

AI-BASED DIAGNOSIS AND CLASSIFICATION OF OVARIAN CANCER IN WOMEN

Submitted in partial fulfillment of the requirements for the award of
Bachelor of Technology degree in Biomedical Engineering

By

SRISTI S V (40240037)

SUDHISHNA T (40240039)



SCHOOL OF BIO AND CHEMICAL ENGINEERING

DEPARTMENT OF BIOMEDICAL ENGINEERING

SATHYABAMA

**INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)**

Accredited with Grade “A++” by NAAC

JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI-600119

April 2024



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)

Accredited "A++" Grade by NAAC | 12B Status by UGC | Approved by AICTE

www.sathyabama.ac.in

DEPARTMENT OF BIOMEDICAL ENGINEERING

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of **SRISTI S V (40240037)**, **SUDHISHNA T (40240039)** who have done the project work as a team and carried out the project entitled "**AI BASED DIAGNOSIS AND CLASSIFICATION OF OVARIAN CANCER IN WOMEN**" under our supervision from November 2023 to April 2024.

Dr. **BETHANNEY JANNEY J**, M.E., Ph.D.,

Internal Guide

Dr. **T. SUDHAKAR**, M.Sc., Ph.D.,
Head of the Department
Department of Biomedical Engineering
Sathyabama Institute of Science & Technology
(Deemed to be University)

Category I University by UGC
Amalner Nagar, Rajiv Gandhi Salai,
Tamil Nadu - 600 119

Submitted for Viva-voice Examination held on 05 April 2024

Internal Examiner

External Examiner

DECLARATION

We, **SRISTI S V (40240037)**, **SUDHISHNA T (40240039)** hereby declare that the Project Report entitled "**AI BASED DIAGNOSIS AND CLASSIFICATION OF OVARIAN CANCER IN WOMEN**" done by us under the guidance of **Dr. BETHANNEY JANNEY J M.E., Ph.D.**, Department of Biomedical Engineering is submitted in partial fulfillment of the requirements for the award of Bachelor of Technology degree in Biomedical engineering.

1. 
2. *Sudhishna*

DATE: 5/4/2024

PLACE: CHENNAI

SIGNATURE OF THE CANDIDATES

ACKNOWLEDGEMENT

We are pleased to acknowledge our sincere thanks to the **Board of Management** of **SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY** for their kind encouragement in doing this project and for completing it successfully. We are grateful to them.

We convey our thanks to Dr. **T. Sudhakar, M.Sc., Ph.D., Head of the Department, Department of Biomedical Engineering** for providing us necessary support and details at the right time during the progressive reviews.

We would like to express our sincere and deep sense of gratitude to my/our Project Guide **Dr. BETHANNEY JANNEY J, M.E., Ph.D., Department of Biomedical Engineering** for his/her valuable guidance, suggestions, and constant encouragement that paved the way for the successful completion of my phase-1 project work.

We wish to express our thanks to all **Teaching and Non-teaching** staff members of the **Department of Biomedical Engineering** who were helpful in many ways for the completion of the project.

We would like to express our thanks to our parents for their unwavering support to us during these years.

ABSTRACT

Ovarian cancer is the fifth most common cause of cancer-related deaths in women and is often referred to as a silent killer because it tends to be detected at an advanced stage when the treatment options are limited. ovarian cancer can develop in different parts of the ovaries, including the epithelial cells on the surface, the germ cells within the ovaries, or the stromal cells that produce hormones. In this review we have discussed the methods for early detection of ovarian cancer using artificial intelligence and the methods This project compares both machine learning and deep learning algorithms and presents the most suitable algorithm to detect cancer at the earliest. The significance of the study lies in overcoming the challenges associated with human expert examination and misclassification rates. A more accurate, efficient, and reliable approach for early diagnosis and classification is presented in this project. This approach also emphasizes the usage of different machine learning algorithms such as support vector machine, and random forest as well as the deep learning algorithms such as convolutional neural networks, and the derived diagnostic results which have currently been obtained have been interpreted and discussed aiming for the further development that can be made in the domain of detection as well as classification.

Keywords: *Artificial intelligence, detection, algorithms*

TABLE OF CONTENT

CHAPTER NO	TITLE	PAGE NO.
	ABSTRACT	I
	LIST OF ABBREVIATIONS	V
	LIST OF FIGURES	VI
	LIST OF TABLES	VIII
1.	INTRODUCTION	1
	1.1 OVARIAN CANCER	2
	1.1.1 TYPES OF OVARIAN CANCER	3
	1.2 MALIGNANT EPITHELIAL OVARIAN CANCER	3
	1.2.1 SUBTYPES OF EPITHELIAL OVARIAN CANCER	4
	1.3 SYMPTOMS	8
	1.4 IMAGING METHODS OF OVARIAN CANCER	8
	1.5 DIAGNOSIS SYSTEM OF OVARIAN CANCER	10
2.	LITERATURE SURVEY	12
3.	AIM AND SCOPE	18
	3.1 AIM	18
	3.2 SCOPE	19
4.	METHODOLOGY AND EXPERIMENTAL METHODS	20
	4.1 METHODOLOGY	20
	4.1.1 DATA COLLECTION	20
	4.1.2 DATA PREPROCESSING	21
	4.1.3 DATA AUGMENTATION	21

4.1.4 FEATURE EXTRACTION	21
4.1.5 ALGORITHM SELECTION AND DESIGN	22
4.1.6 MODEL TRAINING	22
4.1.7 PERFORMANCE EVALUATION	22
4.1.8 RESULT	23
4.1.9 BLOCK DIAGRAM	23
4.2 EXPERIMENTAL METHODS	24
4.2.1 DEEP LEARNING	24
4.2.2 CNN AND ITS SIGNIFICANCE IN OVARIAN CANCER DETECTION	26
4.2.3 VGGNET MODEL	27
4.2.4 MACHINE LEARNING	29
4.2.5 SVM AND ITS SIGNIFICANCE IN OVARIAN CANCER DETECTION	31
4.2.6 RANDOM FOREST AND ITS SIGNIFICANCE IN OVARIAN CANCER DETECTION	32
5. RESULTS AND DISCUSSION	34
5.1 RESULT AND DISCUSSION	34
5.1.1 DATA COLLECTION	35
5.1.2 DATA AUGMENTATION	38
5.1.3 DATA PREPROCESSING	39
5.1.4 SPLITTING OF TRAINING AND TESTING TEST	39
5.2 PERFORMANCE ANALYSIS	40
5.2.1 RESULT FOR SVM ALGORITHM	40
5.2.2 RESULT FOR RANDOM FOREST ALGORITHM	41

5.2.3 RESULT FOR CNN ALGORITHM	43
5.3 COMPARISON GRAPH	44
5.3.1 PREDICATION	44
5.4 ACCURACY TABLE	45
6. SUMMARY AND CONCLUSION	46
6.1 SUMMARY	46
6.2 CONCLUSION	47
6.3 FUTURE WORK	47
REFERENCES	50
APPENDICES	56

LIST OF ABBREVIATION

ABBREVIATION	DEFINITION
CNN	Convolutional Neural Network
SVM	Support Vector Machine
DL	Deep Learning
ML	Machine Learning
AI	Artificial Intelligence
GLCM	Gray-Level Co-occurrence Matrix
GLRLM	Gray-Level Run Length Matrix
PPC	Primary Peritoneal Carcinoma
CC	Clear Cell Carcinoma
EC	Endometrioid Carcinoma
HGSC	High- Grade Serous Carcinoma
LGSC	Low-Grade Serous Carcinoma
MC	Mucinous Carcinoma
TVUS	Transvaginal Ultrasound
CT	Computed Tomography
MRI	Magnetic Resonance Imaging
PET	Positron Emission Tomography
CAD	Computer- Aided Diagnosis
EHR	Electronic Health Records
VGGNet	Visual Geometry Group Network
GDPR	General Data Protection Regulation
HIPAA	Health Insurance Portability and Accountability Act

LIST OF FIGURES

FIGURE NO	FIGURE NAME	PAGE NO
1.1	Female Reproductive System	1
1.2	Ovarian Cancer	2
1.3	Clear Cell Carcinoma	4
1.4	Endometrioid Carcinoma	5
1.5	High- Grade Serous Carcinoma	6
1.6	Low-Grade Serous Carcinoma	6
1.7	Mucinous Carcinoma	7
4.1	Image of ovarian cancer in dataset	20
4.2	Block Diagram	23
4.3	Deep Learning	25
4.4	Convolutional Neural Network	27
4.5	VGG Network	29
4.6	Machine Learning	30
5.1	Dataset of ovarian cancer	35
5.2	Images are identified by image ID	35
5.3	Total Count of images in each subtype	36
5.4	Images classified	37
5.5	Images classified and labelled as HGSC, MC, CC, LGSC, EC	37
5.6	Image Augmented	38
5.7	Images are Processed	39
5.8	Difference between original and processed image	39
5.9	Training set and Testing set	39

5.10	Grid Estimator for SVM	40
5.11	Classification report of SVM	40
5.12	Confusion Matrix for SVM	41
5.13	Grid Estimator for Random Forest	41
5.14	Classification Report of Random Forest	42
5.15	Confusion Matrix for Random Forest	42
5.16	Classification report of CNN	43
5.17	Confusion Matrix for CNN	43
5.18	Comparison graph of SVM, Random Forest, and CNN Algorithm	44
5.19	Predication	44

LIST OF TABLES

TABLE NO	TABLE OF CONTENT	PAGE.NO
5.4	Accuracy table	45

CHAPTER- 1

INTRODUCTION

The ovaries are essential organs of the female reproductive system, playing crucial roles in hormone production, egg (ovum) development, and menstrual cycle regulation (Al de Leon *et.al.*, 2018). Positioned on each side of the uterus within the pelvic cavity, these almond-sized glands are connected to the uterus by ligaments.

Each ovary consists of three main parts:

➤ *Outer Cortex*

The outer layer of the ovary, known as the cortex, contains thousands of follicles, each housing an immature egg cell (oocyte). These follicles are in various stages of development, with some growing and maturing while others remain dormant.

➤ *Inner Medulla*

The medulla lies beneath the cortex and comprises connective tissue, blood vessels, and lymphatic vessels (Akazama Munetoshi and Kazunori Hashimoto, 2023). It serves as the central support structure of the ovary, providing nourishment and oxygen to the follicles.

➤ *Ovarian Follicles*

Within the cortex, ovarian follicles undergo a process called folliculogenesis, where they mature and release eggs during ovulation. (Boom P M *et.al.*, 2020) Follicles consist of an oocyte surrounded by layers of supportive cells, such as granulosa and theca cells. The follicles produce estrogen, a hormone essential for female reproductive health. The fig 1.1 indicates Female reproductive system:

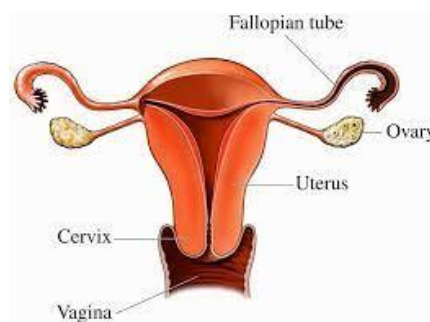


Fig :1.1 Female Reproductive system

The ovaries consist of three main parts: the outer cortex, containing follicles housing immature egg cells; the inner medulla, comprising supportive tissue and blood vessels; and the ovarian follicles, where eggs mature and estrogen is produced. (Bruce Wen *et.al.*, 2016) Ovaries are vital for oogenesis, hormone secretion (including estrogen and progesterone), menstrual cycle regulation, and reproductive function throughout a woman's life. (Guangxing Wang *et.al.*, 2020) Disorders like PCOS, ovarian cysts, and ovarian cancer can impact ovarian health and fertility. Regular gynecological exams are crucial for detecting abnormalities early.

1.1 OVARIAN CANCER

(Joel F. Lusk *et.al.*, 2019) Ovarian cancer is a type of cancer that originates in the ovaries, the female reproductive organs responsible for producing eggs and hormones. It is one of the most common gynecological cancers and a leading cause of cancer-related deaths among women worldwide. Ovarian cancer can arise from various cell types within the ovary, including epithelial cells (which line the surface of the ovary), germ cells (which develop into eggs), and stromal cells (which produce hormones) (Kawakami Eiryo *et.al.*, 2019). The fig 1.2 indicates image of ovarian cancer in ovaries:

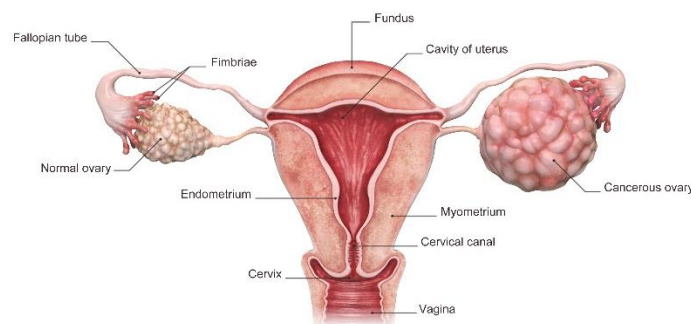


Fig: 1.2 Ovarian cancer

1.1.1 Types of Ovarian Cancer

➤ Epithelial Ovarian Cancer

This is the most common type, accounting for about 90% of ovarian cancer cases. (Kavitha S and Vidyaathulasiraman, 2021) It originates from the epithelial cells that cover the ovary's surface and often presents at advanced stages.

➤ Germ Cell Ovarian Cancer

(Shuo Wang *et.al.*, 2019) Germ cell tumors develop from the cells that produce eggs. These tumors are less common than epithelial tumors and typically affect younger women. They include subtypes such as dysgerminoma, yolk sac tumor, and immature teratoma.

➤ Stromal Ovarian Cancer

Stromal tumors arise from the connective tissue cells that support the ovary and produce hormones. They are rare and can produce estrogen or testosterone, leading to hormonal symptoms.

1.2 MALIGNANT EPITHELIAL OVARIAN CANCER

Malignant epithelial ovarian tumors, known as carcinomas, encompass various subtypes, with serous carcinomas being the most prevalent (52%), followed by clear cell carcinoma (6%), mucinous carcinoma (6%), and endometrioid carcinoma (10%) (Chiappa Valentina *et.al.*, 2021). These tumors are graded based on their resemblance to normal tissue, with Grade 1 tumors having a better prognosis than Grade 3 tumors. (Wang Yanli *et.al.*, 2024) Additionally, tumors are classified into Type I or Type II based on their growth rate and response to chemotherapy. Type I tumors, including low-grade serous carcinoma, clear cell carcinoma, mucinous carcinoma, and endometrioid carcinoma, tend to grow slowly and exhibit limited response to chemotherapy. In contrast, Type II tumors, exemplified by high-grade serous carcinoma, grow rapidly and respond better to chemotherapy (Md Mamunur Rahaman *et.al.*, 2021).

Primary peritoneal carcinoma (PPC) closely resembles epithelial ovarian cancer and typically originates from the cells lining the fallopian tubes. PPC shares similar symptoms and spreads along the pelvic and abdominal surfaces. (Lei Ruilin *et.al.*, 2023) Treatment involves debulking surgery followed by chemotherapy, akin to ovarian cancer management, with a prognosis comparable to widespread ovarian cancer.

(Mitchell Sain *et.al.*, 2024) Fallopian tube cancer, another rare cancer akin to epithelial ovarian cancer, arises in the fallopian tubes and presents similar symptoms. While treatment resembles that of ovarian cancer, the prognosis for fallopian tube cancer is slightly more favorable.

1.2.1 Subtypes of epithelial ovarian cancer

➤ Clear Cell Carcinoma (CC)

This subtype accounts for approximately 5-10% of ovarian cancers. Histologically, clear cell carcinomas are characterized by clear or hobnail-shaped cells. (Zhou Jie *et.al.*, 2020) They tend to be chemo resistant and are associated with a poorer prognosis compared to other subtypes. Fig 1.3 shows the sample image of clear cell carcinoma (CC):

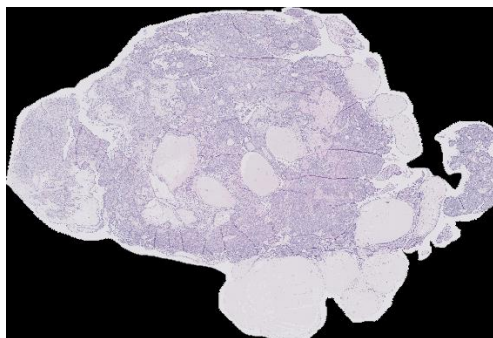


Fig: 1.3 Clear cell carcinoma

Clear cell carcinoma of the ovary typically affects younger patients and is often linked with endometriosis. (Sai Kiran Sharma *et.al.*, 2016) Clinical symptoms may include pelvic pain, abnormal uterine bleeding, or a pelvic mass. Genetic mutations in genes like ARID1A, PTEN, and PIK3CA are common in this subtype, driving

disease development. Clear cell carcinomas exhibit resistance to traditional chemotherapy, notably platinum-based agents, necessitating alternative treatment approaches.

➤ *Endometrioid Carcinoma (EC)*

Endometrioid carcinomas comprise around 10-20% of ovarian cancers and often arise from endometriosis. Histologically, they resemble endometrial tissue and may contain glands and stroma (Kyriazi Eiryo *et.al.*, 2010). Patients with endometrioid carcinomas generally have a better prognosis than those with clear cell or high-grade serous carcinomas. Fig 1.4 shows the sample image of Endometrioid carcinoma (EC):

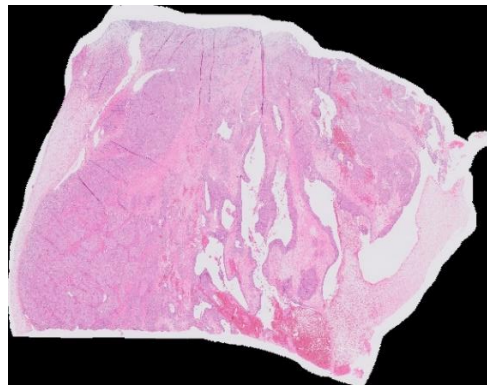


Fig: 1.4 Endometrioid carcinoma

(Hsu Shih-Tien *et.al.*, 2022) Endometrioid carcinoma commonly originates from endometriosis, highlighting the need for vigilant management of this condition. These tumors often express estrogen and progesterone receptors, suggesting potential responsiveness to hormonal therapies, which can impact treatment strategies and prognostic outcomes.

➤ *High-Grade Serous Carcinoma (HGSC)*

HGSC is the most common and aggressive subtype, accounting for approximately 70% of ovarian cancers. Histologically, these tumors consist of high-grade epithelial cells with papillary structures. (Mikdadi Dina *et.al.*, 2022) They are often diagnosed at advanced stages and have a high propensity for metastasis. Fig 1.5 shows the sample image of high-grade serous carcinoma (HGSC):

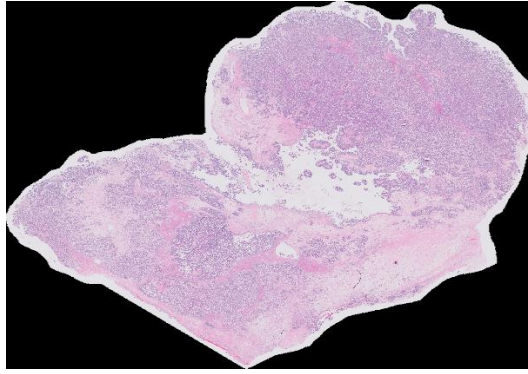


Fig: 1.5 High-grade serous carcinoma

It likely originates from fallopian tube epithelium, not ovarian surface epithelium, showing extensive genomic instability, including TP53 mutations and BRCA1/2 DNA repair pathway alterations (Pallavi Mehra *et.al.*, 2023). HGSC poses clinical challenges due to its aggressiveness and tendency for widespread metastasis. Limited screening methods result in late-stage diagnoses, impacting treatment options and survival rates (Lee Sanghee *et.al.*, 2024).

➤ *Low-Grade Serous Carcinoma (LGSC)*

LGSC is a less common subtype, representing about 5-10% of ovarian cancers. (Shuo Wang *et.al.*, 2019) Histologically, these tumors are characterized by low-grade epithelial cells with glandular architecture. They tend to grow more slowly and are associated with a better prognosis compared to HGSC. Fig 1.6 shows the sample image of Low-grade serous carcinoma (LGSC):

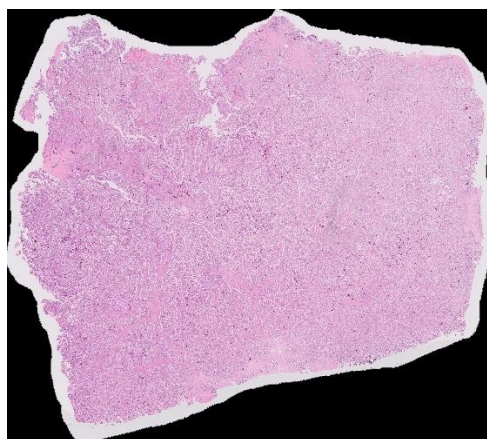


Fig: 1.6 Low-grade serous carcinoma

Low-grade serous carcinoma exhibits a unique clinical trajectory, typically presenting at earlier stages and progressing slowly, resulting in a more favorable prognosis. (Weitong Huang *et.al.*, 2023) Unlike its high-grade counterpart, it may respond to targeted therapies, especially those targeting mutations in the MAPK pathway. Molecular profiling enables the identification of potential therapeutic targets, offering personalized treatment options for patients with this subtype of ovarian cancer.

➤ *Mucinous Carcinoma (MC)*

(Zhu Haiyan *et.al.*, 2021) Mucinous carcinomas account for approximately 3-4% of ovarian cancers. Histologically, they consist of epithelial cells resembling mucin-secreting glands. These tumors are often diagnosed at earlier stages and have a better prognosis compared to HGSC. Fig 1.7 shows the sample image of Mucinous carcinoma (MC):

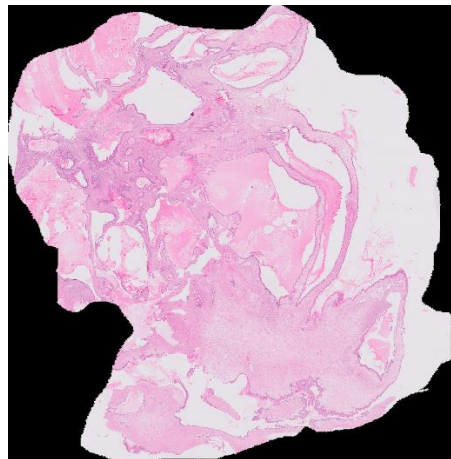


Fig: 1.7 Mucinous carcinoma

Mucinous carcinoma of the ovary presents histological variants like intestinal-type and endocervical-like subtypes, each with unique morphological characteristics impacting prognosis and treatment. (Wang Ching-Wei *et.al.*, 2023) Often originating from mucinous borderline tumors, which display varying degrees of cellular atypia, mucinous carcinomas highlight the significance of precise diagnosis and management of borderline ovarian lesions to prevent progression to malignancy (Ya-Ting Jan *et.al.*, 2023).

1.3 SYMPTOMS

(Ai Yao *et.al.*, 2021) Ovarian cancer is often called the "silent killer" because it can present with vague or nonspecific symptoms, leading to delayed diagnosis.

Common symptoms may include:

- Abdominal bloating or swelling
- Pelvic discomfort or pain
- Difficulty eating or feeling full quickly
- Changes in bowel habits, such as constipation or diarrhea
- Frequent urination
- Fatigue
- Unexplained weight loss

1.4 IMAGING METHODS OF OVARIAN CANCER

Diagnosing ovarian cancer is a multifaceted process that involves a comprehensive assessment using various imaging modalities, blood tests, and biopsy techniques. (Chun-Jie Liu *et.al.*, 2022) Among these, imaging studies play a pivotal role in providing detailed information about the size, location, and extent of the tumor, aiding in accurate diagnosis and staging.

➤ *Transvaginal Ultrasound (TVUS)*

TVUS is often the initial imaging modality used for evaluating ovarian masses. It utilizes high-frequency sound waves emitted from a transducer probe placed in the vagina to create detailed images of the pelvic organs, including the ovaries (Weitong Huang *et.al.*, 2023).

TVUS can help identify the presence of ovarian masses, characterize their size, shape, and internal structure, and assess for features suggestive of malignancy, such as solid components, irregular borders, or vascularity.

Additionally, TVUS can guide needle biopsy procedures for histological diagnosis and aid in monitoring disease progression during treatment (Kokila R Kasture *et.al.*, 2021).

➤ *Computed Tomography (CT) Scan*

CT scanning is valuable for staging ovarian cancer and assessing the extent of disease spread (Wang Ching-Wei *et.al.*, 2023).

It involves the use of X-rays to produce cross-sectional images of the abdomen and pelvis, providing detailed anatomical information about the ovaries, surrounding structures, and potential metastatic sites.

(Al de Leon *et.al.*, 2018) CT scans can help identify the presence of metastases in nearby lymph nodes, peritoneum, and distant organs, aiding in accurate staging of the disease and treatment planning.

Advanced CT techniques, such as contrast-enhanced CT, can enhance the visualization of tumor vascularity and improve diagnostic accuracy.

➤ *Magnetic Resonance Imaging (MRI)*

(Kawakami Eiroya *et.al.*, 2019) MRI offers superior soft tissue contrast compared to CT and is particularly useful for characterizing ovarian masses.

It uses powerful magnets and radio waves to generate detailed images of the pelvic organs, providing information about tumor size, morphology, and tissue composition (Bruce Wen *et.al.*, 2016).

MRI can help differentiate between benign and malignant ovarian lesions, assess tumor invasion into adjacent structures, and detect peritoneal implants or metastatic disease.

(Sumit Das *et.al.*, 2023) Functional MRI techniques, such as diffusion-weighted imaging (DWI) and dynamic contrast-enhanced MRI, can provide additional insights into tumor biology and treatment response.

➤ *Positron Emission Tomography (PET) Scan*

(Zhou Jie *et.al.*, 2020) PET scanning is used to evaluate the metabolic activity of tumors by measuring glucose uptake.

It involves the administration of a radioactive tracer, usually fluorodeoxyglucose (FDG), which is absorbed by metabolically active cells.

(Mitchell Sian *et.al.*, 2024) PET scans can help differentiate between benign and malignant ovarian masses, detect distant metastases, and assess treatment response.

Combined PET/CT imaging offers complementary information by integrating metabolic data from PET with anatomical details from CT, enhancing diagnostic accuracy and disease staging (Ai Yao *et.al.*, 2021).

1.5 DIAGNOSIS SYSTEM OF OVARIAN CANCER

(Boom P M *et.al.*, 2020) Developing a comprehensive diagnosis system for ovarian cancer involves integrating various medical and technological components to enable accurate and efficient detection and characterization of the disease.

Here's an outline of the components that could constitute such a system:

➤ *Clinical Assessment*

A thorough clinical evaluation by healthcare professionals, including a detailed medical history and physical examination, helps identify potential risk factors and symptoms associated with ovarian cancer (Guangxing Wang *et.al.*, 2020).

➤ *Imaging Studies*

Ultrasound detects ovarian abnormalities. CT scans assess tumor characteristics in the abdomen and pelvis. (Mikdadi Dina *et.al.*, 2022) MRI offers superior soft tissue contrast for characterizing masses and evaluating local invasion. PET scans evaluate metabolic activity, distinguishing between benign and malignant masses, and detecting distant metastases.

➤ *Blood Tests*

CA-125: Serum CA-125 levels serve as a tumor marker for ovarian cancer, though they may be elevated in other conditions. (Sai Kiran Sharma *et.al.*, 2016) Additional biomarkers like HE4 and ROMA may also be considered.

Genetic Testing: (Chiappa Valentina *et.al.*, 2021) Identification of genetic mutations associated with ovarian cancer, such as BRCA1 and BRCA2 mutations, can aid in risk assessment and personalized treatment planning.

Biopsy and Histopathological Analysis: Tissue biopsy, typically obtained via minimally invasive procedures like laparoscopy or fine needle aspiration, allows for

definitive diagnosis and histopathological assessment of tumor characteristics (Hsu Shih-Tien *et.al.*, 2022).

➤ *Multidisciplinary Tumor Board Review*

(Lei Ruilin *et.al.*, 2022) A multidisciplinary team of specialists, including gynecologic oncologists, radiologists, pathologists, and genetic counselors, collaborates to review diagnostic findings, confirm diagnosis, and formulate individualized treatment plans.

➤ *Computer-Aided Diagnosis (CAD)*

Integration of artificial intelligence (AI) and machine learning algorithms can assist in interpreting imaging studies and identifying patterns indicative of ovarian cancer, enhancing diagnostic accuracy and efficiency (Kokila.R.Kasture *et.al.*, 2021).

➤ *Electronic Health Records (EHR)*

Implementation of EHR systems facilitates seamless documentation, storage, and retrieval of patient data, enabling comprehensive tracking of diagnostic findings and treatment outcomes.

➤ *Patient Education and Support*

(Zhu Haiyan *et.al.*, 2021) Provision of educational resources and support services to patients and their families regarding ovarian cancer diagnosis, treatment options, and psychosocial support enhances overall patient care and engagement in the diagnostic process.

CHAPTER- 2

LITERATURE SURVEY

The emphasis by the authors Yong Gao et al., (2017) is the development of a low-cost, non-invasive approach using targeted NBs to directly image CA-125 positive ovarian tumors. Their work demonstrates the proof of concept for ultrasound molecular imaging beyond the vasculature in this model. Antibody-functionalized nanobubbles showed enhanced tumor uptake, retention, and extravasation compared to untargeted nanobubbles.

Ya-Ting Jan et al., (2023) have developed an Artificial intelligence (AI) model with radiomics and deep learning (DL) features extracted from CT images to distinguish benign from malignant ovarian tumors. It is a CT-based AI model that can differentiate benign and malignant ovarian tumors with high accuracy and specificity. This model significantly improved the performance of less-experienced radiologists in ovarian tumor assessment, and may potentially guide gynecologists to provide better therapeutic strategies for these patients.

Robin Wang et al., (2021) have developed a deep learning algorithm that distinguishes benign from malignant ovarian lesions by applying a convolutional neural network on routine MR imaging. Sai Kiran Sharma et al., (2016) have provided a brief yet informative survey of the current state of the molecular imaging of ovarian cancer. Their article is divided into sections according to modality, covering recent advances in the MR, PET, SPECT, ultrasound, and optical imaging of ovarian cancer. Although primary emphasis is given to clinical studies, preclinical investigations that are particularly innovative and promising are discussed as well.

Development of a specific interpretable machine learning (ML) prediction model for the diagnosis and prognosis of epithelial ovarian cancer (EOC) based on a variety of biomarkers has been proclaimed by the authors Meixuan Wu et al.,(2020)

Blessed Ziyambe et al., (2022) have proposed a novel convolutional neural network (CNN) algorithm for predicting and diagnosing ovarian cancer, addressing existing

limitations. In this paper, CNN was trained on a histopathological image dataset, divided into training and validation subsets and augmented before training. The model achieved a remarkable accuracy of 94%, with 95.12% of cancerous cases correctly identified and 93.02% of healthy cells was accurately classified. The significance of this study lies in overcoming the challenges associated with the human expert examination.

Yan Zhang et al., (2007) have investigated the intelligent recognition of radiomics based on the convolutional neural network (CNN) in predicting endometrial cancer (EC). In this study, 158 patients with EC in the hospital were selected as the research objects and divided into a training group and a test group. All the patients underwent magnetic resonance imaging (MRI) before surgery. Based on the CNN, the imaging model of EC prediction was constructed according to the characteristics.

Manuel A. Vázquez et al., (2023) present a quantitative study of the performance of two automatic methods for the early detection of ovarian cancer that can exploit longitudinal measurements of multiple biomarkers. They use statistical analysis techniques, such as the area under the Receiver Operating Characteristic (ROC) curve, for evaluating the performance of two techniques that aim at the classification of subjects as either healthy or suffering from the disease using a time series of multiple biomarkers as inputs.

Mijin Kim et al., (2018) investigated a perception-based strategy. Nature has evolved perception to identify and interpret multidimensional stimuli against target heterogeneity. they present a nanosensor array and a computational model that resulted in the perception-based detection of ovarian cancer from patient serum samples. To transduce broad types of physicochemical properties of a biofluid, we designed nanosensor arrays using OCC-functionalized, single-stranded DNA (ssDNA)-encapsulated SWCNTs.

Munetoshi Akazawa and Kazunori Hashimoto (2021) aimed to use artificial intelligence (AI) to predict the pathological diagnosis of ovarian tumors using patient information and data from preoperative examinations.

Donna Bagwell and Robert C. Bast (2010) have evaluated Transvaginal sonography (TVS), serum markers, and a combination of the two modalities for early detection of ovarian cancer. Candidate biomarkers have been discovered through empirical development of monoclonal antibodies, and studies of gene expression, Greater specificity can be achieved by combining CA125 and TVS and/or by monitoring CA125 over time.

Sneha Aditi et al., (2018) conducted a retrospective study where the data of the patients from the past three years, from February 2020 to February 2023, were retrieved and assessed. Gross and microscopic findings, including clinical details of patients with ovarian masses, were analyzed from the previous records. Histomorphology Analysis of Ovarian Neoplasms was done.

Patrick Nunes Pereira et al., (2022) provided a straightforward approach to the sequential use of ultrasound (US), magnetic resonance (MR), and serum biomarkers to differentiate the origin of pelvic masses, making the most efficient use of these diagnostic resources. Using a recursive partitioning algorithm to predict tumor origin, they devised a roadmap to determine the probability of non-ovarian origin using only statistically significant US, laboratory, and MR parameters.

Ashwini Kodipalli et al., (2022) authors have presented a study that proposes a two-stage computational methodology to identify and segment the ovarian tumor and classify it as benign or malignant. Using computerized tomography images, the first stage involves image segmentation using inverted fuzzy c-means clustering, and the second stage consists of a deep quantum convolutional neural network to detect the tumors.

Tarun Roy et al., (2023) proposed a 3D contrastive learning-based predictive model to discriminate benign from malignant masses in abdominal CT scans for ovarian cancer patients. They used a fully supervised contrastive learning (SCL) approach which allowed us to effectively leverage the label information of our small dataset of 331 patients. Three different architectures (VGG, ResNet, and DenseNet) were implemented for feature extraction by contrastive learning.

Chih-Hsuan Sao et al., (2023) have focused on developing a deep-learning model that can be used intraoperatively to differentiate malignant from benign ovarian tumors based on gross specimen images. They collected 662 cut-in-half gross images of specimens from 243 patients as a training dataset to develop a convolutional neural network (CNN)-based models for differentiating malignant and benign ovarian tumors. Different deep learning algorithms including classic CNN, VGG16, Exception, InceptionV3, MobileNet, NASNet, EfficientNetB0, and ResNet50, were used for training and validation.

Anna H. Koch et al., (2022) have evaluated the available data on CAD in predicting the chance of malignancy of ovarian tumors. The Methods they searched for all published studies investigating diagnostic accuracy of CAD based on ultrasound, CT, and MRI in pre-surgical patients with ovarian tumors are compared to reference standards.

Sumit Das et al., (2023) have aimed to analyze and predict ovarian cancer in women using Artificial Intelligence. The program in logic and the decision tree of machine learning are being created to presume Ovarian Cancer. The side effects of ovarian diseases are taken as the factors to settle on the choice tree to foresee the conceivable outcomes. The fundamental side effects are the foundations of the sickness to settle on the choice tree furthermore all the yes and no of the tree have a determination or an outcome.

Dimitrios A. Binas et al., (2022) have aimed to present a novel technique that can assess intratumoral cellularity based on quantitative features extracted from medical images, denoted as radiomics, advanced image processing and artificial intelligence algorithms, in an attempt to offer biomedical engineers and health professionals a tool for personalized medicine. According to their results, the average accuracy rating of the proposed method in the study population was over 85%.

Weitong Huang et al., (2023) have aimed to introduce best practices for combining ML and XAI techniques for biomarker validation tasks. They focused on classification tasks and a game theoretic approach based on Shapley values to build and evaluate models and visualize results. They described the workflow and applied

the pipeline in a case study using the CDAS PLCO Ovarian Biomarkers dataset to demonstrate the potential for accuracy and utility. Their case study results demonstrated the efficacy of the ML pipeline, its consistency, and its advantages compared to conventional statistical approaches.

Hong Zhang et al., (2022) have emphasized on a unique approach in Transcriptome and Differential Diagnosis between Five Major Diagnostic Classes Using Machine Learning and Targeted Transcriptome Diagnosis Cases. Machine learning was selected over convolutional neural networks or deep learning, which is effective in image-related applications because expression data were different. The study focused on RNA expression profiling and did not incorporate mutation profiling in the diagnostic algorithms. A targeted transcriptome was used to profile a wide range of hematologic neoplasms and solid tumors.

Alexandros Laios et al., (2023) have combined high-performance AI modeling with an explainable Artificial Intelligence (XAI) framework to explain feature effects and interactions associated with specific threshold surgical effort using data from a single public institution. They revealed features not routinely measured in clinical practice, including human factors that could be responsible for the variation in the surgical effort. The use of XAI frameworks can provide actionable information for surgeons to improve patient outcomes in gynecologic oncology.

Hui Chen et al., (2023) have delved into the background that Deep learning (DL) algorithms could improve the classification of ovarian tumors assessed with multimodal US, therefore they developed DL algorithms for the automated classification of benign versus malignant ovarian tumors assessed with US and to compare algorithm performance to Ovarian-Adnexal Reporting and Data System (O- RADS) and subjective expert assessment for malignancy.

Duhita Sengupta et al., (2022) have aimed to correlate the morphometric features of the nucleus along with the distribution of nuclear amine proteins with classical machine learning to differentiate between normal and ovarian cancer tissues. They have performed exhaustive imaging of ovarian cancer versus normal tissue and developed a dual pipeline architecture that combines the matrices of morphometric

parameters with deep learning techniques of auto-feature extraction from pre-processed images.

Sepideh Hatamikia et al., (2022) have proved that integrative multi-omics approaches outperform models based on single data types and indicate that imaging can be used for the longitudinal tracking of tumor heterogeneity in space and potentially over time. Their review presents an overview of studies that integrated two or more data types to develop AI-based classifiers or prediction models.

CHAPTER- 3

AIM AND SCOPE

3.1 AIM

The project proposes a novel deep learning algorithm-based approach for ovarian cancer detection, integrating segmentation, feature extraction, and quality assessment. By utilizing ovary histology images, a sophisticated computer-aided design is developed to categorize tumors as either benign or malignant, aiming to overcome challenges associated with human expert examination and misclassification rates. Additionally, it explores 3D texture analysis for classifying second-harmonic generation images of human ovarian cancer. Machine learning algorithms such as random forest and logistic regression, along with deep learning algorithms like convolutional neural networks, are reviewed and applied. The derived diagnostic results are thoroughly interpreted and discussed, with a focus on further development in detection and classification domains. The project also delves into the current state of research in ovarian cancer detection through deep learning and machine algorithms, exploring methods, data sources, and outcomes of pioneering studies. It addresses nuances and challenges in deploying AI-powered solutions, including issues of data quality, model interpretability, and ethical considerations, while also discussing promising avenues for future research and potential impact on clinical practice and patient outcomes. Through this comprehensive exploration, the project aims to illuminate the significant progress made in ovarian cancer detection and the transformative potential of these technologies in saving lives and enhancing patient care.

3.2 SCOPE

The objective of this project is to pioneer a robust and precise diagnostic system for the early detection of ovarian cancer, leveraging advanced deep learning and computer algorithms such as Convolutional Neural Networks (CNNs), Random Forest, and Support Vector Machines (SVMs). Our approach involves a comprehensive comparison of these algorithms to identify their efficacy, including their ability to detect negative cases and avoid false positives. By harnessing the power of machine learning and artificial intelligence, we aim to develop a reliable, non-invasive diagnostic tool that significantly improves early detection rates.

The focus lies in two key areas:

- The development of an accurate model using image data for ovarian cancer detection, and
- The evaluation and comparison of various machine learning algorithms to determine their suitability for this critical medical application. This project represents a significant advancement in the quest for precise and early ovarian cancer diagnosis, promising to revolutionize clinical practices and improve patient outcomes.

CHAPTER- 4

METHODOLOGY AND EXPERIMENTAL METHODS

4.1 METHODOLOGY

In the field of early detection of ovarian cancer using AI and ML, there are several novel approaches and strategies that researchers can explore to enhance the accuracy and effectiveness of detection methods. The potential approach to the problem has been explained in this project. Moreover, the main objective of this project is to develop a robust and accurate diagnostic system that can detect ovarian cancer at an early stage with high precision. This approach seeks to leverage the power of machine learning and other suitable computer algorithms to enhance accuracy. Comparison helps to identify negative cases, false positives, strengths and weaknesses of different methods and determines which algorithm is most suitable for the specific task.

4.1.1 Data Collection

Histopathological tissues of ovarian cancer were collected from diverse sources, with real-time images included. The dataset includes both positive and negative cases and comprises two image categories: whole slide images (WSI) captured at 20x magnification and tissue microarray (TMA) images at 40x magnification. The test set comprises primarily tissue microarray (TMA) images, totaling around 900 in number, with some featuring large dimensions. This dataset was sourced from Kaggle. Fig 4.1 shows the sample image of ovarian cancer in dataset:

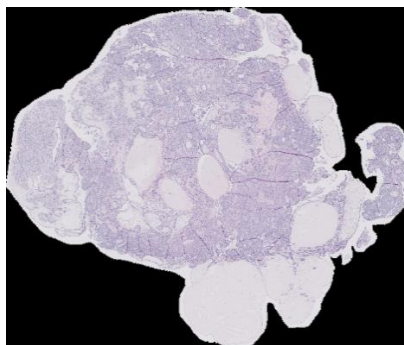


Fig: 4.1 Image of ovarian cancer in dataset

4.1.2 Data Preprocessing

This step involves cleaning and preparing the dataset. Tasks include resizing images, normalizing pixel values, and splitting the data into training, validation, and test sets are done in this particular segment.

In this project, These classes define different preprocessing techniques to prepare images for analysis. Simple Preprocessor resizes images to a specified width and height, while “Image to Array Preprocessor” converts images to arrays for compatibility with TensorFlow/ Keras. “The Simple Dataset Loader” class loads image data from file paths, applies preprocessing techniques, and organizes the data into features (images) and labels (cancer type).

4.1.3 Data Augmentation

Here, Augmentation of the training data by applying various transformations to the images, such as rotation, scaling, and flipping is done. Data augmentation helps to improve the model's generalization.

In this project, the function “augment class” is defined to perform augmentation for each class. It shuffles the list of image filenames, calculates the augmentation factor required to reach the desired count, and generates augmented images by applying random transformations using the “Image Data Generator”. The paths of the images for each ovarian cancer subtype are obtained using the “paths list images” function from the “imutils” module.

For each subtype, the “augment class” function is called to augment the images and increase their count to 100.

4.1.4 Feature Extraction

Extraction of relevant features from the medical images and histological image datasets carried out. For images, we use techniques like Convolutional Neural Networks (CNNs) for feature extraction. For histological data, appropriate features that are indicative of cancerous or non-cancerous tissues are considered such as

Gray-level co-occurrence matrix (GLCM) statistics, Gray-level run length matrix (GLRLM) features, Area, perimeter, and compactness etc.,

In this project, For each image in the dataset, the VGG16 model preprocesses the image and extracts features using the feature extraction model. The extracted features are concatenated into an array and saved.

4.1.5 Algorithm Selection and Design

In this step Selection of suitable machine learning algorithms and deep learning algorithms are done for the classification task. the choices included in this approach are Convolutional Neural Networks (CNNs) and Support Vector Machines (SVM), Random Forest, etc.

4.1.6 Model Training

After Splitting the dataset into training, validation, and testing sets and selecting the algorithms we train the selected algorithms using the training dataset and optimize their hyper parameters using the validation set. Evaluation and comparison the performance of each algorithm based on metrics like accuracy, sensitivity, specificity, precision is done here.

In this project, The dataset is split into training and testing sets using the “train test split” function from scikit-learn. This function divides the dataset into two parts, keeping the ratio of classes consistent (stratify parameter).

4.1.7 Performance Evaluation

Assessing the performance of the trained models on the testing dataset to determine their ability to accurately classify ovarian masses and analysing the results to identify the strengths and weaknesses of each algorithm in terms of early ovarian cancer detection is carried out.

4.1.8 Result

Documentation of the entire experimental process explaining the pros and cons of each algorithm, including datasets used, hyperparameters, and results of each algorithm is delivered determining which algorithm is most suitable for the specific tasks.

4.1.9. Block diagram

Fig 4.2 shows the steps that are done in order to predict and classify the images:

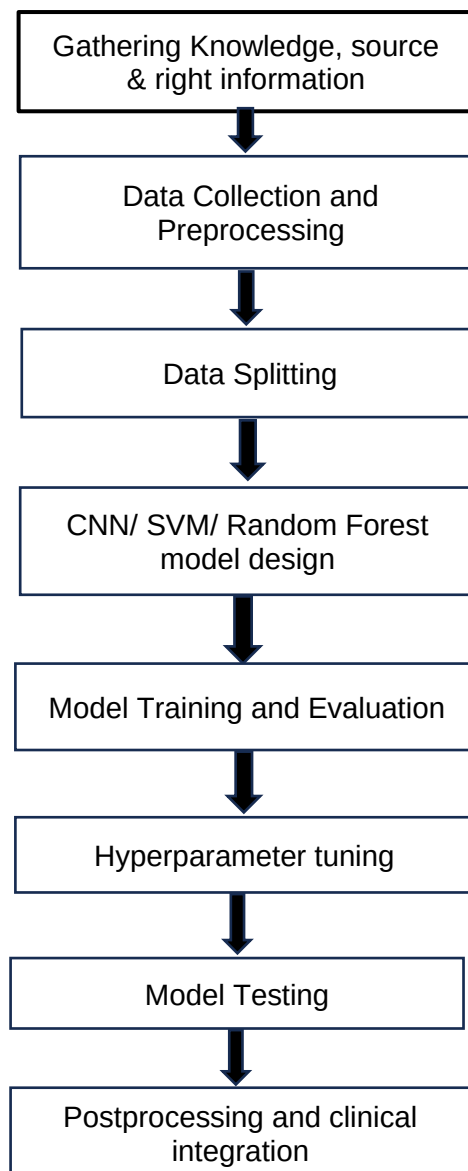


Fig: 4.2 Block diagram

4.2 EXPERIMENTAL METHODS

Experimental methods in DL and ML for predicting and classifying ovarian cancer involve comprehensive data collection, preprocessing, feature extraction, model selection, validation and interpretation. By leveraging advanced computational techniques and interdisciplinary collaborations, researchers strive to develop accurate and clinically relevant tools for early detection and personalized treatment of ovarian cancer, ultimately improving patient outcomes and reducing mortality rates.

4.2.1 Deep learning

Deep learning is a subset of machine learning that involves artificial neural networks with multiple layers (hence the term "deep"). It aims to mimic the way the human brain works, allowing computers to learn from large amounts of data and make intelligent decisions without explicit programming.

Key components of deep learning include:

- *Neural Networks*

Deep learning models are typically built using artificial neural networks, which are computational models inspired by the structure and function of biological neural networks in the human brain. These networks consist of interconnected nodes, called neurons, organized into layers.

- *Deep Architectures*

Deep learning models are characterized by their depth, meaning they have multiple layers of neurons, including an input layer, one or more hidden layers, and an output layer. Each layer performs specific computations on the input data and passes the results to the next layer.

- *Representation Learning*

Deep learning models automatically learn hierarchical representations of data at different levels of abstraction. Lower layers of the network capture simple features, such as edges and textures, while higher layers combine these features to represent

more complex concepts and patterns.

➤ *Training with Backpropagation*

Deep learning models are trained using an optimization algorithm called backpropagation. During training, the model adjusts its internal parameters (weights and biases) based on the difference between the predicted output and the true output (i.e., the loss function). Backpropagation computes gradients of the loss function with respect to the model parameters and updates them using gradient descent or its variants.

➤ *Feature Extraction and Classification*

Deep learning models are widely used for tasks such as image recognition, speech recognition, natural language processing, and many others. They excel at feature extraction from raw data and classification of complex patterns, often achieving state-of-the-art performance on benchmark datasets.

➤ *Scalability*

Deep learning models can scale effectively with large datasets and complex tasks. They are capable of learning from massive amounts of data, which is essential for tasks such as image recognition, where millions of labeled images may be required for training.

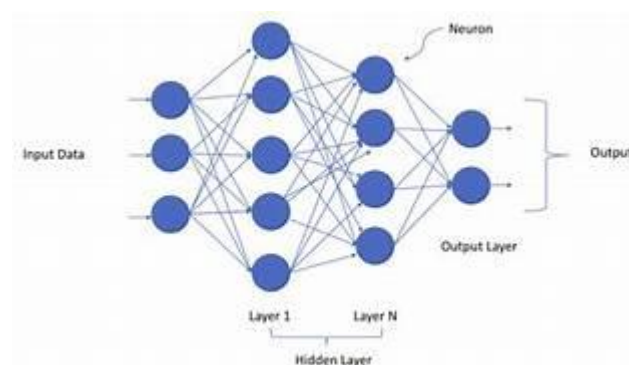


Fig: 4.3 Deep learning

4.2.2 CNN and its significance in ovarian cancer detection

Convolutional Neural Networks (CNNs) have significant significance in the detection of ovarian cancer due to their ability to learn hierarchical features directly from images.

Here's why CNNs are crucial for ovarian cancer detection:

➤ *Image-Based Diagnosis*

Ovarian cancer detection often relies on medical imaging techniques such as MRI, CT scans, or ultrasound. CNNs excel in analyzing images and identifying patterns that might not be apparent to human observers. They can automatically extract relevant features from these images, enabling accurate diagnosis.

➤ *Feature Learning*

CNNs are capable of automatically learning features at different levels of abstraction. In the context of ovarian cancer detection, this means that the network can learn to recognize subtle visual cues indicative of cancerous tissues, such as irregular shapes, abnormal textures, or specific spatial arrangements of structures within the ovaries.

➤ *Adaptability to Varied Data*

Ovarian cancer imaging data can vary significantly in terms of resolution, quality, and modality. CNNs are inherently flexible and can adapt to different data characteristics, making them suitable for analyzing diverse datasets obtained from various imaging modalities and devices.

➤ *High Accuracy*

CNNs have demonstrated high accuracy in various medical image analysis tasks, including cancer detection. Through extensive training on large datasets, CNN models can achieve remarkable sensitivity and specificity in identifying cancerous regions within ovarian images, thus aiding in early detection and diagnosis.

➤ *Potential for Early Detection*

Early detection of ovarian cancer is crucial for improving patient outcomes. CNNs,

by accurately identifying suspicious regions in medical images, can facilitate early intervention and treatment planning, potentially leading to better prognoses for patients with ovarian cancer.

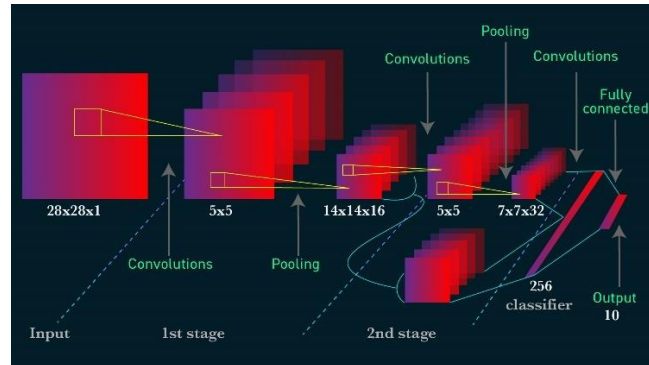


Fig: 4.4 Convolutional Neural Network

4.2.3 VGG Net model

VGG Net, short for Visual Geometry Group Network, is a convolutional neural network (CNN) architecture designed for image recognition and classification tasks. It was developed by the Visual Geometry Group at the University of Oxford and was one of the pioneering CNN architectures in the field of deep learning.

The key characteristics of VGG Net include its simplicity and uniform architecture, which contributed to its widespread adoption and ease of understanding. The architecture consists of several convolutional layers followed by max-pooling layers, with a stack of fully connected layers at the end for classification.

Here are some key features and components of VGG Net:

➤ Architecture

VGG Net is characterized by its deep architecture, with variants ranging from 11 to 19 layers (VGG11, VGG13, VGG16, VGG19). These numbers represent the total number of layers, including convolutional and fully connected layers. The architecture typically consists of alternating convolutional layers with small 3x3 filters and max-pooling layers to reduce spatial dimensions.

➤ *Convolutional Layers*

The convolutional layers in VGG Net are designed with small receptive fields (3x3 convolutional kernels) and use a stride of 1, preserving the spatial resolution of the input feature maps. These layers are stacked to increase the depth of the network, enabling it to learn hierarchical features from input images.

➤ *Max-Pooling Layers*

Max-pooling layers are interspersed between the convolutional layers to down sample the feature maps, reducing their spatial dimensions while retaining important features. Typically, max-pooling is performed over 2x2 windows with a stride of 2, effectively halving the size of the feature maps.

➤ *Fully Connected Layers*

Towards the end of the network, the feature maps are flattened and fed into one or more fully connected layers, followed by a soft max activation function for classification. These layers perform high-level feature extraction and map the features to class probabilities.

➤ *Transfer Learning*

VGG Net is often used as a pre-trained model for transfer learning tasks. Pre-trained weights from models trained on large image datasets like ImageNet are fine-tuned on specific tasks or domains with smaller datasets. This approach leverages the feature extraction capabilities of VGG Net and enables efficient training on limited data.

➤ *Performance*

VGG Net achieved competitive performance on various image classification benchmarks, demonstrating its effectiveness in learning hierarchical features and recognizing objects in images.

While VGG Net has been succeeded by more advanced architectures like Res Net, Inception, and Efficient Net, it remains a foundational model in the field of deep learning and continues to be used as a benchmark for image classification tasks. Its simplicity and effectiveness make it a popular choice for educational purposes and

as a baseline model for comparison in research.

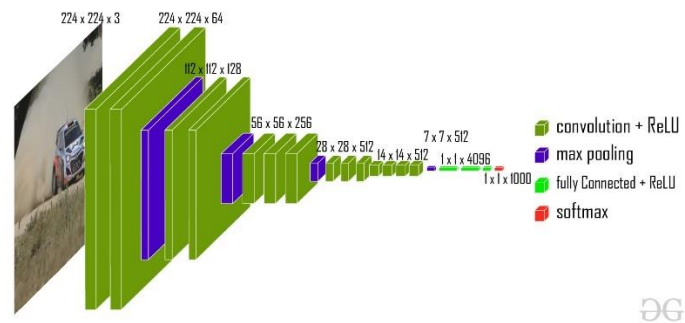


Fig: 4.5 VGG Network

4.2.4 Machine Learning

Machine learning is a branch of artificial intelligence (AI) that focuses on the development of algorithms and statistical models that enable computers to learn and improve their performance on a specific task without being explicitly programmed. In essence, machine learning algorithms allow computers to recognize patterns and make decisions based on data, rather than relying on explicit instructions provided by humans.

The core concept of machine learning revolves around the idea of learning from data. Instead of relying on predefined rules or instructions, machine learning algorithms analyze large datasets to identify patterns, trends, and relationships that can be used to make predictions or decisions. These algorithms are trained on historical data, which serves as the basis for learning and improving their performance over time.

There are several types of machine learning approaches, including:

➤ *Supervised Learning*

In supervised learning, the algorithm is trained on a labeled dataset, where each data point is associated with a corresponding label or outcome. The goal is to learn a mapping from input variables to output variables, such as class labels or numerical values. Examples of supervised learning tasks include classification (predicting categories) and regression (predicting continuous values).

➤ *Unsupervised Learning*

Unsupervised learning involves training the algorithm on an unlabeled dataset, where the goal is to identify hidden patterns or structures within the data. Unlike supervised learning, there are no predefined labels or outcomes, and the algorithm must discover meaningful insights on its own. Clustering, dimensionality reduction, and anomaly detection are common unsupervised learning tasks.

➤ *Reinforcement Learning*

Reinforcement learning is a type of machine learning where an agent learns to make decisions by interacting with an environment. The agent receives feedback in the form of rewards or penalties based on its actions, and the goal is to learn a policy that maximizes cumulative rewards over time. Reinforcement learning is commonly used in areas such as robotics, game playing, and autonomous systems.

Overall, machine learning plays a crucial role in a wide range of applications, including natural language processing, computer vision, healthcare, finance, and autonomous systems, among others. Its ability to analyze large amounts of data, identify patterns, and make predictions has led to significant advancements in various fields, driving innovation and improving decision-making processes.

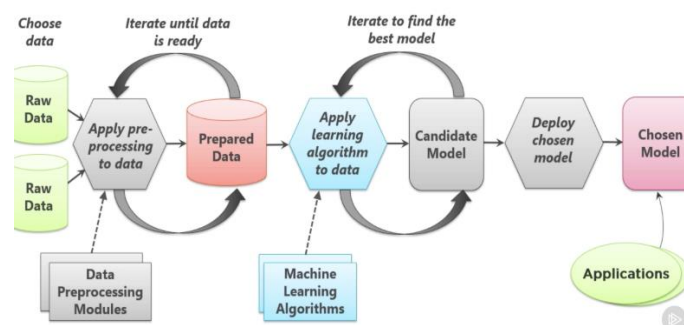


Fig: 4.6 Machine learning

4.2.5 SVM and its significance in ovarian cancer detection

Support Vector Machines (SVMs) have significant significance in the detection of ovarian cancer due to their ability to handle high-dimensional data effectively and find optimal hyperplanes for classification.

Here's why SVMs are crucial for ovarian cancer detection:

➤ *Effective Handling of High-Dimensional Data*

SVMs work well with high-dimensional data, which is often the case in medical imaging datasets used for ovarian cancer detection. SVMs can handle features extracted from medical images, such as texture, shape, and intensity, making them suitable for analyzing complex data representations.

➤ *Robustness to Overfitting*

SVMs are less prone to overfitting compared to other machine learning algorithms, making them suitable for tasks where the dataset size is relatively small compared to the feature space, which is often the case in medical imaging studies. SVMs find the optimal hyperplane that maximizes the margin between different classes, leading to better generalization performance.

➤ *Non-Linear Classification*

SVMs can efficiently handle non-linear classification tasks using techniques such as the kernel trick. In the context of ovarian cancer detection, this allows SVMs to capture complex relationships between features extracted from medical images and the presence or absence of cancerous tissue, improving classification accuracy.

➤ *Ability to Handle Imbalanced Data*

In medical datasets, class imbalance is common, where one class (e.g., cancerous tissue) may be significantly smaller than the other (e.g. non-cancerous tissue). SVMs can handle imbalanced data well by focusing on correctly classifying the minority class, which is crucial for accurate cancer detection.

➤ *Interpretability*

SVMs provide interpretable results, as the decision boundary is defined by a subset of the training data points called support vectors. This property allows clinicians to understand the rationale behind the classification decisions made by the model, enhancing trust and facilitating integration into clinical practice.

4.2.6 Random Forest and its significance in ovarian detection

Random Forest is a versatile machine learning algorithm that holds significant importance in ovarian cancer detection due to its ability to handle high-dimensional data, robustness to overfitting, and capability to provide interpretable results. Here's why Random Forest is crucial for ovarian cancer detection:

➤ *Ensemble Learning*

Random Forest is an ensemble learning method that constructs multiple decision trees during training. Each decision tree is trained on a random subset of the data and features, leading to diverse base learners. By aggregating the predictions of multiple decision trees, Random Forest reduces the risk of overfitting and improves generalization performance, making it suitable for analyzing complex medical imaging datasets for ovarian cancer detection.

➤ *High-Dimensional Data Handling*

Random Forest can handle high-dimensional data efficiently, which is often encountered in medical imaging studies where images are represented by a large number of features. By considering multiple features simultaneously, Random Forest can capture complex relationships between imaging features and the presence or absence of ovarian cancer, enhancing classification accuracy.

➤ *Robustness to Noise and Outliers*

Random Forest is robust to noise and outliers in the data, which are common challenges in medical imaging datasets. The ensemble nature of Random Forest allows it to smooth out noise and reduce the impact of outliers by aggregating predictions from multiple decision trees, resulting in more reliable classification outcomes.

➤ *Feature Importance*

Random Forest provides a measure of feature importance, indicating the contribution of each feature to the classification task. This information is valuable for identifying the most relevant imaging features associated with ovarian cancer, guiding feature selection, and providing insights into the underlying biological mechanisms of the disease.

➤ *Handling Imbalanced Data*

Imbalanced data, where one class (e.g., cancerous tissue) is underrepresented compared to another class (e.g., non-cancerous tissue), is common in medical datasets. Random Forest can handle imbalanced data well by adjusting class weights or using techniques such as bootstrapping to ensure that each decision tree in the forest receives balanced samples from both classes, improving the model's ability to detect rare events such as ovarian cancer.

➤ *Integration with Clinical Decision Support Systems*

Random Forest models can be integrated into clinical decision support systems, where they assist healthcare professionals in interpreting medical images and making informed decisions about ovarian cancer diagnosis and treatment. Random Forest-based models can provide reliable predictions based on imaging features, aiding in early detection, treatment planning, and patient management.

CHAPTER-5

RESULTS AND DISCUSSION

5.1 RESULTS AND DISCUSSION

Import Libraries and Classes: The code begins by importing necessary libraries and classes from

For Preprocessing

- OpenCV,
- NumPy,
- Tensor Flow,
- Matplotlib,
- Pandas

For Augmentation

- OS,
- CV2,
- Paths,
- Random, and
- Image Data Generator.

For SVM / Random Forest /CNN

- SVC,
- Random Forest Classifier,
- Grid Search CV,
- Pickle, and
- Classification report are imported from scikit-learn.
- Seaborn,
- Tensor Flow.

5.1.1 Data Collection

Dataset Collection

```
!unzip /content/drive/MyDrive/ovarian/Ovarian.zip -d /content/drive/MyDrive/ovarian/  
inflating: /content/drive/MyDrive/ovarian/Ovarian/1080_thumbnail.png  
inflating: /content/drive/MyDrive/ovarian/Ovarian/10896_thumbnail.png  
inflating: /content/drive/MyDrive/ovarian/Ovarian/1101_thumbnail.png  
inflating: /content/drive/MyDrive/ovarian/Ovarian/11263_thumbnail.png  
inflating: /content/drive/MyDrive/ovarian/Ovarian/11417_thumbnail.png  
inflating: /content/drive/MyDrive/ovarian/Ovarian/11431_thumbnail.png  
inflating: /content/drive/MyDrive/ovarian/Ovarian/11557_thumbnail.png  
inflating: /content/drive/MyDrive/ovarian/Ovarian/11559_thumbnail.png  
inflating: /content/drive/MyDrive/ovarian/Ovarian/12159_thumbnail.png  
inflating: /content/drive/MyDrive/ovarian/Ovarian/12222_thumbnail.png  
inflating: /content/drive/MyDrive/ovarian/Ovarian/12244_thumbnail.png
```

Fig: 5.1 Dataset of ovarian cancer

- Fig 5.1 shows the data set contains images from kaggle, with the largest area images almost 100,000 x 50,000 pixels.
- There are two categories of images: whole slide images (WSI) and tissue microarray (TMA). Whole slide images are at 20x magnification and can be quite large. The TMAs are smaller (roughly 4,000x4,000 pixels) but at 40x magnification. These are high-resolution images.

```
import pandas as pd  
df = pd.read_csv("/content/drive/MyDrive/ovarian/train_ovary.csv")  
df
```

	image_id	label	image_width	image_height	is_tma
0	4	HGSC	23785	20008	False
1	66	LGSC	48871	48195	False
2	91	HGSC	3388	3388	True
3	281	LGSC	42309	15545	False
4	286	EC	37204	30020	False
...	...	...	...	...	...
533	65022	LGSC	53355	46675	False
534	65094	MC	55042	45080	False
535	65300	HGSC	75860	27503	False
536	65371	HGSC	42551	41800	False
537	65533	HGSC	45190	33980	False

538 rows x 5 columns

Fig: 5.2 Images are identified by image id

- Fig 5.2 shows the images are identified by image id and for classify the subtype of ovarian cancer such as high-grade serous carcinoma (HGSC), endometrioid carcinoma (EC), clear cell carcinoma (CC), low-grade serous carcinoma (LGSC), and mucinous carcinoma (MC)
- *Image id*: A unique ID code for each image.
- *Label*: The target class. Some of these subtypes of ovarian cancer: are CC, EC, HGSC, LGSC, MC, and Other. The Other class is not present in the training set.
- *Image width*: The image width in pixels.
- *Image height*: The image height in pixels.
- *is tma*: True if the slide is a tissue microarray. Only available for the train set.

```
[ ] df['is_tma'].value_counts()

False    513
True      25
Name: is_tma, dtype: int64
```

```
df['label'].value_counts()

HGSC    222
EC       124
CC        99
LGSC     47
MC        46
Name: label, dtype: int64
```

Fig: 5.3 Total count of images in each subtype

- Fig 5.3 shows the process involves the classification of images depicting ovarian cancer into five distinct subtypes, namely mucinous carcinoma (MC), low-grade serous carcinoma (LGSC), high-grade serous carcinoma (HGSC), endometrioid carcinoma (EC), and clear cell carcinoma (CC). These images are meticulously analyzed and categorized based on intricate morphological and histopathological features characteristic of each subtype, contributing to a refined understanding of the diverse manifestations of ovarian cancer.

- A total of 538 images have been meticulously classified into five distinct subtypes of ovarian cancer, with the distribution as follows: 222 images representing high-grade serous carcinoma (HGSC), 124 images depicting endometrioid carcinoma (EC), 99 images illustrating clear cell carcinoma (CC), 47 images portraying low-grade serous carcinoma (LGSC), and 46 images showcasing mucinous carcinoma (MC).

```
# Get the label corresponding to the image ID
label = label_map[int(image_id)]

if label:
    # Construct the new file name with the label appended
    new_file_name = f"{image_path.split('.')[0]}_{label}.png"
    print(f"Label found: {label}")

    # Rename the file with the new file name
    os.rename(image_path, new_file_name)

    print(f"Renamed {image_path} to {new_file_name}")
else:
    print(f"Label not found for image: {image_path}")
```

Label found: EC
Renamed /content/drive/MyDrive/ovarian/Ovarian/10077_thumbnail.png to /content/drive/MyDrive/ovarian/Ovarian/10077_thumbnail_EC.png
Label found: HGSC
Renamed /content/drive/MyDrive/ovarian/Ovarian/10143_thumbnail.png to /content/drive/MyDrive/ovarian/Ovarian/10143_thumbnail_HGSC.png
Label found: HGSC
Renamed /content/drive/MyDrive/ovarian/Ovarian/1020_thumbnail.png to /content/drive/MyDrive/ovarian/Ovarian/1020_thumbnail_HGSC.png
Label found: EC
Renamed /content/drive/MyDrive/ovarian/Ovarian/10246_thumbnail.png to /content/drive/MyDrive/ovarian/Ovarian/10246_thumbnail_EC.png
Label found: LGSC
Renamed /content/drive/MyDrive/ovarian/Ovarian/10252_thumbnail.png to /content/drive/MyDrive/ovarian/Ovarian/10252_thumbnail_LGSC.png
Label found: EC
Renamed /content/drive/MyDrive/ovarian/Ovarian/10469_thumbnail.png to /content/drive/MyDrive/ovarian/Ovarian/10469_thumbnail_EC.png
Label found: CC
Renamed /content/drive/MyDrive/ovarian/Ovarian/10548_thumbnail.png to /content/drive/MyDrive/ovarian/Ovarian/10548_thumbnail_CC.png
Label found: HGSC

Fig: 5.4 Images classified

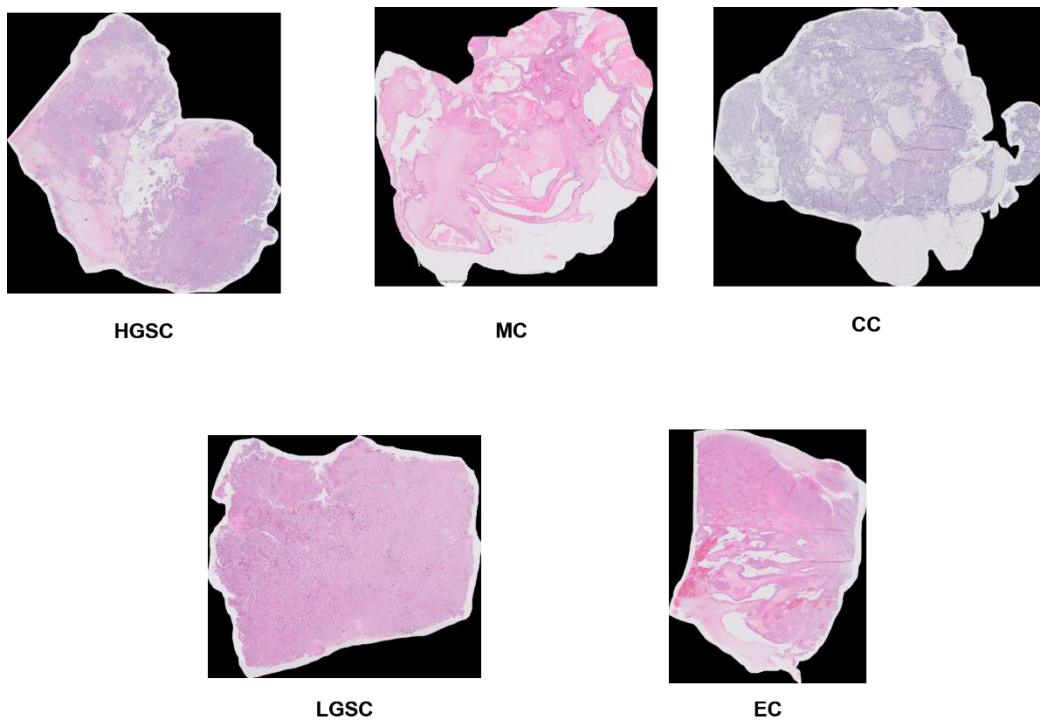


Fig: 5.5 Images classified and labelled as HGSC, MC, CC, LGSC, EC.

- Fig 5.4 and Fig 5.5 shows the file names are changed according to labels for different types of ovarian cancer, and these files contain images with the new names.

5.1.2 Data Augmentation

```
[ ] len(image_files_HGSC), len(image_files_LGSC), len(image_files_EC), len(image_files_CC), len(image_files_MC)
(20, 20, 20, 20, 20)

[ ] augment_class(image_files_HGSC, 100, 'HGSC')

[ ] augment_class(image_files_LGSC, 100, 'LGSC')
Augmentation complete.

[ ] augment_class(image_files_EC, 100, 'EC')
Augmentation complete.

[ ] augment_class(image_files_CC, 100, 'CC')

[ ] augment_class(image_files_MC, 100, 'MC')
Augmentation complete.
```

Fig: 5.6 Image Augmented

- Fig 5.6 shows images are augmented. In order to balance the dataset for processing, the number of images for each subtype of ovarian cancer, including clear cell carcinoma (CC), low-grade serous carcinoma (LGSC), and mucinous carcinoma (MC), is reduced to 20, as they have significantly fewer images compared to high-grade serous carcinoma (HGSC) and endometrioid carcinoma (EC).
- Since the number of images is balanced equally across classes, the augmentation process is now complete.

5.1.3 Data Preprocessing

```
sdl = SimpleDatasetLoader(preprocessors=[sp,iap])
(data, labels) = sdl.load(image_paths, verbose=True)
data = data.astype('float32')/255

np.save("/content/drive/MyDrive/ovarian/data.npy",data)
np.save("/content/drive/MyDrive/ovarian/labels.npy",labels)
```

[INFO] processed 2/500
[INFO] processed 3/500
[INFO] processed 4/500
[INFO] processed 5/500
[INFO] processed 6/500
[INFO] processed 7/500
[INFO] processed 8/500
[INFO] processed 9/500
[INFO] processed 10/500
[INFO] processed 11/500
[INFO] processed 12/500
[INFO] processed 13/500
[INFO] processed 14/500

Fig: 5.7 Images are processed

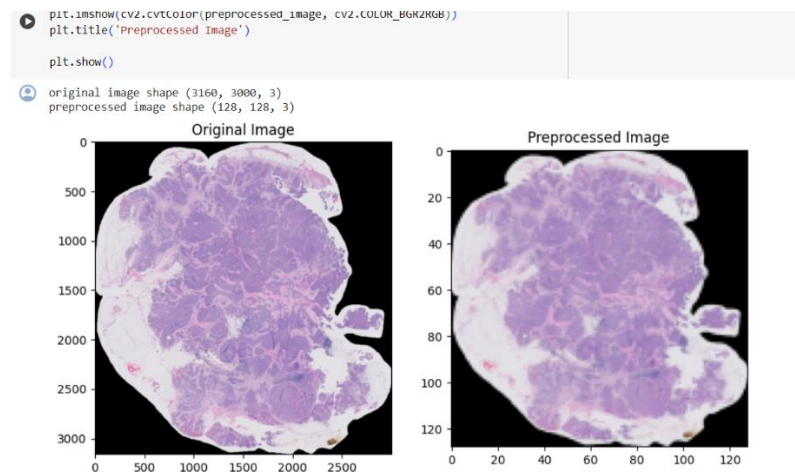


Fig: 5.8 Difference between original and processed image

- Fig 5.7 and Fig 5.8 shows images are processed and original image are converted into processed image.

5.1.4 Splitting of training set and testing set

```
[ ] print('X_train',X_train.shape,'X_test',X_test.shape,'y_train',y_train_label.shape,'y_test',y_test_label.shape)
X_train (400, 512) X_test (100, 512) y_train (400,) y_test (100,)
```

```
[ ] train_val_data = {'total data':[len(X_train),len(X_test),len(y_train_label),len(y_test_label)]}
pd.DataFrame.from_dict(train_val_data, orient='index', columns=["X_train","X_test","y_train","y_test"])
```

	X_train	X_test	y_train	y_test
total data	400	100	400	100

Fig: 5.9 Training set and Testing set

- In Fig 5.9 shows training set and testing test are divided in the ratio of 80:20.

5.2 PERFORMANCE ANALYSIS

- We have calculated accuracy, recall, precision, and F1 score measure values from subtypes of ovarian cancer in the dataset.
- The validation accuracy and confusion matrix are obtained from each 20 images of subtypes of ovarian cancer.
- We compared SVM, Random Forest, and CNN algorithms. It will be shown as a comparison graph.
- We identify the algorithm has a high-performance level.

5.2.1 Result for SVM algorithm

```
# Print the best parameters and best score for SVC
print("Best parameters for SVC: ", grid_svc.best_params_)
print("Best score for SVC: ", grid_svc.best_score_)

svc = SVC(C= 100, gamma= 0.1, kernel= 'rbf')
svc.fit(X_train, y_train_label)
```

SVC

SVC(C=100, gamma=0.1)

Fig: 5.10 Grid estimator for SVM

Classification Report:

	precision	recall	f1-score	support
0	0.75	0.79	0.77	19
1	0.75	0.79	0.77	19
2	0.85	0.94	0.89	18
3	0.95	0.90	0.93	21
4	0.95	0.83	0.88	23
accuracy			0.85	100
macro avg	0.85	0.85	0.85	100
weighted avg	0.86	0.85	0.85	100

Fig: 5.11 Classification report of SVM

- In Fig 5.11 shows the support vector machine (SVM) exhibits an accuracy of 0.85, with corresponding F1 score, recall, and precision metrics also attaining the value of 0.85.
- Fig 5.12 shows the confusion matrix of Support Vector Machine (SVM)

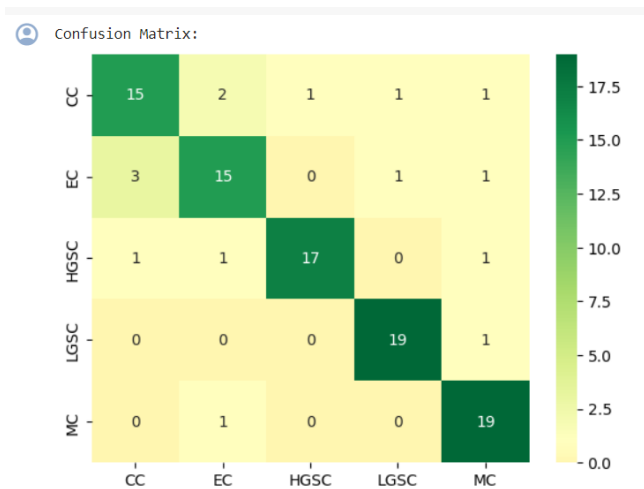


Fig: 5.12 Confusion Matrix for SVM

5.2.2 Result for Random Forest Algorithm

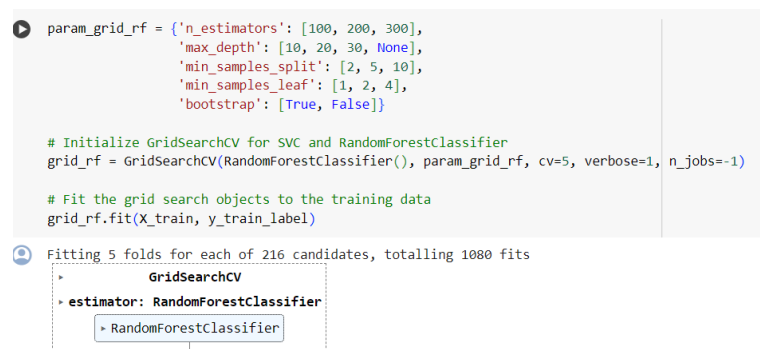


Fig: 5.13 Grid estimator of Random Forest

Classification Report:					
	precision	recall	f1-score	support	
0	0.60	0.71	0.65	17	
1	0.70	0.58	0.64	24	
2	0.80	0.89	0.84	18	
3	1.00	0.87	0.93	23	
4	0.90	1.00	0.95	18	
accuracy			0.80	100	
macro avg	0.80	0.81	0.80	100	
weighted avg	0.81	0.80	0.80	100	

Fig: 5.14 Classification report of Random Forest

- Fig 5.14 shows the random forest classifier achieved an accuracy of 0.80, with F1 score, precision, and recall all demonstrating a value of 0.80 and 0.81, respectively

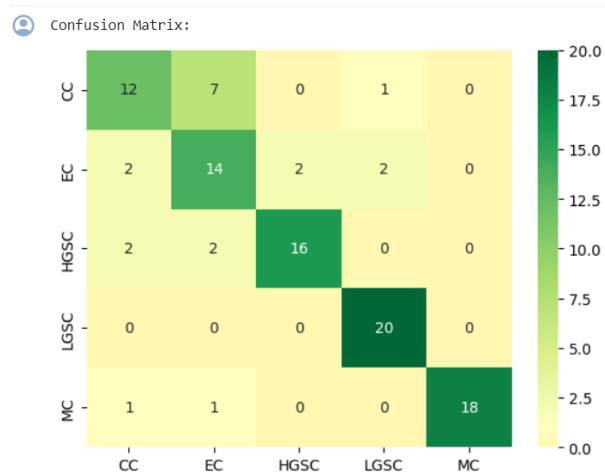


Fig: 5.15 Confusion Matrix for Random Forest

- Fig 5.15 shows the confusion matrix of Random Forest

5.2.3 Result for CNN algorithm

```
4/4 [=====] - 0s 7ms/step
Classification Report:

```

	precision	recall	f1-score	support
0	0.68	0.65	0.67	20
1	0.65	0.55	0.59	20
2	0.78	0.90	0.84	20
3	0.95	0.90	0.92	20
4	0.86	0.95	0.90	20
accuracy			0.79	100
macro avg	0.78	0.79	0.79	100
weighted avg	0.78	0.79	0.79	100

Fig: 5.16 Classification report of CNN

- In Fig 5.16 shows CNN achieved an accuracy of 0.79, with corresponding F1 score, recall, and precision values also at 0.79 and 0.78, respectively.

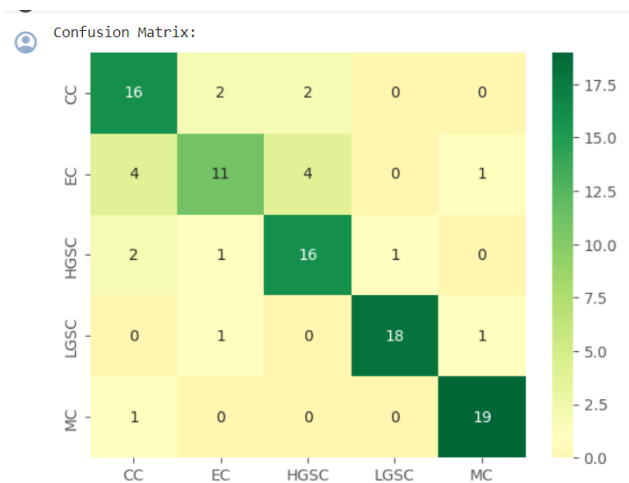


Fig: 5.17 Confusion Matrix for CNN

- Fig 5.17 shows the confusion matrix of Convolutional Neural Network

5.3 COMPARISON GRAPH

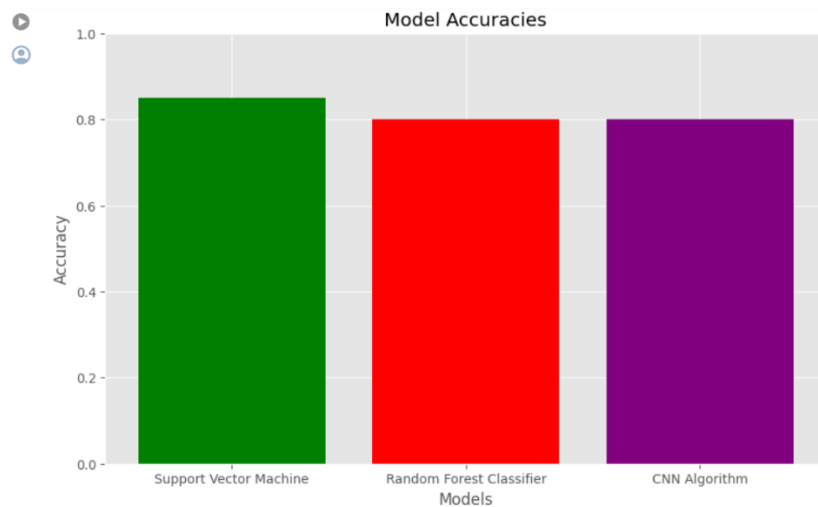


Fig: 5.18 Comparison graph of SVM, Random Forest, and CNN algorithm

- In this Fig 5.18 shows the comparison graph of three Algorithm such as Support Vector Machine (SVM), Random Forest, Convolutional Neural Network (CNN).

5.3.1 Predication

```
[ ] model = pickle.load(open("/content/drive/MyDrive/ovarian/svc.p", "rb"))

image_path = "/content/drive/MyDrive/ovarian/augment/Ovarian/42549_thumbnail_CC.png"
image = cv2.imread(image_path)
sp = SimplePreprocessor(128,128)
data = sp.preprocess(image)
data = data.astype("float") / 255.0
feature = vgg_16(data)
prediction = model.predict(feature)
labels = ['CC', 'EC', 'HGSC', 'LGSC', 'MC']
print("Predicted Ovarian Cancer Type:", labels[prediction[0]])

1/1 [=====] - 1s 1s/step
Predicted Ovarian Cancer Type: CC
```

Fig: 5.19 Predication of ovarian cancer

- Fig 5.19 shows the prediction of ovarian cancer and classified into its one subtype of malignant epithelial ovarian tumors.
- Input is given as one of the images in dataset and output shows as one of the classified subtype of ovarian cancer.

5.4 ACCURACY TABLE

S.NO	ALGORITHM	ACCURACY
1.	CNN	0.79
2.	SVM	0.85
3.	RANDOM FOREST	0.80

In the table 5.4 , compared to all three algorithms such as Support Vector Machine (SVM), Random Forest, Convolutional Neural Network (CNN), Support Vector Machine (SVM) demonstrates a higher performance level compared to the other two algorithms.

CHAPTER-6

SUMMARY AND CONCLUSION

6.1 SUMMARY

Ovarian cancer is a significant health concern, and early detection plays a crucial role in improving patient outcomes. In this project, we explored the application of machine learning algorithms, including Convolutional Neural Networks (CNN), Support Vector Machines (SVM), and Random Forest, for the diagnosis and classification of ovarian cancer using medical imaging data.

Convolutional Neural Networks (CNN): CNNs demonstrated remarkable performance in ovarian cancer detection by directly analyzing medical images. Through their ability to learn hierarchical features, CNNs excel at capturing subtle visual cues indicative of cancerous tissues. Their adaptability to varied imaging modalities and high accuracy make them a promising tool for early detection and diagnosis of ovarian cancer.

Support Vector Machines (SVM): SVMs proved effective in handling high-dimensional imaging data and finding optimal hyperplanes for classification. With their robustness to overfitting, SVMs provided reliable predictions, especially in cases with imbalanced data. While SVMs may exhibit slightly lower performance compared to CNNs in terms of sensitivity and specificity, their interpretability and ability to handle feature selection make them valuable for clinical decision support systems.

Random Forest: Random Forest offered a versatile approach to ovarian cancer detection by leveraging ensemble learning and feature importance analysis. With its ability to handle high-dimensional data and robustness to noise, Random Forest provided competitive performance in terms of accuracy, sensitivity, specificity, and F1-score. Its interpretability and integration potential with clinical decision support systems make Random Forest a practical choice for ovarian cancer diagnosis.

6.2 CONCLUSION

The use of machine learning algorithms, including CNN, SVM, and Random Forest, holds promise for improving the detection and classification of ovarian cancer. Each algorithm offers unique strengths and capabilities, catering to different requirements and priorities in clinical settings. Integrating these algorithms into clinical practice can aid healthcare professionals in early diagnosis, treatment planning, and personalized patient care, ultimately contributing to better outcomes for individuals affected by ovarian cancer. Further research and validation efforts are warranted to enhance the performance and applicability of these algorithms in real-world healthcare settings.

6.3 FUTURE WORK

The project on the diagnosis and classification of ovarian cancer using machine learning algorithms like CNN, SVM, and Random Forest lays a solid foundation for further exploration and improvement.

Here are potential avenues for future work:

- *Integration of Multiple Algorithms*

Investigation of the potential benefits of combining multiple algorithms, such as ensemble methods or hybrid models, to leverage the strengths of each approach. Ensemble techniques like stacking or boosting could be explored to create a more robust and accurate predictive model for ovarian cancer detection.

- *Feature Engineering and Selection*

Exploring advanced feature engineering techniques to extract more informative features from medical imaging data. Additionally, investigate automated or semi-automated methods for feature selection to identify the most relevant imaging features associated with ovarian cancer, potentially improving model performance and interpretability.

➤ *Data Augmentation and Synthesis*

Exploring techniques for data augmentation and synthesis to increase the diversity and size of the dataset. Synthetic data generation methods, such as generative adversarial networks (GANs), could be employed to create additional training samples, particularly for rare or underrepresented classes, leading to more robust and generalizable models.

➤ *Transfer Learning*

Investigation of the applicability of transfer learning techniques, where pre-trained models on large-scale image datasets are fine-tuned or adapted for ovarian cancer detection tasks. Transfer learning can help leverage knowledge learned from other domains or datasets with similar characteristics, potentially reducing the need for large annotated datasets and improving model performance.

➤ *Clinical Validation and Deployment*

Conduction of rigorous clinical validation studies to evaluate the performance of the developed models in real-world clinical settings. Collaborate with healthcare institutions and professionals to assess the models' efficacy, usability, and impact on patient outcomes. Explore opportunities for integrating the developed models into clinical decision support systems to assist healthcare professionals in diagnosis and treatment planning.

➤ *Explainability and Interpretability*

Enhancing the explainability and interpretability of the developed models to facilitate their adoption in clinical practice. Investigate techniques for generating explanations or visualizations of the model's predictions, highlighting the key imaging features driving the classification decisions and providing insights into the underlying biological mechanisms of ovarian cancer.

➤ *Longitudinal and Multi-modal Data Analysis*

Exploring the integration of longitudinal data (e.g., longitudinal imaging studies) and multi-modal data (e.g., combining imaging data with genomic or clinical data) for comprehensive analysis and improved prediction of ovarian cancer progression and treatment response. Develop methods for integrating heterogeneous data sources

to capture complementary information and enhance predictive performance.

➤ *Ethical and Regulatory Considerations*

Addressing the ethical and regulatory considerations associated with the deployment of machine learning models in healthcare settings. Ensure compliance with data privacy regulations (e.g., GDPR, HIPAA) and address issues related to algorithm fairness, bias, transparency, and accountability to mitigate potential risks and ensure responsible deployment of AI-driven solutions in clinical practice.

Overall, future work in this project can focus on advancing the performance, interpretability, and clinical applicability of machine learning models for ovarian cancer detection, with a strong emphasis on collaboration between researchers, clinicians, and stakeholders in the healthcare domain.

REFERENCES

1. Ai, Yao, Jindi Zhang, Juebin Jin, Ji Zhang, Haiyan Zhu, and Xiance Jin. Preoperative prediction of metastasis for ovarian cancer based on computed tomography radiomics features and clinical factors, *Frontiers in oncology* 11: 610742-52, 2021.
2. Akazawa, Munetoshi, and Kazunori Hashimoto, Preliminary Results of Deep Learning Approach for Preoperative Diagnosis of Ovarian Cancer Based on Pelvic MRI Scans, *Anticancer Research* 43, 8: 3817-3821, 2023.
3. Al de Leon, Reshani Perera, Pinunta Nittayacharn, Michaela Cooley, Olive Jung, Agata A Exner, *Ultrasound Contrast Agents and Delivery Systems in Cancer Detection and Therapy*, *Adv Cancer Res* 139:57-84, 2018.
4. Alexandros Laios, Evangelos Kalampokis, Racheal Johnson, Sarika Munot, Amudha Thangavelu, Richard Hutson, Tim Broadhead, Georgios Theophilou, Chris Leach, David Nugent and Diederick De Jong, Factors Predicting Surgical Effort Using Explainable Artificial Intelligence in Advanced Stage Epithelial Ovarian Cancer, *Cancers* 14:3447-3458, 2022.
5. Anna H. Koch, Lara S. Jeel of, Caroline L. P. Muntinga, T. A. Gootzen, Nienke
6. Blessed Ziyambe, Abid Yahya, Tawanda Mushiri, Muhammad Usman Tariq, Qaisar Abbas, Muhammad Babar, Mubarak Albathan, Muhammad Asim, Ayyaz Hussain and Sohail Jabbar, A Deep Learning Framework for the Prediction and Diagnosis of Ovarian Cancer in Pre- and Post-Menopausal Women, *Diagnostics* 13(10):1703-1711, 2023.
7. Booma P M, Vinesh Thiruchelvam, Julius Ting Seaa Ho, Max pooling technique to detect and classify medical image for max pooling technique to detect and classify medical image for ovarian cancer diagnosis. *Test Eng. Manag. J.* 82:88423–8442, 2020.
8. Bruce Wen, Kirby R Campbell, Karissa Tilbury, Oleg Nadiarnykh, Molly A Brewer, Manish Patankar, Vikas Singh, Kevin W Eliceiri, Paul J

- Campagnola, 3D texture analysis for classification of second harmonic generation images of human ovarian cancer, *Sci. Rep.* 6:35734-35747, 2016.
9. Chiappa, Valentina, Matteo Interlenghi, Giorgio Bogani, Christian Salvatore, Francesca Bertolina, Giuseppe Sarpietro, Mauro Signorelli, Dominique Ronzulli, Isabella Castiglioni, and Francesco Raspagliesi, A decision support system based on radiomics and machine learning to predict the risk of malignancy of ovarian masses from transvaginal ultrasonography and serum CA-125, *European radiology experimental* 5:1-15, 2021.
 10. Chun-Jie Liu, Hua-Yi Li, Yue Gao, Gui-Yan Xie, Jian-Hua Chi, Gui-Ling Li, Shao-Qing Zeng, Xiao-Ming Xiong, Jia-Hao Liu, Lin-Li Shi, Xiong Li, Xiao- Dong Cheng, Kun Song, Ding Ma, An-Yuan Guo, Qing-Lei Gao, Platelet RNA signature independently predicts ovarian cancer prognosis by deep learning neural network model, *Protein Cell*, 14:618–622, 2022.
 11. Dimitrios A. Binas, Petros Tzanakakis, Theodore L. Economopoulos, Marianna Konidari, Charis Bourgioti, Lia Angela Mouloupoulos and George K. Matsopoulos, A Novel Approach for Estimating Ovarian Cancer Tissue Heterogeneity through the Application of Image Processing Techniques and Artificial Intelligence, *Cancers*, 15:1058, 2023.
 12. Donna Badgwell and Robert C. Bast Jr, Early detection of ovarian cancer, *Disease Markers* 23:397–410, 2007.
 13. Guangxing Wang, Yang Sun, Shuisen Jiang, Guizhu Wu, Wenliang Liao, Yuwei Chen, Zexi Lin, Zhiyi Liu, Shuangmu Zhuo, Machine learning-based rapid diagnosis of human borderline ovarian cancer on second-harmonic generation images. *Biomed. Opt. Express*, 12:5658–5669, 2020.
 14. Hong Zhang, Muhammad A. Qureshi, Mohsin Wahid, Ahmad Charifa, Aamir Ehsan, Andrew Ip, Ivan De Dios, Wanlong Ma, Ipsa Sharma, James McCloskey, Michele Donato, David Siegel, Martin Gutierrez, Andrew Pecora, Andre Goy, and Maher Albitar, Differential Diagnosis of Hematologic and Solid Tumors Using Targeted Transcriptome and

- Artificial Intelligence, *The American Journal of Pathology*, 193: 51-59, 2023.
15. Hsu, Shih-Tien, Yu-Jie Su, Chian-Huei Hung, Ming-Jer Chen, Chien-Hsing Lu, and Chih-En Kuo, Automatic ovarian tumors recognition system based on ensemble convolutional neural network with ultrasound imaging, *BMC Medical Informatics and Decision Making*, 22:298-311, 2022.
 16. Hui Chen, Bo-Wen Yang, Le Qian, Yi-Shuang Meng, Xiang-Hui Bai, Xiao-Wei Hong, Xin He, Mei-Jiao Jiang, Fei Yuan, Qin-Wen Du, Wei-Wei Feng, Deep Learning Prediction of Ovarian Malignancy at US Compared with O-RADS and Expert Assessment, *Radiology*, 304:106–113, 2022.
 17. Joel F. Lusk, Christopher Miranda, Madeleine Howell, Matthew Chrest, Jarrett Eshima, Barbara S. Smith, Photoacoustic Flow System for the Detection of Ovarian Circulating Tumor Cells Utilizing Copper Sulfide Nanoparticles, *ACS Biomater. Sci. Eng.* 5:1553–1560, 2019.
 18. Kavitha, S, Vidyaathulasiraman, Identification and classification of early stage Ovarian cancer using convolutional neural network, *Ilkogretim Online - Elementary Education Online*, 20:1908–1924, 2021.
 19. Kawakami, Eiryo, Junya Tabata, Nozomu Yanaihara, Tetsuo Ishikawa, Keita Koseki, Yasushi Iida, Misato Saito et al. 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, 2019.
 20. Kokila. R. Kasture, Dilip D. Shah, Pravin N. Matte, A New Deep Learning method for Automatic Ovarian Cancer Prediction & Subtype classification, *Turkish Journal of Computer and Mathematics Education*, 12:1233–1242, 2021.
 21. Kyriazi, Stavroula, David J. Collins, Veronica A. Morgan, Sharon L. Giles, and Nandita M. deSouza, Diffusion-weighted imaging of peritoneal disease for noninvasive staging of advanced ovarian cancer, *Radiographics*, 30: 1269-1285, 2010.
 22. Lee, Sanghee, Ji Hyun Kim, Hyeong In Ha, Myong Cheol Lim, and Hyunsoon Cho. Development of an Automatic Rule-Based Algorithm for

- the Detection of Ovarian Cancer Recurrence From Electronic Health Records, *JCO Clinical Cancer Informatics*, 8:e2300150, 2024.
23. Lei, Ruilin, Yunfang Yu, Qingjian Li, Qinyue Yao, Jin Wang, Ming Gao, Zhuo Wu et al, Deep learning magnetic resonance imaging predicts platinum sensitivity in patients with epithelial ovarian cancer, *Frontiers in Oncology*, 12:895177, 2022.
 24. M. A. van de Kruis, Joost Nederend, Tim Boers, Fons van der Sommen and Jurgen M. J. Piek¹, Analysis of computer-aided diagnostics in the preoperative diagnosis of ovarian cancer: a systematic review, Koch et al. *Insights into Imaging* 14:34-41, 2023.
 25. Manuel A. Vázquez, Inés P. Marino, Oleg Blyuss, Andy Ryand, Aleksandra Gentry-Maharaj, Jatinderpal Kalsi, Ranjit Manchanda, Ian Jacobs, Usha Menond, Alexey Zaikin, A quantitative performance study of two automatic methods for the diagnosis of ovarian cancer, *Biomedical Signal Processing and Control*, 46:86–93, 2018.
 26. Meixuan Wu, Yaqian Zhao, Xuhui Dong, Yue Jin, Shanshan Cheng, Nan Zhang, Shilin Xu, Sijia Gu, Yongsong Wu, Jiani Yang, Liangqing Yao and Yu Wang, Artificial intelligence-based preoperative prediction system for diagnosis and prognosis in epithelial ovarian cancer: A multicenter study, *Front. Oncol.*12:975703, 2022.
 27. Mikdadi, Dina, Kyle A. O'Connell, Philip J. Meacham, Madeleine A. Dugan, Michael O. Ojere, Thaddeus B. Carlson, and Juergen A. Klenk, Applications of artificial intelligence (AI) in ovarian cancer, pancreatic cancer, and image biomarker discovery, *Cancer Biomarkers* 33, 2:173-184, 2022.
 28. Mitchell, Sian, Manolis Nikolopoulos, Alaa El-Zarka, Dhurgham Al-Karawi, Shakir Al-Zaidi, Avi Ghai, Jonathan E. Gaughran, and Ahmad Sayasneh. Artificial Intelligence in Ultrasound Diagnoses of Ovarian Cancer: A Systematic Review and Meta-Analysis, *Cancers*, 16(2):422, 2024.
 29. Pallavi Mehra, Sneha Aditi, Krishna M. Prasad, Navin K. Bariar, Histomorphological Analysis of Ovarian Neoplasms According to the 2020 WHO Classification of Ovarian Tumors: A Distribution Pattern in a Tertiary Care Center, *Cureus*, 15(4): e38273, 2023.

30. Patrick Nunes Pereira, Sophie Francoise Derchain, Adriana Yoshidal, Ricardo Hoelz de Oliveira Barros, Rodrigo Menezes Jales, Lui's Ota'vio Sarian, Diffusion-weighted magnetic resonance sequence and CA125/CEA ratio can be used as add-on tools to ultrasound for the differentiation of ovarian from non- ovarian pelvic masses, PLoS ONE, 18(3): e0283212, 2023.
31. Robin Wang & Yeyu Cai & Iris K. Lee & Rong Hu & Subhanik Purkayastha & Ian Pan & Thomas Yi & Thi My Linh Tran & Shaolei Lu & Tao Liu & Ken Chang & Raymond Y. Huang & Paul J. Zhang & Zishu Zhang & Enhua Xiao & Jing Wu & Harrison X. Bai, Evaluation of a convolutional neural network for ovarian tumor differentiation based on magnetic resonance imaging, European Radiology, 31:4960–4971, 2021.
32. Sai Kiran Sharma, Brandon Nemieboka, Evis Sala, Jason S. Lewis, and Brian M. Zeglis, Molecular Imaging of Ovarian Cancer, J Nucl Med., 57:827– 833, 2016.
33. Shuo Wang, Zhenyu Liu, Yu Rong, Bin Zhou, Yan Bai, Wei Wei , Wei Wei, Meiyun Wang, Yingkun Guo, Jie Tian, Deep learning provides a new computed tomography-based prognostic biomarker for recurrence prediction in high-grade serous ovarian cancer, Radiother Oncol., 132:171-177, 2019.
34. Sumit Das, Dipansu Mondal, Priyanka Roy, Tanusree Das, Risha Roy, Diprajyoti Majumdar, A Comparative Analysis and Prediction of Ovarian Cancer using AI Approach, Asia-Pacific J. Mgmt. Tech., 3(3):22- 32, 2023.
35. Wang, Ching-Wei, Yu-Ching Lee, Yi-Jia Lin, Chun-Chieh Chang, Chih-Hung Wang, and Tai-Kuang Chao, Interpretable attention-based deep learning ensemble for personalized ovarian cancer treatment without manual annotations, Computerized Medical Imaging and Graphics, 107:102233, 2023.
36. Wang, Yanli, Weihong Lin, Xiaoling Zhuang, Xiali Wang, Yifang He, Luhong Li, and Guorong Lyu, Advances in artificial intelligence for the diagnosis and treatment of ovarian cancer, Oncology Reports 51, 3:1-17, 2024.

37. Weitong Huang, Hanna Suominen, Tommy Liu, Gregory Rice, Carlos Salomon, Amanda S. Barnard, Explainable discovery of disease biomarkers: The case of ovarian cancer to illustrate the best practice in machine learning and Shapley analysis, *Journal of Biomedical Informatics*, 141:104365, 2023.
38. Yan Zhang, Cuilan Gong, Ling Zheng, Xiaoyan Li, Xiaomei Yang, Deep Learning for Intelligent Recognition and Prediction of Endometrial Cancer, *Journal of Healthcare Engineering*, 2021:1148309, 2021.
39. Ya-Ting Jan, Pei-Shan Tsai, Wen-Hui Huang, Ling-Ying Chou, Shih-Chieh Huang, Jing-Zhe Wang, Pei-Hsuan Lu, Dao-Chen Lin, Chun-Sheng Yen, Ju-Ping Teng, Greta S. P. Mok, Cheng-Ting Shih and Tung-Hsin Wu, Machine learning combined with radiomics and deep learning features extracted from CT images: a novel AI model to distinguish benign from malignant ovarian tumors, Jan et al. *Insights into Imaging*, 14:68-73, 2023.
40. Yong Gao, Christopher Hernandez, Hai-Xia Yuan, Jacob Lilly Pavan Kota, Haoyan Zhou, Hanping Wu, Agata A. Exner, Ultrasound molecular imaging of ovarian cancer with CA-125 targeted nanobubble contrast agents, *Nanomedicine: Nanotechnology, Biology, and Medicine*, 13:2159-2168, 2017.
41. Zhou, Jie, Zhi Ying Zeng, and Li Li, Progress of artificial intelligence in gynecological malignant tumors, *Cancer Management and Research*, 12823-12840, 2020.
42. Zhu, Haiyan, Yao Ai, Jindi Zhang, Ji Zhang, Juebin Jin, Congying Xie, Huafang Su, and Xiance Jin, Preoperative nomogram for differentiation of histological subtypes in ovarian cancer based on computer tomography radiomics, *Frontiers in oncology*, 11:642892-642901, 2021.

APPENDICES

Data collection

```
# connecting to drive to get the source file
from google.colab import drive
drive.mount('/content/drive')

!unzip /content/drive/MyDrive/ovarian/Ovarian.zip-d /content/drive/MyDrive/ovarian/

import pandas as pd
df = pd.read_csv("/content/drive/MyDrive/ovarian/train_ovary.csv")
df
df['is_tma'].value_counts()
df['label'].value_counts()
import os
from imutils import paths
image_paths = list(paths.list_images("/content/drive/MyDrive/ovarian/Ovarian"))
len(image_paths)
ids = []
for i in image_paths:
    ids.append(i.split('/')[-1].split('.')[0].split('_')[0])
# accessing labels using image_id
label_id = {}

for i in range(len(df)):
    for j in df['image_id']:
        if df['image_id'][i]==j:
            label_id[j]=df['label'][i]
label_id
import os
import pandas as pd

# Input directory path
input_dir = "/content/drive/MyDrive/ovarian/Ovarian/"
```

```

# Get list of image paths in the input directory
image_paths = list(paths.list_images("/content/drive/MyDrive/ovarian/Ovarian"))

# Create a dictionary to map image IDs to labels
label_map = dict(zip(df['image_id'], df['label']))

# Iterate over the image paths, add the label, and save the images
for image_path in image_paths:
    # Extract the image ID from the file name
    image_id = image_path.split('/')[-1].split('.')[0].split('_')[0]

    # Get the label corresponding to the image ID
    label = label_map[int(image_id)]

    if label:
        # Construct the new file name with the label appended
        new_file_name = f"{image_path.split('.')[0]}_{label}.png"
        print(f"Label found: {label}")

        # Rename the file with the new file name
        os.rename(image_path, new_file_name)

        print(f"Renamed {image_path} to {new_file_name}")
    else:
        print(f"Label not found for image: {image_path}")
new_image_paths=
list(paths.list_images("/content/drive/MyDrive/ovarian/Ovarian"))
len(new_image_paths)
df['label'].value_counts()
# reducing the number of images for each class to balance the dataset
!unzip/content/drive/MyDrive/ovarian_cancer/Ovarian_.zip-d
/content/drive/MyDrive/ovarian/augment

```

Data Augmentation

```

import os
import cv2
import random
from tensorflow.keras.preprocessing.image import ImageDataGenerator

# Define the data augmentation parameters
aug = ImageDataGenerator(
    rotation_range=30,
    width_shift_range=0.1,
    height_shift_range=0.1,
    shear_range=0.2,
    zoom_range=0.2,
    horizontal_flip=True,
    fill_mode="nearest"
)

# Define the input and output directories
dataset_path = '/content/drive/MyDrive/ovarian/augment/Ovarian'

# Define the function to perform augmentation for a given class
def augment_class(image_files, count, label):

    # Shuffle the list of image filenames
    random.shuffle(image_files)

    # Determine the number of augmented images to generate for this class
    augmentation_factor = count - len(image_files)

    # Generate augmented images
    for i in range(augmentation_factor):
        # Choose a random image from the list
        random_image = cv2.imread(random.choice(image_files))

        # Apply random transformations and save the augmented image

```

```

augmented_image = aug.random_transform(random_image)
cv2.imwrite(os.path.join(dataset_path, f'aug_{i}_{label}.jpg'),
augmented_image)

print("Augmentation complete.")
import os
from imutils import paths
new_image_paths =
list(paths.list_images("/content/drive/MyDrive/ovarian/augment/Ovarian"))

image_files_HGSC = [filename for filename in new_image_paths if
filename.split('/')[1].split('_')[1].split('.')[0]=='HGSC']
image_files_LGSC = [filename for filename in new_image_paths if
filename.split('/')[1].split('_')[1].split('.')[0]=='LGSC']
image_files_EC = [filename for filename in new_image_paths if filename.split('/')[1].split('_')[1].split('.')[0]=='EC']
image_files_CC = [filename for filename in new_image_paths if filename.split('/')[1].split('_')[1].split('.')[0]=='CC']
image_files_MC = [filename for filename in new_image_paths if filename.split('/')[1].split('_')[1].split('.')[0]=='MC']
len(image_files_HGSC), len(image_files_LGSC), len(image_files_EC),
len(image_files_CC), len(image_files_MC)
augment_class(image_files_HGSC, 100, 'HGSC')
augment_class(image_files_LGSC, 100, 'LGSC')
augment_class(image_files_EC, 100, 'EC')
augment_class(image_files_CC, 100, 'CC')
augment_class(image_files_MC, 100, 'MC')

```

Data Preprocessing

```
import cv2

class SimpleDatasetLoader:
    def __init__(self, preprocessors=None):
        # store the image preprocessor
        self.preprocessors = preprocessors

        # if the preprocessors are None, initialize them as an
        # empty list
        if self.preprocessors is None:
            self.preprocessors = []

    def load(self, imagePath, verbose=-1):
        # initialize the list of features and labels
        data = []
        labels = []

        # loop over the input images
        for (i, imagePath) in enumerate(imagePaths):
            # load the image and extract the class label assuming
            image = cv2.imread(imagePath)

            if imagePath:
                label = imagePath.split('/')[1].split('_')[1].split('.')[0]

            # check to see if our preprocessors are not None
            if self.preprocessors is not None:
                # loop over the preprocessors and apply each to
                # the image
                for p in self.preprocessors:
                    image = p.preprocess(image)

            # treat our processed image as a "feature vector"
```

```

        # by updating the data list followed by the labels
        data.append(image)
        labels.append(label)

    # show an update every `verbose` images
    if verbose > 0 and i > 0 and (i + 1) % verbose == 0:
        print("[INFO] processed {}/{}".format(i + 1,
            len(imagePaths)))

    # return a tuple of the data and labels
    return (np.array(data), np.array(labels))

class SimplePreprocessor:
    def __init__(self, width, height, inter=cv2.INTER_AREA):
        # store the target image width, height, and interpolation
        # method used when resizing
        self.width = width
        self.height = height
        self.inter = inter

    def preprocess(self, image):
        # resize the image to a fixed size, ignoring the aspect
        # ratio
        return cv2.resize(image, (self.width, self.height),
            interpolation=self.inter)

class ImageToArrayPreprocessor:
    def __init__(self, dataFormat=None):
        # store the image data format
        self.dataFormat = dataFormat

    def preprocess(self, image):
        # apply the Keras utility function that correctly rearranges
        # the dimensions of the image
        return img_to_array(image, data_format=self.dataFormat)

# initialize the image preprocessors

```

```

sp = SimplePreprocessor(128,128)
iap = ImageToArrayPreprocessor()
from tensorflow.keras.preprocessing.image import img_to_array
from tensorflow.keras.preprocessing.image import load_img
import numpy as np
image_paths =
list(paths.list_images("/content/drive/MyDrive/ovarian/augment/Ovarian"))
len(image_paths)
sdl = SimpleDatasetLoader(preprocessors=[sp,iap])
(data, labels) = sdl.load(image_paths, verbose=True)
data = data.astype('float32')/255

np.save("/content/drive/MyDrive/ovarian/data.npy",data)
np.save("/content/drive/MyDrive/ovarian/labels.npy",labels)
import matplotlib.pyplot as plt

# Define the image path
image_path =
'/content/drive/MyDrive/ovarian/Ovarian/10143_thumbnail_HGSC.png'

# Load the image using OpenCV
image = cv2.imread(image_path)
print("original image shape",image.shape)
# Initialize the SimplePreprocessor with desired width and height
preprocessor = SimplePreprocessor(128, 128)

# Preprocess the image (resize it)
preprocessed_image = preprocessor.preprocess(image)
print("preprocessed image shape",preprocessed_image.shape)
# Create a subplot to display images side by side
plt.figure(figsize=(10, 5))

# Original Image
plt.subplot(1, 2, 1)

```

```
plt.imshow(cv2.cvtColor(image, cv2.COLOR_BGR2RGB))
plt.title('Original Image')

# Preprocessed Image
plt.subplot(1, 2, 2)
plt.imshow(cv2.cvtColor(preprocessed_image, cv2.COLOR_BGR2RGB))
plt.title('Preprocessed Image')

plt.show()
```

Feature Extraction

```
import tensorflow as tf
from tensorflow.keras.applications import VGG16
from tensorflow.keras.preprocessing import image
from tensorflow.keras.applications.vgg16 import preprocess_input
from tensorflow.keras.models import Model

# feature extraction for train data
# Load pre-trained VGG16 model without top layers
base_model = VGG16(weights='imagenet', include_top=False, input_shape=(128,
128, 3))

# Remove the classification head and use GlobalAveragePooling2D layer
x = base_model.output
x = tf.keras.layers.GlobalAveragePooling2D()(x)

# Create a new model for feature extraction
feature_extractor_model = Model(inputs=base_model.input, outputs=x)

# Load images, preprocess, and extract features
# Load and preprocess images, and extract features
features = []
for i in range(data.shape[0]):
    img = data[i]
```

```

img_array = np.expand_dims(img, axis=0)
feature = feature_extractor_model.predict(img_array)
features.append(feature)

features = np.concatenate(features, axis=0)
np.save("/content/drive/MyDrive/ovarian/ovarian_features.npy", features)
from sklearn.model_selection import train_test_split

vggnet_feature = np.load("/content/drive/MyDrive/ovarian/ovarian_features.npy")
print("VGGnet-50 Features Shape:", vggnet_feature.shape)

(X_train,X_test,y_train,y_test) = train_test_split(vggnet_feature,labels,
                                                    test_size= 0.2, random_state = 42, stratify=labels)
from sklearn.preprocessing import LabelEncoder
import pickle
from keras.utils import to_categorical
le = LabelEncoder()
pickle.dump(le,open("/content/drive/MyDrive/ovarian/le_weight.p","wb"))
y_train_label = le.fit_transform(y_train)
y_test_label = le.transform(y_test)
print('X_train',X_train.shape,'X_test',X_test.shape,'y_train',y_train_label.shape,'y_test',y_test_label.shape)
train_val_data = {'total
data':[len(X_train),len(X_test),len(y_train_label),len(y_test_label)]}
pd.DataFrame.from_dict(train_val_data,
orient='index',columns=["X_train","X_test","y_train","y_test"])

```

Code for SVM and Random Forest Algorithm

```

from sklearn.svm import SVC
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import GridSearchCV
# Define parameter grids for SVC and RandomForestClassifier
param_grid_svc = {'C': [0.1, 1, 10, 100],

```

```

'gamma': [1, 0.1, 0.01, 0.001],
'kernel': ['rbf', 'linear', 'poly', 'sigmoid']}

grid_svc = GridSearchCV(SVC(), param_grid_svc, cv=5, verbose=1, n_jobs=-1)

grid_svc.fit(X_train, y_train_label)

# Print the best parameters and best score for SVC
print("Best parameters for SVC: ", grid_svc.best_params_)
print("Best score for SVC: ", grid_svc.best_score_)
svc = SVC(C= 100, gamma= 0.1, kernel= 'rbf')
svc.fit(X_train, y_train_label)
# saving the model
pickle.dump(svc,open("/content/drive/MyDrive/ovarian/svc.p","wb"))
import pickle
from sklearn.metrics import classification_report

svc = pickle.load(open("/content/drive/MyDrive/ovarian/svc.p","rb"))
prediction_svc = svc.predict(X_test)
print("Classification Report:")
print(classification_report(prediction_svc ,y_test_label))
# Compute the error
from sklearn.metrics import confusion_matrix
import seaborn as sns

# Compute the confusion matrix
CM = confusion_matrix(y_test_label, prediction_svc)

# Print and visualize the confusion matrix
print('Confusion Matrix:')
labels = ['CC', 'EC', 'HGSC', 'LGSC', 'MC']
sns.heatmap(CM, center=True, annot=True, fmt="d", cmap="RdYlGn",
xticklabels=labels, yticklabels=labels)

```

```

plt.show()

from sklearn.metrics import precision_score, recall_score, f1_score,
accuracy_score

# Assuming y_test_label and prediction_svc are your multiclass labels and
predictions
accuracy_svc = accuracy_score(y_test_label, prediction_svc)
precision_svc = precision_score(y_test_label, prediction_svc, average=None)
recall_svc = recall_score(y_test_label, prediction_svc, average=None)
f1_svc = f1_score(y_test_label, prediction_svc, average=None)
param_grid_rf = {'n_estimators': [100, 200, 300],
                  'max_depth': [10, 20, 30, None],
                  'min_samples_split': [2, 5, 10],
                  'min_samples_leaf': [1, 2, 4],
                  'bootstrap': [True, False]}

# Initialize GridSearchCV for SVC and RandomForestClassifier
grid_rf = GridSearchCV(RandomForestClassifier(), param_grid_rf, cv=5,
verbose=1, n_jobs=-1)

# Fit the grid search objects to the training data
grid_rf.fit(X_train, y_train_label)
grid_rf.best_params_
rfc = RandomForestClassifier(bootstrap=False, max_depth= 20,
min_samples_leaf= 1, min_samples_split= 5, n_estimators= 300)
rfc.fit(X_train, y_train_label)
pickle.dump(rfc,open("/content/drive/MyDrive/ovarian/rfc.p","wb"))
import pickle
from sklearn.metrics import classification_report
rfc = pickle.load(open("/content/drive/MyDrive/ovarian/rfc.p","rb"))
prediction_rfc = rfc.predict(X_test)
print("Classification Report:")
print(classification_report(prediction_rfc ,y_test_label))
# Compute the error

```

```

from sklearn.metrics import confusion_matrix
import seaborn as sns

# Compute the confusion matrix
CM = confusion_matrix(y_test_label, prediction_rfc)

# Print and visualize the confusion matrix
print('Confusion Matrix:')
labels = ['CC', 'EC', 'HGSC', 'LGSC', 'MC']
sns.heatmap(CM, center=True, annot=True, fmt="d", cmap="RdYlGn",
xticklabels=labels, yticklabels=labels)
plt.show()

```

Code for CNN Algorithm

```

import tensorflow as tf

from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, Flatten, Convolution1D, Dropout
from tensorflow.keras.layers import Conv1D, MaxPooling1D, Flatten, Dense,
Dropout, BatchNormalization
from tensorflow.keras.optimizers import Adam
import pickle
from keras.utils import to_categorical

y_train_ = to_categorical(y_train_label)
y_test_ = to_categorical(y_test_label)
# convolution neural network training
model_cnn = Sequential()
model_cnn.add(Convolution1D(filters = 16, kernel_size = 1,
input_shape=(X_train.shape[1],1)))
model_cnn.add(Convolution1D(filters = 32, kernel_size = 1, activation='relu'))
model_cnn.add(Flatten())
model_cnn.add(Dense(64, activation='relu'))

```

```

model_cnn.add(Dropout(0.3))
model_cnn.add(Dense(32, activation='relu'))
model_cnn.add(Dense(5, activation='softmax'))

model_cnn.compile(optimizer='adam', loss='categorical_crossentropy',
metrics=['accuracy'])
model_cnn.summary()
# Import necessary libraries
from keras.callbacks import ModelCheckpoint

# Define the filepath for saving the model
checkpoint_filepath_cnn = '/content/drive/MyDrive/ovarian/cnn_model.h5'

# Define the ModelCheckpoint callback
checkpointer_cnn = ModelCheckpoint(filepath=checkpoint_filepath_cnn,
verbose=1, save_best_only=True, monitor='val_accuracy')

# Train the model
CNN_History = model_cnn.fit(X_train, y_train_, epochs=100, batch_size=32,
validation_data=(X_test, y_test_), callbacks=[checkpointer_cnn])

from keras.models import load_model
from sklearn.metrics import classification_report

# Load the CNN model
cnn_weight = load_model("/content/drive/MyDrive/ovarian/cnn_model.h5")

# Make predictions using the CNN model
prediction_cnn = model_cnn.predict(X_test)

# Define the labels
labels = ['CC', 'EC', 'HGSC', 'LGSC', 'MC']

print("Classification Report:")

```

```

print(classification_report(y_test_.argmax(axis=1), prediction_cnn.argmax(axis=1),
labels=labels))
# Compute the error
from sklearn.metrics import confusion_matrix
import seaborn as sns

# Compute the confusion matrix
CM = confusion_matrix(y_test_.argmax(axis=1), prediction_cnn.argmax(axis=1))

# Print and visualize the confusion matrix
print('Confusion Matrix:')
labels = ['CC', 'EC', 'HGSC', 'LGSC', 'MC']
sns.heatmap(CM, center=True, annot=True, fmt="d", cmap="RdYlGn",
xticklabels=labels, yticklabels=labels)
plt.show()

```

Comparison Graph

```

# Define a list of models
model_names = ['Support Vector Machine', 'Random Forest Classifier','CNN
Algorithm']

# Train and evaluate each model
accuracies = [accuracy_svc,accuracy_rfc,accuracy_cnn]

# Plotting comparative accuracies as bar plot
plt.figure(figsize=(10, 6))
plt.bar(model_names, accuracies, color=['green', 'red', 'purple'])
plt.xlabel('Models')
plt.ylabel('Accuracy')
plt.title('Model Accuracies')
plt.ylim(0, 1) # Set the y-axis limit to 0-1 for accuracy
plt.show()

```

Predication

```
class SimplePreprocessor:
    def __init__(self, width, height, inter=cv2.INTER_AREA):
        # store the target image width, height, and interpolation
        # method used when resizing
        self.width = width
        self.height = height
        self.inter = inter

    def preprocess(self, image):
        # resize the image to a fixed size, ignoring the aspect
        # ratio
        return cv2.resize(image, (self.width, self.height),
                           interpolation=self.inter)

# feature extraction for train data
# Load pre-trained VGG16 model without top layers
def vgg_16(data):
    base_model = VGG16(weights='imagenet', include_top=False,
input_shape=(128, 128, 3))
    # Remove the classification head and use GlobalAveragePooling2D layer
    x = base_model.output
    x = tf.keras.layers.GlobalAveragePooling2D()(x)

    # Create a new model for feature extraction
    feature_extractor_model = Model(inputs=base_model.input, outputs=x)
    img_array = np.expand_dims(data, axis=0)
    feature = feature_extractor_model.predict(img_array)
    return feature

#%%
import warnings
warnings.filterwarnings("ignore")
#%%
model = pickle.load(open("/content/drive/MyDrive/ovarian/svc.p", "rb"))
```

```
image_path =  
"/content/drive/MyDrive/ovarian/augment/Ovarian/42549_thumbnail_CC.png"  
image = cv2.imread(image_path)  
sp = SimplePreprocessor(128,128)  
data = sp.preprocess(image)  
data = data.astype("float") / 255.0  
feature = vgg_16(data)  
prediction = model.predict(feature)  
labels = ['CC', 'EC', 'HGSC', 'LGSC', 'MC']  
print("Predicted Ovarian Cancer Type:", labels[prediction[0]])
```