Journal of Intelligent Systems and Internet of Things
Vol. 17, No. 02, PP. 238-249, 2025
DOI: https://doi.org/10.54216/JISIoT.170215

Multi-View Feature Learning Approach in Deep Learning Model for Improving Endometrial Cancer Detection from Medical Images

Karthick Natarajan[1,*], Nithya Palanisamy[2]

[1]Research Scholar, Department of Computer Science, PSG College of Arts and Science, Coimbatore- 641014, Tamilnadu, India

[2]Associate Professor and Head , Department of Networking and Mobile Application, PSG College of Arts and Science, Coimbatore- 641014, Tamilnadu, India

Emails: karthickphd.vlbjcas@gmail.com; nithi.selva@gmail.com

Abstract

An accurate diagnosis of Endometrial Cancer (EC) is crucial for gynecologists, as different types may require specific treatments. Radiomics, a quantitative method, can help analyze and quantify image heterogeneity, aiding in lesion diagnosis. Previous research introduced a Transformer-based Semantic-Aware U-Net with Deep Endometrial Cancer Prediction (TSA-UNet-DeepECP) to segment and classify EC stages in Magnetic Resonance Imaging MRI scans. However, the heterogeneous properties of input scans can affect the DeepECP model's performance. Hence, this study presents the TSA-UNet with an Improved DeepECP model (TSA-UNet-IDeepECP) for EC stage classification. This IDeepECP model incorporates a multi-view learning approach, combining local 2D MRI image information with global 3D MRI image information. First, the endometrium MRI scans are collected, augmented, and segmented using the TSA-UNet model. Various Deep Learning (DL) models, one for 2D and one for 3D, are fed the segmented images. In contrast to the 3D view model, which collects global information from 3D MRI images, the 2D view model primarily recovers local features from 2D MRI data. The multi-view DeepECP model is trained using these combined characteristics. A Fully Connected (FC) layer and the softmax classifier are used for classifying EC stages using the combined features. When compared to traditional models, a TSA-UNet-IDeepECP model achieves better performance in EC detection from MRI images.

Received: January 27, 2025 Revised: March 30, 2025 Accepted: June 05, 2025

Keywords: Endometrial cancer; MRI; TSA-UNet; DeepECP; Heterogeneity; Multi-view learning

1. Introduction

One of the most prevalent malignant tumors affecting the female reproductive system, namely Endometrial Cancer (EC), is on the rise as the population ages and more people become overweight. The majority of cases occur in the endometrium, leading to the term "EC." Early-stage EC has a favorable prognosis, but advanced-stage EC has a poor prognosis due to metastasis [1-2]. In 2024, an estimated 13,250 uterine cancer deaths occurred in the US, with an annual increase of 3.0% [3]. Uterine corpus cancer ranks ninth among women's cancers in the US, accounting for 3.4% of new cases. Uterine cancer has a relative survival rate of 80.8% after 5 years, albeit it varies by lymph node stage at diagnosis [4]. In India, there are thousands of new cases annually, with an incidence rate of 2.5 to 3 per 100,000 women [5]. Depending on how far the tumor has spread, the International Federation of Gynecology and Obstetrics (FIGO) classifies EC as either stage I or stage IV [6]. Tumor invasion into the myometrium is less than half in Stage IA and half or greater in Stage IB. Stage IA has a 5-year survival rate of 90–

96%, and Stage IB has a 78–87% following surgery. According to [7], the survival rate drops to 66.0% when myometrial invasion is greater than 50% and 92.6% when it is less than 50%. Radiation treatment and a lymphadenectomy may be necessary for high-risk patients, although a hysterectomy may be sufficient for low-risk patients [8]. It is more common for Asian women to acquire EC at a younger age and in a more advanced stage compared to other populations [9]. An automated predictive framework for early EC staging is needed to improve diagnostic efficiency and inform treatment recommendations. Local EC staging is best done with MRI or contrast-enhanced dynamic MRI. The ESUR guidelines recommend MRI for determining disease severity in newly diagnosed patients [10], but radiologists' interpretations may differ based on their expertise.

In recent years, Deep Learning (DL) has become a popular Computer-Aided Diagnosis (CAD) method in image recognition. DL can automatically identify target areas when trained on large datasets [11]. Medical image processing widely uses this technology for various applications [12], like tumor classification, cancer diagnosis, cancer recurrence prediction, medical image segmentation, etc. Some DL studies have also focused on evaluating EC. Urushibara et al. [13] assessed the effectiveness of a Convolutional Neural Network (CNN) in diagnosing EC from MRI scans. The CNN performed well in distinguishing healthy and tumor images but had difficulties differentiating EC from other benign lesions. Healthcare systems have introduced transfer-learning models to address this issue. However, in medical imaging, challenges arise due to variations in input image features between training and testing datasets. Furthermore, the scarcity of training images for sagittal sections compared to axial MRI sections has an impact on model training efficiency.

In response to these issues, a new DeepECP model has been developed that accurately identifies and categorizes various stages of EC from MRI scans. It uses a Feature Entanglement Generative Adversarial Network (FE-GAN) architecture to generate synthetic MRI images and employs a DL-based multi-modal fusion approach to fuse MRI sequences from multiple modalities. To capture feature vectors from multi-modality MRI sequences, such as axial, coronal, and sagittal, it employs three CNN-LSTM networks with FC layers. An attention technique combines the feature vectors; a softmax classifier classifies the EC stages into Stages I, II, and III. By using a huge volume of MRI sequences, this approach increases training efficiency and accuracy in recognizing EC phases. However, MRI scan segmentation is required for the exact identification and classification of EC phases. The U-Net model is commonly utilized in medical image segmentation, despite its limitations in long-range associations. In addition, transformers with self-attention systems offer a unique approach despite their low-level localization power limitations.

A new model dubbed Transformer-based Semantic-Aware U-Net (TSA-UNet) with DeepECP has been created to precisely partition and classify several stages of EC in MRI images in order to solve this issue. The TSA-UNet paradigm integrates the transformer and U-Net ideas, which comprise encoder, fusion, and decoder modules. A CNN-transformer hybrid model, the encoder gathers multi-level bitemporal features and encodes tokenized sequences to retrieve high-level information. To improve feature distinction and lower semantic differences, the Transformer unit applies Semantic Guidance Attention (SGA). An Adaptive Cross-Fusion (ACF) unit combines bitemporal traits. A Cascaded Upsampling Decoder (CUD) simultaneously integrates low-level CNN and high-level transformer characteristics to precisely identify and segregate various stages of EC in MRI. The DeepECP model then classifies the segmented images. The performance of the DeepECP model, on the other hand, can be compromised by varied qualities of input scans.

Hence, this manuscript introduces the TSA-UNet-IDeepECP model for classifying stages of EC. The IDeepECP model utilizes a multi-view learning strategy to effectively utilize both local and global data from MRI scans. The model's goal is to enhance EC stage categorization by integrating 2D and 3D MRI scans and using separate DL methods for each view. The main contributions of this study include:

1. First, endometrial MRI scans are collected and split into two sets: one for training and one for testing. Before the TSA-UNet model for MRI segmentation is utilized to extract the Region Of Interest (ROI), a training set is augmented by the FE-GAN.
2. Subsequently, the development of the IDeepECP multi-view learning model involves utilizing various DL frameworks to train MRI images from diverse views. The features extracted from these different views are combined and inputted into the softmax layer for efficient classification of different stages of EC.

Here is the structure of the manuscript: a literature review is included in Section 2. An explanation of the TSA-UNet-IDeepECP model is presented in Section 3, and its test results are presented in Section 4. Section 5 summarizes the study and suggests future research.

2. Literature Survey

This section discusses recent studies on using AI frameworks to identify and categorize EC from MRI scans. Mao et al. [14] created an automated early EC diagnosis system using MRI scans. They utilized semantic segmentation with the U-Net architecture to identify uterine and lesion areas. The classification was based on the lesion-to-uterine area ratio, using threshold values from the ROC curve. Nonetheless, limited patient cases and the absence of data from healthy persons led to poor classification performance. Tao et al. [15] assessed the performance of shallow CNN, ResNet, and optimized ResNet in identifying EC from MRI features. Due to the limited sample size, the recognition accuracy was low. Additionally, the experimental data could not fully eliminate the influence of subjective factors. Yang et al. [16] combined classical radiomics and DL features from MRI scans to differentiate between low- and high-risk EC patients. However, missing clinical data led to subpar diagnostic accuracy. This single-center retrospective study may not generalize well to multiple centers due to data heterogeneity.

Liu et al. [17] created an MRI-based radiomics model using a Multi-Variate Logistic Regression (MVLR) to differentiate between benign and malignant endometrial lesions. On the other hand, lacking important characteristics connected to inherent diversity restricted the classification performance. Estimating Myometrial Infiltration (MI) depth, Xiong et al. [18] developed a multi-step DL algorithm to classify endometrial cancer on sagittal T2-Weighted Images (T2WI). They applied an attention U-Net and an SSD-based recognition system to segment MRI data. An elliptical fitting technique was then applied to produce a uterine cavity contour, which guided the MI depth for categorization. The performance, though, was affected by the varied characteristics of pictures from many sources. Meng et al. [19] evaluated the predictive usefulness of MRI-based radiomics for early-stage EC risk classification and staging. They predicted using a Multi-Layer Perceptron (MLP), filtered 3D radiomic characteristics from MRI scans using Pearson's correlation coefficient, and retrieved them. Variations in MRI scanning settings and manual tumor segmentation could create biases and affect prediction accuracy.

Cui et al. [20] offered a framework for the automatic segmentation of atypical endometrial hyperplasia, EC, uterine leiomyoma, and endometrial polyps on MRI. Combining components like the pixel interaction, classification re-activation map, inter-pixel connection network, or Deeplab v3+ units helped this two-stage model to increase segmentation accuracy. On the other hand, cases with irregular shapes showed slightly poorer performance. Their tiny sample size and the absence of usage of other datasets further contributed to problems with generalizability. Contrasting differences by risk category and molecular classification, Bae et al. [21] investigated MRI findings in endometrial cancer patients. By examining historical MRI data, researchers sought links between imaging features and genetic markers. The study was retrospective, and some patients lacked full MRI data, which were constraints.

The literature review indicates the existence of a research gap in the use of AI to identify and classify EC using MRI images due to variation in the input images. Existing studies do not comprehensively tackle challenges like the heterogeneous image sources, uneven image quality, varying lesion shapes, and feature heterogeneity. The main idea behind current multi-modal approaches in medical imaging is to capture and merge distinct characteristics from various image modalities, like T1-weighted MRI, T2-weighted MRI, and Contrast-Enhanced (CE) T1 MRI. Nevertheless, they also face difficulties in acquiring heterogeneity or feature variations of different images from a single MRI modality.

Thus, this paper proposes the novel TSA-UNet-IDeepECP model that extracts local features (2D MRI images) and global features (3D MRI images) of MRI images with similar modalities, such as T1, T2, and CE-T1. Then, it merges both local and global features to learn a complementary feature representation across various modalities to accurately classify ECs. By leveraging multiple views within similar modalities, this model is an effective method to handle feature heterogeneity compared to conventional multi-modal methods. It also provides improved robustness to image resolution or orientation variation without the use of multi-modal strategies.

3. Proposed Methodology

This part of the paper presents the TSA-UNet-IDeepECP model to segment and categorize ECs from MRI data. A simplified version of this study is shown in Figure 1.

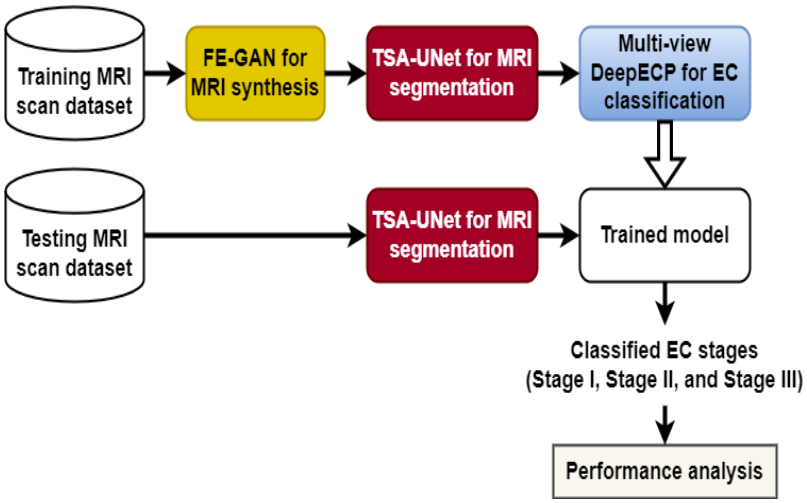


Figure 1. Schematic Representation of this Study

3.1 FE-GAN Model for MRI Synthesis

The FE-GAN algorithm creates synthetic MRI scans to increase the training dataset. It is made up of a generator ($\mathcal{G}$) and a discriminator ($\mathcal{D}$). While the discriminator separates synthetic from actual scans using a CNN, the generator builds realistic MRI pictures using an encoder-decoder network.

3.2 TSA-UNet Model for MRI Segmentation

The TSA-UNet model segmented MRI scans using encoder, fusion, and decoder modules. A CNN extracted semantic characteristics, while a Transformer modelled context correlations in the high-level feature space. Upsampling Transformer features during decoding can improve performance but may lead to loss of detail due to the smaller size of high-level features than the actual image. To address this, lower-level CNN features are introduced. The Transformer uses SGA to reduce data, bridge semantic gaps, and improve feature fusion. ACF combines CNN and Transformer features to fuse two bitemporal traits. CUD resizes fused features, and a convolution (conv) layer segments ROIs during decoding.

3.3 Improved DeepECP Model for EC Classification

The segmented ROIs are fed into two separate networks: a 3D model using ResNet18 for global feature learning and a 2D model using CNN-LSTM for local feature learning. Each model is explained in the following section. The IDeepECP model comprises two DL frameworks, each focusing on different views for EC classification, as depicted in Figure 2. The model hyperparameters are listed in Table 1.

Table 1: Hyperparameters for TSA-UNet-IDeepECP Model

Hyperparameters	**Range**
Learning rate for 2D model	0.0001
Learning rate for 3D model	0.001
Loss function	Cross-entropy
Optimizer	Adam
Batch dimension	64
Epochs	250
Dropout rate	0.5
Momentum	0.99
Weight decay	0.0005

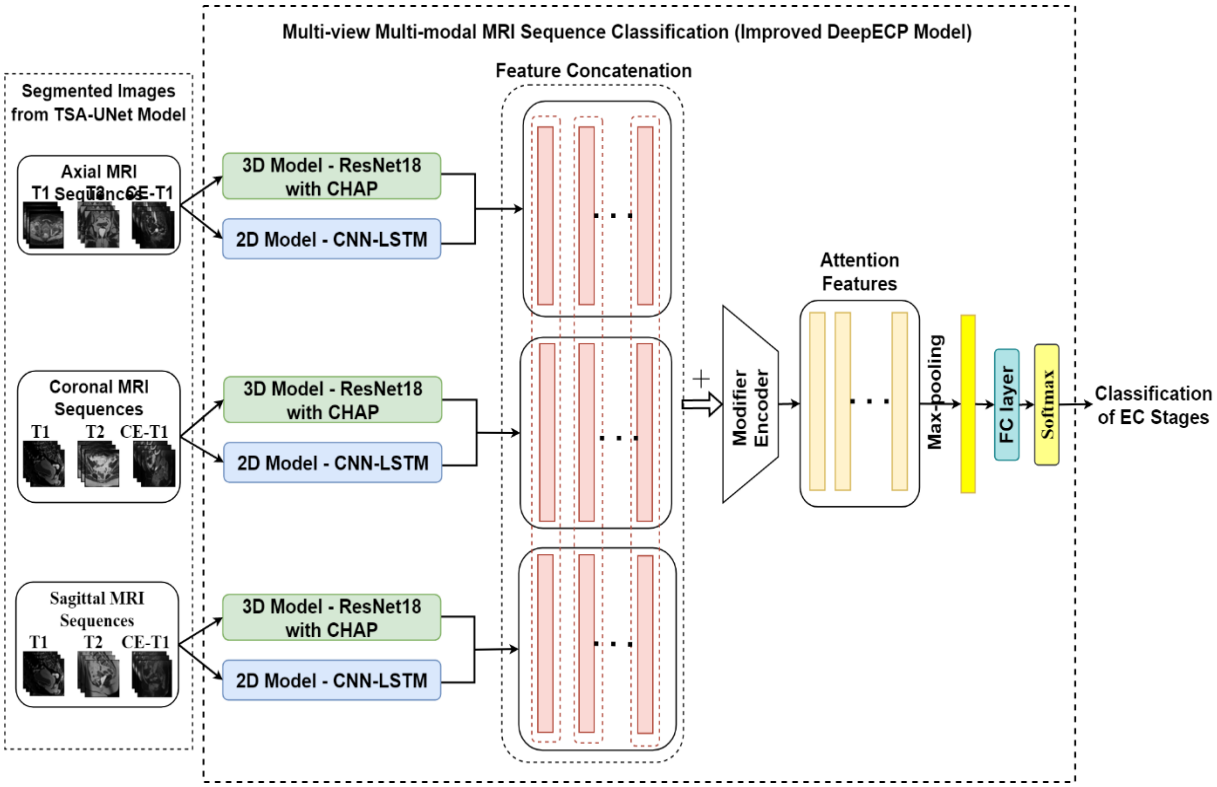


Figure 2. Structure of Improved DeepECP Model for EC Classification

3.3.1 3D View Model for Global Feature Learning

The ROIs segmented from the TSA-UNet are used as input for the 3D global view of the entire endometrium. To extract comprehensive global features related to EC, 3D CNNs are employed. The challenge lies in effectively learning useful features. With the increasing depth of CNN models since AlexNet's introduction in 2012, the issue of gradient vanishing during training has become more noticeable. Residual Network (ResNet) implements identity shortcuts to circumvent multiple layers in order to resolve this matter. The resultant output x_l of the layer l is as follows if the layer $l-1$ is linked to layer l:

$$x_l = H_l(x_{l-1}) + x_{l-1} \quad (1)$$

The non-linear transformation function $H(\cdot)$ in Eq. (1) includes the convolution operation, Rectified Linear Unit (ReLU), and Batch Normalization (BN). By incorporating shortcut connections, ResNet streamlines deep network training. For EC-related data, ResNet employs linear activation for identity mapping. In contrast, non-linear activation is employed to eliminate irrelevant data, thereby preventing the loss of valuable information. In this way, ResNet effectively mitigates network degradation caused by excessive depth, thereby allowing the model to acquire more features that are sophisticated. The ResNet18 is chosen as the base model for this study due to its shallow and computationally efficient structure for extracting complex features compared to the ResNet50 using different MRI sequences.

Therefore, it is used to learn endometrium ROIs in a 3D view of T1, T2, and CE-T1 MRI images for extracting EC-related global features. To further enhance the model's ability to learn 3D features comprehensively, dilated convolutions are added to the standard ResNet. By incorporating dilated convolutions, the model can achieve a larger Receptive Field (RF) beyond the limitations of a $3 \times 3 \times 3$ convolution kernel. If the expansion rate is r and the convolution kernel is k, the RF of the convolution with dilation can be determined by

$$RF = [k + (k-1)(r-1)]^3, r > 1 \quad (2)$$

RF of k^3 is used for ordinary convolution when $r = 1$. When $r = 2$, dilated convolution increases the RF to 25 without adding parameters. However, misuse of dilated convolutions can lead to a mesh effect. To prevent this, a Cascaded Hierarchical Atrous Pyramid (CHAP) module is used. A structure map for $n \times n \times n \times f$ is generated by applying a $1 \times 1 \times 1$ segmentation to the input. Subsequently, the hierarchical features are combined as a feature map of the same size. Figure 3. illustrates the pyramid of RFs in a hierarchical structure. In the second

structure, the original RFs are enlarged by passing feature maps using an individual route generated by the root layer and an extension layer. It generates new ones that exceed the original range. The number of newly generated scopes increases as the dilated rates for the root and expansion layers are increased.

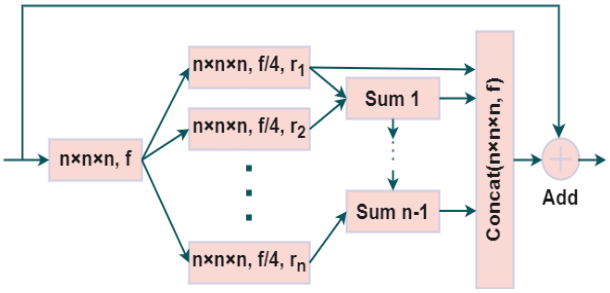


Figure 3. CHAP Module

For instance, the output of a hierarchy with its radius collection of (4,6,8,10,12,14,16,18,20,22,24,26) that is passed through the path with dilated rates of (3,1) is an output with a radius array of (8,10,12,14,16,18,20,22,24,26,28,30), which introduces two new radius values of 28 and 30. Changing the path to dilated rates of (3,3) adds a new radius of 32. $r_1, r_2, \ldots, r_n$ are different combinations of dilation rates in the CHAP module in Figure 3. To guarantee high-quality output, the CHAP module uses the residual operation to combine data from several RFs prior to the residual operation. By replacing one layer in the standard ResNet18 model with the CHAP module, this model can better learn 3D global features from endometrium MRI images. The topology of the 3D model that was created with ResNet18 through the CHAP module is shown in Figure 4.

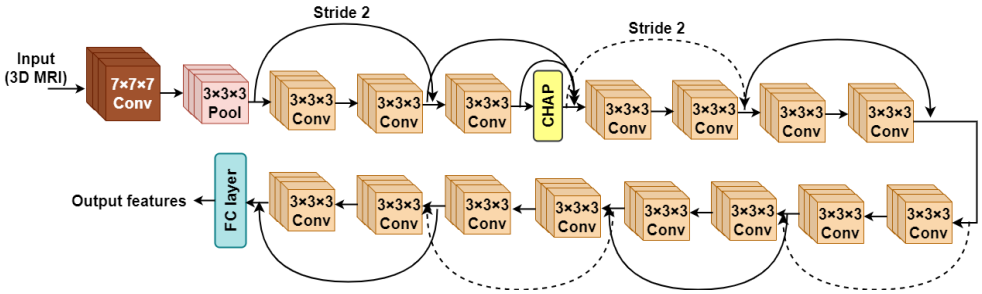


Figure 4. Structure of ResNet18 with CHAP Module for Global Feature Learning

3.3.2 2D View Model for Local Feature Learning

The 2D local view makes use of the CNN-LSTM network to extract the spatiotemporal features of MRI images. There are three separate parts to it, and they handle various MRI modalities. To learn local spatiotemporal information, each branch uses three convolutional layers next to an LSTM layer, as shown in Figure 5.

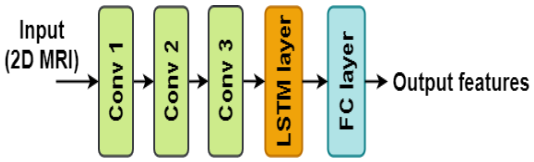


Figure 5. Structure of CNN-LSTM Network for Local Feature Learning

3.3.3 Fusion of Global and Local Features

Both 2D and 3D models undergo different training processes. A modifier encoder with attention strategy to learn relevant lesion characteristics merges the features from each model, such as global and local. Lastly, the EC stages are classified using the FC layers, which are followed by the softmax function. The IDeepECP model integrates 2D and 3D models for multi-view learning, allowing it to acquire both global and local data gathered from 3D

images. Thus, the TSA-UNet-IDeepECP model can handle input heterogeneity and accurately classify various stages of EC.

3.4 Potential Integration of TSA-UNet-IDeepECP with IoT Systems for Real-Time Diagnosis

For real-time diagnosis, the proposed TSA-UNet-IDeepECP model can be integrated with Internet of Things (IoT) and edge devices. In IoT-based clinical systems, edge computing is utilized to process MRI scans instantly for local decision support. Cloud computing is utilized to execute DL models for global decision support regarding EC categorization because of its computational power and scalability. Initially, MRI scans from different patients are gathered and pre-processed at the local edge devices within clinics/hospitals. This system accelerates the image collection process. These scans are then sent to cloud storage systems, where the TSA-UNet-IDeepECP model is implemented for EC categorization. Thus, the edge and cloud-based deployment of this proposed model can provide high-performance computing capabilities, enabling the TSA-UNet-IDeepECP model to effectively handle large-scale high-resolution MRI scans and deliver clinicians with precise outcomes.

4. Experimental Results

This section evaluates the effectiveness of the TSA-UNet-IDeepECP model and compares it with previous models, including CNN [13], Optimized ResNet [15], MVLR [17], and MLP [19]. The proposed and existing models were implemented in Python and tested on two distinct datasets described below.

4.1 Dataset Details

This study utilizes two publicly available archives to access MRI scans of both healthy individuals and patients with EC.

1. The Cancer Genome Atlas Uterine Corpus Endometrial Carcinoma (TCGA-UCEC) database [22] is a valuable scholarly resource that connects tumor features to gene sequences derived from medical scans of TCGA patients. TCIA stores imaging data, while the Genomic Data Commons (GDC) website stores medical, chromosomal, and clinical information. Scientists can analyze the TCGA/TCIA datasets using patient IDs to study connections between the tissue genome, radiography traits, and clinical outcomes. Globally, TCGA samples, with approximately 500 samples per tumor category, are collected to yield diverse image files with varying scanner modalities, manufacturers, and acquisition procedures. There are 65 individuals, 226 investigations, 912 sequences, 75,829 images, and four different types of imaging (CT, CR, MRI, and PET) in this collection.
2. Participants in the National Cancer Institute's Clinical Proteomic Tumour Analysis Consortium UCEC (CPTAC-UCEC) study are included in the CPTAC-UCEC dataset [23]. This group does genetic and proteomic research to figure out how tumors start at the molecular level. By making the pathology and MRI scans of CPTAC patients available to the public, TCIA has allowed researchers to examine tumor features in connection with proteomic, inherited, and clinical data. All types of tumor imaging data are included in the CPTAC-cancer group of the TCIA set. The MRI scans are derived from standard diagnostic scans performed on patients before clinical treatment, resulting in heterogeneous databases due to variations in scanner modalities, manufacturers, and acquisition procedures. There is a total of 250 subjects, 103 investigations, 1655 series, and 153,199 scans in this dataset, which includes (MRI/PET/CT) scans.

In this multi-model categorization, the input for separate RGB channels is composed of T1, T2, as well as the CE-T1 weighted MRI sequences (as illustrated in Figure 6.). Once these datasets are augmented by the FE-GAN, the TCGA-UCEC dataset has 1811 training images and 1169 test images, whereas the CPTAC-UCEC includes 3326 training images and 1690 test images.

4.2 Evaluation Metrics

Accuracy: It measures the fraction of properly detected EC stages to the overall images tested:

$$Accuracy = \frac{True\ Positive\ (TP) + True\ Negative\ (TN)}{TP + TN + False\ Positive\ (FP) + False\ Negative\ (FN)} \quad (3)$$

TP is the total +ve images properly recognized, TN is the total -ve images properly recognized, FP is the total -ve images imperfectly recognized as +ve, and FN is the total +ve images

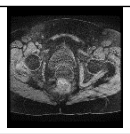
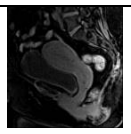
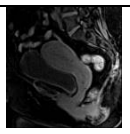
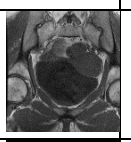
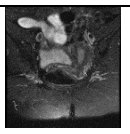
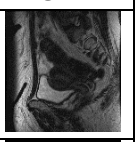
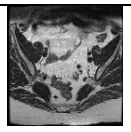
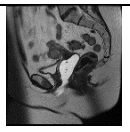
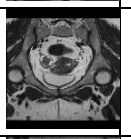
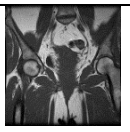
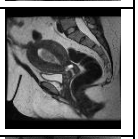
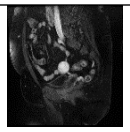
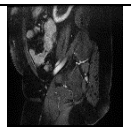
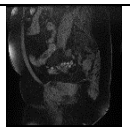
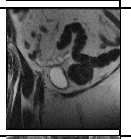
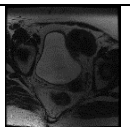
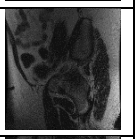
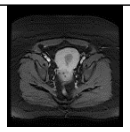
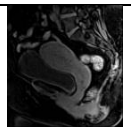
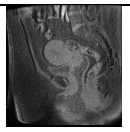
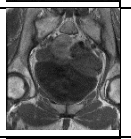
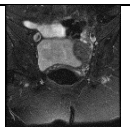
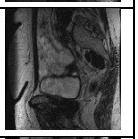
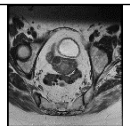
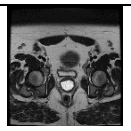
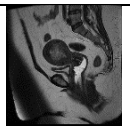
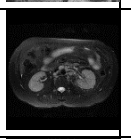
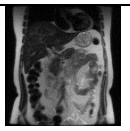
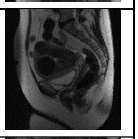
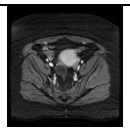
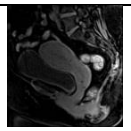
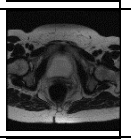
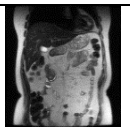
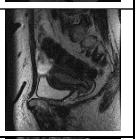
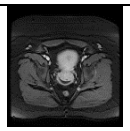
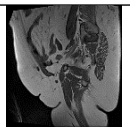
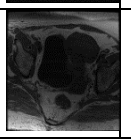
EC Stages	Modality	CPTAC			TCGA		
		Axial	***Coronal***	***Sagittal***	***Axial***	***Coronal***	***Sagittal***
Stage 1A	T1						
	T2						
	CE-T1						
Stage 1B	T1						
	T2						
	CE-T1						
Stage 2	T1						
	T2						
	CE-T1						
Stage 3	T1						
	T2						
	CE-T1						

Figure 6. Sample MRIs for Various Stages of EC from CPTAC and TCGA Archives

Precision: Precision is computed by

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

Recall: Recall is computed by

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

F1 score: It is calculated as:

$$F1 = \frac{2 \times Precision \times Recall}{Precision+Recall} \quad (6)$$

Inference time: It is the time taken to process a single image.

4.3 Performance Analysis

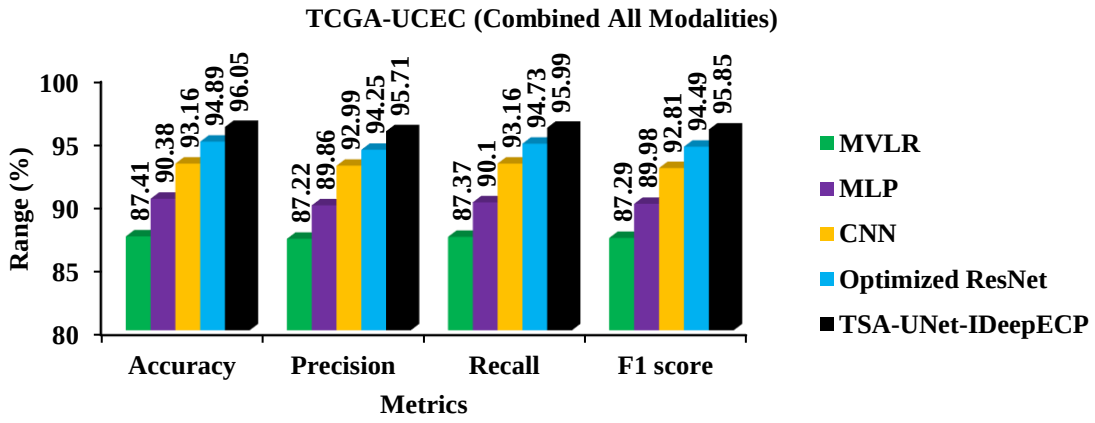


Figure 7. Tests on the TCGA-UCEC Dataset with Different EC Detection Models

Figure 7. examines various EC detection models using the TCGA-UCEC dataset. The TSA-UNet-IDeepECP model improves accuracy by 9.88%, 6.27%, 3.1%, and 1.22% compared to the MVLR, MLP, CNN, and Optimized ResNet models, respectively. The precision is 9.73%, 6.51%, 2.93%, and 1.55% higher than the other models, respectively. The recall is 9.87%, 6.54%, 3.04%, and 1.33% higher than other models, respectively. Similarly, the F1 score is 9.81%, 6.52%, 3.28%, and 1.44% greater than other models, respectively. This suggests that the proposed TSA-UNet-IDeepECP model enhances detection performance compared to conventional models by utilizing multi-view learning instead of single-view.

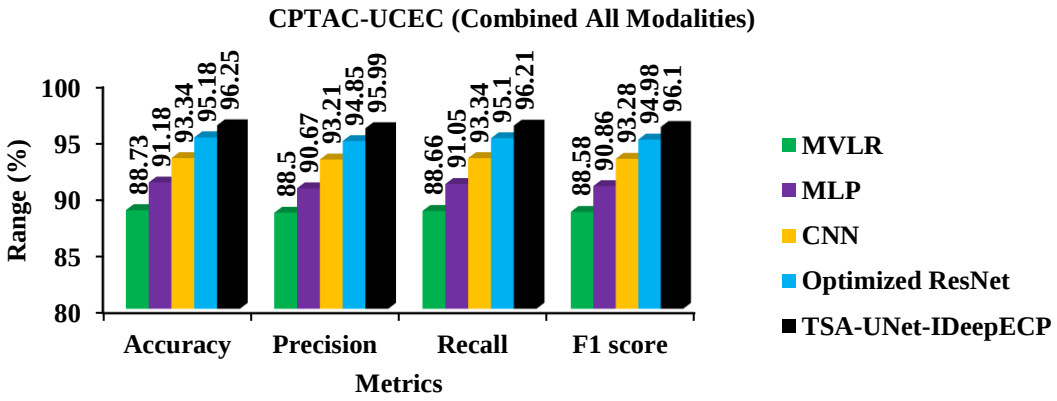


Figure 8. Comparing Different EC Detection Methods on the CPTAC-UCEC Database

Figure 8 examines various EC detection methods with the help of the CPTAC-UCEC database. The TSA-UNet-IDeepECP model shows improvements in the accuracy of 8.48%, 5.56%, 3.12%, and 1.12% compared to the MVLR, MLP, CNN, and optimized ResNet models respectively. The precision is also higher by 8.46%, 5.87%, 2.98%, and 1.2% compared to other models, respectively. Additionally, the recall is increased by 8.52%, 5.67%, 3.07%, and 1.17% compared to other models, respectively. Likewise, the F1 score is 8.49%, 5.77%, 3.02%, and 1.18% higher than the MVLR, MLP, CNN and optimized ResNet models, respectively. These findings imply that the TSA-UNet-IDeepECP model improves detection performance by using multi-view learning rather than single-view techniques.

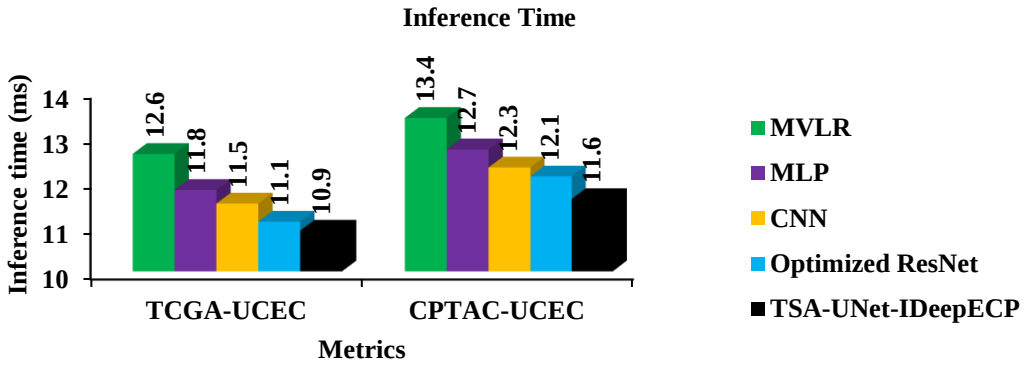


Figure 9. Inference Time for Proposed and Existing Models on Two Distinct Datasets

Figure 9 shows the inference time of various EC detection models on the two datasets. The TSA-UNet-IDeepECP on the TCGA-UCEC dataset reduces the inference time by 13.49%, 7.63%, 5.22%, and 1.8% compared to the MVLR, MLP, CNN, and optimized ResNet, respectively. For CPTAC-UCEC dataset, it reduces the inference time by 13.43%, 8.66%, 5.69%, and 4.13% compared to the same models. Thus, it is concluded that the TSA-UNet-IDeepECP model can be clinically acceptable for EC detection faster with higher accuracy.

4.4 Discussion

Though this TSA-UNet-IDeepECP performs well on the TCGA-UCEC and CPTAC-UCEC datasets, it has certain limitations.

- First, it has dataset bias due to solely using these datasets. MRI scans in these datasets are heterogeneous, but they might not fully reflect patient demographic diversity in real-time settings. This can still affect the generalizability of the model to larger population data. Therefore, additional research efforts may target different datasets on a larger, demographically diverse population to enhance model resilience in various healthcare settings.
- Second, this TSA-UNet-IDeepECP model categorizes Stages I-III of EC, whereas Stage IV cases are not evaluated due to a lack of MRI scans in this stage. This still limits the model's applicability to categorize advanced EC stages. Thus, future work could extend this model to categorize Stage IV EC cases to improve performance across various kinds of EC stages.

5. Conclusion

The TSA-UNet-IDeepECP model is presented in this paper for EC stage classification using a multi-view learning technique. A mix of 3D and 2D views of MRI scans allows the model to capture both local and global aspects of EC. A CNN-LSTM network learns the local information, while dilated convolutions and CHAP on ResNet18 recover the global information. A modifier encoder and attention strategy layers are used to combine and further improve the characteristics from both perspectives. The FC layer with a softmax function is applied for the final classification of EC. ECP model, based on the experimental findings, shows a greater accuracy rate of 96.05% for the TCGA dataset and 96.25% for the CPTAC dataset, thereby demonstrating it is better than earlier models. The TSA-UNet-IDeepECP model, based on the experimental findings, shows a higher accuracy rate of 96.05% for the TCGA dataset and 96.25% for the CPTAC dataset, indicating it is better than prior models to accurately categorize EC from MRI data.

Funding: "This research received no external funding"

Conflicts of Interest: "The authors declare no conflict of interest."

References

[1] Markowska, A. Chudecka-Głaz, K. Pityński, W. Baranowski, J. Markowska, and W. Sawicki, "Endometrial cancer management in young women," *Cancers*, vol. 14, no. 8, pp. 1–13, 2022.

[2] T. M. Kuhn, S. Dhanani, and S. Ahmad, "An overview of endometrial cancer with novel therapeutic strategies," *Curr. Oncol.*, vol. 30, no. 9, pp. 7904–7919, 2023.

[3] R. L. Siegel, A. N. Giaquinto, and A. Jemal, "Cancer statistics, 2024," *CA: A Cancer J. Clin.*, vol. 74, no. 1, pp. 12–49, 2024.

[4] *Cancer of the Endometrium - Cancer Stat Facts*. SEER. [Online]. Available: https://seer.cancer.gov/statfacts/html/corp.html

[5] *Endometrial cancer statistics*. World Cancer Research Fund International. [Online]. Available: https://www.wcrf.org/cancer-trends/endometrial-cancer-statistics/, Jun. 26, 2024.

[6] M. Menendez-Santos, C. Gonzalez-Baerga, D. Taher, R. Waters, M. Virarkar, and P. Bhosale, "Endometrial Cancer: 2023 Revised FIGO Staging System and the Role of Imaging," *Cancers*, vol. 16, no. 10, pp. 1–24, 2024.

[7] J. C. Kasius, J. M. Pijnenborg, K. Lindemann, D. Forsse, J. van Zwol, G. B. Kristensen, and F. Amant, "Risk stratification of endometrial cancer patients: FIGO stage, biomarkers and molecular classification," *Cancers*, vol. 13, no. 22, pp. 1–19, 2021.

[8] G. F. Ma, G. L. Lin, S. T. Wang, Y. Y. Huang, C. L. Xiao, J. Sun, and L. B. Xiang, "Prediction of recurrence-related factors for patients with early-stage cervical cancer following radical hysterectomy and adjuvant radiotherapy," *BMC Women's Health*, vol. 24, no. 1, p. 81, 2024.

[9] R. Johnson, C. I. Liao, C. Tian, M. T. Richardson, K. Duong, N. Tran, and J. K. Chan, "Uterine cancer among Asian Americans–disparities & clinical characteristics," *Gynecol. Oncol.*, vol. 182, pp. 24–31, 2024.

[10] M. Palmér, Å. Åkesson, J. Marcickiewicz, E. Blank, L. Hogström, M. Torle, and H. Leonhardt, "Accuracy of transvaginal ultrasound versus MRI in the PreOperative Diagnostics of low-grade Endometrial Cancer (PODEC) study: a prospective multicentre study," *Clin. Radiol.*, vol. 78, no. 1, pp. 70–79, 2023.

[11] M. N. Yeasmin, M. Al Amin, T. J. Joti, Z. Aung, and M. A. Azim, "Advances of AI in image-based computer-aided diagnosis: A review," *Array*, vol. 23, pp. 1–23, 2024.

[12] Y. A. Kadhim, M. U. Khan, and A. Mishra, "Deep learning-based computer-aided diagnosis (cad): applications for medical image datasets," *Sensors*, vol. 22, no. 22, pp. 1–21, 2022.

[13] Urushibara, T. Saida, K. Mori, T. Ishiguro, K. Inoue, T. Masumoto, and T. Nakajima, "The efficacy of deep learning models in the diagnosis of endometrial cancer using MRI: a comparison with radiologists," *BMC Med. Imag.*, vol. 22, no. 1, pp. 1–14, 2022.

[14] W. Mao, C. Chen, H. Gao, L. Xiong, and Y. Lin, "A deep learning-based automatic staging method for early endometrial cancer on MRI images," *Front. Physiol.*, vol. 13, pp. 1–12, 2022.

[15] J. Tao, Y. Wang, Y. Liang, and A. Zhang, "Evaluation and monitoring of endometrial cancer based on magnetic resonance imaging features of deep learning," *Contrast Media Mol. Imag.*, vol. 2022, no. 1, pp. 1–9, 2022.

[16] J. Yang, Y. Cao, F. Zhou, C. Li, J. Lv, and P. Li, "Combined deep-learning MRI-based radiomic models for preoperative risk classification of endometrial endometrioid adenocarcinoma," *Front. Oncol.*, vol. 13, pp. 1–10, 2023.

[17] J. Liu, S. Li, H. Lin, P. Pang, P. Luo, B. Fan, and J. Yu, "Development of MRI-based radiomics predictive model for classifying endometrial lesions," *Sci. Rep.*, vol. 13, no. 1, pp. 1–10, 2023.

[18] L. Xiong, C. Chen, Y. Lin, W. Mao, and Z. Song, "A computer-aided determining method for the myometrial infiltration depth of early endometrial cancer on MRI images," *Biomed. Eng. Online*, vol. 22, no. 1, pp. 1–17, 2023.

[19] H. Meng, Y. F. Sun, Y. Zhang, Y. N. Yu, J. Wang, J. N. Wang, and X. P. Yin, "Predicting risk stratification in early-stage endometrial carcinoma: significance of multiparametric MRI radiomics Model," *J. Imag. Inform. Med.*, vol. 37, no. 1, pp. 1–11, 2024.

[20] Y. M. Cui, H. L. Wang, R. Cao, H. Bai, D. Sun, J. X. Feng, and X. F. Lu, “The Segmentation of multiple types of uterine lesions in magnetic resonance images using a sequential deep learning method with image-level annotations,” *J. Imag. Inform. Med.*, vol. 37, no. 1, pp. 1–12, 2024.

[21] H. Bae, S. E. Rha, H. Kim, J. Kang, and Y. R. Shin, “Predictive value of magnetic resonance imaging in risk stratification and molecular classification of endometrial cancer,” *Cancers*, vol. 16, no. 5, pp. 1–17, 2024.

[22] *TCGA-UCEC*. The Cancer Imaging Archive (TCIA). [Online]. Available: https://www.cancerimagingarchive.net/collection/tcga-ucec/, 2024.

[23] *CPTAC-UCEC*. The Cancer Imaging Archive (TCIA). [Online]. Available: https://www.cancerimagingarchive.net/collection/cptac-ucec/, 2024.