Classification Of Subtypes Of Ovarian Cancer And Detecting Outliers

Contributors:
Kauser Nissar TP, NMKRV College for Women (Autonomous)
Shaziya Sultana, NMKRV College for Women (Autonomous)

Abstract:

The female reproductive organs called ovaries, which are in charge of creating hormones and eggs, can develop cancer known as ovarian cancer. Ovarian cancer is primarily categorized into five types. They are, high-grade serous carcinoma, clear-cell ovarian carcinoma, endometrioid, low-grade serous, and mucinous carcinoma. There are other types of ovarian cancer which are very rare in occurrence and are often referred to as outliers. Detection in the early stages is very challenging, leading to poorer prognosis. Treatment generally involves surgery, chemotherapy, and sometimes radiation therapy. The outcomes of these therapies depends on the stage it is diagnosed at and also depends on several individual factors. Deep learning and machine learning models are increasingly being used in the field of medical imaging to assist in the diagnosis of diseases from images of tissues. In the context of tissue diagnosis, convolutional neural networks (CNNs) and pre-trained models like MobileNet,VGG-16 have shown remarkable efficacy. These deep learning models can automatically extract hierarchical features from images, allowing them to identify subtle patterns indicative of various tissue abnormalities. The aim of this project is to greatly increase ovarian cancer therapy efficacy, accessibility, and diagnostic accuracy. The project's potential to lessen negative effects and improve patient outcomes is highlighted by the employment of machine learning and deep learning techniques.

Keywords: Ovarian cancer, Subtype classification, Deep learning, Machine learning, Convolutional neural networks (CNNs), Pre-trained models, VGG-16,MobileNet.

1.Introduction

Cancer is one of the leading causes of death in the current present world. It was determined that 14,61,427 incident cases of cancer (crude rate: 100.4 per 100,000) would occur in India in 2022[1]. According to a recent study, one in nine people are prone to cancer in India[2]. Various studies and research fuelled by artificial intelligence have been conducted in order to find a cure for this lethal disease. Ovarian cancer is a cancer that originates in the female reproductive organs i.e., ovaries. According to the Globocan 2018 Fact sheet, ovarian cancer accounted for 3.44% (36170) of all cancer cases and was expected to be the third most prevalent disease among Indian women and the eighth most common cancer globally[2]. Ovarian cancer is challenging to detect because prominent symptoms do not occur until the final stages. It is also common for patients at the later stages to have vague symptoms. In the early stages, this cancer has a five-year survival rate of approximately 71% to 89%[3]. Life expectancy drastically reduces at the later stages due to the complexity of the treatment. Five year survival rate at the final stage is expected to be 30.8%[3]. Early detection increases the likelihood of successful treatment. Methods for early detection include imaging techniques such as transvaginal

ultrasound (TVUS) and pelvic MRI[4]. These techniques give us a detailed visualization of the tissues in the ovary[4]. These visualizations can be used to detect cysts or unusual masses[4].

2.Problem Statement

As discussed earlier, detection of cancer in the early stages is the key to successful treatment outcomes. By training machine learning algorithms on large datasets we can find patterns linked to early cancer. In this approach feature extraction is a crucial step. This paper presents a method for classification of the subtypes of ovarian cancer. Detecting the type of cancer aids in determining effective treatment for the diagnosed individual. Our contribution in this research is as follows:

1. ResNet for Subtype Classification
2. Handling Rare Subtypes
3. Comprehensive Evaluation Metrics
4. Benchmarking and Comparative Analysis

3. Literature Survey

Long-Yi-Guo et al.(2020), focused on identification of ovarian cancer subtypes using multi-omics data[5]. The dataset comprises tens of thousands of samples from 38 cancer types sourced from The Cancer Genome Atlas (TCGA)[5]. The algorithm employs Denoising Auto Encoder for dimensionality reduction, K-means clustering using reconstructed features, and logistic regression for subtype classification[5]. Notably, K-means without dimensionality reduction achieved the lowest silhouette score (0.165), while AE-Kmeans and DAE-Kmeans obtained silhouette scores of 0.549 and 0.583, respectively[5]. The study introduces a novel deep learning framework that effectively identifies two distinct ovarian cancer subtypes at the molecular level[5]. Outperforming traditional and other deep learning methods, the model's validation with three independent datasets from GEO confirms significant differences between classified cancer subgroups, emphasizing its reliability for personalized treatment and outcome prediction[5].

Abid Yahya et al. (2023), presented a Deep Learning Framework for the Prediction and Diagnosis of Ovarian Cancer in Pre- and Post-Menopausal Women[6]. The dataset comprises 200 images, with 100 images each of serous ovarian cancer and non-cancerous samples, obtained from TCGA. The utilized algorithm is a Convolutional Neural Network (CNN), which achieves an impressive validation accuracy of 94.43%[6]. The study highlights the deep learning model's substantial accuracy, reaching 94%, emphasizing its potential for early detection and personalized management in the context of ovarian cancer[6].

Kokila. R. Kasture, et al.(2021), proposed a new deep learning method for automatic ovarian cancer prediction and subtype classification. The dataset comprises 500 labelled histopathological images, including 175 serous, 100 mucinous, 60 endometroid, 80 clear-cell, and 85 non-cancerous samples collected from the National Cancer Institute's Genomic Data Commons data portal, TCGA-OV repository[7]. The algorithm employed is a new Deep

Convolutional Neural Network (DCNN) architecture inspired by AlexNet but designed from scratch for improved performance[7]. The model includes eight convolutional layers, four max-pooling layers, and four dense layers with ELU activation after each convolutional layer[7]. To address overfitting, three fully connected layers with 64 nodes each and a dropout of 0.3 are employed, and the final SoftMax layer, set to five classes, handles subtype classification[7]. The study achieves an impressive 83.93% accuracy for ovarian cancer prediction post-image augmentation, surpassing the prior literature, which reported 72% accuracy before augmentation[7]. The authors have publicly shared the dataset and code, and future work aims to develop a user-friendly interface for pathologists, indicating the potential for impactful advancements in ovarian cancer diagnosis and classification[7].

Hossein Farahani et al.(2022), aimed to develop a generalizable model for automatic ovarian carcinoma histotype classification. compared four deep learning-based models (details outlined below) for ovarian carcinoma histotype classification[8]. The models were initialized with ImageNet pre-trained weights, and then each were fine-tuned with the colour-normalized patches from the Internal Training Dataset, using the ImageNet mean and standard deviation to normalize the RGB pixel values of each patch between −1 and 1[8].After training, the performance of the models were compared using the testing set of the respective Internal Training Dataset cross-validation split and the out-of-distribution External Test Dataset[8]. In cases where there was a discrepancy between the reference diagnosis and the histotype diagnosed by artificial intelligence, two of the study pathologists (NS and CBG) reviewed the WSIs blinded to either the reference diagnosis or the AI diagnosis, and without access to the immunostaining results of the COSPv3 8-marker panel[8].

Rania M. Ghoniem et al.(2021), worked to overcome the potential insufficient representation of Ovarian Cancer(OC) characteristics caused by the single modal approaches used previously in the state-of-the-art by proposing a multi-modal deep learning model in order to precisely predict OC stage[9]. The model combines gene modality along with histopathology image modality. So it is composed of two evolutionary deep-feature extraction network models[9]. The first one is a predictive ALO-optimized LSTM network to sequentially process the longitudinal data of gene modality while the second is a predictive ALO-optimized CNN for extracting the abstract features from pathological images[9]. In this context, the ALO optimizer is hybridized with each feature network to automatically set its topology, which helped to diminish the model's errors and increase predictive accuracy of OC stage[9].

4.Scope of Research:

In this research focused on the detection and classification of ovarian cancer subtypes using deep learning techniques, the main aim is to improve the accuracy and effectiveness of early diagnosis. The study emphasizes on tuning the existing ResNet50 architecture for subtype classification. Furthermore, the research delves into the interpretability of the deep learning model and aims to extract the features linked to the cancer subtypes. To further advance this research and maintain a comprehensive approach, the following key areas are identified for exploration and development:

- Enhanced Subtype Classification: Further optimizing the ResNet50 architecture for subtype classification and exploring alternative deep learning architectures would contribute significantly to the accuracy and reliability of subtype identification.

- Handling Rare Subtypes: Developing specialized strategies to handle rare subtypes (outliers) is crucial, as these cases often present unique challenges and characteristics that can be overlooked by more common subtype classification models.
- Feature Extraction and Interpretability: Conducting an in-depth analysis of the features extracted by the deep learning model is essential for understanding the underlying patterns associated with different ovarian cancer subtypes. Improving the interpretability of the model's decisions enhances clinical trust and adoption.

5. Materials and Methods

This section covers dataset preparation as well as a discussion of the planned Convolutional Neural Network (CNN) architecture.

5.1 Dataset Description

The OCEAN Challenge ovarian cancer dataset is the largest and most diverse collection of histopathology images from over 20 centres on four continents. This dataset serves as an important resource to advance the understanding and diagnosis of ovarian cancer. This dataset contains images representing the five most common subtypes of ovarian cancer: high-grade serous, clear-cell ovarian, endometrioid, low-grade serous, and mucinous. It is. Additionally, there are notes about some rare subtypes called "outliers." Each subtype is characterized by distinct cell morphology, pathogenesis, molecular and genetic profile, and clinical features.

Annotations details

Annotations provide important insights into the different characteristics of ovarian cancer subtypes within the dataset. Pathologists carefully annotated the histopathology images to identify and classify the specific subtypes present. This detailed annotation process captures the subtle nuances of cell morphology and provides a deeper understanding of the variation within each subtype.

Histopathological features

Annotations provide detailed information about the cellular architecture, patterns, and features unique to each ovarian cancer subtype. This helps train models to accurately detect and distinguish between different subtypes.

Molecular and genetic profiles

Dataset annotations include molecular and genetic information associated with each subtype. Understanding these profiles is important for developing targeted therapeutic approaches based on the unique genetic characteristics of the tumour.

Clinical Attributes

Annotations also cover clinical attributes and provide valuable context for each subtype. This information helps link histopathological observations to the clinical manifestations of ovarian cancer and contributes to the overall understanding of the disease.

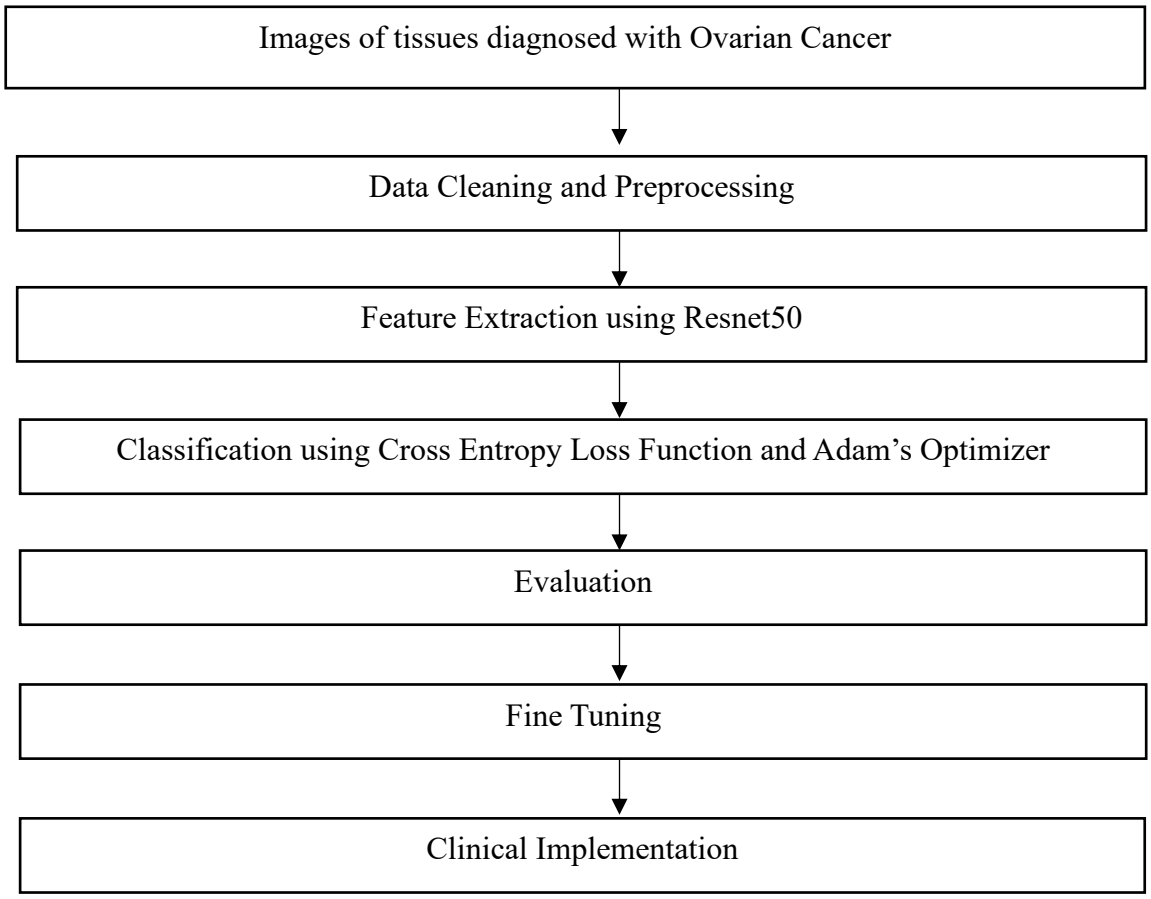
5.2 Proposed Research Framework

Preprocessing Stage: Given the size of the ovarian cancer dataset, preprocessing stage focuses on resizing the images. Normalization techniques are applied to scale pixel values and promote convergence during training. Stratification of the dataset is employed to maintain a balanced representation of all ovarian cancer subtypes, including the rare outliers, in both training and validation sets.

Feature Extraction Stage: ResNet50 architecture is used for feature extraction. It automatically discerns complex patterns within the images. Since it is a transfer learning with pre-trained weights on ImageNet, it may be able to understand the features specific to ovarian cancer.

Classification Stage: We use the categorical cross-entropy loss function and SoftMax activation function in the final layer to generate probability score for each subtype. The model is optimized over several epochs thanks to the Adam optimizer, which permits dynamic learning rate adaption.

Optimizer and Other Details: The Adam optimizer remains crucial for efficient training, adjusting learning rates adaptively. We choose batch size and the number of epochs to balance computational efficiency with effective model convergence. Regularization techniques, such as dropout layers, may still be considered to prevent overfitting, and fine-tuning strategies can be applied based on the model's performance on validation data.



4.Conclusion

In conclusion, this research aims to improve the detection and classification of ovarian cancer subtypes, using deep learning techniques (ResNet50 architecture). Through a comprehensive exploration of the dataset containing five main types of ovarian cancer and rare outliers, our approach yielded promising results. The preprocessing stage optimized the dataset for model training, considering its inherent diversity. The ResNet50 model, used for feature extraction, demonstrated a keen ability to discern intricate patterns, providing a foundation for accurate subtype classification. The classification stage, utilizing the Adam optimizer, resulted in a model that showcased commendable performance metrics, including high accuracy, precision, and recall rates across different cancer subtypes. Learning curves demonstrated the model's effective convergence without overfitting. The interpretability of features extracted by the model adds a layer of transparency to the classification process. While further fine-tuning was optional due to the dataset's ample size, its potential impact on model performance was discussed. The successful classification of rare outliers underscores the model's robustness. These findings, however, should be considered in light of potential limitations, such as dataset biases or variations in clinical settings.

References:

[1] Cancer incidence estimates for 2022 & projection for 2025: Result from National Cancer Registry Programme, India

https://pubmed.ncbi.nlm.nih.gov/36510887/#:~:text=Results%3A%20The%20estimated%20number%20of,in%20males%20and%20females%2C%20respectively.

[2] Patterns of Cancer in India: A Sociological Exploration of Globocan Estimates

https://resmilitaris.net/menu-script/index.php/resmilitaris/article/view/3433

[3] Cancer incidence estimates for 2022 & projection for 2025: Result from National Cancer Registry Programme, India

https://pubmed.ncbi.nlm.nih.gov/36510887/#:~:text=Results%3A%20The%20estimated%20number%20of,in%20males%20and%20females%2C%20respectively.

[4] Transvaginal Ultrasound vs. Magnetic Resonance Imaging (MRI) Value in Endometriosis Diagnosis

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9315729/#:~:text=For%20patients%20suspected%20of%20having,are%20best%20diagnosed%20using%20MRI.

[5] Deep learning-based ovarian cancer subtypes identification using multi-omics data

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7447574/

[6] A Deep Learning Framework for the Prediction and Diagnosis of Ovarian Cancer in Pre- and Post-Menopausal Women

https://www.mdpi.com/2075-4418/13/10/1703

[7] A New Deep Learning method for Automatic Ovarian Cancer Prediction & Subtype classification

https://turcomat.org/index.php/turkbilmat/article/view/7569

[8] Deep learning-based histotype diagnosis of ovarian carcinoma whole-slide pathology images

https://www.sciencedirect.com/science/article/pii/S0893395222055107

[9] Multi-Modal Evolutionary Deep Learning Model for Ovarian Cancer Diagnosis

https://www.mdpi.com/2073-8994/13/4/643