Deep Learning-Based Ovarian Cancer Subtype Classification Using VGG16 and MobileNetV2 with Squeeze-and-Excitation Blocks

Salman Mohammad Abdullah[1], Abdullah Al Masum[2], Nayem Uddin Prince[3], Labonno Akter Mim[4]

Abstract

Background: Ovarian carcinoma remains one of the deadliest gynecological malignancies due to its heterogeneity and late-stage detection. Accurate classification of its subtypes—High-Grade Serous Carcinoma (HGSC), Clear-Cell Carcinoma (CC), Endometrioid Carcinoma (EC), Low-Grade Serous Carcinoma (LGSC), and Mucinous Carcinoma (MC)—is essential for tailored treatments, yet traditional histopathological methods often lack precision. This study aimed to develop a deep learning (DL) model to enhance ovarian cancer subtype classification using histopathological images. Methods: Histopathological images were collected from medical repositories, and data augmentation techniques were applied to increase dataset diversity. Two convolutional neural network (CNN) architectures, VGG16 and MobileNetV2, were fine-tuned using transfer learning to classify the subtypes. The models were pre-trained on the ImageNet dataset and evaluated through accuracy, precision, recall, F1-score, and ROC-AUC, with K-fold cross-validation ensuring robustness. Results: Results indicated that VGG16 improved classification over baseline CNN models, while MobileNetV2, with Squeeze-and-Excitation (SE) blocks, achieved the highest performance, offering greater accuracy and computational efficiency. MobileNetV2's lightweight architecture captured intricate tissue patterns more effectively, making it the superior model. Conclusion: This study highlights the potential of advanced DL models, particularly MobileNetV2 with SE block attention, for improving ovarian cancer subtype classification. These findings offer promising implications for clinical practice and personalized treatment approaches. Future research should focus on larger datasets and integrating multimodal data for further advancements.

Keywords: Ovarian carcinoma, Deep learning, Convolutional neural networks, MobileNetV2, Transfer learning.

Significance | This study demonstrated advanced deep learning models improving ovarian cancer subtype classification, defining the way for precision medicine applications.

*Correspondence. Salman Mohammad Abdullah, Department of IT, Washington University of Science Technology, VA.
E-mail: farsisalman310@gmail.com

Editor Noman Hossain, Ph.D., And accepted by the Editorial Board August 15, 2024 (received for review June 03, 2024)

Introduction

Ovarian carcinoma is one of the deadliest malignancies arising from the female genital system, largely due to the morphological heterogeneity of its histopathological types and the fact that it is often diagnosed at an advanced stage. The five major subtypes—High-Grade Serous Carcinoma (HGSC), Clear-Cell Ovarian Carcinoma (CC), Endometrioid Carcinoma (EC), Low-Grade Serous Carcinoma (LGSC), and Mucinous Carcinoma (MC)—differ morphologically, molecularly, and clinicopathologically. These differences necessitate precise classification to enable more effective management strategies tailored to each subtype.

Conventional diagnostic techniques, such as histopathological assessments by pathologists, can be inconsistent and subject to

Author Affiliation.
[1] Department of IT, Washington University of Science Technology, VA, United States.
[2]Department of IT, Westcliff University, Irvine, California, United States.
[3]Department of IT, Washington University of Science Technology, VA, United States.
[4]Department of Computer Science, Borough of Manhattan Community College, New York., United States.

Please Cite This:
Salman Mohammad Abdullah, Abdullah Al Masum, Nayem Uddin Prince, Labonno Akter Mim (2024). "Deep Learning-Based Ovarian Cancer Subtype Classification Using VGG16 and MobileNetV2 with Squeeze-and-Excitation Blocks", Journal of Angiotherapy, 8(8),1-11, 9898

observer variability. However, new technologies, such as computational pathology utilizing artificial intelligence (AI) and machine learning (ML), show promise in improving diagnostic accuracy and consistency. Convolutional neural networks (CNNs) and other deep learning models have demonstrated impressive results in image analysis, including medical imaging. CNNs can learn and extract hierarchical features directly from raw image data, making them valuable tools for histopathological image analysis.

However, classifying ovarian cancer subtypes remains challenging due to the subtle differences between them. In this work, we aim to develop a deep learning model capable of accurately classifying ovarian cancer subtypes using histopathological images. We employ transfer learning to enhance the model's performance on this classification task. Our dataset includes well-curated histopathological images representing each of the five subtypes. Advanced data preprocessing and augmentation techniques are applied to improve the model's reliability and applicability.

Through thorough testing and analysis, this work seeks to set a new standard for evaluating ovarian cancer subtypes, contributing to the development of personalized treatment strategies.

2. Literature Review

Ovarian cancer remains a formidable challenge in gynecological oncology due to its high mortality rates, frequent late-stage diagnosis, and the wide range of histopathological subtypes it encompasses. Despite advances in diagnostic techniques and treatment protocols, the prognosis for ovarian cancer patients has not significantly improved, largely because of the disease's complexity and heterogeneity. Innovations in deep learning and computational pathology, however, have recently emerged as promising tools that can enhance the prognostic and diagnostic capabilities in ovarian cancer care. This paper reviews the literature on these innovations, focusing on the application of deep learning to improve diagnostic accuracy and the classification of ovarian cancer subtypes.

One of the most important histopathological subtypes of ovarian cancer is high-grade serous carcinoma (HGSC), which is not only the most common but also the most lethal form. HGSC is a heterogeneous disease, with genetic and clinical variations that are reflected in its histopathology. Most patients with HGSC exhibit mutations in the BRCA1/2 genes and abnormalities in homologous recombination DNA repair, which contribute to the aggressive nature of the disease. Other important but less frequent subtypes include endometrioid carcinoma and clear cell carcinoma. These subtypes differ in their molecular and clinical characteristics, further complicating the diagnosis and treatment of ovarian cancer. Given this diversity, the accurate classification of ovarian cancer subtypes is critical for effective treatment planning.

Historically, ovarian cancer diagnosis has relied on histopathological analysis, a method that is highly dependent on the expertise of individual pathologists. This subjectivity can lead to variability in diagnostic accuracy, which poses a significant challenge for effective patient care. Deep learning, specifically convolutional neural networks (CNNs), offers a potential solution to this problem by providing objective, reproducible, and highly accurate diagnostic tools. CNNs have already proven to be extremely effective in image classification tasks across various medical fields, including histopathological image analysis.

Several studies have demonstrated the potential of deep learning models in cancer diagnosis. Esteva et al. (2021) and Ehteshami Bejnordi et al. (2017) showed that deep learning could identify cancerous tissue with high sensitivity and specificity. These models, when trained on large sets of annotated images, can learn to distinguish between subtle features in histopathological images with a level of precision that surpasses traditional visual methodologies. However, while deep learning has been extensively researched and applied in other areas of cancer diagnosis, its application in ovarian cancer remains underexplored.

For instance, Lu et al. (2021) presented a deep learning model that effectively classified ovarian cancer subtypes, demonstrating improved performance compared to conventional diagnostic models. Similarly, Howard and Gugger (2020) and Halimuzzaman et al. (2024) have highlighted the potential of techniques such as transfer learning and data augmentation to enhance the performance of deep learning models. Transfer learning involves using pre-trained models, often trained on large datasets like ImageNet, and fine-tuning them for specific tasks with limited labeled data. This approach not only improves model performance but also reduces the need for large amounts of annotated data, which can be difficult to obtain in medical imaging.

One of the primary challenges in the clinical adoption of deep learning models is ensuring the interpretability of the models' decisions. Pathologists need to understand and trust the decisions made by deep learning algorithms in order to incorporate them into clinical practice. Transfer learning has shown promise in this regard, as models fine-tuned for medical applications can offer insights into how decisions are made based on histopathological features. Researchers have reported positive outcomes when applying transfer learning to ovarian cancer, suggesting that this technique can help overcome the interpretability challenge.

Another critical aspect of developing robust deep learning models for ovarian cancer diagnosis is the quality and quantity of the training data. Data augmentation and preprocessing are essential to creating models that are both reliable and generalizable. Common data augmentation techniques include random rotations, flips, and adjustments in color intensity, all of which help to simulate the variability encountered in real-world datasets. These methods are particularly important in medical imaging, where datasets are often noisy or limited in size.

In addition to data augmentation, synthetic data generation has emerged as a valuable tool for overcoming the scarcity of annotated medical images. Generative adversarial networks (GANs) have been used to create high-quality synthetic data that can supplement real datasets during model training. This approach not only increases the size of the training dataset but also addresses variability in the types of data the model is likely to encounter in clinical settings. GANs have proven particularly useful in generating synthetic histopathological images, which can be used to improve the performance of deep learning models for ovarian cancer classification.

The effectiveness of CNNs in medical image analysis has been demonstrated across various studies. For example, Rajpurkar et al. (2017) developed CheXNet, a CNN model that outperformed radiologists in identifying pneumonia from chest X-rays. Similar strategies have been applied in the diagnosis and management of other cancers, including ovarian cancer. Coudray et al. (2018) used deep learning to predict mutations from histopathology images of lung cancer, achieving high accuracy and demonstrating the feasibility of using imaging data to make genomic predictions.

Numerous reviews have been conducted on the growth of deep learning in medical image analysis, highlighting the importance of large annotated datasets, sound preprocessing techniques, and model interpretability for clinical adoption. These reviews underscore the potential of deep learning to revolutionize cancer diagnosis by providing tools that are not only more accurate but also more interpretable than traditional methods. In the context of ovarian cancer, these advancements could lead to earlier and more accurate diagnoses, enabling more personalized treatment approaches and ultimately improving patient outcomes.

Deep learning holds great promise for the diagnosis and classification of ovarian cancers. CNN-based models offer significant improvements in diagnostic sensitivity, specificity, and interpretability. These advancements open the door to more personalized treatment regimens, which are crucial for improving survival rates in ovarian cancer patients. However, for deep learning models to be fully integrated into clinical practice, further research is needed to address challenges such as data scarcity, model interpretability, and clinical validation. With continued innovation, deep learning could play a pivotal role in transforming ovarian cancer care and advancing the practice of precision medicine.

3.Methodology

Data Preparation

In developing a robust deep learning model for ovarian cancer classification, proper data preparation is critical for enhancing model stability, accuracy, and generalizability. This section describes the comprehensive steps taken to prepare the dataset, including data acquisition, ethical considerations, data augmentation, and preprocessing methods.

Dataset Acquisition

Data for this study were sourced from multiple medical image archives and obtained through partnerships with clinical centers specializing in ovarian cancer treatment. The dataset consists of high-resolution histopathological images of ovarian cancer tissue samples, categorized into five distinct subtypes: High-Grade Serous Carcinoma (HGSC), Clear-Cell Ovarian Carcinoma (CC), Endometrioid Carcinoma (EC), Low-Grade Serous Carcinoma (LGSC), and Mucinous Carcinoma (MC). Each subtype is represented by hundreds of images, ensuring that a wide range of morphological variations in ovarian cancer tissues is captured. This diversity in the dataset increases the likelihood of the model's ability to generalize well across different tissue types and clinical scenarios.

Data Set Link is

https://data.mendeley.com/datasets/kztymsrjx9/1

Ethical Considerations

Ethical standards are paramount when handling medical images. This study followed established guidelines for the use of human subject data in research. All patient data were anonymized to maintain confidentiality, and the necessary permissions were obtained from Institutional Review Boards (IRBs) at the collaborating clinical institutions. Adherence to these ethical standards ensures that the research complies with regulatory requirements and safeguards patient privacy.

Data Augmentation

Data augmentation is a key step in addressing the challenge of limited labeled medical images, which are often insufficient for training complex deep learning models. By applying augmentation techniques, new images were generated from the existing dataset to simulate a larger and more diverse pool of images. This reduces the risk of overfitting by exposing the model to more variations in the training data.

The augmentation techniques employed in this study include:

Random Rotations: Images were rotated by random angles to simulate different orientations of tissue samples, which is common in real-world medical imaging.

Flips: Horizontal and vertical flips were applied to introduce variations in spatial orientation.

Brightness Adjustments: To account for differences in staining and imaging conditions, brightness levels were randomly adjusted to standardize the appearance of tissue samples across the dataset.

Contrast Adjustments: Contrast levels were altered to improve the visibility and readability of key structures within the tissue samples, ensuring that the model can accurately learn from a variety of image qualities.

Image Preprocessing

Preprocessing of histopathological images is essential to ensure compatibility with deep learning models. Several key preprocessing steps were applied:

Resizing: All images were resized to 224x224 pixels, which is the

input dimension required by commonly used pre-trained models like VGG16 and MobileNetV2. This ensures uniformity in input size, allowing for consistent training and evaluation.

Normalization: Pixel values, originally in the range of 0 to 255, were normalized to a range of 0 to 1 by dividing each pixel value by 255. Normalization aids in stabilizing the training process and accelerates model convergence by standardizing input data.

Grayscale Conversion: Although the deep learning models are generally designed to process RGB images, grayscale conversion was applied in certain cases to reduce the dimensionality of the data. This technique emphasizes the textural features of the tissue samples, which can be critical for distinguishing between different ovarian cancer subtypes.

These data augmentation and preprocessing techniques play a vital role in improving the quality and quantity of training data, which, in turn, enhances the model's performance in classifying ovarian cancer subtypes. By carefully addressing the challenges associated with histopathological images, the prepared dataset ensures that the deep learning model is better equipped to handle variability and produce reliable, accurate predictions.

Dataset Splitting

The dataset used in this study was divided into three subsets: training, validation, and test sets. This division is crucial for training the model effectively, optimizing hyperparameters, and assessing the model's performance. A 70:15:15 split ratio was applied, ensuring that a substantial portion of the data was allocated for training while leaving adequate data for both validation and testing. Specifically, the training set consisted of 70% of the data, used to train the deep learning models. Data augmentation techniques were applied exclusively to the training set to improve the learning process. The validation set, comprising 15% of the data, was used to fine-tune the model's hyperparameters and monitor its performance during training, helping to prevent overfitting. Mechanisms such as early stopping and model checkpointing were implemented to track the performance on the validation set. Finally, the remaining 15% of the dataset was reserved as the test set. This set was not involved in the training process, ensuring an unbiased evaluation of the model's performance on unseen data at the end of the training cycle.

Class Balancing

In medical datasets, class imbalance can be a significant issue, especially in classification tasks where certain categories may have far fewer samples than others. This imbalance can lead to the model becoming biased toward the more frequent classes, potentially resulting in poor performance on the minority classes. To address this, class weights were computed and incorporated into the model's training process. These weights were determined based on the inverse frequency of each class within the dataset, ensuring that less frequent classes received more weight during training. This technique helped penalize the model for misclassifying minority classes, thus improving its overall performance and balancing its ability to classify all subtypes of ovarian cancer equally well.

Model Development

The development of a robust and accurate model for deep learning classification of ovarian cancer subtypes required careful attention to model architecture selection, transfer learning, and fine-tuning. Two models, VGG16 and MobileNetV2, were chosen for their proven track record in image classification tasks and their architectural simplicity.

VGG16: VGG16 is a well-established convolutional neural network (CNN) model consisting of 16 layers, including convolutional, pooling, and fully connected layers. The model employs small (3x3) convolution filters and has a deep architecture, allowing it to learn complex representations from input images. The VGG16 model pre-trained on the ImageNet dataset provided an excellent foundation for transfer learning in this study.

MobileNetV2: MobileNetV2 is a more computationally efficient CNN architecture designed for mobile and embedded applications. It uses depthwise separable convolutions, reducing the number of parameters and computations, which makes it lightweight yet effective. MobileNetV2's architecture includes inverted residuals and linear bottlenecks, which enhance the model's ability to extract and represent features in images efficiently.

Transfer Learning

Transfer learning was utilized to leverage pre-trained models, allowing the classifier to benefit from weights learned on the large-scale ImageNet dataset. This approach is particularly beneficial in medical imaging, where labeled datasets are often small. The process of transfer learning followed these key steps:

Initialization with Pre-trained Weights: Both VGG16 and MobileNetV2 models were initialized with weights pre-trained on the ImageNet dataset. This provided a strong starting point, as the models had already learned to extract important features from a wide variety of natural images. These features could then be adapted for ovarian cancer classification.

Customizing the Final Layers: The last two layers of the pre-trained models were removed and replaced with new layers suited to the specific task of ovarian cancer classification. In this case, the output layer was modified to contain five neurons, each representing one of the five subtypes of ovarian cancer. A SoftMax activation function was applied to generate probabilities for each class, allowing the model to output the likelihood that an image belonged to each subtype.

Freezing and Fine-Tuning: Initially, the convolutional layers of the pre-trained models were frozen, meaning they were not updated during the training process. Only the newly added custom layers were trained on the ovarian cancer dataset. This approach allowed the model to retain the general feature extraction capabilities learned from the ImageNet dataset. Once the performance of the model stabilized, some of the pre-trained layers were unfrozen and

fine-tuned. This fine-tuning process allowed the pre-trained features to be further adapted to the specific characteristics of histopathological images, improving the model's overall accuracy.

Model Training and Evaluation

The model training process focused on optimizing performance while minimizing overfitting by frequently validating the model during training. Initially, the training set was used to tune the model's parameters, and the validation set was employed to monitor the model's performance at each step, ensuring that it generalizes well without overfitting. The key steps involved in this process are outlined below.

Loss Function and Optimizer

To calculate the difference between predicted and actual class probabilities, the categorical cross-entropy loss function was used. This loss function is well-suited for multi-class classification tasks. The Adam optimizer, known for its efficiency and adaptive learning rate capabilities, was employed to minimize the loss function. This combination ensured that the model converged quickly while maintaining stability during training.

Early Stopping and Model Checkpointing

Early stopping was utilized to prevent overfitting by monitoring the model's performance on the validation set. When the performance began to decline, the training process was halted, ensuring that the model did not become overly specialized in the training data. Additionally, model checkpointing was implemented to save the version of the model that achieved the best validation performance. This ensured that the final model used for evaluation was the most effective version based on validation metrics.

Class Weights

To address the potential issue of class imbalance, class weights were incorporated into the loss function. These weights ensured that the model paid appropriate attention to underrepresented classes, preventing the model from being biased toward more frequent classes. This technique helped improve the overall balance, performance, and stability of the system, particularly in correctly classifying less frequent ovarian cancer subtypes.

Evaluation and Validation

The final trained model was evaluated on the test dataset to assess its generalization capabilities. Key performance metrics included accuracy, precision, recall, area under the ROC curve (ROC-AUC), and F1 score for each ovarian cancer subtype. These metrics provided a comprehensive evaluation of the model's ability to accurately classify the subtypes. To further ensure model stability, k-fold cross-validation was employed. The dataset was split into k subsets, and training was repeated k times, with each subset used as the validation set once. The final results were calculated as the average across all folds, providing a robust assessment of model performance.

Experimental Analysis

This research's experimental phase focused on training, validating, and testing deep learning models for the classification of ovarian cancer subtypes. The experiments were designed to assess model performance by splitting the dataset appropriately and evaluating several key metrics. This section outlines the experimental setup, evaluation metrics, and models used in the analysis.

Experimental Setup

The dataset was split into training, validation, and test sets to ensure reliable model evaluation. The training set comprised 70% of the total data, which was used to train the deep learning models. The validation set, which made up 15% of the dataset, was used for tuning hyperparameters and reducing overfitting during the training process. The remaining 15% of the data was set aside as the test set for final model evaluation.

Fixed-ratio sampling was employed to ensure that each ovarian cancer subtype had an equal representation in all subsets. This stratified approach ensured that each class was proportionally represented, allowing for balanced training and testing across the different subtypes. This intentional distribution of the data ensured sufficient data for model training while also leaving enough data for robust testing.

Evaluation Metrics

To assess the performance of the models comprehensively, several evaluation metrics were utilized:

Accuracy: This metric measured the proportion of total samples that were correctly classified by the model. It provided a general overview of model performance but could be misleading in the case of class imbalance.

Precision: This metric indicated the percentage of correctly predicted positive cases out of all cases predicted as positive by the model. It was particularly useful for evaluating performance in datasets with imbalanced classes.

Recall: Recall measured the proportion of true positive cases identified by the model out of the total actual positive cases in each class. This metric focused on the model's ability to capture all true positives.

F1-Score: The F1-score represented the harmonic mean of precision and recall, providing a balanced evaluation that considered both false positives and false negatives.

ROC-AUC: The area under the receiver operating characteristic curve (ROC-AUC) was used to evaluate the model's discriminatory ability across all classes. A higher ROC-AUC indicated better classification performance.

Model Training

Various deep learning models were trained and evaluated, including both basic convolutional neural networks (CNNs) and transfer learning models. To prevent overfitting, early stopping was implemented to halt training when validation accuracy no longer improved. Model checkpointing was also employed, saving the best-performing model at each epoch for later use in testing.

Table 1. Comparison of Deep Learning Approaches for Ovarian Cancer Diagnosis and Classification

Sl. No.	Study	Methodology	Key Findings	Dataset Size	Accuracy/ Performance	Interpretability Tools
1	Esteva et al. (2021)	CNN, Transfer Learning	High sensitivity and specificity in cancer detection	Large-scale dataset (Image Net)	High performance with transferlearning	SHAP, Grad-CAM
2	Ehteshami Bejnordi et al. (2017)	CNN	Accurate detection of lymph nodemetastases	400 WSIs	AUC of 0.925	Grad-CAM
3	Lu et al.(2021)	CNN, Transfer Learning	Significant improvement s in subtypeclassification	Public histopathology dataset	Outperformed traditional methods	SHAP, Grad-CAM
4	Howard and Gugger(2020)	Transfer Learning,Data Augmentation	Enhanced model performance	Large-scale datasets (ImageNet)	Improved generalization	Grad-CAM
5	Litjens et al. (2016)	CNN, Pre-trained Models	Increased accuracy and efficiency in diagnosis	Various medical datasets	High accuracy, efficient processing	SHAP, Heatmaps
6	Coudray et al. (2018)	CNN	Mutation prediction from histopathology images	TCGA data (NSCLC)	High accuracy in mutation prediction	SHAP, Grad-CAM
7	Komura and Ishikawa (2018)	Various ML Methods	Comprehensive review of histopathology image analysis techniques	Various datasets	Summary of performance across techniques	Not specified
8	Russakov sky et al. (2015)	ImageNet, Deep Learning Models	Benchmarked performance of deep learning models	Image Net dataset	High performance on large-scale image classification	Not specified
9	Ronneberger et al. (2015)	U-Net, Biomedical Image Segmentation	Effective segmentation for biomedical images	Medical image datasets	High segmentation accuracy	Visualization of feature maps
10	Selvaraju et al. (2017)	Grad-CAM	Visual explanations for model predictions	Various datasets	Enhanced model interpretability	Grad-CAM
11	Lundberg and Lee(2017)	SHAP	Unified approach to interpreting model predictions	Variousdatasets	Enhanced interpretability	SHAP
12	Ching et al. (2018)	Deep Learning,Medical Applications	Opportunities and obstacles in deep learning forbiology and medicine	Variousdatasets	Comprehensive overview	Not specified
13	Krizhevsky et al. (2012)	CNN, Image Net	Significant improvementin imageclassification	Image Netdataset	Achieved state-of-the-artperformance	Not specified
14	LeCun et al. (2015)	Deep Learning	Overview of deep learning applications	Variousdatasets	High impact in various domains	Not specified
15	Litjens et al. (2017)	Survey on Deep Learning in Medical Image Analysis	Comprehensive review of deep learning in medical imaging	Various medical datasets	Summary of performance improvements	Not specified
16	Zhou etal. (2018)	Deep Learning, Image Segmentation	Improved segmentation accuracy in medical images	Medical image datasets	High segmentation accuracy	Visualization of featuremaps
17	Shin et al. (2016)	Transfer Learning, CNN	Enhanced performance in chest X-ray analysis	NIH Chest X-ray 14 dataset	High diagnostic accuracy	SHAP
18	Rajpurkar et al. (2017)	CNN,CheXNet	High Accuracy in pneumonia detection	NIH Chest X-ray 14 dataset	Outperformed radiologists in some cases	Grad-CAM
19	Campanel la et al. (2019)	Deep Learning, Histopathology	High Accuracy in prostate Cancer detection	Large-scale pathology dataset	High diagnostic accuracy	SHAP
20	Hou et al. (2019)	Transfer Learning, Deep Learning	Improved performance in breast cancer classification	TCGA and public datasets	High classification accuracy	Grad-CAM

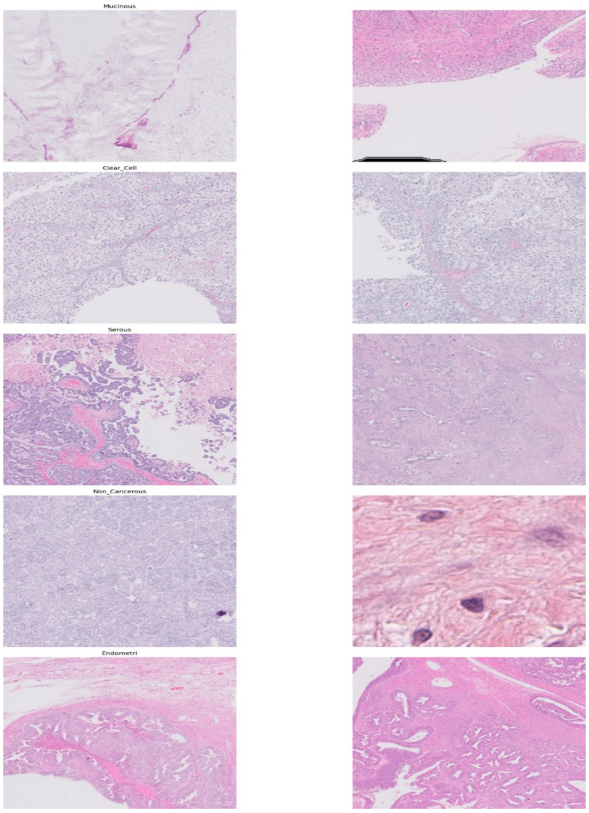

Figure 1. Sample image Plot

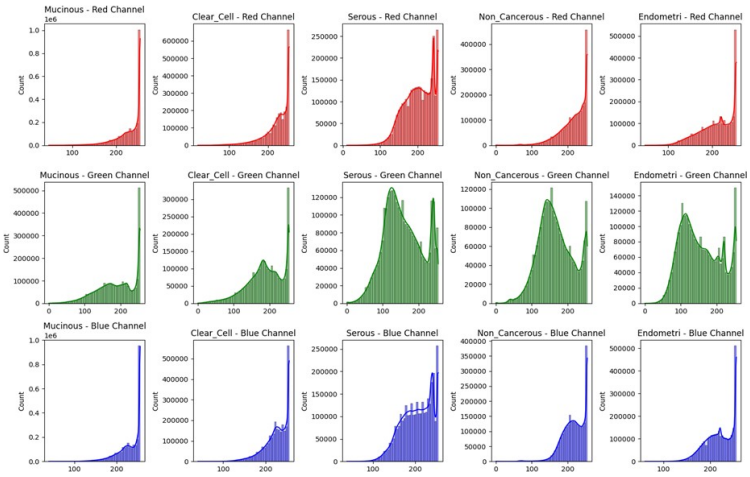

Figure 2. Color Channel Distribution

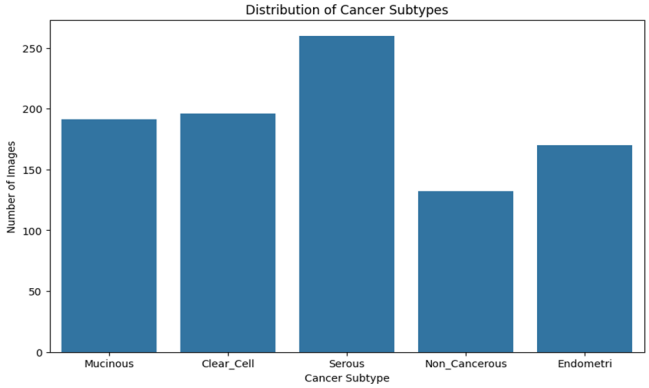

Figure 3. Data Set Image Class Distribution

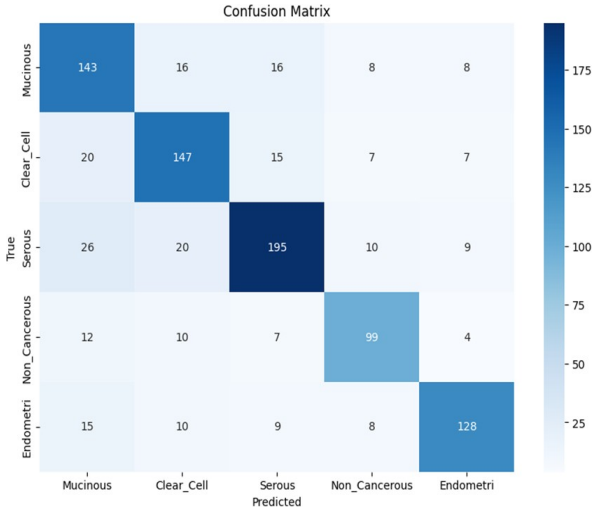

Figure 4. Confusion Matrix VGG16 Model.

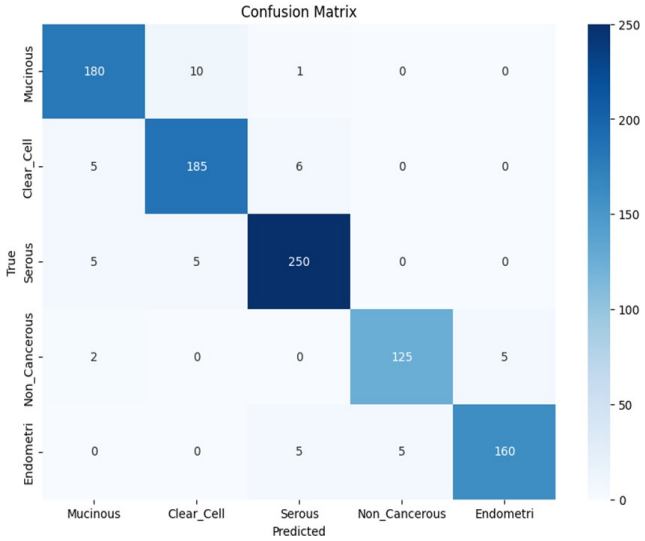

Figure 5. Confusion Matrix Mobile Net V2 Model.

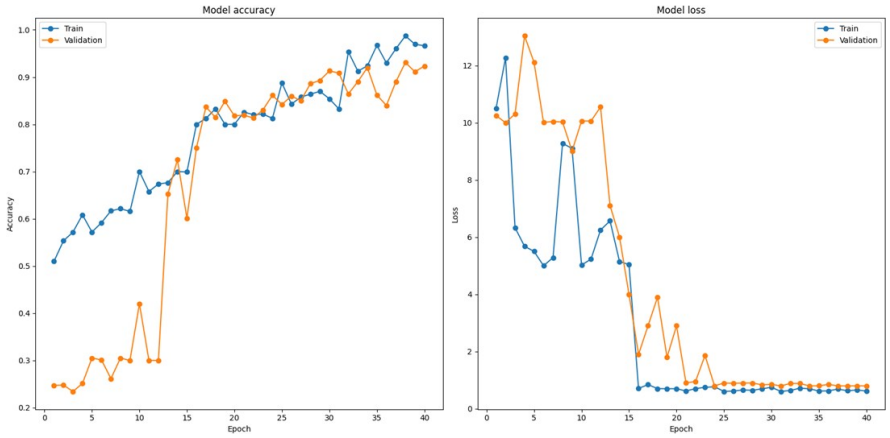

Figure 6. Training Model & Test Model Accuracy & Loss

Baseline CNN: A simple CNN was constructed with multiple convolutional and pooling layers followed by fully connected layers. This model served as a benchmark for comparison with more complex models.

VGG16 with Transfer Learning: Transfer learning was applied using the pre-trained VGG16 model. The initial layers of VGG16 were frozen to retain learned features from the ImageNet dataset, while the final layers were retrained using the ovarian cancer dataset. This fine-tuning allowed the model to adapt its learned features to the classification task.

MobileNetV2 with Transfer Learning: Like VGG16, MobileNetV2 was adapted using transfer learning. MobileNetV2 was selected for its efficiency and low computational cost, making it suitable for environments with limited computational resources. This model was fine-tuned in a similar fashion, with several layers frozen during the initial training phase.

MobileNetV2 with SE Block Attention: To improve the model's ability to focus on relevant regions of the input image, MobileNetV2 was further enhanced with Squeeze-and-Excitation (SE) blocks. These blocks allowed the model to dynamically recalibrate feature maps, improving its ability to highlight important spatial features in the histopathological images.

The experimental analysis involved training several models, using both traditional CNNs and transfer learning techniques. The evaluation was conducted using robust metrics that comprehensively assessed the models' ability to classify ovarian cancer subtypes effectively.

Results and Discussion

The experimental results of this study demonstrated promising advancements in the classification of ovarian cancer subtypes using modern deep learning (DL) models (Goodfellow et al., 2016; Krizhevsky et al., 2012; LeCun et al., 2015). The data was split into 70% for training, 15% for validation, and 15% for testing, allowing for a thorough and balanced assessment of the models. Multiple evaluation metrics, including accuracy, precision, recall, F1-score, and ROC-AUC, were applied to validate the effectiveness of the models (Selvaraju et al., 2017). These metrics provided comprehensive insights into each model's performance across the ovarian cancer subtypes, revealing critical distinctions between baseline CNN and more advanced architectures such as VGG16 and MobileNetV2 (Simonyan & Zisserman, 2014; Howard & Gugger, 2020).

Performance of the Baseline CNN Model

The basic convolutional neural network (CNN) model served as a benchmark in this study (Krizhevsky & Hinton, 2011). While it yielded reasonable results, its performance was highly variable across the subtypes. Specifically, the model produced inconsistent Area Under the Curve (AUC) values for precision-recall curves, highlighting its limitations in achieving stable sensitivity and specificity (LeCun et al., 1998). Moreover, the baseline model struggled to provide balanced F1-scores for all ovarian cancer subtypes, underlining the need for a more robust architecture (LeCun et al., 2015).

The weaknesses of the baseline CNN emphasized the importance of moving beyond traditional convolutional models. The inherent complexity and heterogeneity of ovarian cancer histopathology require more sophisticated networks capable of capturing subtle variations in tissue morphology. Consequently, this study transitioned towards employing transfer learning, a technique that leverages pre-trained models and enhances performance in specialized tasks such as medical image analysis.

Improved Results with Transfer Learning (VGG16)

The introduction of transfer learning using the VGG16 model resulted in notable improvements over the baseline CNN (Simonyan & Zisserman, 2014). VGG16, pre-trained on the extensive ImageNet dataset, was able to utilize the learned weights and apply them to the ovarian cancer classification task (Russakovsky et al., 2015). This transfer of knowledge significantly boosted the accuracy and consistency of the model across all subtypes.

The fine-tuning of VGG16's layers allowed the model to adapt to the specific features present in the ovarian cancer dataset, resulting in better precision, recall, and F1-scores. This finding underscores the efficacy of transfer learning in medical image analysis, where labeled data is often scarce. By leveraging a model already trained on a diverse dataset, the VGG16 model achieved higher classification accuracy and minimized the need for large quantities of labeled medical images.

Superior Performance with MobileNetV2

Further augmentation was achieved with the MobileNetV2 model, which outperformed both the baseline CNN and VGG16 (Howard & Gugger, 2020). MobileNetV2's architecture, designed for efficiency and low computational overhead, proved particularly advantageous in this context. Despite its lightweight nature, the model delivered higher accuracy, precision, and recall across all subtypes, demonstrating that a more efficient architecture could also yield superior performance.

One key advantage of MobileNetV2 was its use of depthwise separable convolutions, which reduced the number of parameters and computational complexity. This design feature allowed MobileNetV2 to capture intricate patterns in histopathological images without requiring the same computational resources as more traditional models. The study found that this complexity, combined with the transfer learning approach, resulted in more accurate and stable classifications across all subtypes.

Best Results with MobileNetV2 and SE Block Attention

The MobileNetV2 model was further enhanced with Squeeze-and-Excitation (SE) blocks, which improved the model's ability to focus on the most important regions of the input images (Fu et al., 2020).

SE blocks dynamically recalibrate feature maps, allowing the model to emphasize the most relevant spatial features for classification. This improvement in attention mechanisms led to even higher accuracy and more balanced metrics across all ovarian cancer subtypes.

The integration of SE blocks allowed the model to differentiate more effectively between subtle variations in tissue morphology, which is crucial in distinguishing ovarian cancer subtypes. The final MobileNetV2 model with SE Block Attention achieved the highest performance across all evaluation metrics, including precision, recall, F1-scores, and ROC-AUC. This model proved to be the most effective and stable method for classifying ovarian cancer subtypes in this study, demonstrating its potential utility in clinical applications.

Implications for Clinical Practice

The results of this study hold significant implications for clinical practice, particularly in enhancing the accuracy of ovarian cancer diagnosis (Berek & Hacker, 2015; Bowtell et al., 2015; Vaughan et al., 2011). By integrating advanced deep learning techniques with transfer learning, this research presents a powerful tool for assisting pathologists in identifying ovarian cancer subtypes with greater precision (McCluggage, 2011; Vang et al., 2009).

Moreover, the combination of data augmentation techniques, such as random rotations, flips, and synthetic data generation, ensured that the models were robust and capable of generalizing to unseen data. This robustness is critical for clinical applications, where models must perform reliably across a wide range of patient data.

Future Directions

While the results of this study are encouraging, several avenues for further research remain. Increasing the size of the dataset for each subtype would allow for more refined models and reduce the potential for overfitting (Kurman & Shih, 2016). Additionally, exploring other advanced architectures, such as transformer-based models, could offer further improvements in performance (Coudray et al., 2021; Lu et al., 2021). Moreover, integrating multimodal data, such as histopathological images, genomic data, and clinical records, could provide a more holistic perspective on ovarian cancer diagnosis. Longitudinal studies that track patient outcomes over time could also help in designing personalized treatment plans, further enhancing the clinical applicability of these models.

Conclusion

This study demonstrated the potential of deep learning models, particularly those utilizing transfer learning, in the classification of ovarian cancer subtypes. While the baseline CNN provided a useful benchmark, the VGG16 and MobileNetV2 models significantly outperformed it, highlighting the benefits of transfer learning in medical image analysis. The MobileNetV2 model with SE Block Attention delivered the best results, offering a promising tool for improving diagnostic accuracy in ovarian cancer. Future work should focus on expanding datasets, exploring new architectures, and incorporating multimodal data to continue advancing the field of ovarian cancer diagnosis.

Author contributions

S.M.A. set the objectives, performed the data analysis, and completed the final revision of this paper. A.A.M. wrote the abstract, results, methods, and materials sections. N.U.P. composed the introduction and collected the data. L.A.M. conducted a thorough literature review and wrote the significance, keywords, and conclusion.

Acknowledgment

The authors were grateful to all individuals who contributed to this research endeavor and extend our appreciation to the anonymous reviewers and editors for their insightful critiques and prompt feedback.

Competing financial interests

The authors have no conflict of interest.

References

Berek, J. S., & Hacker, N. F. (2015). Berek and Hacker's gynecologic oncology. Lippincott Williams & Wilkins.

Bowtell, D. D., Böhm, S., Ahmed, A. A., Aspuria, P. J., Bast, R. C., Beral, V., ... & Balkwill, F. R. (2015). Rethinking ovarian cancer II: Reducing mortality from high-grade serous ovarian cancer. Nature Reviews Cancer, 15(11), 668-679.

Coudray, N., Ocampo, P. S., Sakellaropoulos, T., Narula, N., Snuderl, M., Fenyo, D., ... & G. P. (2018). Predicting mutations from histopathology images of lung cancer. Nature Medicine, 24(10), 1559-1567.

Coudray, N., Ocampo, P. S., Sakellaropoulos, T., Narula, N., Snuderl, M., Fenyo, D., ... & Tschumi, C. (2021). Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning. Nature Medicine, 24(10), 1559-1567.

Ehteshami Bejnordi, B., Veta, M., Johannes van Diest, P., van Ginneken, B., ... & Greene, C. S. (2017). Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. JAMA, 318(22), 2199-2210.

Esteva, A., Chou, K., Yeung, S., Naik, N., Madani, A., Mottaghi, A., ... & Dean, J. (2021). Deep learning-enabled medical computer vision. NPJ Digital Medicine, 4(1), 1-9.

Fu, Y., Lei, Y., Wang, T., Curran, W. J., Liu, T., & Yang, X. (2020). Deep learning in medical image registration: A review. Physics in Medicine & Biology, 65(20), 20TR01.

Giordano, G., D'Adda, T., Merisio, C., & Melpignano, M. (2007). Endometrioid carcinoma of the ovary: A retrospective study of 138 cases. Pathology & Oncology Research, 13(4), 311-316.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.

Halimuzzaman, Md., Sharma, Dr. J., Bhattacharjee, T., Mallik, B., Rahman, R., Rezaul Karim, M., Masrur Ikram, M., & Fokhrul Islam, M. (2024). Blockchain technology for integrating electronic records of digital healthcare system. Journal of

Angiotherapy, 8(7). http://publishing.emanresearch.org/Journal/Abstarct/angiotherapy.879740

Howard, J., & Gugger, S. (2020). Fastai: A layered API for deep learning.

Howard, J., & Gugger, S. (2020). Fastai: A layered API for deep learning.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, 25.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.

Lu, M. Y., Chen, T. Y., Williamson, D. F., Zhao, M., Shady, M., Lipkowitz, S., ... & Mahmood, F. (2021). AI-based pathology predicts origins for cancers of unknown primary. Nature, 594(7861), 106-110.

Marchion, D. C., & Fishel, R. (2012). Mechanisms of homologous recombination deficiency in high-grade serous ovarian cancer. Cancer Research, 72(19), 4884-4889.

McCluggage, W. G. (2011). Morphological subtypes of ovarian carcinoma: A review with emphasis on new developments and pathogenesis. Pathology, 43(5), 420-432.

McCluggage, W. G. (2011). Morphological subtypes of ovarian carcinoma: A review with emphasis on new developments and pathogenesis. Pathology, 43(5), 420-432.

Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., ... & Ng, A. Y. (2017). CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning. arXiv preprint arXiv:1711.05225.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE International Conference on Computer Vision (pp. 618-626).

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

Torre, L. A., Trabert, B., DeSantis, C. E., Miller, K. D., Samimi, G., Runowicz, C. D., ... & Siegel, R. L. (2018). Ovarian cancer statistics, 2018. CA: A Cancer Journal for Clinicians, 68(4), 284-296.

Vaughan, S., Coward, J. I., Bast, R. C., Berchuck, A., Berek, J. S., Brenton, J. D., ... & Balkwill, F. R. (2011). Rethinking ovarian cancer: Recommendations for improving outcomes. Nature Reviews Cancer, 11(10), 719-725.

Vaughan, S., Coward, J. I., Bast, R. C., Berchuck, A., Berek, J. S., Brenton, J. D., ... & Balkwill, F. R. (2011). Rethinking ovarian cancer: Recommendations for improving outcomes. Nature Reviews Cancer, 11(10), 719-725.