

Empowering Ovarian Cancer Subtype Classification with Parallel Swin Transformers and WSI Imaging

Lubna ALkahla
Department of Software
Ninevah University, Iraq
lubna.thanoon@uoninevah.edu.iq

Jwan Saeed
Department of Information Technology
Duhok Polytechnic University, Iraq
jwan.najeeb@dpu.edu.krd

Maher Hussein
Computer Center
University of Telafer, Iraq
maher.k.hussein@uotelafer.edu.iq

Abstract: Ovarian cancer constitutes a notable proportion of cancer-related mortalities among women. The diagnostic classification of ovarian cancer subtypes has demonstrated complexity, characterized by limited concordance among pathologists. Vision Transformer (ViT) models have emerged as the predominant architecture in numerous computer vision applications, encompassing tasks such as image classification and cancer detection. Their success stems primarily from their capacity to integrate global contextual information through self-attention mechanisms during the learning process. However, the key issue with ViT is its compatibility with high-res images. Computation grows quadratically with image size, resulting in a large number of tokens and significant computational demands for self-attention. Swin Transformer (Swin-T) addresses this challenge by introducing two main concepts: hierarchical feature mapping and windowed attention transformation. This work presents a parallel implementation of Swin Transformers (Swin-Ts) that leverages the powerful feature extraction capabilities and aimed at classifying five subtypes within ovarian cancer utilizing Whole Slide Imaging (WSI) and it yielded average precision, recall, and F1-score metrics of 0.958, 0.964, and 0.96 correspondingly. The findings show that the proposed parallel Swin-Ts reduce the misclassification errors and improve medical image analysis robustness. Additionally, the suggested technique is promising for accurate and efficient ovarian carcinoma subtype categorization, with possible applicability to other cancers. Future research will integrate other data sources and validate the technique in various clinical contexts.

Keywords: Ovarian cancer, Swin-transformer, sub-type classification, histology image analysis.

Received July 1, 2024; accepted October 10, 2024
<https://doi.org/10.34028/iajit/21/6/5>

1. Introduction

Ovarian cancer represents a substantial contributor to mortality within the spectrum of cancer afflictions affecting women. Statistics from 2021 reveal that ovarian cancer was responsible for approximately 40% of fatalities associated with malignancies affecting the female genital system [4]. Ovarian cancer encompasses five primary subtypes and the process of diagnosing and categorizing ovarian cancer has also been demonstrated to present challenges, characterized by limited to moderate consensus among pathologists. Therefore, conventional approaches to classifying ovarian cancer subtypes typically depend on histological examination, a process susceptible to subjectivity and constrained by tissue accessibility [10]. Medical imaging, crucial in cancer detection, includes various modalities like X-rays, MRI, and CT scans. However, Whole Slide Imaging (WSI) stands out in histopathological analysis. Unlike other techniques, WSI digitizes entire histological slides at high resolution, enabling detailed cellular analysis. WSI offers unparalleled spatial resolution for accurate tumor identification and margin assessment. Furthermore, it is digital form can easily kept, shared, and analyzed. As a result of using these

advanced algorithms, the diagnoses are more accurate and effective in terms of the workflow [3]. It is, therefore, diagnostic tool number one in cancer of the ovary. Examining pathology slides in the field of pathology involves detecting or diagnosing illnesses through the study of microscopic slide images. These images show a wide variety in features and patterns, and one of the most significant differences concerning the more conventional natural images is their scale. It is a typical workflow in that a pathologist reviews a slide image for types of tissue, where she iteratively moves to various regions to identify regions of tissue types. This is an iterative process for observing individual details of tissues, but one can zoom out to observe broader patterns created from groups of tissues [3]. The computer-aided analysis of images allows for the opening of thousands of research opportunities in microscopic image processing. Moreover, the automatic and precise detection and classification of cancer are essential in improving treatment modalities and the betterment of patient prognoses [19].

With the advancements in machine and deep learning techniques, Convolutional Neural Networks (CNNs) have long been the predominant method for cancer detection [1, 15, 22]. Wu *et al.* [20] utilized a deep CNN

based on AlexNet to automatically diagnose different types of ovarian cancer from cytological images, achieving an accuracy rate of 78.20%. Hwangbo *et al.* [6] created machine learning models to predict platinum sensitivity in the High-Grade Serous Ovarian Cancer Carcinoma (HGSOC) patients. The Logistic Regression (LR)-based model, utilizing six variables, performed best in identifying platinum-resistant cases with an Area Under Curve (AUC) of 0.741. Sengupta *et al.* [16] introduced an innovative deep hybrid learning framework integrating XGBoost, a variant of Random Forest (RF), and a CNN. Their objective was to establish associations between nuclear morphological features and nuclear lamin protein distribution for distinguishing between normal and ovarian cancer tissues. Their approach yielded impressive results, achieving a flawless test AUC score of 1. However, their model was limited by its small sample size, and its reliance on samples from a single institution serving a specific patient demographic. Similarly, Sun *et al.* [17] used a graph convolutional network to rebuild the gene characteristics utilizing by considering both gene features and the structure of the network. Then applied a boosting technique to forecast which genes might be linked to ovarian cancer susceptibility. Consequently, our approach attained a strong AUC score of 0.7541. An Inception V3 deep learning model was developed by Liu *et al.* [11] for predicting ovarian cancer response to platinum-based chemotherapy, relying solely on histopathology images and achieved an AUC of 0.846 ± 0.009 . Another deep learning strategy has been developed by Farahani *et al.* [4] for ovarian carcinoma histotype classification based on histological features. The approach is generalizable even on externally stained test sets and has potential for informing histotype diagnosis and supporting histotype-specific ovarian cancer treatment by training four AI algorithms based on deep CNNs to automatically classify hematoxylin and eosin-stained whole slide images. The best-performing model achieved an accuracy of 80.97% on the test data. As an advanced deep learning technique, the transformer network has emerged as a prominent performer in the field of computer vision. It is particularly noteworthy that the Vision Transformer (ViT) [2] has demonstrated the superiority of a pure attention-based model over CNN-based models, yielding better results. The ability of ViTs to aggregate instance-level features for fine-grained classification tasks was demonstrated by Gao *et al.* [5]. The instance based Vision Transformer (i-ViT) framework was applied by the authors of this study to classify papillary renal cell carcinoma subtypes. The results indicate that ViT models could be promising candidates for summarizing patch-level features (similar to their “instances”) to WSI-level metrics. Furthermore, develop a machine learning-based strategy was proposed for improving ovarian carcinoma histotype diagnosis. Four different architectures were trained and

evaluated on an external dataset, including 948 WSIs. The proposed models achieved a mean slide-level diagnostic concordance of $80.97 \pm 0.03\%$ using a One-Stage Deep Transfer Learning Network (1STL). The model outperformed DeepMIL, VarMIL21, and a two-stage deep transfer learning network. The color normalization strategy was used to overcome color inconsistencies in H and E images. In the same context, a combination of multi-scale CNN and transformer model was proposed by Zhou *et al.* [21] to directly extract features. Pathological image features were selected using Elastic-Net and then merged with clinical information. Survival prediction was conducted using Support Vector Classifier (SVC), RF, and XGBoost via cross-validation. Results indicated RF model outperformed SVC in survival prediction, with XGBoost being less effective. More recently, the Swin Transformer (Swin-T) [12] stands out as a prominent and remarkable architecture. It focuses solely on modeling local relationships within each stage, progressively diminishing the width and height of the feature map while enlarging the receptive field. Consequently, it serves as a versatile foundational framework for various primary downstream visual tasks. Thus, recent advancements in deep learning, notably with Swin_T, have improved ovarian carcinoma subtype classification by better analyzing image context. Integrating histopathological images with molecular and clinical data has increased accuracy and personalized treatment. Challenges with consistent model performance across datasets highlight the need for ongoing research.

This paper introduces a distinct approach aimed at harnessing the capabilities of the Swin-T models for the identification of the five different subtypes of ovarian cancer. The key contributions of this work are:

- Harnessing the power of Swin-Ts in a parallel configuration to get precise differentiation of the five unique subtypes of ovarian cancer, and significantly enhance the model performance.
- Demonstrating improvements over traditional classification (CNN and ViT) methods.
- The proposed multi-Swin_Ts model consistently surpasses the prior methodologies in the classification of ovarian cancer across various datasets.

The remainder of this paper is structured as follows: section 2 provides a brief literature review of related work. Section 3 outlines the proposed methodology. Section 4 presents and discusses the findings. Finally, section 5 concludes the work.

2. Material and Methodology

2.1. Dataset

The dataset focused on ovarian carcinoma, the most lethal type of female reproductive system cancer, is

accessible through Kaggle which called ovarian-cancer-classification dataset [9]. It includes data on the five most common subtypes: High-Grade Serous Carcinoma (HGSC), Clear-Cell Ovarian Carcinoma (CC), Endometrioid (EC), Low-Grade Serous Carcinoma (LGSC), and Mucinous Carcinoma (MC). Each subtype is characterized by distinct cellular morphologies, etiologies, molecular and genetic profiles, and clinical attributes, which are crucial for subtype-specific treatment approaches. The depicted examples of various WSIs in Figure 1 show diverse characteristics and color profiles. These images pose unique challenges distinct from conventional image recognition tasks.

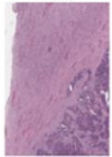

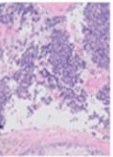

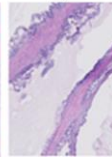
Sub-Type	HGSD	CC	EC	LGSC	MC
Simple					
#WSI Train	120K	5579	7421	2908	3276
#WSI Test	1188	551	733	287	323

Figure 1. The dataset description.

2.2. Data Pre-Processing and Sampling

The WSI images were pre-processed via images resizing and normalization, to prepare the data for model training. Data sampling is then performed to separate training, validating and testing sets. The validation data was selected as 10% of the training dataset. Let D denote the dataset, which is divided into D_{train} , D_{val} , and D_{test} then,

$$D = D_{train} \cup D_{test}, D_{train} \cap D_{test} = \emptyset \quad (1)$$

$$D_{val} = 0.1 \cdot D_{train}, D_{train} \cap D_{val} = \emptyset \quad (2)$$

For the ovarian carcinoma dataset, which has a varying number of samples across different subtypes, down sampling can be a handy technique to balance the

training dataset. Down sampling basically means decrease the number of data instances in the majority class to be equal to that of the minority one hence achieving an even data distribution for training. By so doing, each type within the training dataset will have an equal representation with other classes as indicated by the minimum original class size LGSC. The rationale for choosing down sampling over other techniques, such as oversampling or synthetic data generation, was to avoid introducing potential bias or noise from synthetic samples, which could compromise the model's ability to generalize. By balancing the dataset in this manner, equal representation of each subtype during training was ensured, enhancing the model's ability to learn from all classes effectively.

2.3. Swin Transformer

The Swin-T, characterized by its purely transformer-based architecture as shown in Figure 2, is increasingly recognized as a versatile foundational framework applicable across a spectrum of tasks. With an increase in network depth, the prior ViT maintains consistent down sampling operations to generate cohesive feature maps devoid of segmentation, as depicted in Figure 3-a). Conversely, the Swin-T employs a hierarchical architecture of maps, as depicted in Figure 3-b), mirroring the hierarchical structure found in CNNs. During the initialization phase, the input images are divided into non-overlapping patches facilitated by a patch splitting module. Subsequently, through successive transformer layers, the adjoining patches are progressively amalgamated. By employing non-overlapping windows for self-attention computation, the computational complexity shifts from a quadratic to a linear paradigm.

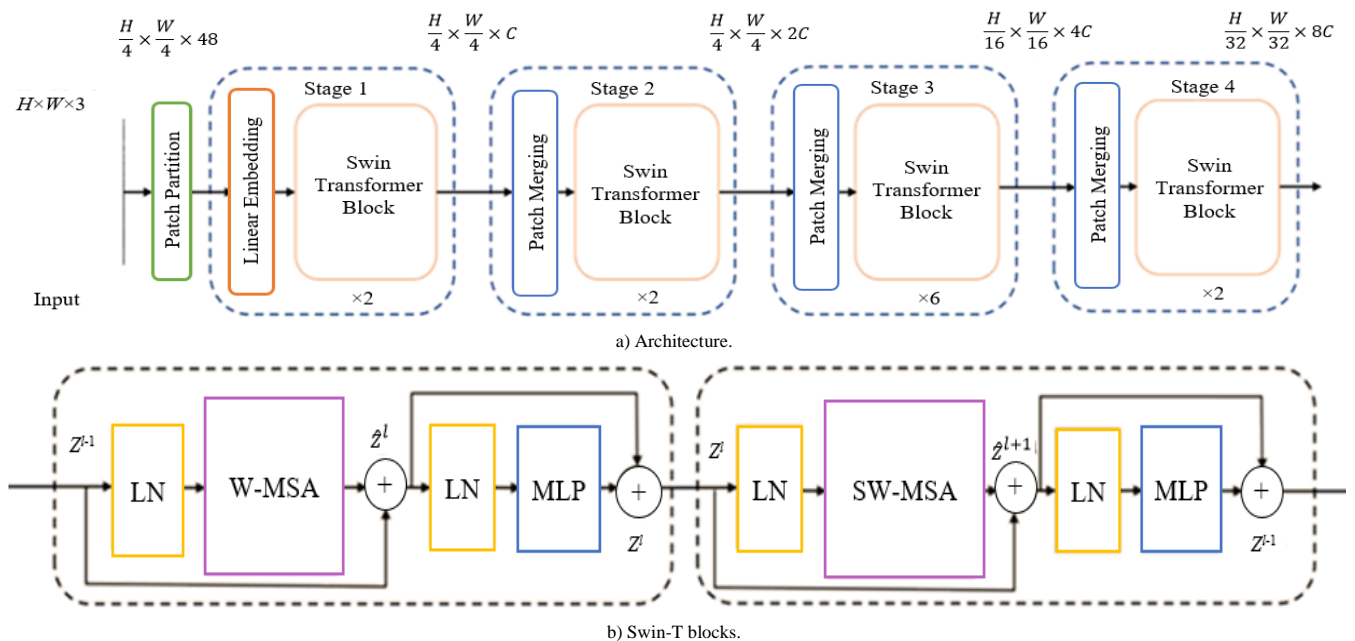


Figure 2. The general structure of the Swin-T.

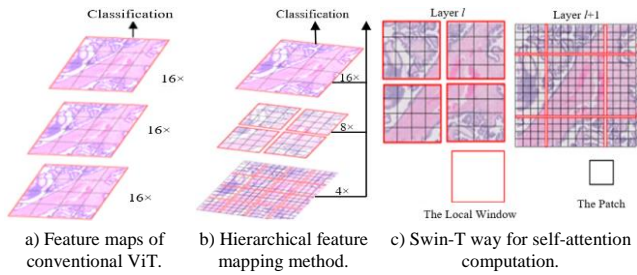


Figure 3. Comparing feature maps: ViT vs. Swin-T.

However, such partitioning would lead to a reduction in the interconnectedness of each individual window [12]. In tackling this particular issue, the Swin-T adopts the shifted window strategy, a method illustrated in Figure 3-c). In layer L (left), a standard window partitioning method is employed, where self-attention is calculated within each window. In the subsequent layer layer $L+1$ (right), the partitioning of the windows is shifted, leading to the formation of new windows. The self-attention calculation in the new windows extends beyond the boundaries of the previous windows in layer L , establishing connections between them [8]. The incorporation of shifted windows facilitates the integration of information across each window within the preceding layer, thus notably augmenting the model's receptive field. This aspect serves as a pivotal point of difference from conventional transformer structures. The graphical representation of the Swin-T structure is depicted in Figure 2-a). Comprising four distinct stages dedicated to the extraction of feature maps, each stage is characterized by the inclusion of Swin-T blocks. The dual components encompassing patch partition and linear embedding function analogously to the consolidation of patches. The operational mechanism of the Patch Merging module bears resemblance to the pooling layer of the CNNs, as it effectively in the above samples of input images.

2.4. The Blocks of the Swin Transformer

In contrast to the traditional Multi-head Self-Attention (MSA) module, the Swin-T block is structured around the concept of shifted windows (SW-MSA) [13]. There are two sequential Swin-T blocks. Each block consists of four components: a Layer Normalization (LN) layer, a MSA module, a residual connection, and a 2-layer Multi-Layer Perceptron (MLP) as shown in Figure 2-b). The Window-based Multi-head Self-Attention (W-MSA) module and the SW-MSA module are utilized in two consecutive transformer blocks, respectively. Using this window partitioning strategy, the procedure for computing the feature map across successive Swin-T blocks is detailed below:

$$\hat{z}^l = W - MSA(LN(z^{l-1})) + z^{l-1} \quad (3)$$

$$z^l = MLP(LN(\hat{z}^l)) + z^{l-1} \quad (4)$$

where \hat{z}^l represents the output features of the SW-MSA module for block l , and z^l denotes the output features

of the MLP module for the same block. The global context learning of feature representations is facilitated by the interaction between windows achieved through shifting. In the Shifted Window Multi-head Self-Attention (SW-MSA) mechanism, the window arrangement is adjusted towards the upper-left corner of the image. This adjusts the window configuration, allowing each window to consist of multiple sub-windows while maintaining the same number of patches [7]. The output of SW-MSA can be expressed as:

$$\hat{z}^{l+1} = W - MSA(LN(z^l)) + z^l \quad (5)$$

$$z^{l+1} = MLP(LN(\hat{z}^{l+1})) + \hat{z}^{l+1} \quad (6)$$

In a traditional ViT block, each token is compared to all other tokens, resulting in a quadratic increase in computational demands as the image resolution grows. Assuming each window has dimensions of M by M , the input images with height h and width w will be divided into windows of $h/M \times w/M$. The computational intricacy of the MSA is formulated as:

$$\Omega_{MSA} = 4hwc^2 + 2(hw)^2 C \quad (7)$$

while the computational intricacy of the W-MSA) is delineated as:

$$\Omega_{W-MSA} = 4hwc^2 + 2M^2hwC \quad (8)$$

In this context, let C denote the dimensionality. When M is predetermined to a constant size (default set to 7), the computational complexity of W-MSA correlates linearly with the multiplication of h and w .

3. Methodology

The attention mechanisms have been employed in various capacities in recent years; however, the emergence of transformers represents new neural network backbone that prominently leverage attention, particularly self-attention. The proposed algorithm utilizes five Swin-T models operating in parallel to classify five common ovarian carcinoma subtypes. Each Swin-T model is dedicated to classifying a specific subtype, enhancing the overall accuracy and efficiency of the classification process. The reason behind utilizing five parallel Swin-T models instead of a single multi-class model is essential due to the dataset's complexity, which includes five distinct ovarian cancer subtypes. Each subtype has unique features that are challenging for one model to differentiate effectively. By dedicating a separate model to each subtype, we enhance classification accuracy and reduce misclassification by allowing each model to specialize in its specific subtype. The workflow begins with WSI going via pre-processing and data sampling phase to create training and testing sets. The training set is used to train the Swin-T models, each focusing on one specific of the subtypes. In the parallel configuration, each Swin-T model processes its respective input images through a series of stages, including patch partitioning, linear

embedding, and Swin-T blocks. The models learn to extract and represent features specific to their assigned subtype, thereby improving classification performance. The outputs from the Swin-T models are then aggregated to form the final classification results. This parallel approach allows for specialized feature extraction and reduces the complexity of multi-class classification, leading to higher precision and recall rates, as reflected in the results. The use of multiple Swin-T models ensures that each subtype is classified with a high degree of accuracy, demonstrating the robustness and effectiveness of the proposed methodology. Figure 4 illustrates the overall workflow of the proposed methodology, while Algorithm (1)

provides the corresponding pseudocode. The hyperparameters of model training were optimized by implementing the grid search method, setting key parameters. This optimizations were crucial for enhancing model performance. Accordingly, the image dimensions of 224×224 pixels, the window size was specified as 7×7 , and the training spanned 100 epochs. Adam served as the optimizer with a learning rate of 0.0001. Furthermore, the patch size was defined as 4×4 , while ReLU function acted as the activation. Lastly, the output function was Sigmoid for the final predictions. The training hyperparameter configuration that utilized for the multi-Swin ViT models is provided in in Table 1.

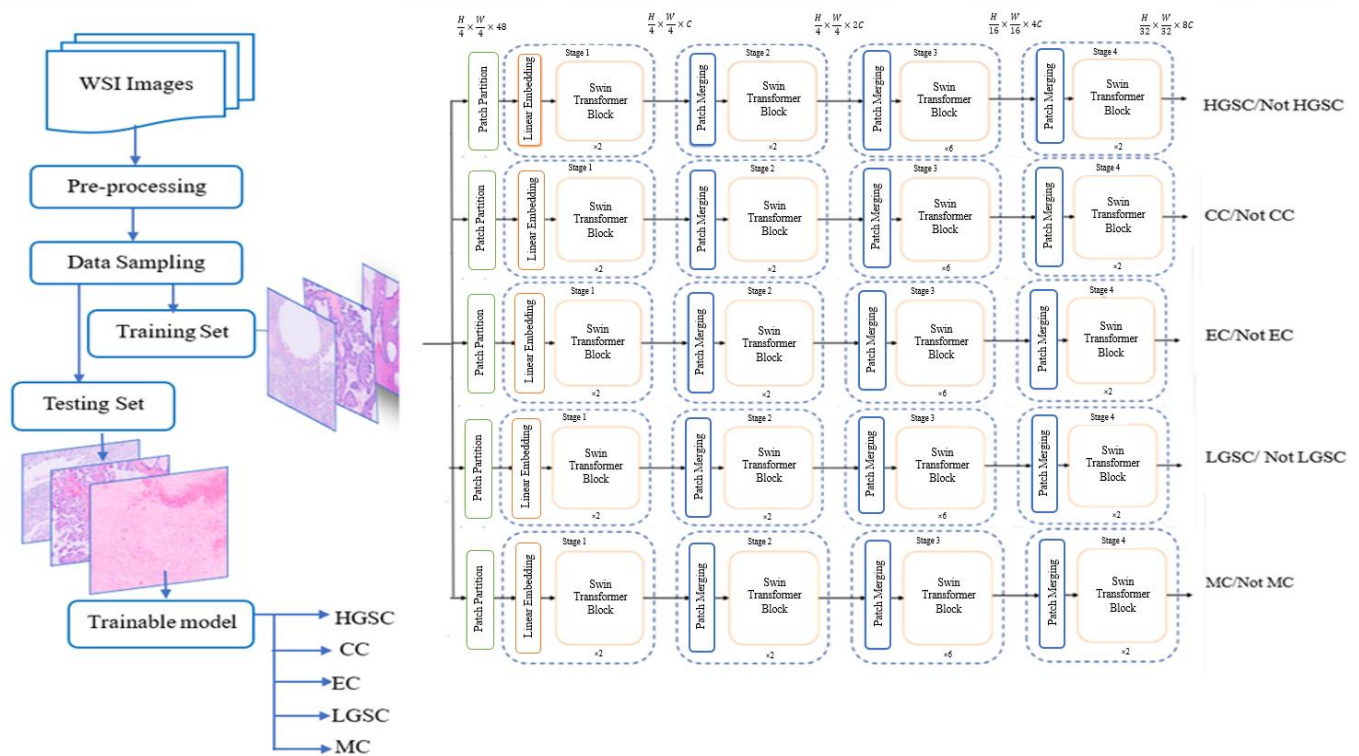


Figure 4. The general workflow of the proposed parallel Swin-transformer.

Algorithm 1: The Pseudocode for Parallel Swin-T Classification Algorithm.

#Step 1. Load and preprocess the dataset.

```
function load_and_preprocess_data():
    WSI_images=load_WSI_images()
    pre_images=preprocess_images(WSI_images)
    training_set, testing_set = data_sampling(pre_images)
    return training_set, testing_set
```

#Step 2. Initialize Swin-T models.

```
function initialize_swin_t_models():
    swin_t_HGSC=Swin_T1()
    swin_t_CC=Swin_T2()
    swin_t_EC=Swin_T3()
    swin_t_LGSC=Swin_T4()
    swin_t_MC=Swin_T5()
    return [swin_t_HGSC, swin_t_CC, swin_t_EC,
    swin_t_LGSC, swin_t_MC]
```

#Step 3. Train Swin-T models.

```
function train_models(training_set, swin_t_models):
```

```
    for model, subtype_images in zip(swin_t_models,
    training_set):
```

```
        #Each subset D_train(i) is used to train the corresponding
        Swin-T model M_i,
```

```
        #where i indicates the specific carcinoma subtype. The
        training process involves
```

```
        #optimizing the model parameters  $\theta_i$  by minimizing the loss
        function L:
```

```
        # $\theta_i^* = \operatorname{argmin}_{\theta_i} L(M_i(D_{\text{train}}(i); \theta_i))$ 
        model.train(subtype_images)
```

```
# Step 4. Test Swin-T models and classify subtypes.
```

```
function test_and_classify(testing_set, swin_t_models):
```

```
    results=[]
```

```
    for model, subtype_images in zip(swin_t_models,
    testing_set):
```

```
        predictions=model.predict(subtype_images)
        results.append(predictions)
```

```
    return results
```

```

# Step 5. Aggregate and evaluate results.
function aggregate_and_evaluate(results):
    combined_results=aggregate_results(results)
    evaluation_metrics=evaluate(combined_results)
    return evaluation_metrics

#Main function
function main():
    #Step 1. Load and preprocess data.
    training_set, testing_set=load_and_preprocess_data()
    #Step 2. Initialize Swin-T models.
    swin_t_models=initialize_swin_t_models()
    #Step 3. Train Swin-T models.
    train_models(training_set, swin_t_models)
    #Step 4. Test Swin-T models and classify subtypes.
    results=test_and_classify(testing_set, swin_t_models)
    #Step 5. Aggregate and evaluate results.
    evaluation_metrics=aggregate_and_evaluate(results)
    #Output evaluation metrics
    print(evaluation_metrics)

#Execute the main function
main()

```

Table 1. The configuration of the training hyperparameters.

Hyperparameters	Setting
Image size	224×224
Window size	7×7
Epoch	100
Optimizer	Adam
Learning rate	0.0001
Patch size	4×4
Activation function	ReLU
Output function	Sigmoid

4. Results and Discussion

This work was implemented using Python 3.9.10 on a high-performance system featuring an Intel Core i7-11800H CPU, an NVIDIA GeForce RTX 3060 GPU, and 16GB of DDR4 RAM. This hardware and software configuration provided a robust and efficient environment for the training and evaluation of the parallel Swin-T models, facilitating rigorous and high-quality analysis.

4.1. Metrics for Evaluating Performance

The assessment of the suggested approach utilizes common classification performance indicators, such as precision, recall, and the F1-score. Here, True Positive (TP) stands for the number of true positives, which are the instances accurately recognized. FN refers to the number of false negatives, meaning the instances that were missed. FP indicates the number of false positives, which are the instances wrongly classified as positive when they are actually negative. Precision is calculated as the fraction of TP instances among all instances predicted to be positive [14, 18], and it can be formulated as:

$$\text{Precision} = TP / (TP + FP) \quad (9)$$

The proportion of positive samples that are accurately predicted as positive is known as recall, and it can be

represented as follows:

$$\text{Recall} = TP / (TP + FN) \quad (10)$$

The F1-score represents the balanced assessment of the classifier's performance, calculated as the harmonic mean of precision and recall:

$$F1 - \text{Score} = \frac{2 \times ((\text{precision} * \text{recall}))}{(\text{precision} + \text{recall})} \quad (11)$$

Additionally, the accuracy of the model is defined as the ratio of correctly predicted instances to the total instances in the data:

$$\text{Accuracy} = (TP + TN) / (TP + FP + TN + FN) \quad (12)$$

4.2. The Performance of the Proposed Methodology

To evaluate the performance of the proposed methodology, the ovarian carcinoma Image Samples dataset was utilized. The sample of the comparison between training and validation accuracy and loss per epoch is displayed in Figure 5. The accuracy plot demonstrates that both training and validation accuracy consistently increase before stabilizing within a specific range. Conversely, in the loss graph, both loss curves initially trend towards zero, but after several epochs, the validation loss begins to diverge from the training loss.

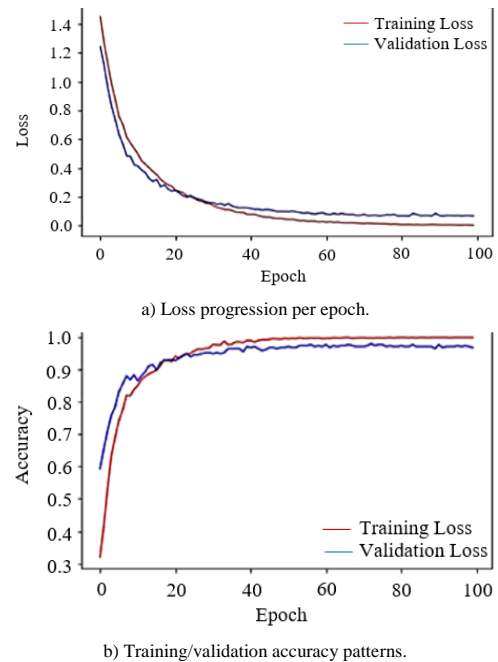


Figure 5. The proposed model training progression.

The evaluation of the effectiveness of the proposed methodology in classifying the five most prevalent subtypes of ovarian carcinoma was conducted using the ovarian carcinoma Image Samples dataset. This assessment utilized confusion matrices and the metrics of precision, recall, and F1-score, which collectively demonstrate the performance of the classification approach. The confusion matrices presented in Figure 5 for both the parallel Swin-Ts configuration Figure 6-a)

and the single Swin-T configuration Figure 6-b) offer valuable insights into the classification accuracy and the distribution of errors for each subtype. While Table 2 shows the performance evaluation metrics of the proposed methodology.

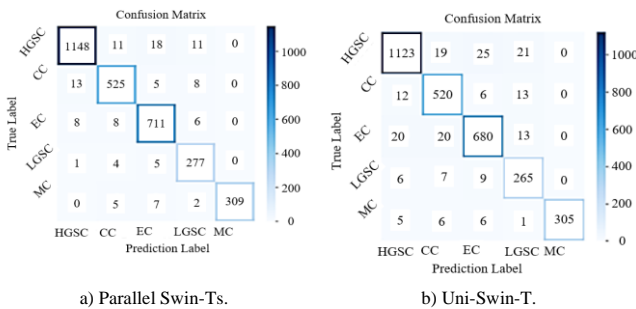


Figure 6. Confusion matrices.

Table 2. The performance evaluation metrics of the proposed methodology.

	Cancer subtype	Precision	Recall	F1-score
Parallel Swin-Ts	HGSC	0.98	0.97	0.97
	CC	0.95	0.95	0.95
	EC	0.95	0.97	0.96
	LGSC	0.91	0.97	0.94
MC	1	0.96	0.98	
Mean		0.958	0.964	0.96
Uni_Swin-T	HGSC	0.96	0.95	0.95
	CC	0.91	0.94	0.93
	EC	0.94	0.93	0.93
	LGSC	0.85	0.92	0.88
MC	1	0.94	0.97	
Mean		0.932	0.9326	0.932

For the parallel Swin-Ts setting, the precision on the HGSC subtype is 0.98, making it very accurate in which cases are true positive. The recall is a bit lower, at 0.97, showing that the model nicely catches positive instances. The F1 score, balancing the two, similarly presents high figures, leading to a value of 0.97, which leaves good model performance. The MC subtype, meanwhile, achieved the best performance metrics of 1 for precision and 0.96 for recall, resulting in an F1-score of 0.98. As such, the parallel model of Swin-Ts delivers a high level of effectiveness in correctly diagnosing MC cases with minimum incidences of false positives and false negatives. The EC subtype exhibited vital performance metrics, with a precision of 0.95 and a recall of 0.97, giving an F1-score of 0.96. In contrast, the LGSC subtype displayed a relatively reduced precision of 0.91 but a good recall of 0.97, providing an F1-score of 0.94.

The lower performance with LGSC compare to other classes can be due to several aspects, such as the complexity and variability of this subtype. These factors introduce difficulties in both the histopathological and molecular characteristics, distinguishing it from other subtypes. Furthermore, the limited labeled data for LGSC may have reduced the model's ability to generalize effectively. This means that, though the model shows a suitable identification of the actual cases of LGSC, it compares a more elevated frequency of

wrong identifications with the rest of the histological subtypes, especially in the form of false positives. As such, using parallel Swin-Ts, mean precision, recall, and F1-score of 0.958, 0.964, and 0.960, respectively, were derived.

The Uni-Swin-T approach, on the other hand, gives slight decreases in the performance measures. Notably, precision values reflected a decline in the LGSC of 0.85 against a recall of 0.92 and F1 of 0.88. It shows a low performance score when Uni-Swin-T correctly identifies the LGSC cases but with a higher rate of false positives and negatives. The Swin-T model can further reduce the values of recall (0.93) and F1-score (0.93) of the EC subtype than the parallel Swin-Ts. The values of precision and recall of the CC and HGSC subtypes are slightly less under the single Swin-Ts than under the parallel, resulting in a somewhat lower F1 score. This makes the average precision, recall, and F1-score over a number of the evaluation methodologies turn out to be approximately 0.932, 0.9326, and 0.932, respectively.

Overall, the parallel Swin-Ts configuration outperforms the Uni-Swin-T configuration in most subtypes, especially in terms of precision and recall for LGSC and EC subtypes. The superior performance of proposed methodology can be attributed to its enhanced ability to capture and distinguish intricate patterns within the ovarian carcinoma subtypes, thereby reducing misclassification errors and improving the overall robustness of the classification model.

On the other hand, the resource demands associated with running multiple parallel Swin-T models present a potential limitation, as they may lead to increased processing times and higher requirements for memory and computational power. Despite these challenges, the methodology has demonstrated effective performance with significant improvements in classification accuracy. The notable gains achieved highlight the method's potential value and effectiveness, even when faced with resource constraints.

4.3. Comparing State-of-the-Art Performance Evaluations

It is clear from Table 3 which demonstrates through a detailed comparative analysis with current methods that the proposed approach highlights significant strengths and potential improvements over established benchmarks. The effectiveness of the proposed methodology lies in its adoption of the parallel Swin-Ts that demonstrates significant advantages over traditional models. By reducing memory overhead and enhancing scalability, the methodology optimizes resource utilization, ensuring effective management of complex datasets. Furthermore, the utilization of shift window mechanism enables the model to adapt to various spatial relationships within images, thereby efficiently capturing intricate patterns and dependencies across different regions.

Table 3. The state-of-the-art and performance comparison.

Method	Image type	#Images	Data sampling		Performance %	
			Train	Test	Acc.	F1-score
AlexNet [20]	Cytological	20328	18295	2033	78.20	NA
RF [21]	H and E slides	16594	12 224	4370	94.6	94
CNN-CAE [8]	Ultrasound	1613	1,289	648	90.12	NA
ISTL [4]	WSI	1008	948	60	80.97	78.41
Ours	WSI	17,622	14,540	3,082	96.8	0.96

*Acc: Accuracy; NA: Not available.

5. Conclusions

A distinct methodology has been proposed via utilizing parallel Swin-T models for the subtype's classification of the ovarian carcinoma. Using the power of feature extraction and attention skills in Swin-Ts, this work proved mass improvement in performance classification in all subtypes, validated on an ovarian carcinoma WSI dataset. More importantly, it showed that the parallel Swin-Ts configuration outperformed the single Swin-T model in average precision, recall, and F1-score values of 0.958, 0.964, and 0.96, respectively. Notably, this parallel configuration achieved the best F1 scores for the MC and HGSC subtypes, proving that this is a practical approach to dedicating a specified model for each subtype. Although the LGSC subtype reached the worst performance compared to other ones and is more challenging in terms of this class accurate classification, it also benefited from this approach most, showing the improvement of such metrics as recall and F1-score to be 0.91, 0.97, and 0.94, respectively. This could be remedied by the parallel part, which allows for embedded directed learning and classification for each subtype and helps reduce misclassification errors. This not only enhances the classification process but also shows high potential in the application of transformer-based models in medical image analysis. Therefore, the proposed parallel Swin-Ts workflow shows promise for accurate and efficient classification of ovarian carcinoma subtypes. Future studies can extend this by integrating more data sources and applying the same methodology to other cancer types to make it more applicable and valuable in computational pathology.

References

[1] Chen L., Qiao C., Wu M., Cai L., Yin C., Yang M., Sang X., and Bai W., "Improving the Segmentation Accuracy of Ovarian-Tumor Ultrasound Images Using Image Inpainting," *Bioengineering*, vol. 10, no. 2, pp. 184, 2023. DOI:10.3390/bioengineering10020184

[2] D'Ascoli S., Touvron H., Leavitt M., Morcos A., Biroli G., and Sagun L., "ConViT: Improving Vision Transformers with Soft Convolutional Inductive Biases," *Journal of Statistical Mechanics*, vol. 2022, no. 11, pp. 1-27, 2022. DOI: 10.1088/1742-5468/ac9830

[3] Fahim T., Alam F., and Ahmmed K., "OVANet:

Dual Attention Mechanism Based New Deep Learning Framework for Diagnosis and Classification of Ovarian Cancer Subtypes from Histopathological Images," *IEEE Access*, vol. 12, pp. 131942-131953, 2024. <https://ieeexplore.ieee.org/document/10679974>

[4] Farahani H., Boschman J., Farnell D., Darbandsari A., Zhang A., Ahmadvand P., Jones S., Huntsman D., Kobel M., Gilks C., Singh N., and Bashashati A., "Deep Learning-based Histotype Diagnosis of Ovarian Carcinoma Whole-Slide Pathology Images," *Modern Pathology*, vol. 35, no. 12, pp. 1983-1990, 2022. DOI:10.1038/s41379-022-01146-z

[5] Gao Z., Hong B., Zhang X., Li Y., Jia C., Wu J., Wang C., Meng D., and Li C., "Instance-based Vision Transformer for Subtyping of Papillary Renal Cell Carcinoma in Histopathological Image," in *Proceedings of the Medical Image Computing and Computer Assisted Intervention*, Strasbourg, pp. 299-308, 2021. DOI:10.1007/978-3-030-87237-3_29

[6] Hwangbo S., Kim S., Kim J., Eoh K., Lee C., Kim Y., Suh D., Park T., and Song Y., "Development of Machine Learning Models to Predict Platinum Sensitivity of High-Grade Serous Ovarian Carcinoma," *Cancers*, vol. 13, no. 8, pp. 1875, 2021. DOI: 10.3390/cancers13081875

[7] Jiang Y., Zhang Y., Lin X., Dong J., Cheng T., and Liang J., "SwinBTS: A Method for 3D Multimodal Brain Tumor Segmentation Using Swin Transformer," *Brain Sciences*, vol. 12, no. 6, pp. 1-15, 2022. DOI: 10.3390/brainsci12060797

[8] Jung Y., Kim T., Han M., Kim S., Kim G., Lee S., and Choi Y., "Ovarian Tumor Diagnosis Using Deep Convolutional Neural Networks and a Denoising Convolutional Autoencoder," *Scientific Reports*, vol. 12, pp. 1-10, 2022. DOI:10.1038/s41598-022-20653-2

[9] Kaggle, Ovarian-Cancer-Classification-Iamvishaldubey, <https://www.kaggle.com/code/charu31/ovarian-cancer-classification-iamvishaldubey/input>, Last Visited, 2024.

[10] Koshiyama M., Matsumura N., and Konishi I., "Subtypes of Ovarian Cancer and Ovarian Cancer Screening," *Diagnostics*, vol. 7, no. 1, pp. 12, 2017. DOI:10.3390/diagnostics7010012

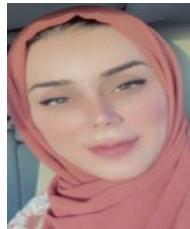
[11] Liu Y., Lawson B., Huang X., Broom B., and Weinstein J., "Prediction of Ovarian Cancer Response to Therapy Based on Deep Learning Analysis of Histopathology Images," *Cancers*, vol. 15, no. 16, pp. 1-13, 2023. DOI:10.3390/cancers15164044

[12] Liu Z., Lin Y., Cao Y., Hu H., Wei Y., Zhang Z., Lin S., and Guo B., "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows," *IEEE/CVF International Conference*

- on Computer Vision, Montreal, pp. 9992-10002, 2021. DOI:10.1109/iccv48922.2021.00986
- [13] Liu Z., Tan Y., He Q., and Xiao Y., "SwinNet: Swin Transformer Drives Edge-Aware RGB-D and RGB-T Salient Object Detection," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 7, pp. 4486-4497, 2022. DOI: 10.1109/tcsvt.2021.3127149
- [14] Madhukar B., Bharathi S., and Ashwin M., "Classification of Breast Cancer using Ensemble Filter Feature Selection with Triplet Attention Based Efficient Net Classifier," *The International Arab Journal of Information Technology*, vol. 21, no. 1, pp. 17-31, 2024. DOI: 10.34028/iajit/21/1/2
- [15] Narayanan L., Krishnan S., and Robinson H., "A Hybrid Deep Learning Based Assist System for Detection and Classification of Breast Cancer from Mammogram Images," *The International Arab Journal of Information Technology*, vol. 19, no. 6, pp. 965-974, 2022. DOI:10.34028/iajit/19/6/15
- [16] Sengupta D., Ali S., Bhattacharya A., Mustafi J., Mukhopadhyay A., and Sengupta K., "A Deep Hybrid Learning Pipeline for Accurate Diagnosis of Ovarian Cancer Based on Nuclear Morphology," *PLoS One*, vol. 17, no. 1, pp. 1-20, 2022. DOI: 10.1371/journal.pone.0261181
- [17] Sun K., Sun L., Zhao Y., Chen Y., Hao X., Liu H., Liu X., and Chen J., "XGBG: A Novel Method for Identifying Ovarian Carcinoma Susceptible Genes Based on Deep Learning," *Frontiers in Oncology*, vol. 12, pp. 1-7, 2022. DOI:10.3389/fonc.2022.897503
- [18] Tummala S., Kim J., and Kadry S., "BreaST-Net: Multi-Class Classification of Breast Cancer from Histopathological Images Using Ensemble of Swin Transformers," *Mathematics*, vol. 10, no. 21, pp. 1-15, 2022. DOI: 10.3390/math10214109
- [19] Wang S., Liu Z., Rong Y., Zhou B., Bai Y., Wei W., Wei W., Wang M., Guo Y., and Tian J., "Deep Learning Provides a New Computed Tomography-based Prognostic Biomarker for Recurrence Prediction in High-Grade Serous Ovarian Cancer," *Radiotherapy and Oncology*, vol. 132, pp. 171-177, 2019. DOI:10.1016/j.radonc.2018.10.019
- [20] Wu M., Yan C., Liu H., and Liu Q., "Automatic Classification of Ovarian Cancer Types from Cytological Images Using Deep Convolutional Neural Networks," *Bioscience Reports*, vol. 38, no. 3, 2018. DOI: 10.1042/bsr20180289
- [21] Zhou R., Zhao B., Ding H., Fu Y., Li H., Wei Y., Xie J., Chen C., Yin F., and Huang D., "Survival Prediction of Ovarian Serous Carcinoma Based on Machine Learning Combined with Pathological Images and Clinical Information," *AIP Advances*, vol. 14, no. 4, pp. 1-13, 2024. DOI:10.1063/5.0196414
- [22] Zhuang H., Li B., Ma J., Monkam P., Qian W., and He D., "An Attention-based Deep Learning Network for Predicting Platinum Resistance in Ovarian Cancer," *IEEE Access*, vol. 12, pp. 41000-41008, 2024. <https://ieeexplore.ieee.org/document/10472495>



Lubna ALkahlia is a lecturer at the University of Nineveh, College of Information Technology, Software Department. She specializes in Artificial Intelligence, holding a Ph.D. in this field. Her academic and research pursuits are centered around advancing the domain of Artificial Intelligence, contributing to both theoretical and applied aspects of the discipline.



Jwan Saeed received a B.Sc. degree in Computer Science from Mosul University, in 2002 and the M.Sc. degree in computer science from Duhok University, in 2010. Currently, she has a Ph.D. degree from the Technical Informatics College of Akre, Duhok Polytechnique University. Her research interests include the Applications of Artificial Intelligence, Computer Vision and Pattern Recognition.



Maher Hussein completed his Bachelor's degree in Computer Science from Al-Hadbaa University College in Mosul in 2008. He earned a Master's degree in the same field from the University of Mosul in 2013. His academic journey culminated in a Ph.D. in Computer Science with a specialization in Artificial Intelligence. His research interests Encompass Artificial Intelligence, Pattern Recognition, and Computer Vision.