Acta Biomed 2024; Vol. 95, N. 5: e2024176 DOI: 10.23750/abm.v95i5.16407

Original article

A robust AI-pipeline for ovarian cancer classification on histopathology images

Haitham Kussaibi[1], Elaf Alibrahim[2], Eman Alamer[2], Ghadah Al hajji[2], Shrooq Alshehab[2], Zahra Shabib[2], Noor Alsafwani[1], Ritesh G. Menezes[1]*

[1]Department of Pathology, College of Medicine, Imam Abdulrahman bin Faisal University, Dammam, Saudi Arabia; [2]College of Medicine, Imam Abdulrahman bin Faisal University, Dammam, Saudi Arabia

Abstract. *Background and aim:* Ovarian cancer is the leading cause of gynecological cancer deaths due to late diagnosis and high recurrence rates. While histopathological analysis is the gold standard for diagnosis, artificial intelligence (AI) models have shown promise in accurately classifying ovarian cancer subtypes using histopathology images. Herein, we introduce an end-to-end AI pipeline for automated identification of epithelial ovarian cancer (EOC) subtypes based on histopathology images and evaluate its performance compared to pathologists' diagnoses. *Methods:* A dataset of over 2 million image tiles from 82 whole slide images (WSIs) of major EOC subtypes (clear cell, endometrioid, mucinous, and serous) was curated from public and institutional sources. A convolutional neural network (ResNet50) was used to extract features, which were then introduced to 2 classifiers (NN and LightGBM) to predict cancer subtypes. *Results:* Both AI classifiers achieved patch-level accuracy (97-98%) on a test set. Furthermore, adding a class-weighted cross-entropy loss function to the pipeline showed better discriminative performance among subtypes. *Conclusions:* AI models trained on histopathology images can accurately classify EOC subtypes, potentially assisting pathologists and reducing subjectivity in ovarian cancer diagnosis. (www.actabiomedica.it)

Key words: ovarian cancer, artificial intelligence, histopathology images, deep learning, whole slide images

Introduction

Ovarian cancer (OC) is the 3rd most common gynecological cancer worldwide (1) and the 7th in the Kingdom of Saudi Arabia (KSA) among females (Saudi Council) (2). Due to the late stage of diagnosis and the high chance of recurrence (70%) (3), OC outranks other gynecological cancers in terms of mortality rate. Despite all the advanced developments in medical technology in diagnosis and treatment, OC is still discovered in the late stages, which ultimately affects the treatment options and prognosis. Clinically, ultrasound and tumor biomarkers are used as the conventional diagnostic tools to investigate suspected cases of OC (4). However, histopathological study is the cornerstone to confirm the diagnosis (5). According to the World Health Organization (WHO), ovarian cancers are classified histologically into epithelial, germ cell, and sex-cord stromal tumors, in which epithelial ovarian cancer (EOC) stands out with the highest incident rate (6). In 2018, the Global Cancer Observatory (GLOBOCAN) reported that EOC accounts for 4.3% of cancer mortality annually, making EOC the most lethal gynecological cancer worldwide (7). EOC is a heterogeneous disease that consists of five histological subtypes: High-Grade Serous Carcinoma (HGSC), Low-Grade Serous Carcinoma (LGSC), Mucinous Carcinoma (MC), Endometrioid Carcinoma (EC), and Clear Cell Carcinoma (CCC) (8). Each subtype exhibits unique histological morphology, molecular biology, pathogenesis, and clinical behavior (6). Medical industries are interested in

implementing Artificial Intelligence (AI) in different fields. Multiple studies were conducted respecting AI recruitment in active patient care (9). According to Aliza Becker (10), AI usage in the medical field could be within one or more categories: 1. For research purposes targeting pathology and/or treatment efficacy; 2. To minimize possible complications, 3. To maximize patient care during ongoing treatment or procedure, and 4. In evaluating disease risk and predicting prognosis or treatment efficacy. Myszczynska et al. shed light on machine learning utilization in diagnosing and treating neurodegenerative diseases (11). The AI models reviewed in the earlier study can analyze various clinical parameters such as neuroimages, electroencephalograms, and even cognitive test performance, suggesting an era of the machine learning revolution (11). The application of AI is becoming a trend in pathology settings, particularly in diagnosing cancer (12). Dolezal et al. emphasized the significant role of deep learning in analyzing digitized histopathological slides as it alleviates the workload on pathologists and detects features that could be easily missed (13). Deep learning (DL), a subtype of AI, may achieve human-level accuracy by abstracting the original data layer by layer, obtaining different levels of abstract features, and using them for detection, classification, or segmentation (14). DL could predict clinical biomarkers, gene expression patterns, survival outcomes, and pathogenic mutations from traditional histoimages (12). Convolutional neural networks (CNN) are the most prominent model of DL, preferred in image recognition and classification. The Residual network (ResNet) is a type of CNN with wide image recognition and classification applications in medical and non-medical trials (15). In 2022, Rahman et al. worked with ResNet50 to measure its efficiency in diagnosing breast cancer by reading and classifying mammographic images. ResNet50 scored 93% in accuracy, with less effort and more time effectiveness (9). ResNet50 has been used in histopathology image detection, such as breast and lung cancer diagnosis, with 92% and 97.49% accuracy, respectively (16). Several studies have discussed the use of AI in the diagnosis, classification, and prognosis of ovarian cancer. In 2018, researchers utilized Alexnet, a deep convolutional neural network (DCNN), to classify EOC into four subtypes: serous, mucinous, endometrioid, and clear cell, based on cytology (17). The study involved 87 Hematoxylin-Eosin (H&E) stained tissue samples comprising 81312 images, including augmented images. The DCNN shall consist of five conventional layers, three max-pooling, and two fully reconnected layers. Aiming to improve network accuracy, this study used two input sets: the original slide set and the other with slide augmentation through image rotation and edge sharpening filters. The classification accuracy was only 72.76% in total, with endometrioid being the lowest in accuracy (64.53%). Data augmentation proved its role in increasing total accuracy by 5.44%. Yet, the study has some limitations as the model incorrectly classified endometrioid as serous, mucinous as clear cell, and clear cell as mucinous with an error rate of 15.11%, 12.64%, and 11.39%, respectively (17). In 2022, a study was conducted to evaluate the performance of four deep learning algorithms in histotype classification of the four main subtypes of epithelial ovarian cancer. Two datasets were collected from different centers with a total of 948 H&E-stained whole-image slides, and each set was used for either training or testing only. The final diagnosis of the model is referenced to the diagnosis of expert pathologists. The performance of the four models was compared, and the best performer was selected based on the mean slide-level diagnostic concordance. Results showed a mean slide-level diagnostic concordance of 80.97 ± 0.03% in one of the models, with an overall performance of 0.7722 (Cohen's Kappa). In case of discrepancy between the four models, two pathologists were asked to review the slides blindly (18). A retrospective cohort study was conducted to classify samples into either borderline ovarian tumor (SBOT) or high-grade serous ovarian cancer (HGSOC) (19). The study included 30 cases from the institutional pathology system database, with equal representation of each type. Researchers used a Support Vector Machine (SVM) classifier as an AI model to sort the sample based on 41 cellular features. By using Groovy scripts and QuPath, pathologists annotated the slides manually into stromal or tumor cells. Final results revealed an accuracy of 86.4%-89.1% and 85.4%-90.8% for HGSOC and SBOT, respectively. Moreover, the overall SVM accuracy reached up to 90.5%-90.7%. One of

the significant features that played a crucial role in differentiation is Eosin OD staining intensity. Despite the high accuracy rate achieved, it is still strenuous for AI models to approximate human capacity. Multiple reasons for misclassification were observed, such as histological artifacts and under/over cellular segmentation (19). Two studies used a convolutional neural network (CNN) to predict HGSOC response to platinum-based therapy and the prognosis accordingly (20,21). Liu et al. aimed to identify the morphological features that can enhance the decision-making towards a treatment plan and patient survival rate (20). Researchers utilized 248 samples from The Cancer Genome Atlas (TCGA); 208 samples were used to train inception V3, and the remaining 40 were used for testing. To limit model bias, specialized pathologists selected regions of interest (ROI) from original slides to be cropped into small tiles to fit the model input size. Finally, they tagged the tiles as resistant or sensitive to chemotherapy. Moreover, training was repeated 16 times with parameter adjustments to ensure model generalizability. Another measure was used to size up and down the number of resistance or sensitive cases to assess if it would affect the model accuracy. Eventually, 85% of patients were categorized correctly, with a specificity of 90% and sensitivity of 73%. Furthermore, researchers stated that age, grade, and stage were not significantly associated with chemotherapy response. The main weakness of this study is the online data source, small validation, sample size, limited generalizability as it depends on TCGA only, and the data used mainly were high-grade tumors, which decreased the clinical heterogeneity (20). Whereas Laury et al. used samples of patients diagnosed and treated for high-grade extrauterine serous carcinoma at HUS Helsinki University Hospital between 1982 and 2013 (21). Around 205 whole slide images were used to train the CNN, and 22 slides were used to test the model. The model could identify 18 out of 22 testing slides with a sensitivity of 73% and a specificity of 91%. The positive predictive value was 89%, with an overall accuracy of 82%. The main weakness of this study was that molecular testing results were not available for all samples as the test was not routinely done (21). In a retrospective cohort study conducted in 2020, Tanabe et al. used the CNN model AlexNet to analyze ovarian cancer patients' serum glycopeptides by converting glycopeptides expression into a 2-dimensional (2D) barcode (22). A sample of 97 patients diagnosed with early-stage EOC was collected, of which 60% were used for training and 40% for validation. Two sets of barcodes were prepared with different alignments based on liquid chromatography elution time and principal component analysis, resulting in area under the curve (AUC): 0.881 and 0.851, respectively. Cancer Antigen 125 (CA125) and Human epididymis 4 (HE4) information were added to the barcode to enhance the diagnostic performance of the multicolored model further. This resulted in increased diagnostics accuracy to 95%. Nonetheless, an Area Under the Receiver Operating Characteristic curve of almost 100% brings the model generalizability into question. Thus, extending training samples is a must to eliminate possible over-fitting (22). Another observational cohort study aimed to predict the mutation of Breast Cancer (BRCA 1/2) genes in ovarian cancer patients from H&E tissue using a CLAM-based approach (23). A sample of 664 ovarian cancer patients was involved; 464 were utilized for the training process and 132 for validation. The study showed disappointing results with low accuracy, reaching 62.9% and 87.9% specificity with a 16.7% sensitivity, which suggests that phenotype may not be strongly related to genotype (23). Other fields of research regarding AI and ovarian histopathological slide images included a retrospective cohort study that aimed to assess drug-related ovarian toxicity (24). The study aimed to create a deep learning algorithm that can quantify corpora lutea (CLs) in H&E-stained ovarian tissue with similar accuracy to the pathologist's assessment. The study was conducted on the ovarian tissue of female Sprague-Dawley rats, and the sample was trained and validated using RetinaNet, the deep-learning algorithm of choice in this study. Due to the structural variability of rodent ovarian tissue, enumeration of CLs and evaluation of ovarian toxicity can be time-consuming and challenging. Therefore, an automated method to quantify CLs would potentially decrease the workload for pathologists. Overall, drug-related ovarian toxicity was evident in the form of either increase or decrease in the number of CLs. In addition, mistakes in counting CLs by RetinaNet were similar to the pathologist's

variation, with a mean absolute difference of 0.57 + 0.096 between the pathologists and RetinaNet. Further training of the model with a further increase in sample size would improve the model's accuracy (24). A systematic review published in 2023 included 45 studies aiming to review the application of AI in pathological images for either diagnosis or prognosis of ovarian cancer (25). In these 45 studies, 80 models were used, including convolutional neural networks and their different architectures. CNN was the most commonly used (41/80), followed by support vector machine (SVM) (10/80), and random forest (6/80). Additionally, in most of the studies (18/45), the data source used for ovarian pathology slides is The Cancer Genome Atlas (TCGA), and other studies used their data (12/45). Most researchers have used segmentation to determine the region of interest for tumor tissue. Most published research (nearly 37 papers) showed a high or unclear risk of bias due to limited data, such as the number of included patients and images (25). Another systematic review published in 2022 included 39 studies (4). Of these studies, 7 used pathological images, 13 were based on serum tumor markers, and 19 used high-throughput data. The review aimed to objectively assess AI algorithms' application in diagnosing and predicting ovarian cancer prognosis. The 7 studies that used pathological images to diagnose ovarian cancer showed specificity higher than 90%. However, there is a wide gap between the predictive performance of the current AI model and the clinicians' experience, especially in prognostic prediction, due to the narrow data set (4). In Saudi Arabia, AI in diagnostic radiology is the most studied and, therefore, most advanced in the health intelligence field (26). Nonetheless, studies on AI applications in histopathology have arisen recently, including but not limited to Alahmadi et al. who proposed an integrating framework of Vision Transformer (ViT) and Model-agnostic Explanations (LIME) aiming to detect and diagnose ovarian cancer based on histopathological images (27). Data has been collected from 20 different medical centers. Introduced models exhibit outstanding performance in both training and testing sets. However, there was a slight decline in performance from the training phase to the testing phase, which reemphasized the need for further work and model enhancement. The highest accuracy, precision, and recall levels were observed in the ViT-Large-P32-384 model with 98.86%, 97.18%, and 97.06%, respectively. Despite all effort and work in diagnostic AI models locally and globally, there are still gaps that need to be filled and areas to be improved (27). Therefore, further studies are essential to fuel AI utilization in histopathology to rise to the advancement of other health fields in KSA. In this work, we present a complete end-to-end AI pipeline for the automated identification of epithelial ovarian cancer (EOC) subtypes. This pipeline is developed and trained using a dataset of digital histopathological images representing the major EOC subtypes, and its performance is evaluated against pathologists' diagnoses.

Methods

Dataset collection and pre-processing

Sixty-four public cases, including (20x) whole slide images (WSIs), were retrieved from the Cancer Imaging Archive (CIA) and grouped into four subtypes of epithelial ovarian cancer (clear cell carcinoma (CCC), endometrioid carcinoma (EC), mucinous carcinoma (MC), and serous carcinoma (SC) (28). In-house cases were collected from the archive of the pathology department at King Fahad University Hospital (KFUH) between 2014 and 2023, ending up with 18 Hematoxylin and Eosin (H&E) stained slides. The cases were also classified into the four most common epithelial ovarian carcinoma subtypes (CCC, EC, MC, and SC) (Figure 1). Experienced pathologists who assigned slide-level labels and tumor annotation reviewed and confirmed the diagnosis. The collected slides were scanned and digitized at (20x) power in our facility by (Ventana I-Scan) to obtain 18 whole slide images in tiff format.

Statement on waiver of consent

Consent for this study has been waived due to its retrospective nature. The patients involved in the study were impossible to track. Additionally, the data in this study were retrieved from archived microscopic slides and contained no patient identification, ensuring the subjects' anonymity.

Both public and in-house cases were divided into training (80%) and evaluation (20%) subsets.

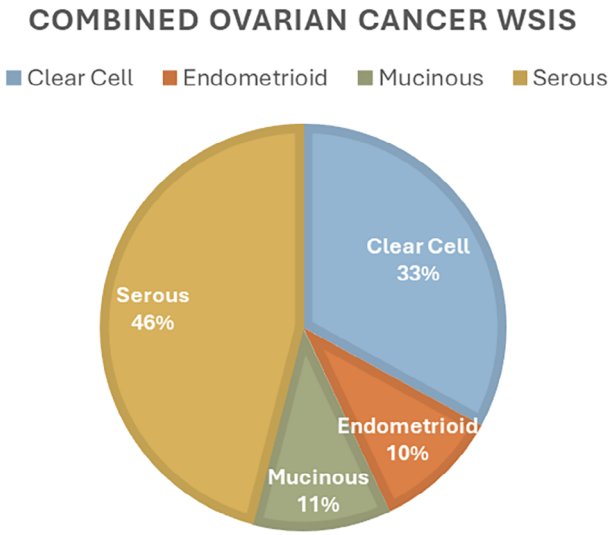

Figure 1. Distribution of the different classes in the combined WSIs dataset.

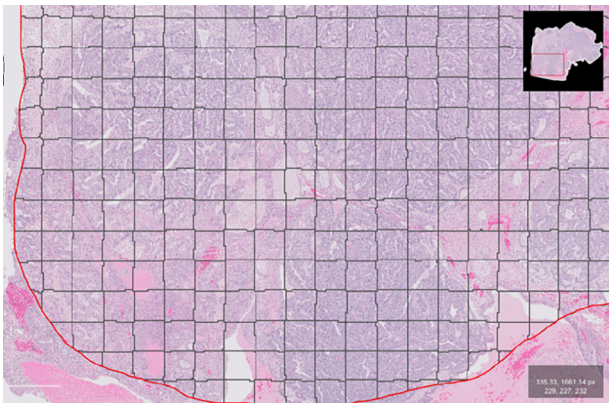

Figure 2. Tiles extracted from the annotated tumor region.

Extract tiles from the WSIs

The WSIs were huge (thousands of pixels), making processing them entirely impossible, as the expected input size for the majority of deep learning models is 224 pixels. Therefore, all the public and in-house WSIs were cropped into tiles of size (224 x 224 pixels) (Figure 2). Using QuPath software version 0.5 (29), pathologists first annotated tumor regions of interest (ROIs) and then extracted tiles from those ROIs by a customized script.

Pre-processing techniques

Considering the variation in color between the different WSIs, image normalization and color standardization are essential to normalize the obtained patches before introducing them to the model. For that purpose, we adopted the following Torchvision function: (transforms.Normalize(mean=[0.485, 0.456, 0.406], std=[0.229, 0.224, 0.225])).

Network architectures and training loop

Considering the large size of the dataset, representative features from the resulting tiles were extracted and then used to train a classifier rather than training the network directly on the images. This approach tremendously reduces the consumption of computing resources and the time needed to train the network.

CNN extractor architecture

ResNet50, a CNN model, was used for feature extraction. The model is pre-trained on ImageNet. It has been frozen after removing its classifier head to preserve pre-learned weight.

Training process

Two classifiers (NN-based and lightGBM) were selected for experimental training on the extracted features. The NN-based classifier was configured with specific parameters: an Adam optimizer with a learning rate of 0.001 and a scheduler, a batch size of 128, 21 epochs with early stopping, and loss functions utilizing CrossEntropy and Class-weighted CrossEntropy. Additionally, cross-validation was incorporated during the training phase. On the other hand, the lightGBM model was tuned with the following parameters: multiclass objective, a learning rate of 0.05, and a multi_error metric. Early stopping was applied after ten rounds, whereby training ceased if validation scores failed to improve for ten consecutive rounds.

PyTorch was used as a Deep-learning framework with the following experimental setup: GPU (NVIDIA RTX3060 12GB), RAM 32GB, CPU i7 13th gen.

Performance evaluation metrics

Four metrics were used to assess network performance (accuracy, precision, recall, and F1 score) along with confusion matrices for class discrimination:

- Accuracy: percentage of correct predictions
- Precision (specificity): accuracy of positive predictions.
- Recall (sensitivity): the ability to identify all positive instances.
- F1-score: harmonic mean of precision and recall
- Confusion matrices for detailed analysis of classification errors

Results

Extracted tiles

Using QuPATH, over 2 million tiles/patches were extracted from the collected and annotated WSIs (public and in-house) (Table 1).

Network prediction performance

Both networks (lightGBM and NN) achieved comparable results, with NN achieving slightly higher performance (Table 2).

However, the discriminative strength among classes was much better when using class-weighted NN (NN-cw) (Table 3).

In our study, the overall misclassification was in the following order from the highest to the lowest error rate: 1. EC misclassified as SC, 2. MC misclassified as CCC, 3. SC misclassified as EC, and 4. CCC misclassified as MC (Table 4).

Table 1. Dataset details.

	CCC	EC	MC	SC	Total
Public WSIs	25	5	6	28	64
Public tiles	750k	90k	110k	850k	1.800k
KFHU WSIs	2	3	3	10	18
KFHU tiles	21k	37k	37k	164k	259k

Table 2. Evaluation metrics: Network classification performance.

		Precision					Recall					F1 score				
	Acc.	Avg	CCC	EC	MC	SC	Avg	CCC	EC	MC	SC	Avg	CCC	EC	MC	SC
lightGBM	0.97	**0.95**	0.98	0.91	0.94	0.97	**0.91**	0.99	0.79	0.9	0.99	**0.93**	0.98	0.84	0.92	0.98
NN	0.98	**0.96**	0.99	0.9	0.96	0.99	**0.95**	0.99	0.9	0.95	0.99	**0.95**	0.99	0.9	0.95	0.99
NN-cw	0.97	**0.91**	1	0.74	0.94	0.99	**0.97**	0.99	0.97	0.97	0.96	**0.94**	0.99	0.84	0.95	0.98

Abbreviation: Acc.: Accuracy

Table 3. Confusion matrices of the three classification networks' performance.

		NN	NN-cw	LGBM	NN	NN-cw	LGBM	NN	NN-cw	LGBM	NN	NN-cw	LGBM
True Labels	**CCC**	99.02%	98.76%	98.68%	0.03%	0.07%	0.08%	0.68%	1.01%	0.73%	0.27%	0.16%	0.52%
	EC	0.06%	0.01%	0.26%	86.47%	92.89%	78.66%	3.10%	3.18%	0.97%	10.37%	3.93%	20.11%
	MC	3.13%	1.25%	8.61%	0.13%	0.21%	0.69%	96.58%	98.48%	89.95%	0.16%	0.05%	0.75%
	SC	0.05%	0.04%	0.14%	0.70%	1.82%	0.78%	0.37%	0.40%	0.19%	98.89%	97.75%	98.89%
		CCC			**EC**			**MC**			**SC**		
		Predicted Labels											

Table 4. Misclassification error rate.

	LightGBM	NN	NN-cw
EC misclassified as SC	20.11%	10.37%	3.93%
MC misclassified as CCC	8.61%	3.13%	1.25%
SC misclassified as EC	0.78%	0.70%	1.82%
CCC misclassified as MC	0.73%	0.68%	1.01%

Discussion

Histopathology image examination can be challenging and time-consuming for pathologists. An accurate diagnosis is essential to selecting a proper treatment and predicting prognosis, particularly with different morphologic features and variations of each subtype of EOC. For this reason, AI can provide a fast and accurate analysis of large and complex microscopic images. This study aimed to construct a full end-to-end AI pipeline that helps pathologists identify EOC subtypes incorporating histopathological images from public and in-house cases. The AI workflow developed in this study demonstrated promising performance in accurately predicting the major subtypes of EOC from WSIs. Utilizing deep learning techniques to extract discriminative features along with LightGBM and NN classifiers resulted in patch-level accuracy rates of 97-98% on the evaluation dataset. These results align with previous studies highlighting the potential of AI models for ovarian cancer classification and diagnosis from histopathology images. These masterful scores aligned with Alahmadi et al. results, with a score of 97.07% (27). The large and diverse dataset of over 2 million image tiles extracted from 82 whole slide images likely contributed to the strong generalization performance achieved by the models.

Although the overall accuracy is promising, it was variable among EOC subtypes. CCC shows a significantly higher F1 score (0.98, 0.99, 0.99) with the three classifiers (lightGBM, NN, and NN-cw), respectively, which is slightly similar to SC (0.98, 0.99, 0.98). In comparison, EC scored the lowest (0.84, 0.90, 0.84) (Table 2). A close look at the confusion matrices (Tables 3, 4) revealed that EC is mostly misclassified as SC, followed by MC, which is misclassified as CCC, which is consistent with Wu et al. findings. Wu et al. denoted these misclassifications to the unclear morphology of some of the samples (17). According to a study by Lim et al., pathological and immunohistochemical features were used to re-evaluate 109 cases of EOC diagnosed by pathologists. Results showed that 30% of cases initially diagnosed as EC were reclassified to SC, emphasizing the overlapping morphology between the two subtypes (30). As illustrated in (Table 4), it is remarkable that all three models misclassify one specific subtype as another (EC as SC and MC as CCC) and vice versa, which strongly implies the intersecting morphological features rather than a model error. As addressed thoroughly by Lim et al., these morphological features include the Confirmatory Endometrioid Features (CEFs) outlined by WHO histological criteria for EC classification. The CEFs mostly associated with EC were: 1. Squamous metaplasia, 2. Endometriosis, 3. Adenofibromatous background, and 4. Borderline endometrioid or mixed Mullerian component. On the other hand, SC is characterized by one or more of the following architectural patterns: solid, cribriform (pseudo-endometrioid), or transitional cell carcinoma-like (30). Additionally, MC and CCC are characterized by gastrointestinal-like epithelium with intracytoplasmic mucin and polygonal cells with clear cytoplasm, respectively. Therefore, an immunohistochemical profile is essential for further differentiation between these two subtypes (7). Another study supporting the previously discussed findings showed that EC was mostly misclassified as SC or MC. This study used four different deep-learning models to classify the four main subtypes of EOC. Some models showed variation in the classification due to the overlapping morphology. In some of the cases, the morphological feature that accounted for the misclassification was the presence of a transitional pattern, which is a pattern that can be detected in both EC and SC. In other cases, the absence of intracellular mucin made differentiating between EC and MC challenging. The discrepancy in classification between the models was reviewed by pathologists, who further supported the diagnosis by the incorporation of an immunohistochemical profile of each subtype (18). However, certain limitations must be recognized. First, the study is constrained by a small in-house sample size. A larger sample would be ideal for more

robustly validating the AI model's accuracy. Due to the limited dataset available from KFUH, including data from additional hospitals or provinces would likely yield more accurate and generalizable results. Second, there is an imbalance in the number of samples across the subtypes. Despite the researcher's efforts to ensure the inclusion of all subtypes, the overrepresentation of SC subtypes was noticeable in this study sample, as SC is the most common EOC subtype in KSA and worldwide (6). The NN-cw model was an attempt to overcome this limitation. When class-weighted techniques were applied during training, the NN classifier showed a better discriminative ability between the different cancer subtypes. Third, a shortage of computer resources restricted data input and analysis.

Despite these limitations, the current paper provides an excellent accuracy level, reaching 97-98%, and encourages more work concerning AI application in general pathology settings and EOC classification. While this study primarily focused on the accuracy of AI models in classifying ovarian cancer subtypes, it is essential to consider their potential to reduce diagnostic times in clinical practice. Once integrated into routine workflows, AI systems can expedite the analysis of large histopathological datasets, providing rapid preliminary results, especially in secondary centers. The practical adoption of AI systems also depends on their cost-effectiveness. AI-assisted diagnostics can potentially reduce the need for second opinions from reference centers, leading to savings in time and costs associated with transport, consultation, and additional testing. In some cases, AI could minimize the need for expensive molecular tests by improving diagnostic accuracy. A formal economic analysis comparing AI-driven workflows with traditional pathology practices will be essential to thoroughly assess AI integration's financial benefits and feasibility, which we plan to address in future work. While the experience of trained pathologists remains unmatched, particularly for complex or ambiguous cases, our findings indicate that AI systems can effectively handle routine diagnostic tasks. By automating the classification of epithelial ovarian cancer subtypes with high accuracy, AI can help pathologists manage their workload more efficiently, allowing them to dedicate more time to complex cases. This collaboration between AI and human expertise can potentially enhance the diagnostic process.

Conclusion

This study proves that AI-based computational pathology techniques can achieve human-level accuracy in ovarian cancer subtype classification from histopathology images. If properly validated, such AI-assisted diagnostic tools could greatly benefit pathology workflow by reducing subjectivity, decreasing inter-observer variability, and increasing efficiency compared to manual review alone. Future research should focus on prospective clinical validation and extend this approach to other cancer types and diagnostic tasks in digital pathology.

Ethics Approval: The Institutional Review Board (IRB) of Imam Abdulrahman bin Faisal University approved the use of patients' clinical information and histopathological samples (IRB number IRB-UGS-2023-01-410, approved on 22/10/2023).

Conflict of Interest: Each author declares that he or she has no commercial associations (e.g. consultancies, stock ownership, equity interest, patent/licensing arrangement, etc.) that might pose a conflict of interest in connection with the submitted article.

Funding: No funding was received for this study.

Code Availability: The complete pipeline code is available on https://github.com/hkussaibi.

Authors Contribution: Conceptualization and Design: HK, ElA, EmA, GA, SA, ZS, NA. Project Management HK, NA. Data Collection HK, ElA, EmA, GA, SA, ZS, NA. Data Analysis HK, ElA, EmA, GA, SA, ZS, NA. Writing and Drafting HK, ElA, EmA, GA, SA, ZS, NA, RM. Critical Review and Feedback HK, NA, RM. Supervision and Mentorship HK, NA, RM. Technical Support HK, RM. Ethical Considerations HK, ElA, RM. Dissemination and Communication HK, RM.

Declaration on the Use of AI: While preparing this work, the authors used ChatGPT to improve language and readability. After using this tool, the authors reviewed and edited the content as needed and took full responsibility for the publication's content.

References

1. Momenimovahed Z, Tiznobaik A, Taheri S, Salehiniya H. Ovarian cancer in the world: epidemiology and risk factors. Int J Womens Health. 2019; 11:287-99. doi: 10.2147/IJWH.S197604.
2. Al-Rawaji A. Cancer Incidence Report 2020 - Kindom of Saudi Arabia, Saudi Health Council, National Cancer Center, Saudi Cancer registry, 2020,1-81.
3. Mikdadi D, O'Connell KA, Meacham PJ, et al. Applications of artificial intelligence (AI) in ovarian cancer, pancreatic cancer, and image biomarker discovery. Cancer Biomark. 2022;33(2):173-84. doi: 10.3233/CBM-210301.
4. Zhou J, Cao W, Wang L, Pan Z, Fu Y. Application of artificial intelligence in the diagnosis and prognostic prediction of ovarian cancer. Comput Biol Med. 2022; 146:105608. doi: 10.1016/j.compbiomed.2022.105608.
5. Orsulic S, John J, Walts AE, Gertych A. Computational pathology in ovarian cancer. Front Oncol. 2022; 12:924945. doi: 10.3389/fonc.2022.924945.
6. Devouassoux-Shisheboran M, Genestie C. Pathobiology of ovarian carcinomas. Chin J Cancer. 2015;34(1):50-5. doi: 10.5732/cjc.014.10273.
7. Ignacio R, Susanna L, Belén Pérez M, Andrés Poveda V, José P. Morphological and molecular heterogeneity of epithelial ovarian cancer: therapeutic implications. Eur J Cancer Suppl. 2020; 15:1-15. doi: 10.1016/j.ejcsup.2020.02.001.
8. Lalwani N, Prasad SR, Vikram R, Shanbhogue AK, Huettner PC, Fasih N. Histologic, molecular, and cytogenetic features of ovarian cancers: implications for diagnosis and treatment. Radiographics. 2011;31(3):625-46. doi: 10.1148/rg.313105066.
9. Rahman H, Naik Bukht TF, Ahmad R, Almadhor A, Javed AR. Efficient breast cancer diagnosis from complex mammographic images using deep convolutional neural network. Comput Intell Neurosci. 2023; 2023:7717712. doi: 10.1155/2023/7717712.
10. Aliza B. Artificial intelligence in medicine: what is it doing for us today? Health Policy Technol. 2019; 8(2):198-205. doi: 10.1016/j.hlpt.2019.03.004.
11. Myszczynska MA, Ojamies PN, Lacoste AMB, et al. Applications of machine learning to diagnosis and treatment of neurodegenerative diseases. Nat Rev Neurol. 2020; 16(8):440-56. doi: 10.1038/s41582-020-0377-8.
12. El-Ghany SA, Azad M, Elmogy M. Robustness fine-tuning deep learning model for cancers diagnosis based on histopathology image analysis. Diagnostics (Basel). 2023; 13(4): 806. doi: 10.3390/diagnostics13040699.
13. Dolezal JM, Kochanny S, Dyer E, et al. Slideflow: deep learning for digital histopathology with real-time whole-slide visualization. BMC Bioinformatics. 2024; 25(1):134. doi: 10.1186/s12859-024-05758-x.
14. Cai L, Gao J, Zhao D. A review of the application of deep learning in medical image classification and segmentation. Ann Transl Med. 2020; 8(11):713. doi: 10.21037/atm.2020.02.44.
15. Taye MM. Theoretical understanding of convolutional neural network: concepts, architectures, applications, future directions. Computation. 2023; 11(3):52. doi: 10.3390/computation11030052.
16. Muhammed T. Automated classification of histopathology images using transfer learning. Artif Intell Med. 2019; 101:101743. doi: 10.1016/j.artmed.2019.101743.
17. Wu M, Yan C, Liu H, Liu Q. Automatic classification of ovarian cancer types from cytological images using deep convolutional neural networks. Biosci Rep. 2018; 38(3): BSR20171150. doi: 10.1042/BSR20180289.
18. Farahani H, Boschman J, Farnell D, et al. Deep learning-based histotype diagnosis of ovarian carcinoma whole-slide pathology images. Mod Pathol. 2022; 35(12):1983-90. doi: 10.1038/s41379-022-01146-z.
19. Jiang J, Tekin B, Guo R, Liu H, Huang Y, Wang C. Digital pathology-based study of cell- and tissue-level morphologic features in serous borderline ovarian tumor and high-grade serous ovarian cancer. J Pathol Inform. 2021; 12:24. doi: 10.4103/jpi.jpi_76_20.
20. Liu Y, Lawson BC, Huang X, Broom BM, Weinstein JN. Prediction of ovarian cancer response to therapy based on deep learning analysis of histopathology images. Cancers (Basel). 2023; 15(16):4085. doi: 10.3390/cancers15164044.
21. Laury AR, Blom S, Ropponen T, Virtanen A, Carpén OM. Artificial intelligence-based image analysis can predict outcome in high-grade serous carcinoma via histology alone. Sci Rep. 2021; 11(1):19165. doi: 10.1038/s41598-021-98480-0.
22. Tanabe K, Ikeda M, Hayashi M, et al. Comprehensive serum glycopeptide spectra analysis combined with artificial intelligence (CSGSA-AI) to diagnose early-stage ovarian cancer. Cancers (Basel). 2020; 12(9):2551. doi: 10.3390/cancers12092373.
23. Nero C, Boldrini L, Lenkowicz J, et al. Deep-learning to predict BRCA mutation and survival from digital H&E slides of epithelial ovarian cancer. Int J Mol Sci. 2022; 23(19):11539. doi: 10.3390/ijms231911326.
24. Hu F, Schutt L, Kozlowski C, Regan K, Dybdal N, Schutten MM. Ovarian toxicity assessment in histopathological images using deep learning. Toxicol Pathol. 2020; 48(2): 350-61. doi: 10.1177/0192623319877871.
25. Breen J, Allen K, Zucker K, et al. Artificial intelligence in ovarian cancer histopathology: a systematic review. NPJ Precis Oncol. 2023; 7(1):83. doi: 10.1038/s41698-023-00432-6.
26. Aljerian N, Arafat M, Aldhubib A, et al. Artificial intelligence in healthcare and its application in Saudi Arabia. Int J Innov Res Med Sci. 2022; 7:666-70. doi: 10.23958/ijirms/vol07-i11/1558.
27. Abdulrahman A. Towards ovarian cancer diagnostics: a vision transformer-based computer-aided diagnosis framework with enhanced interpretability. Results Eng. 2024; 23: 102651. doi: 10.1016/j.rineng.2024.102651.
28. Wang C-W, Chang CC, Lo SC, et al. A dataset of histopathological whole slide images for classification of treatment effectiveness to ovarian cancer (Ovarian Bevacizumab

Response) (Version 2) [Data set]. Cancer Imaging Arch. 2021. doi: 10.7937/TCIA.985G-EY35.
29. Bankhead P, Loughrey MB, Fernández JA, et al. QuPath: open-source software for digital pathology image analysis. Sci Rep. 2017; 7:168-78. doi: 10.1038/s41598-017-17204-5.
30. Lim D, Murali R, Murray MP, Veras E, Park KJ, Soslow RA. Morphological and immunohistochemical reevaluation of tumors initially diagnosed as ovarian endometrioid carcinoma with emphasis on high-grade tumors. Am J Surg Pathol 2016; 40:302-12. doi: 10.1097/pas.0000000000000550

Correspondence:
Received: 10 September 2024
Accepted: 12 October 2024
Dr. Haitham Kussaibi, MD
Department of Pathology, College of Medicine,
Imam Abdulrahman bin Faisal University
King Faisal Road 31441, Dammam, Saudi Arabia
E-mail: hkussaibi@iau.edu.sa
ORCID ID: 0000-0002-9570-0768