and a pointwise convolution. As a consequence, model performance is maintained or even improved while computational complexity and parameter values are significantly reduced. Every entrance flow, middle flow, and exit flow module in the architecture is intended to gradually abstract and collect characteristics from the input pictures. The efficiency and performance of Xception's architecture make it a great choice for a range of image recognition and classification applications. The model's capacity to efficiently train deep networks is increased by utilizing residual connections like to those seen in ResNet.

**3.3.5 EfficientNetB4:** EfficientNetB4, part of the EfficientNet model family, excels in performance with fewer parameters and lower computational costs compared to traditional CNNs. It employs compound scaling to uniformly scale depth, width, and resolution using a set of scaling coefficients. Starting with an EfficientNet design using squeeze-and-excitation (SE) optimized mobile inverted bottleneck MBConv blocks and depthwise separable convolutions, EfficientNetB4 enhances performance by deepening the network, expanding its width, and increasing input image resolution. Pre-trained on ImageNet, it provides a strong foundation for various computer vision tasks. Its final layers, including a global pooling layer and dense layers for classification, enable efficient detection and handling of complex patterns in high-resolution images.

### 3.4 Hardware/ Software Requirements

For the research project "Leveraging Deep Learning for Ovarian Cancer Classification Using Image Data" to be implemented effectively, a number of requirements must be satisfied. These implementation requirements encompass not just hardware and software components, but also activities such as data collection and model evaluation.

- **Hardware Requirements:**

- **Advanced computing resources:** The availability of a computer infrastructure with enough processing power and memory to handle the intricate tasks involved in image analysis and deep learning.
- **Graphics Processing Unit (GPU):** Accelerating the training of deep Convolutional Neural Networks (CNNs) with a GPU is crucial to cut down on the time needed to create a model.

- **Software Requirements**
- **Deep Learning Frameworks:** Use popular deep learning frameworks for CNN model building, training, and assessment, such as PyTorch or TensorFlow.
- **Python Programming Language:** Utilize Python's ubiquity and versatility to build deep learning algorithms and do data analysis.
- **Image Processing Libraries:** In your work, make use of image processing tools like OpenCV for pre-processing and picture augmentation.

**Data Collection:**

- **Diverse and Representative Dataset:** Assemble an extensive collection of medical images that includes both healthy and sick instances in order to ensure the models that are built are stable and applicable.
- **Annotated Dataset**: Ensure that labeled data, clearly annotated to show whether ovarian illness is present or absent, is available for model validation and training.
- **Model Training and Evaluation**
- **Transfer Learning Models:** It is possible to use transfer learning frameworks by utilizing pre-trained models, like those acquired from ImageNet. This approach uses the most recent data to enhance the diagnosis of ovarian illness.
- **Evaluation Metrics:** For both detection and segmentation tasks, establish and use relevant measures, such as accuracy, precision, recall, and F1 score, to evaluate the performance of the model.

**Ethical Considerations:**

- **Hospital confidentiality and data protection:** Compliance with legal and ethical standards to safeguard the confidentiality and integrity of healthcare data utilized for research.
- **Informed consent:** where appropriate, getting the informed consent of the people whose medical images are part of the collection.

**Documentation and Reproducibility:**

- **Code Documentation:** Complete documentation of the implementation code to ensure consistency, repeatability, and collaboration in the future.
- **Version Control:** use version control systems (such as Git) to monitor changes to the codebase and maintain an organized development process.

- The study may be conducted systematically and comprehensively by meeting these implementation requirements, ensuring the correctness of the results and expanding the application of deep CNNs and transfer learning to diagnose ovarian illness.

### 3.5 Project Management and Financial Analysis

Choosing the study topic and title was the first step in our research project and required careful attention. This process took about two weeks. We started by identifying potential research areas, doing some initial research to ensure the topic was relevant and doable, and then finalized the study topic and title. The next step was the literature review, which involved finding and analyzing important papers to better understand our topic. This process took about two months. We searched for relevant academic papers, read and summarized the key points, and organized the information based on themes and relevance to our research. After completing the literature review, we focused on maintaining our dataset, aiming to finish this by March 2023. This involved finding reliable data sources, collecting data, cleaning it to remove errors, and updating the dataset.

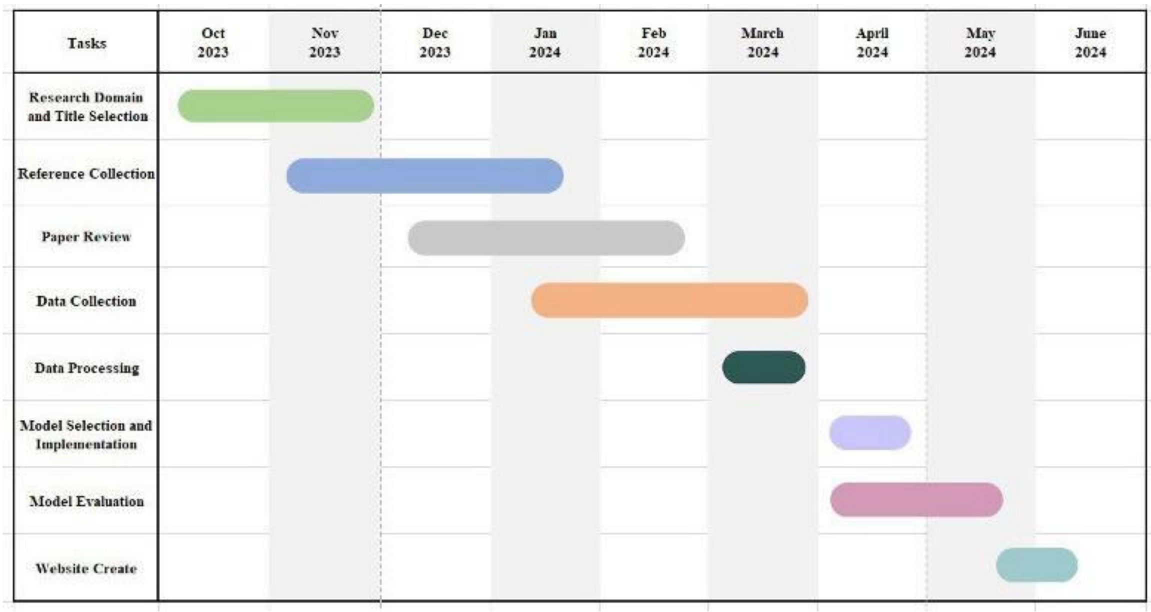


Figure 3.5.1: Gant Chart

Once we had the dataset, we preprocessed and coded the data to prepare it for analysis. This included normalizing and transforming the data and organizing it to be easy to analyze. Next, we implemented our research methods and analyzed the results. This involved applying our chosen methods to the data, analyzing the data to find meaningful insights, and interpreting the results to draw conclusions that matched our research goals.

Several costs were involved in the project. Copying costs for printing materials were estimated at 2,000 TK. Transport costs for traveling to libraries or other places were estimated at 1,750 TK. Domain and hosting costs for a website or database were estimated at 31,500 TK. To gather knowledge, we followed other articles related to our thesis, which cost 10,000 TK. To increase the loading capacity of the PC, we purchased 16 GB RAM, which cost 10,000 TK. For GPQ, the cost was 20,000 TK. The total estimated budget was 73,860 TK, and these expenses were self-funded.

Table 3.5.1: Financial Management

| SL NO | Costing Sections | Cost (BDT) |
|---|---|---|
| **1** | Data Set Collection (transport) | 1750 |
| **2** | Communication | 260 |
| **3** | Internet | 350 |
| **4** | Literature Review | 10000 |
| **5** | GPU | 20000 |
| **6** | RAM (16GB) | 10000 |
| **7** | Domain | 1500 |
| **8** | Hosting | 30000 |
| Contingency: 10% of the adjusted | **Sub Total =** | **73860** |

## 3.6 Summary

The methodology chapter outlines a comprehensive approach to enhancing ovarian cancer diagnosis using advanced AI techniques. Various CNN architectures, including VGG16, VGG19, Xception, EfficientNetB4, and a custom model, are employed for precise feature extraction from medical images. The research utilizes transfer learning and is conducted on Google CoLab with TensorFlow, leveraging high-performance computing resources. Implementation requirements include advanced hardware, relevant software, diverse datasets, and adherence to ethical standards. The project management section details a well-structured timeline and a self-funded budget of 73,860 TK, ensuring thorough documentation and reproducibility.

# CHAPTER 4

# IMPLEMENTATION

## 4.1 Overview

This chapter outlines the implementation of our deep learning-based approach for ovarian cancer subtype classification. Utilizing a dataset from Kaggle, our methodology leverages data pre-processing and augmentation techniques to enhance model accuracy and generalization. We trained multiple CNN architectures, including EfficientNetB4, Xception, VGG16, and VGG19, to classify various ovarian cancer subtypes. The chapter details the experimental setup, model training procedures, and evaluation metrics such as accuracy, precision, recall, and F1-score, demonstrating the efficacy of our approach in advancing medical diagnostics.

## 4.2 Train Model

### A. Datasets

For our research, we utilized a dataset from Kaggle as the primary source of medical imaging data related to ovarian cancer. Kaggle, a renowned platform for machine learning and data science competitions, offers a wide array of well-curated datasets, including those tailored for healthcare research. The selected Kaggle dataset, "Ovarian Cancer Subtype Classification," contains histopathological images of ovarian cancer tissue samples. This dataset was chosen for its comprehensive and diverse collection of images, representing various subtypes of ovarian cancer, including High-Grade Serous Carcinoma (HGSC), Clear-Cell Ovarian Carcinoma (CC), Endometrioid Carcinoma (EC), Low-Grade Serous Carcinoma (LGSC), and Mucinous Carcinoma (MC). While each subtype in the dataset originally contained more than 1,000 images, we selected 1,000 images per class due to computational constraints. This balanced and rich dataset is suitable for training and evaluating deep learning models. By leveraging this dataset, we aimed to enhance the accuracy and generalization of our deep learning models, enabling them to effectively identify and classify the distinct subtypes of ovarian cancer.

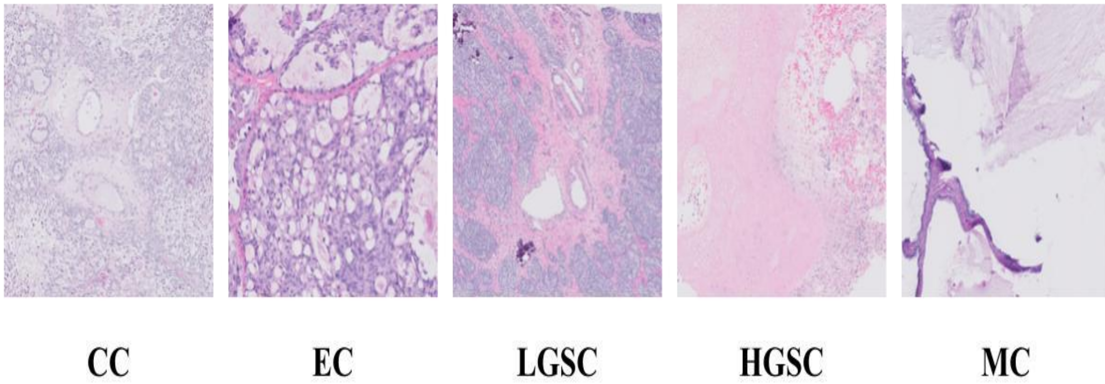


Figure 4.2.1: Ovary Disease Image

The data count plot is shown below:

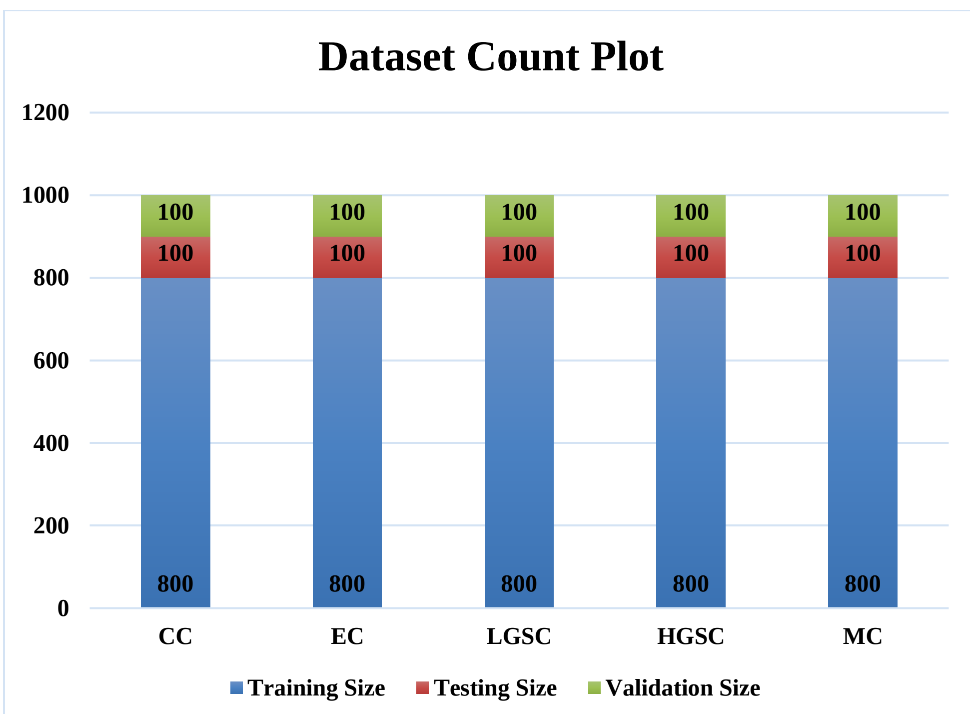


Figure 4.2.2: Ovary Disease Image Count

**B. Process of Experiments**

The initial stage in the recommended architecture is to extract information from the ovary photos. In situations when deep neural networks require more data to train and function more effectively, data augmentation techniques are employed to address the problem. The anticipated technique for the experiment is summarized as follows:

**Image Acquisition:**

ovarian illness Image Database provided the information needed for the model evaluation. In this stage, targeted website images were used. To make sure that the images in the dataset had white backgrounds, a thorough examination was done. Color images are placed against white backgrounds.

**Image Augmentation:**

This is the application of photo enhancement. The act of adding new data to an existing dataset while maintaining the label data is known as picture augmentation. Growing the data collection makes the model more generalizable by enhancing its capacity to handle unknown inputs. Our training data objectives were met by employing data augmentation. However, color and position were improved by using brightness, contrast, saturation, scaling, cropping, flipping, and rotation. Random spins between -20 and 20 degrees, unintentional rotations of 90 degrees, distortion, bending, reversals of the vertical and horizontal planes, skating, and conversion of luminous intensity were some of the data augmentation techniques used. Ten improvements were made to every original picture. Some modifications improve a varied picture.

**Training:**

A CNN learner model is created at this point. The model's classification accuracy was assessed after it was trained on the dataset using EfficientNetB4, Xception, VGG16, and Vgg19. Using Early Stopping call-backs, all representations were trained during 25 epochs (80 iterations). Patience is the number of epochs without progress till training is finished. The rate of learning is 0.00001.

**Classification:**

Xception, VGG16, Vgg19, and EfficientNetB4 neural networks were used in this step to automatically identify cell disorders. The neural network was chosen for categorization due to its proven effectiveness in a variety of real-world applications.

**Results:**

Two parts based on the transfer learning, and proposed methodologies show the experiment outcomes.

### 4.3 Proposed Model

The proposed model is a convolutional neural network (CNN) architecture designed for image classification tasks. It consists of three convolutional blocks, each followed by max-pooling layers to downsample the feature maps. The first block applies a 3x3 convolution with 32 filters, while the second block employs two parallel branches, each with 1x1 and 3x3 convolutions, respectively, and the third block utilizes a similar structure with 1x1 and 5x5 convolutions. The outputs of the second block branches are concatenated before passing to the third block. This architecture allows the model to capture both local and global features effectively. The final layers include a flattening operation, followed by a dense layer with 256 units and ReLU activation, and a dropout layer with a dropout rate of 20% to prevent overfitting. The output layer consists of a softmax activation function, suitable for multi-class classification tasks, with the number of neurons determined by the `class_count` variable. The model is compiled using the Adam optimizer and categorical cross-entropy loss function, with accuracy as the evaluation metric.

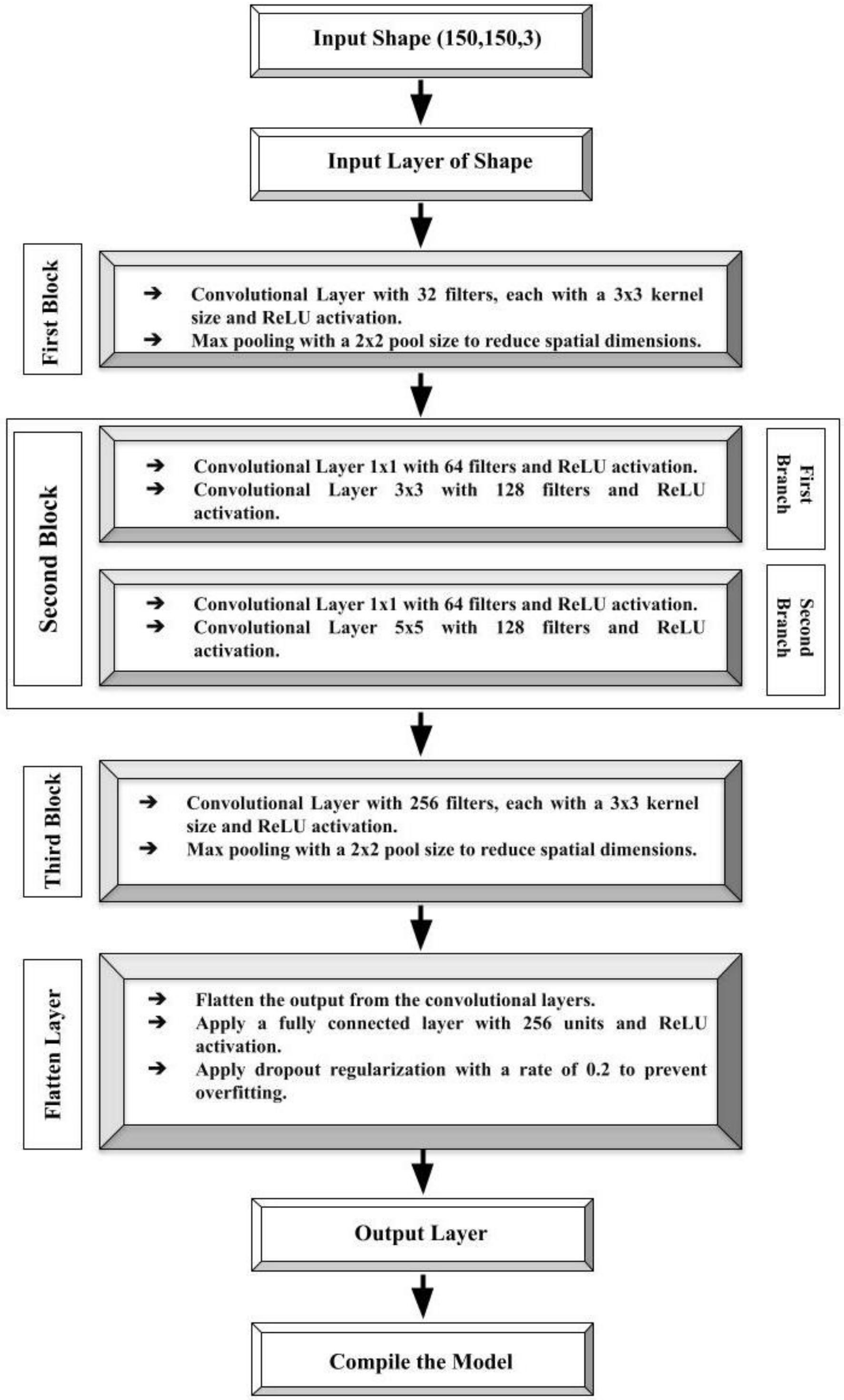


Figure 4.3.1: Flow work of proposed model

### 4.4 Model Evaluation

A variety of criteria are used to assess CNN basis approaches for machine learning classification model performance. The evaluation criteria include the following: accuracy, precision, recall, F1-score, and confusion matrix (CM).

**Accuracy:**

The accuracy of the classification model is one metric. The projected percentage of the model indicates its accuracy. The proportion of correctly categorized images to total samples is known as accuracy. Equation of accuracy.

Accuracy=(TP+TN)/(TP+TN+FP+FN)

**Precision:**

The likelihood of a positive label and the quantity of positive labels make up accuracy. The accuracy of positively anticipated situations is shown by the precision. When FP is more essential than FN, precision is helpful. Mathematically speaking, this equation:

Precision=TP/(TP+FP)

**Recall:**

Recall or sensitivity quantify the quantity of positively predicted occurrences that are accurately classified. When recall is higher than FP, it is a valuable metric. The F1-score is an additional classification accuracy measure that takes recall and precision into

consideration. Since accuracy and recall are harmonic means, the F1 score can offer a thorough comprehension of these two metrics. When Precision and Recall are equal, it performs optimally. To compute it, apply the below formula:

Recall=TP/(TP+FN)

**F-1 Score:**

A recall-precision measure of categorizing accuracy is the F1 score. The F1-score provides a comprehensive view as it is the harmonic mean of Precision and Recall. When accuracy and recall are equal, they are at their best. Method:

F-1 Score=2*(Precision*Recall)/(Precision+Recall)

**Specificity**

Specificity is the proportion of patients with a negative test result who do not have the illness. It is helpful to exclude the majority of people who are not ill using a very precise test. For this reason, a positive outcome on a very specialized test may rule out a particular person's condition. The equation is as follows:

Specificity =TN∕((TN+FP))

The model's capacity to generalize is limited by how closely it can resemble the training set of data. A specific table format known as the confusion matrix (CM) can be used to evaluate the performance of an algorithm. Critical prediction requirements including accuracy, precision, specificity, and memory are highlighted by CM. Confusion matrices are especially helpful since they provide quick evaluations of variables such as TP, FP, TN, and FN. Based on the research questions, the subsequent sections respond to the research questions of this investigation.

# CHAPTER 5

# RESULT AND ANALYSIS

## 5.1 Experimental Setup

To conduct comprehensive investigations, the experimental setup for the study "Leveraging Deep Learning for Ovarian Cancer Classification Using Image Data" uses a carefully thought-out configuration of hardware, software, and data resources. High-performance hardware, such as a powerful graphics processing unit (GPU) and a robust central processor unit (CPU), comprise the computing core. Deep learning frameworks like as TensorFlow or PyTorch are used to create transfer learning models in a Python programming environment. After being assembled from a range of medical photos, the experimental dataset is meticulously preprocessed, standardized, and normalized to offer a trustworthy and representative input for training and evaluating the model. Using deep Convolutional Neural Network (CNN) architectures that have previously undergone training, along with some fine-tuning, is an essential part of the model setup in order to identify ovarian illness. Informed permission and hospital data protection are two ethical challenges that are integrated into the setup to uphold norms of ethics. Throughout the experiment, a great deal of documentation is maintained, including code details, model architecture, hyperparameters, and conclusions, with the goal of encouraging future collaboration and transparency in the field of advanced ovarian illness detection.

## 5.2 Experimental Result

This section compares and contrasts the four pre-trained CNN networks: EffectiveeNetB4, VGG19, VGG16, Xception, and the recommended model. Classification performance comes first. We then look at those models' overall metrics. collecting, justifications, probable causes, and opportunities for better outcomes.

Table 5.2.1: Comparison of Accuracy for ovary Disease Detection

| **Architecture** | **Training Accuracy** | **Model Accuracy** | Precision | Recall | F1-Score |
|---|---|---|---|---|---|
| EfficientNetB4 | 99.12% | 92.4% | 0.92 | 0.92 | 0.92 |
| VGG19 | 99.9% | 92.20% | 0.94 | 0.92 | 0.92 |
| VGG16 | 99.42% | 85.6% | 0.86 | 0.86 | 0.86 |
| Xception | 99.85% | 92.4% | 0.92 | 0.92 | 0.92 |
| Proposed Model | 99.73% | 96% | 0.96 | 0.96 | 0.96 |

A comparison of many pre-trained deep Convolutional Neural Network architectures, including the suggested model, EfficientNetB4, VGG19, VGG16, and Xception, is shown in Table 2. The examination clarifies each CNN architecture's accuracy in identifying ovarian cancer. Remarkably, the findings show that the recommended model performed the best, with an accuracy of over 98%. With a maximum accuracy of 92.4%, EfficientNetB4 and Xception were the models that fared the best. This illustrates their amazing capacity to recognize ovarian illness patterns in clinical images. This little fluctuation in accuracy emphasizes how important it is to carefully consider the CNN architecture when it comes to ovarian illness detection, as it directly affects the diagnostic's reliability and accuracy.

**Training and Validation Accuracy and loss of CNN Networks:**

shows the accuracy of the original model's training and validation, with the accuracy and loss percentages on the y-axis and the number of epochs on the x-axis. The image displays training and validation data well separated, without any overfitting.

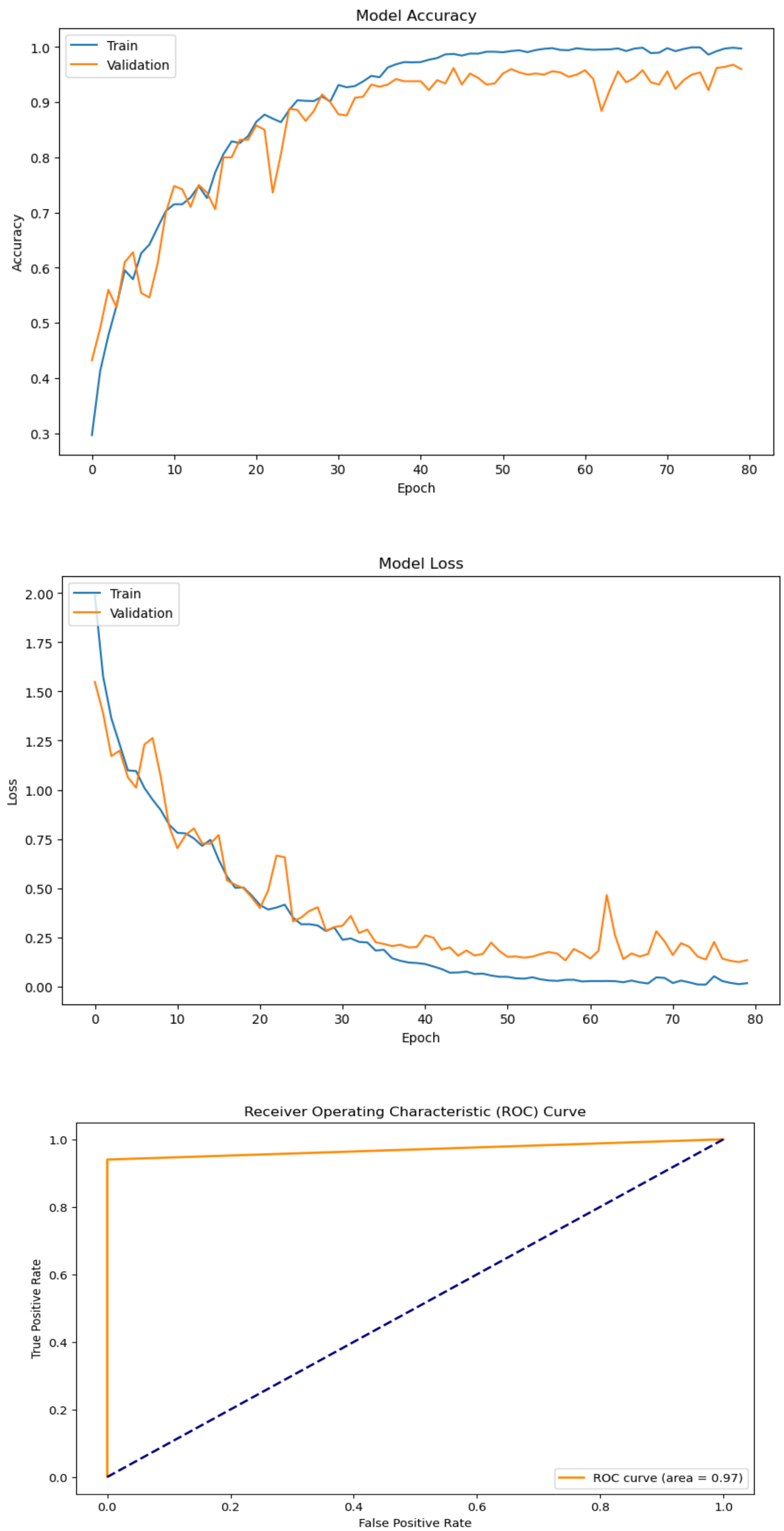


Figure 5.2.1: Proposed model evaluation

**Confusion Matrix of Proposed Model:**

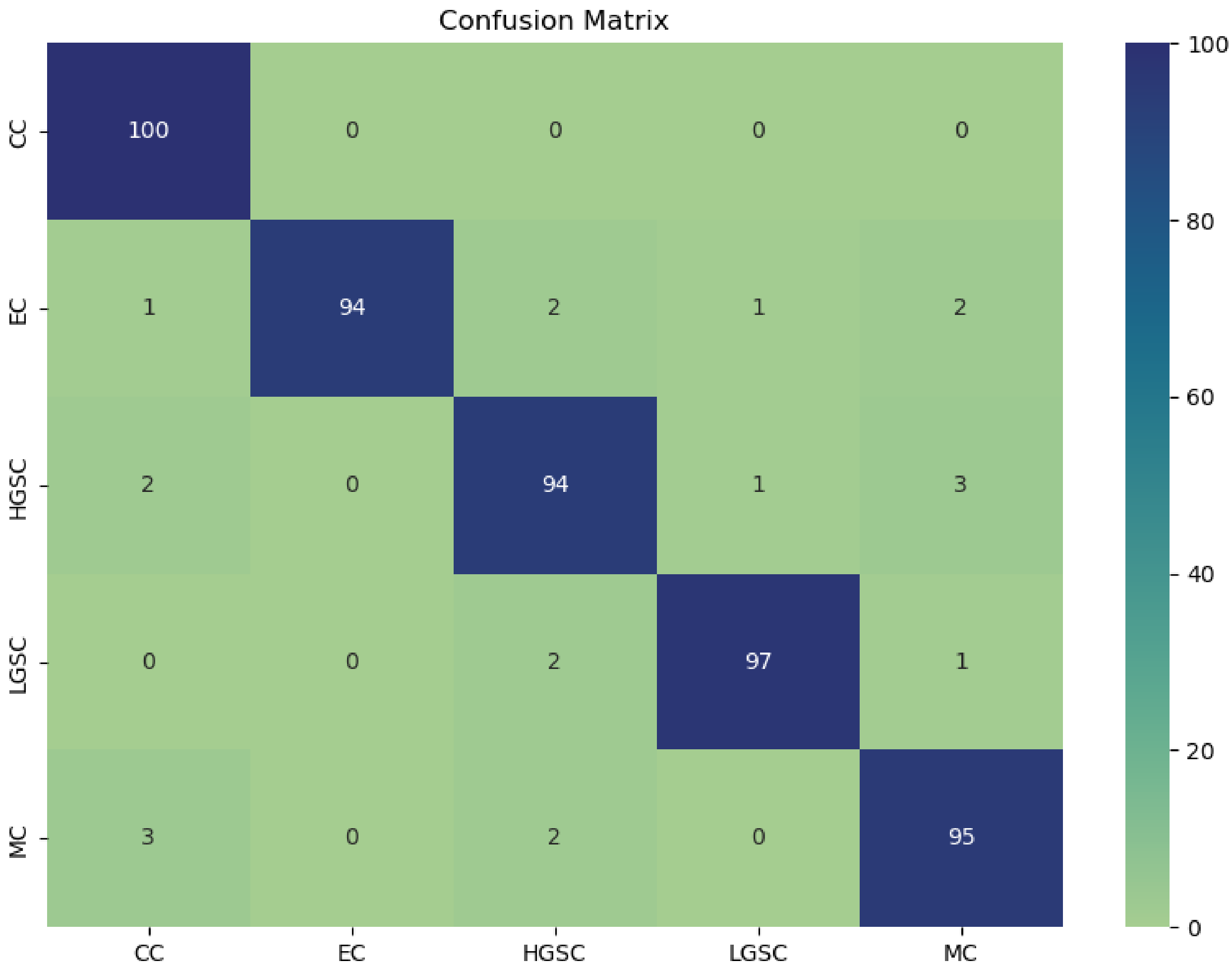


Figure 5.2.2: Confusion Matrices for Proposed Model

In ovarian cancer classification, the confusion matrix is an indispensable tool for evaluating our model's ability to accurately distinguish between different subtypes. It directly compares the model's predicted classifications (e.g., high-grade serous carcinoma) with the actual subtypes present in the test data. By analyzing the distribution of true positives (correctly identified cancers), true negatives (correctly classified healthy cases), false positives (mistakenly classified healthy cases as cancer subtypes), and false negatives (missed cancers), we gain valuable insights into the model's performance for each subtype. This analysis helps us pinpoint areas for improvement, guiding further refinement of the model to achieve more accurate and reliable ovarian cancer classification, ultimately leading to better patient outcome.

## 5.3 Summary

How well deep convolutional neural networks (CNNs) recognize and categorize photos of ovarian disease into two different categories —afflicted and normal—is examined in this comprehensive research. The research employs a large collection of 5000 original images and a creative augmentation approach to get eleven pictures from a single shot. In the area of ovarian illness detection, the study examines the value of original CNNs and transfer learning techniques models. The study specifically looks at segmentation and detection calculations, recognizing them as distinct processes in the diagnostic pipeline. An extensive testing procedure is used to evaluate one proposed model and four well-known CNN-based models—EfficientNetB4, VGG19, VGG16, and Xception—on the five types of ovarian disorders. Surprisingly, the suggested model comes out on top, with an astounding 96% accuracy, as seen in Figure 3.2.3. The study demonstrates that accuracy drops when the input picture is different from the training set of the ImageNet Dataset, which helps to clarify the complicated consequences of transfer learning. It's interesting to note that using various augmentation strategies independently with test sets and varying background noise levels are discovered to be potential reasons of worse performance.

# CHAPTER 6

# IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

## 6.1 Impact on Life

Our research on "Leveraging Deep Learning for Ovarian Cancer Classification Using Image Data" has profound implications for individuals affected by ovarian cancer. The deployment of advanced AI techniques in diagnostics has the potential to revolutionize patient outcomes by providing quicker and more accurate diagnoses. This can lead to earlier interventions, more personalized treatment plans, and ultimately, improved survival rates. Moreover, the psychological benefits of accurate and swift diagnoses cannot be overstated, as they alleviate the anxiety associated with uncertain or delayed results.

## 6.2 Impact on Society & Environment

The impact of using AI-driven diagnostic tools in healthcare is significant. Better diagnosis leads to improved health outcomes, which reduces the strain on healthcare systems. It allows for more efficient use of resources and can lower overall healthcare costs by reducing the need for repetitive and unnecessary procedures. Making advanced diagnostic tools more accessible can help underserved and remote populations receive better medical care, addressing healthcare disparities and promoting equity. On the environmental side, the computational power needed for training and deploying deep learning models is substantial. High-performance computing resources, such as GPUs, consume large amounts of energy, contributing to the carbon footprint. As AI continues to advance, it's crucial to develop and use more energy-efficient algorithms and hardware. Additionally, it's important to balance technological innovation with environmental sustainability by promoting the use of renewable energy sources and energy-saving measures in data centers.

## 6.3 Ethical Aspects

The research project "Leveraging Deep Learning for Ovarian Cancer Classification Using Image Data" is based on an approach that consistently gives ethical considerations top priority. Adhering to legal and ethical criteria is crucial when utilizing medical pictures, since patient privacy and confidentiality must be safeguarded. Informed permission is obtained when needed, which is essential to ethical research practices. Given the research's potential social influence on health care, ethical deployment is necessary to guarantee that advances in diagnostics translate into better personalized treatment without violating human rights. An second manner in which the ethical dimension is articulated is through the facilitation of responsible knowledge sharing, inspection, and reproducibility—all made possible by transparent documentation of the research process. In general, ethical issues are essential to the research's dedication to the welfare of society and the appropriate use of state-of-the-art medical technology.

## 6.4 Sustainability Plan

The research project "Leveraging Deep Learning for Ovarian Cancer Classification Using Image Data" is based on an approach that consistently gives ethical considerations top priority. Adhering to legal and ethical criteria is crucial when utilizing medical pictures, since patient privacy and confidentiality must be safeguarded. Informed permission is obtained when needed, which is essential to ethical research practices. Given the research's potential social influence on health care, ethical deployment is necessary to guarantee that advances in diagnostics translate into better personalized treatment without violating human rights. An second manner in which the ethical dimension is articulated is through the facilitation of responsible knowledge sharing, inspection, and reproducibility—all made possible by transparent documentation of the research process. In general, ethical issues are essential to the research's dedication to the welfare of society and the appropriate use of state-of-the-art medical technology.

## 6.5 Summary

Our research aims to make a significant impact on the diagnosis and treatment of ovarian cancer through the use of advanced deep learning models. While the benefits to individual lives and society are substantial, we are equally mindful of the environmental and ethical implications. By adopting a holistic approach that considers the impact on life, society, the environment, and sustainability, we strive to contribute positively to the field of medical diagnostics and beyond.

# CHAPTER 7

# CONCLUSION AND FUTURE WORK

## 7.1 Conclusions

It is impossible to diagnose ovarian illness without accurately identifying and classifying it in its early stages. This article primarily aims to provide a comprehensive evaluation of Deep Convolutional Neural Networks' (D-CNN) segmentation and ovarian illness detection ability. Our study demonstrates the effectiveness of a CNN-based segmentation technique, particularly on smaller and less complex datasets, highlighting its applicability. Remarkably, within the body of data now available, there are very few studies that are expressly devoted to the diagnosis of ovarian illness. Because of this, our comparison research is extremely valuable and has the potential to significantly improve the treatment of ovarian illness. We investigate how well a number of CNN models—including transfer learning techniques—classify ovarian diseases. The results demonstrate that the five networks—EfficientNetB4, VGG19, VGG16, Xception, and the proposed model—perform more accurately when combined than when used alone. It is crucial to emphasize, however, that despite the framework's overall effectiveness, there were a few instances where the estimates were off. Future studies that look at different strategies like contrast-enhancing or other image-processing approaches are therefore desperately needed.

Moreover, we suggest using picture segmentation before classification to improve CNN models' capacity to extract relevant features. The suggested ensemble is computationally more expensive than known CNN baselines due to the training of five CNN models, but the possible increase in accuracy offsets this expense. Future research might investigate snapshot ensembles as a means of easing the computational load. By providing informative information that might have a considerable influence on clinical practice and inspire future study in the quickly developing field of medical image processing, this work profoundly enhances the diagnostic techniques for ovarian illnesses.

## 7.2 Further Suggested Works

The results of this study open up new possibilities for future research and offer significant new information on the identification and segmentation of ovarian disease using Deep

Convolutional Neural Networks (D-CNN). Examining other image-processing techniques, such as contrast enhancement, is a crucial recommendation for more research in order to solve the erroneous predictions that the ensemble framework occasionally produces. Additionally, segmentation strategies before to classification are proposed as a possible improvement for CNN models in feature extraction. In addition, given the computational complexity of the suggested methodology, future studies ought to look at the use of methods such snapshot ensemble to lighten the strain on computer resources.

While the primary focus of this study was the diagnosis of ovarian illness, further focused research is necessary given the dearth of prior studies in this area. Future studies on ovarian illness detection methods should incorporate novel designs, a range of datasets, and useful clinical applications in order to deepen our understanding of this important area of medical diagnostics. In conclusion, expanding the study scope, investigating efficient ensemble methodologies, and improving image-processing procedures are the research implications for ovarian illness diagnosis. By taking these steps, the area of medical image analysis will advance and diagnosis accuracy and efficiency will rise.

### 7.3 Limitations

During the presentation of advancements in using deep learning to classify ovarian cancer, the research identified several limitations. These limitations include a lack of diversity in the training dataset, which was limited to 1,000 images per class. This limited dataset could lead to biased models affecting diagnostic accuracy across different populations. Additionally, computational limitations restricted the ability to experiment with more sophisticated models. Despite rigorous preprocessing, variations in image quality may still impact model performance. The generalizability of these models to real-world clinical data remains unverified, raising concerns about potential overfitting. Ethical and legal considerations regarding patient data privacy are critical and must be addressed to ensure compliance and protect patient rights. Furthermore, the practical implementation in clinical settings and the interpretability of CNNs pose significant challenges. To address these limitations, future research should focus on incorporating more diverse datasets, enhancing computational resources, and ensuring ethical standards to develop more robust and clinically applicable models

# REFERENCES


[1] J.-P. Yang et al., "Glycyrrhizin ameliorates impaired glucose metabolism and ovarian dysfunction in a polycystic ovary syndrome mouse model," Biol. Reprod., vol. 109, no. 1, pp. 83–96, 2023.

[2] A. Takahashi, L. Hong, and I. Chefetz, "How to win the ovarian cancer stem cell battle: destroying the roots," Canc. Drug Resist., vol. 3, no. 4, pp. 1021–1033, 2020.

[3]

[4] H. Sung et al., "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," CA Cancer J. Clin., vol. 71, no. 3, pp. 209–249, 2021.

[5] Y.-N. Su et al., "Effects of Yulin Tong Bu formula on modulating gut microbiota and fecal metabolite interactions in mice with polycystic ovary syndrome," Front. Endocrinol. (Lausanne), vol. 14, p. 1122709, 2023.

[6] E. Hinchcliff, S. N. Westin, and T. J. Herzog, "State of the science: Contemporary front-line treatment of advanced ovarian cancer," Gynecol. Oncol., vol. 166, no. 1, pp. 18–24, 2022.

[7] G. C. Jayson, E. C. Kohn, H. C. Kitchener, and J. A. Ledermann, "Ovarian cancer," Lancet, vol. 384, no. 9951, pp. 1376–1388, 2014.

[8] I. A. Cree, V. A. White, B. I. Indave, and D. Lokuhetty, "Revising the WHO classification: female genital tract tumours: Revising the WHO classification," Histopathology, vol. 76, no. 1, pp. 151–156, 2020.

[9] J. Prat and FIGO Committee on Gynecologic Oncology, "Staging classification for cancer of the ovary, Fallopian tube, and peritoneum: Abridged republication of guidelines from the international federation of gynecology and obstetrics (FIGO)," Obstet. Gynecol., vol. 126, no. 1, pp. 171–174, 2015.

[10] P. Bedi, S. B. Goyal, A. S. Rajawat, and M. Kumar, "An integrated adaptive bilateral filter-based framework and attention residual U-net for detecting polycystic ovary syndrome," Decision Analytics Journal, vol. 10, no. 100366, p. 100366, 2024.

[11] S. Chen, Y. Chen, L. Yu, and X. Hu, "Overexpression of SOCS4 inhibits proliferation and migration of cervical cancer cells by regulating JAK1/STAT3 signaling pathway," Eur. J. Gynaecol. Oncol., vol. 42, no. 3, p. 554, 2021.

[12] H. Farahani et al., "Deep learning-based histotype diagnosis of ovarian carcinoma whole-slide pathology images," Mod. Pathol., vol. 35, no. 12, pp. 1983–1990, 2022.

[13] S. Nobel et al., "RETRACTED: Modern subtype classification and Outlier Detection using the Attention Embedder to transform ovarian cancer diagnosis," Tomography, vol. 10, no. 1, pp. 105–132, 2024.

[14] M. A. Vázquez et al., “A quantitative performance study of two automatic methods for the diagnosis of ovarian cancer,” Biomed. Signal Process. Control, vol. 46, pp. 86–93, 2018.

[15] M. Akazawa and K. Hashimoto, “Artificial intelligence in ovarian cancer diagnosis,” Anticancer Res., vol. 40, no. 8, pp. 4795–4800, 2020.

[16] B. Ziyambe et al., “A deep learning framework for the prediction and diagnosis of ovarian cancer in pre- and post-menopausal women,” Diagnostics (Basel), vol. 13, no. 10, p. 1703, 2023.

[17] D. J. Ho et al., “Deep Interactive Learning-based ovarian cancer segmentation of H&E-stained whole slide images to study morphological patterns of BRCA mutation,” J. Pathol. Inform., vol. 14, no. 100160, p. 100160, 2023.

[18] Y. Liu, B. C. Lawson, X. Huang, B. M. Broom, and J. N. Weinstein, “Prediction of ovarian cancer response to therapy based on deep learning analysis of histopathology images,” Cancers (Basel), vol. 15, no. 16, 2023.

[19] K. R. Kasture, “A new deep learning method for automatic ovarian cancer prediction & subtype classification,” Turkish Journal of Computer and Mathematics Education (TURCOMAT), vol. 12, no. 12, pp. 1233–1242, 2021.

[20] M. A. Jopek et al., “Deep learning-based, multiclass approach to cancer classification on liquid biopsy data,” IEEE J. Transl. Eng. Health Med., vol. 12, pp. 306–313, 2024.

[21] C. C. Gajjela et al., “Leveraging mid-infrared spectroscopic imaging and deep learning for tissue subtype classification in ovarian cancer,” Analyst, vol. 148, no. 12, pp. 2699–2708, 2023.

[22] N. Abuzinadah et al., “Improved prediction of ovarian cancer using ensemble classifier and SHAPly explainable AI,” Cancers (Basel), vol. 15, no. 24, 2023.

[23] L. K. Hema et al., “Region-based segmentation and classification for ovarian cancer detection using convolution neural network,” Contrast Media Mol. Imaging, vol. 2022, p. 5968939, 2022.

[24] L.-Y. Guo, A.-H. Wu, Y.-X. Wang, L.-P. Zhang, H. Chai, and X.-F. Liang, “Deep learning-based ovarian cancer subtypes identification using multi-omics data,” BioData Min., vol. 13, no. 1, p. 10, 2020.

[25] G. Wadhwa, N. Jayanthi, and M. Mathur, “A deep convolutional neural network approach for detecting malignancy of ovarian cancer using densenet model,” Annals of the Romanian Society for Cell Biology, pp. 4449–4458, 2021.

[26] T.-H. Nguyen, T.-N. Nguyen, and B.-V. Ngo, “A VGG-19 model with transfer learning and image segmentation for classification of tomato leaf disease,” AgriEngineering, vol. 4, no. 4, pp. 871–887, 2022.

[27] B. R. Kavitha, N. Shaffi, M. Mahmud, F. Hajamohideen, and P. Narayanan, “A new model evaluation framework for Tamil handwritten character recognition,” in Proceedings of Trends in Electronics and Health Informatics, Singapore: Springer Nature Singapore, 2023, pp. 93–106.

[28] X. Tian, H. Daigle, and H. Jiang, “Feature detection for digital images using machine learning

algorithms and image processing,” in Proceedings of the 6th Unconventional Resources Technology Conference, 2018.

[29] A. Mehmood, “Efficient anomaly detection in crowd videos using pre-trained 2D convolutional neural networks,” IEEE Access, vol. 9, pp. 138283–138295, 2021.

[30] R. Xu, H. Lin, K. Lu, L. Cao, and Y. Liu, “A forest fire detection system based on ensemble learning,” Forests, vol. 12, no. 2, p. 217, 2021.

# Appendix A

## Course Outcomes, Complex Engineering Problems (EP) and Complex Engineering Activities (EA) Addressing

**Title: Leveraging Deep Learning for Ovarian Cancer Classification Using Image Data**

**Student ID:**


Gazi Adnan

ID: 202-15-3796

Tanvir Mahmud

ID: 202-15-3811


## CO Description for FYDP

| CO | CO Descriptions | PO |
|---|---|---|
| **Phase -I** | | |
| **CO1** | Integrate recently gained and previously acquired knowledge to identify a ovary disease detection problem for the Final Year Design Project (FYDP) | **PO1** |
| **CO2** | Analyze different aspects of the goals in designing a solution for this FYDP | **PO2** |
| **CO3** | Explore diverse problem domains through a literature review, delineate the issues, and establish these goals for the FYDP | **PO4** |
| **CO4** | Perform economic evaluation and cost estimation and employ suitable project management procedures throughout the development life cycle of the FYDP | **PO11** |
| **Phase -II** | | |
| **CO5** | Design and develop technical solutions and system components or processes that meet specified requirements, ensuring compliance with public health and safety standards, as well as considering cultural, socioeconomic, and environmental factors in this FYDP | **PO3** |
| **CO6** | Choose and apply appropriate methodologies, resources, and contemporary engineering and IT technologies to address complex engineering processes, encompassing prediction and modeling, while adhering to relevant constraints in this FYDP | **PO5** |
| **CO7** | Analyze societal, health, safety, legal, and cultural considerations, along with associated responsibilities, in the context of professional engineering practice and the resolution of this problem, employing logical reasoning guided by contextual understanding. | **PO6** |
| **CO8** | Comprehend and evaluate the enduring sustainability and impact of professional engineering endeavors in addressing intricate engineering challenges within social and environmental frameworks. | **PO7** |

| | | |
|---|---|---|
| **CO9** | Implement ethical principles and adhere to professional standards and norms in this FYDP | **PO8** |
| **CO10** | Capable of operating proficiently both individually and as a team member or leader across diverse teams and interdisciplinary settings in this FYDP. | **PO9** |
| **CO11** | Proficiently communicate with the engineering community and broader society regarding complex engineering endeavors, including the ability to comprehend and generate comprehensive reports and design documentation, as well as provide and receive clear instructions throughout this FYDP. | **PO10** |
| **CO12** | Acknowledge the importance of self-directed and life-long learning within the evolving landscape of technology, and possess the readiness and capability to engage in lifelong learning endeavors. | **PO12** |

# Addressing CO (1 to 8), Knowledge Profile (K), Attainment of Complex Engineering Problems (EP), and Attainment of Complex Engineering Activities (EA)

**Addressing CO (1 to 8), Knowledge Profile (K), Attainment of Complex Engineering Problems (EP):**

| SN | EP Definition | Attainment | CO | Justification<br>(with Knowledge Profile) | References |
|---|---|---|---|---|---|
| **1.** | **EP1: Depth of Knowledge required** | Yes | CO1, CO2, CO3, CO5, CO6, CO7 and CO8 | This project demonstrates **fundamental engineering (K3)** principles by employing deep neural networks, data augmentation techniques, and diverse CNN architectures for image processing and classification tasks. The project demonstrates **specialist knowledge (K4)** by conducting transfer learning with advanced CNN architectures like VGG19 and EfficientNetB4, enhancing ovary disease detection accuracy in medical images, crucial for computer-aided diagnosis. | **Page no:** [23-25] **Section:** [4.2] **Page no:** [1-2] **Section:** [1.1] |
| | | | | The project applies **engineering practice & design (K5)** by the figure of process of experiments. The project addresses **engineering** | **Page no:** [15] |

| | | | | | |
|---|---|---|---|---|---|
| | | | | **practice & technology (K6)** by employing CNN model. | **Section:** [3.3] **Page no:** [27] **Section:** [4.3] |
| | | | | This project ensures to **K8 (Research Literature)** by synthesizing insights from recent studies, to advance ovary disease detection using deep learning, showcasing a comprehensive understanding of current methodologies. | **Page no:** [8-10] **Section:** [2.2, 2.3] |
| **2.** | **EP2: Range of Conflicting Requirements** | Yes | CO2, and CO7 | This project addresses **EP-2** by recognizing the hurdles in ovary disease detection, including limitations of traditional methods and the complexities of integrating transfer learning CNNs. Through comparative analysis, it confronts challenges in understanding spatial distributions, offering insights for refining diagnostic methodologies. | **Page no:** [10-11] **Section:** [2.4, 2.5] |
| **3.** | **EP3: Depth of analysis required** | Yes | CO2, and CO6 | This project addresses **EP-3** by meticulously comparing experimental outcomes, highlighting Transfer Learning as the chosen significant solution for enhancing ovary disease detection amidst multiple potential approaches. | **Page no:** [30-33] **Section:** [5.2] |

| | | | | | |
|---|---|---|---|---|---|
| **4.** | **EP4: Familiarity of Issues** | Yes | CO8 | This project's interdisciplinary approach extends beyond computer science and engineering, impacting medical diagnostics in respiratory diseases like ovary disease, contributing to advancements in health and hospital care practices which indicates **EP-4**. | **Page no:** [7-10] **Section:** [2.2, 2.4] |
| **5.** | **EP5: Extends of application codes** | No | CO5 | N/A | N/A |
| **6.** | **EP6: Extends of stakeholders involved and** | No | CO8 | N/A | N/A |
| **7.** | **EP7: Interdependence** | Yes | CO5 | This project's comprehensive approach addresses high-level problems by integrating various components across data collection, statistical analysis, and proposed methodology, ensuring a holistic solution to complex challenges in medical diagnostics which ensures **EP-7**. | **Page no:** [23-27] **Section:** [4.2,4.3] |

**Addressing CO11 with Complex Engineering Activities (EA) [Some or all of the following]:**

| SN | EA Definition | Attainment | CO | Justification | References |
|---|---|---|---|---|---|
| 1. | EA1: Range of resources | Yes | CO11 | Our project utilizes diverse resources such as high-performance computing infrastructure, GPUs, deep learning frameworks, annotated datasets, and ethical considerations to ensure systematic research and contribute to advancements in ovary disease detection through transfer learning CNNs. | **Page no:** [18-20] **Section:** [3.4] |
| 2. | EA2: Level of . | No | | N/A | N/A |
| 3. | EA3: Innovation | No | | N/A | N/A |
| 4. | EA4: Consequences for society and the environment | Yes | | This project contributes to society by improving health care through advanced ovary disease detection methods, while also promoting environmental sustainability by employing efficient computational resources and adhering to ethical guidelines for hospital data privacy. | **Page no:** [35-36] **Section:** [6.1, 6.2, 6.4] |

| | | | | | |
|---|---|---|---|---|---|
| **5.** | **EA-5: Familiarity** | Yes | | This project expands upon existing research by examining a novel approach in ovary disease detection through transfer learning and deep CNNs, demonstrated through preliminary terminologies and a comprehensive comparative analysis, offering new insights into the field. | **Page no:** [7-11] **Section:** [2.2, 2.4] |

**Addressing CO (4, 9, 10, and 12):**

| SN | COs | Attainment | Justification | References |
|---|---|---|---|---|
| **1** | CO4 | Yes | This project addresses **CO4** by integrating effective project management and financial oversight, ensuring meticulous planning, resource allocation, and budget estimation for optimal resource utilization throughout the research lifecycle. | **Page no:** [20-22] **Section:** [3.5] |
| **2** | CO9 | Yes | The project demonstrates adherence to ethical principles by prioritizing hospital privacy, obtaining informed consent, and transparently documenting the research process, ensuring responsible knowledge dissemination and societal well-being through the ethical application of advanced health care technologies which comply with **CO9**. | **Page no:** [36] **Section:** [6.3] |
| **3** | CO10 | No | N/A | N/A |
| **4** | CO12 | Yes | The project's dedication to continuous learning **(CO12)** and adaptation within the dynamic technological landscape is reflected in its comprehensive data collection, rigorous statistical analysis, meticulous methodology development, and thorough experimental results and analysis, showcasing a commitment to staying updated and refining techniques to address modern challenges. | **Page no:** [13-19, 23-29] **Section:** [3.2, 3.3, 3.4, 4.2,4.3,4.4] |

## Ovarian Cancer Classification

ORIGINALITY REPORT

| 16% | 13% | 6% | 7% |
|---|---|---|---|
| SIMILARITY INDEX | INTERNET SOURCES | PUBLICATIONS | STUDENT PAPERS |

PRIMARY SOURCES

| | | |
|---|---|---|
| 1 | dspace.daffodilvarsity.edu.bd:8080<br>Internet Source | 3% |
| 2 | Submitted to Daffodil International University<br>Student Paper | 2% |
| 3 | www.mdpi.com<br>Internet Source | 1% |
| 4 | Submitted to University of Southern Mississippi<br>Student Paper | 1% |
| 5 | A. Jenefa, V. Ebenezer, E. Bijolin Edwin, J. Jerlin Rajan, P Kingston Stanley, M.Roshni Thanka. "Automating MRI-Based Ovarian Cancer Diagnosis with a DCNN*", 2023 International Conference on Sustainable Communication Networks and Application (ICSCNA), 2023<br>Publication | 1% |
| 6 | dugi-doc.udg.edu<br>Internet Source | 1% |
| 7 | Submitted to University of Bolton<br>Student Paper | <1 |

45