Received: May 31, 2023. Revised: June 29, 2023.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Enhancing Endometrial Cancer Detection by Feature Entanglement Image Generator and Multimodal Feature Learning with Attention Mechanism

Karthick Natarajan[1]* **Kamatchi Annappan[2]**

[1]*Department of Computer Science, PSG College of Arts and Science, Coimbatore, Tamilnadu, India*
[2]*Department of Computer Science with Data Analytics, PSG College of Arts and Science, Coimbatore, Tamilnadu, India*
* Corresponding author's Email: karthickphd.vlbjcas@gmail.com

Abstract: Endometrial cancer (EC) rate is rising progressively worldwide so early diagnosis of EC using medical imaging is vital to increase a patient's survival rate. To avoid the misdiagnosis of EC on magnetic resonance imaging (MRI) scans by clinicians, an automated staging model based on deep learning has broadly emerged in medical systems. Many deep learning models like convolutional neural networks (CNNs) with transfer learning schemes are developed to classify normal and cancer patients from a sequence of MRI scans. But such models have poor sensitivity to classify EC stages from other benign lesions, resulting inaccurate diagnoses of EC. Also, the classical transfer learning schemes reduce the accuracy of detecting medical images due to the discrepancies in data distribution between the source and target domains. Hence, this article proposes a novel deep EC prediction (DeepECP) model that involves image synthesis and classification processes. First, a feature entanglement generative adversarial network (FE-GAN) is proposed for MRI synthesis that creates a desired MRI sequence according to the complementary features of multiple MRI modalities. Then, a multi-modal CNN with long short-term memory (LSTM) network followed by the fully connected (FC) layer is developed to extract a sequence of cancer features from multi-modal MRI sequences. Moreover, an attention strategy is used to fuse those extracted features and get a final feature vector, which is given to the softmax function to classify EC stages. Finally, the extensive experiments show that the DeepECP model on the TCGA-UCEC and CPTAC-UCEC datasets reaches 93.2% and 93.3% accuracy in detecting EC stages, respectively compared to the support vector machine (SVM), VGGNet-16, InceptionResNet and CNN models.

Keywords: Endometrial cancer, Medical imaging, Automated staging, Deep learning, GAN, CNN-LSTM.

1. Introduction

Endometrial cancer (EC) is a major cancer in females worldwide [1], with 27.8 new cases per 100,000 females and a 5.1 death rate per 100,000 females annually [2]. It affects 3.1% of females and is projected to increase to 65,950 more cases and 12,550 deaths by 2022. The increasing rate of EC is linked to risk factors like being overweight [3]. Despite the importance of early treatment [4-5], no detection system exists for EC and invasive biopsy processes are still required to identify EC in symptomatic females [6]. The shift towards high-risk EC detection may miss the possibility of identifying EC in asymptomatic females [7]. Low-risk patients typically require a simple hysterectomy, while high-risk patients require adjuvant radiation therapy. An efficient, automated detection model is crucial for early EC staging, increasing diagnostic efficacy and providing valuable data for physicians to suggest appropriate therapies.

The international federation of gynecology and obstetrics (FIGO) classifies EC into four stages: (i) Stage I: cancer that is confined to the endometrium (Stage 1A) and myometrium (Stage 1B); (ii) Stage II: a tumor that has circulated to the cervix; (iii) Stage III: a tumor that has dispersed to the ovaries (Stage 3A), vagina (Stage 3B), and lymph nodes (Stage 3C); (iv) Stage IV: a tumor that has circulated to the urinary bladder (Stage 4A), rectum, or organs situated far from the uterus like pelvis, abdomen

(Stage 4B).

MRI aids in local staging and accurate diagnosis of EC by predicting invasion depth, cervical stroma invasion, and lymph node metastases [9-10]. In 2009, the European society of urogenital radiology (ESUR) released guidelines for EC staging, recommending MRI as the preferred imaging modality for disease severity assessment. However, pathological assessment often differs between radiologists due to the radiologist's knowledge [11].

Recently, deep learning has broadly emerged in computer vision as a novel computer-aided diagnosis model. This model can automatically find the target region after training on a huge quantity of images [12 -16]. But it has poor sensitivity for distinguishing between healthy and tumor images, making it difficult to accurately differentiate EC from benign lesions. To combat this issue, several transfer learning-based models have been applied in healthcare systems, which learn the prediction function in the target domain using prior knowledge from the massive amount of annotated images (such as ImageNet) in the source domain [17-18]. But the input image features are extremely varied between the training (large-scale natural image sets) and test (a small number of MRI scans or other clinical datasets) sets in the medical imaging applications. Due to this fact, standard transfer learning can degrade the detection efficiency in medical imaging because of the significant distribution mismatches between the source and target domains.

Therefore in this manuscript, a novel DeepECP model is proposed, which is a multimodal system to achieve a robust and reliable solution for detecting multiple EC stages (i.e., Stage I, Stage II, and Stage III) from MRI scans. Initially, the FE-GAN model is introduced, which comprises generator and discriminator modules. The generator adopts an encoder-decoder network to create a target MRI modality by learning complementary information from multiple modalities, whereas the discriminator uses CNN to separate the synthetic image from the actual one. After MRI image synthesis, a multi-modal fusion based on deep learning is developed for enhancing the fusion of multi-modality MRI sequences. Three different CNN-LSTM networks followed by FC layers are used for each modality, which learns and captures the feature vectors from different modalities. Such feature vectors are then fused by the attention strategy to create a final feature vector, which is fed to the softmax layer to classify into different stages of EC. Thus, this new DeepECP model increases the training efficiency by using a huge quantity of MRI sequences and the accuracy of identifying EC stages.

The remaining sections are prepared as the following: Section 2 reviews prior studies for EC detection. Section 3 describes the DeepECP and section 4 demonstrates its performance. Section 5 summarizes the findings and future work.

2. Literature survey

A machine-learning technique has been used to accurately diagnose EC based on different stages [19], but accuracy was less for a large-scale dataset. To produce accurate predictions, a sufficient number of patients was needed. CNN model was applied [20] to recognize the myometrial invasion depth of EC using MRI scans. But the accuracy was not effective due to the small dataset. An adapted VGGNet-16 has been presented [21] for classifying ECs in the hysteroscopic samples. But accuracy was less since it did not consider lesions with a lower frequency of occurrence. SVM, random forest, etc., have been used [22] to predict the incidence of premature EC from clinical records. Due to the small database, its prediction performance was not effective.

An improved multi-resolution InceptionResNet model was proposed [23] to estimate EC stages. The mutation of risky chromosomes was also estimated using the H&E slides. But the prediction accuracy was low because of appearing extraneous non-tumor cells, such as myometrium. The YOLOv3 model was suggested [24] to localize the tumor region on EC MRI scans. Such localized areas were fed into the ResNet for recognizing myometrial invasion depth. But precision was less since the number of images was not adequate. Also, it extracted 2D tumor patches that may fail to completely use the volume information and learn the spatial characteristics. An U-net based prediction system was proposed [25] for early EC diagnosis on MRI scans. But, the f-score value was low since the dataset was limited.

2.1 Research gap and contribution

From the literature, it is observed that most studies trained machine and deep learning models on a fewer number of images. This results in less training efficiency and accuracy in detecting different stages of EC. In contrast with those studies, the DeepECP is a novel recognition model suitable for rapid and reliable EC detection from a huge quantity of MRI images with higher accuracy.

3. Proposed methodology

This section briefly explains the DeepECP model. Fig. 1 portrays a pipeline of the presented study, which involves the following major processes:

Received: May 31, 2023. Revised: June 29, 2023.

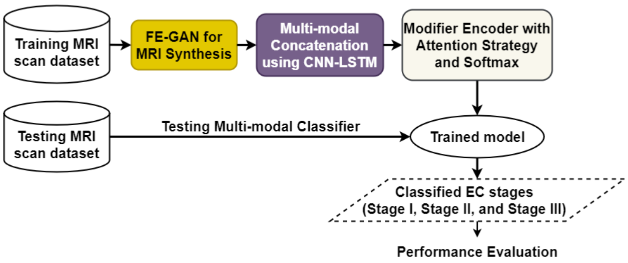
Training MRI scan dataset
FE-GAN for MRI Synthesis
Multi-modal Concatenation using CNN-LSTM
Modifier Encoder with Attention Strategy and Softmax
Testing MRI scan dataset
Testing Multi-modal Classifier
Trained model
Classified EC stages (Stage I, Stage II, and Stage III)
Performance Evaluation

Figure. 1 Pipeline of the presented EC detection model

Table 1. Lists of notations

Notations	Explanation
$\mathcal{G}$	Generator network
$\mathcal{D}$	Discriminator network
C_N	Shared characteristics
N	Number of source modalities
C_F	Concatenated data
Ξ', μ	Mean of explicit data
Δ', σ	Standard variance of explicit data
L_1	Error
γ, β	Scale and offset variables, respectively
x	Input example
X_1, X_2	Input modalities
y	Scan of the actual target modality
F	LFC unit
$\bar{Y}$	Artificial desired modality
$\bar{\bar{Y}}$	Pseudo-target modality
L_{tot}	Total loss function
λ_1, λ_2	Variables
f^t, g^t, o^t	Forget, external and output gates at period t in LSTM
x^t, h^t, s^t	Input, hidden and state vectors
U^k, W^k, b^k	Weights and biases
$tanh$	Hyperbolic tangent activation function
S_1, S_2, S_3	Kernel dimensions in three Conv layers
K_1, K_2, K_3	Channel numbers of three Conv layers

- First, the training MRI scans are given to the FE-GAN for the synthesis process, which creates more synthetic scans resembling the actual one.
- Then, the multi-modal concatenation using CNN-LSTM is performed, which learns various features from the multi-modality MRI scans and concatenates them. Also, the modifier Encoder (ER) with an attention strategy followed by the FC and softmax is learned to get the trained model.
- Later, the trained model is used to classify the test scans into various stages of EC.

Table 1 lists the notations used in this study.

3.1 Dataset description

Two different MRI scan datasets are acquired from the cancer imaging archive (TCIA).

1. The cancer genome atlas uterine corpus endometrial carcinoma (TCGA-UCEC) dataset [26]: An academic group focuses on linking tumor phenotypes to gene sequences by providing medical scans related to participants from TCGA. This allows researchers to examine the TCGA/TCIA datasets for relationships between tissues, radiography traits and clinical results. The images are often collected as part of repetitive treatment rather than exact investigations or therapeutic tests. The dataset is heterogeneous, with four modalities (CT, CR, MRI, & PET), 65 participants, 226 studies, 912 series and 75829 images available.
2. The clinical proteomic tumor analysis consortium UCEC (CPTAC-UCEC) dataset [27]: The CPTAC-UCEC group of subjects uses proteome and genetic studies to know tumor molecular bases. TCIA collects pathological and MRI scans from CPTAC patients, allowing researchers to explore tumor characteristics and match proteomic, genetic and medical information. The CPTAC-cancer category includes imaging from all tumor kinds. MRI scan databases are heterogeneous, with three modalities (CT, MRI and PET), 250 participants, 103 studies, 1655 series and 153199 images.

This multi-modal classification merges T1, T2, and CE-T1 weighted MRI sequences, allowing all sequences to input separate RGB channels.

3.2 MRI scan synthesis using FE-GAN

The architecture of FE-GAN is shown in Fig. 2, which includes the generator $\mathcal{G}$ and the discriminator $\mathcal{D}$. The aim of $\mathcal{G}$ is to create an accurate desired MRI modality, which is similar to the actual scan by using the encoder-decoder network as the support. Also, the aim of $\mathcal{D}$ is to differentiate the synthetic scan from the actual one by using the simple CNN.

In $\mathcal{G}$, the ERs get several MRI scans and extract the abstract characteristics via squeezing the scan resolution and expanding the receptive field, whereas the decoder (DR) obtains such characteristics and recreates the concatenated data with an upsampling. Moreover, the expected MRI modality is created. To capture both shared and explicit data from multiple modalities, two different ERs, namely shared ER (SER), and explicit ER (EER) are utilized for

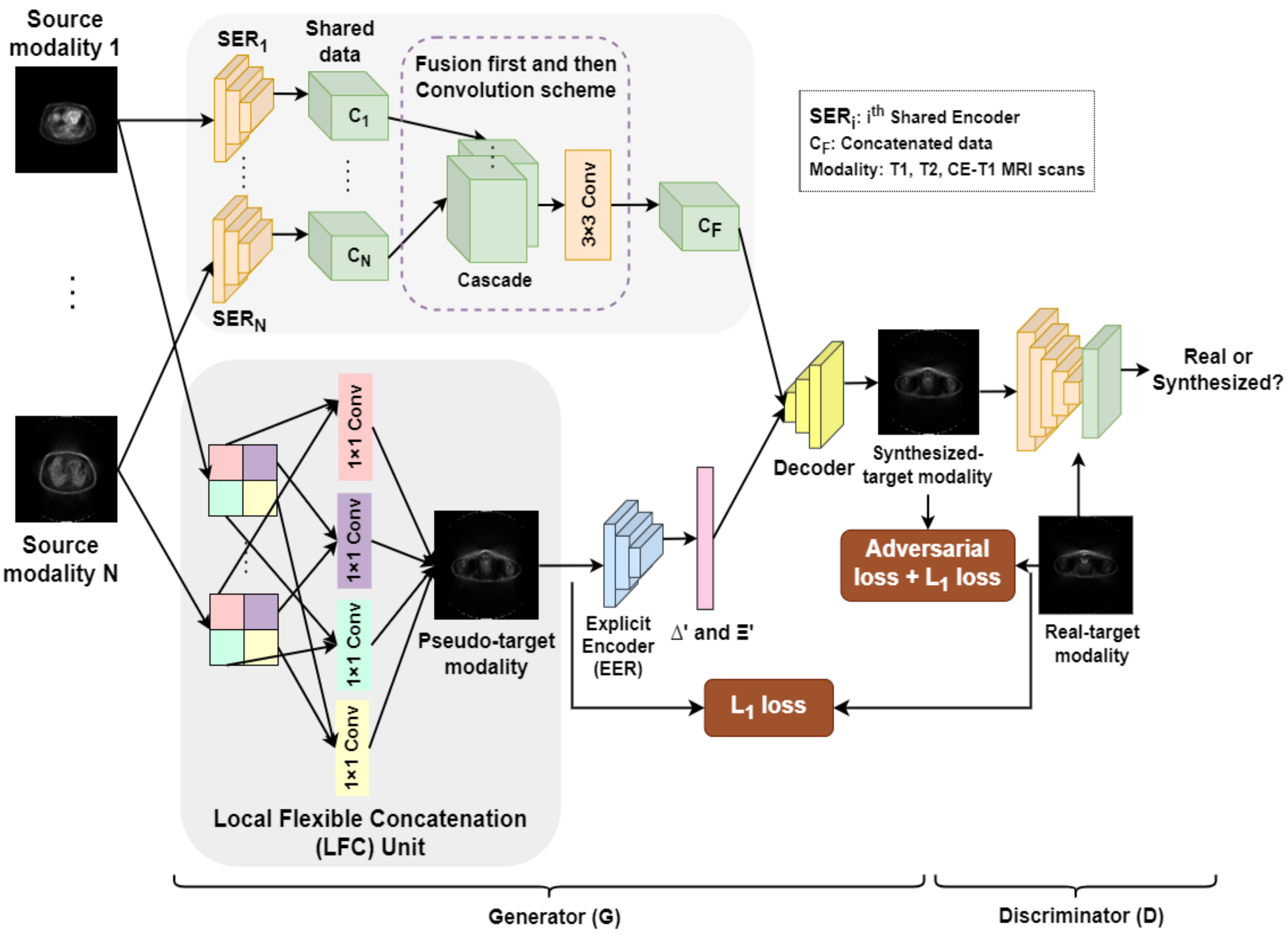

Figure. 2 Architecture of FE-GAN for MRI scan synthesis

efficient feature extraction.

The SER captures high-dimensional shared data related to scan semantics and total pattern, while the EER learns low-dimensional explicit data about the scan's texture. Concatenating shared data improves semantic feature representation. To extract explicit data, the local flexible concatenation (LFC) unit recreates a pseudo-target modality, which is then given to the EER for extraction. The shared and explicit data are concatenated in the decoding link, and the concatenated characteristics are progressively reinstated to the artificial desired modality through upsampling. The entry of $\mathcal{D}$ is a group of scans such as the source MRI scans, and the resultant synthetic target scans. The multi-layer CNN is used to obtain the feature map (Fmap) with a global receptive field, after the tanh activation to estimate whether the scan is actual or artificial.

3.2.1. Generator network

Encoder unit

Shared encoder (SER): Because $\mathcal{G}$ obtains several MRI scans with various contrasts, it is rational to assign an equivalent ER, i.e., SER to capture the shared data of all given modalities, which is similar to the actuality that the amount of SER matches to that of the given modalities as shown in Fig. 2. All SERs share an equal design with distinct variables. Every SER comprises three downsampling units and four residual units (ResUnit). The downsampling unit is developed as a design of convolutional (Conv) – instance normalization (IN) – rectified linear unit (ReLU) activation layers.

Initially, the scan of dimension 256×256 is convolved using a 7×7 kernel with a padding of 3 and a stride of 1 to maximize the amount of channels when maintaining the scale constant. In the succeeding 2 downsampling units, the kernel dimension is assigned as 4×4 with stride 2 and padding 1 to split the Fmap dimension and maximize the amount of channels. The IN is adopted rather than batch normalization to keep the individuality of all modalities. Therefore, it is ensured that the scan stylization is not influenced by the batch dimension.

In the ResUnit, the Padding – Conv – IN – ReLU design is applied to extend the system and capture high-dimensional shared data. Also, the Fmap dimension and the amount of channels stay unaltered in the ResUnit. Based on this design, the result of the

SER is a Fmap of dimension $256 \times 64 \times 64$. In the shared data encoding link, the fusion first and then convolution scheme is applied to concatenate the shared data from different modalities. The attribute channels of the shared data captured by SER are fused and convolved via a Conv layer, squeezed to efficiently merge the data from various modalities. Consider $\{C_1, \ldots, C_N\}$ is the shared characteristics captured by the SER from N source modalities with a dimension of $256 \times 64 \times 64$, and are cascaded along the channel direction, thus creating Fmap with channel of $256 \times N$. Especially, a 3×3 kernel is assigned for convolution with a stride of 1 and a padding of 1 to learn the shared data after cascading. Each channel is weighted and summed in the Conv layer, and the concatenated data C_F with a channel of 256 is acquired. It enhances the channel details only without decreasing the resolution of abstract characteristics at a high degree.

Explicit encoder (EER): The EER is developed as a system with smaller quantity of layers because the shallow network layers can capture low-dimensional explicit data associated with the edges and scan textures. It is efficient with five units designed as Conv–ReLU to capture the explicit data to all modalities. Additionally, the IN layer is eliminated due to its poor performance in restoring the actual features and standard variances of the explicit data. The pseudo-target scan of dimension 256×256 is fed to the initial Conv with a 7×7 kernel with stride 1 and padding 3, which is followed by four units whose kernel dimension is assigned as 4×4 with stride 2 and padding 1. Then, all channels of the feature segment comprising explicit data are squeezed into a real value with a global mean pooling. Moreover, the feature channel size is decreased to 8 with a 1×1 Conv kernel. Using the final three linear layers, the mean (Ξ') and standard variance (Δ') of the explicit data are acquired, which are used for concatenation with the shared data in the DR.

Local flexible concatenation (LFC) unit

The LFC unit is designed before EER to create the pseudo-target modality for extracting the explicit data during the test phase. The multi-modal scans are split into many scan segments to efficiently utilize the data across modalities. As depicted in Fig. 2, the areas from multiple modalities with a similar color define that they relate to a similar position of the real scan. The 1×1 Conv with 2 channels is implemented for areas at the similar position, with various Conv kernels for various areas. Afterward, the scan segments after convolution can be merged based on their earlier relative locations, thus creating the pseudo-target modality for successive explicit encoding. Observe that the 1×1 Conv gives an end-to-end multi-modal concatenation strategy, i.e., all modalities are weighted by the Conv variables to reveal the roles of various modalities to various areas. Also, end-to-end learning minimizes the amount of variables to be trained in contrast to the multi-channel cascade method. Moreover, the L_1 error is taken as the objective function to fine-tune the model and create a highly accurate desired modality with sharper details.

Decoder unit

The DR obtains the concatenated shared data including the average and standard variances of the explicit data to create the desired modality. A flexible IN (FIN) layer is designed to concatenate the two kinds of data efficiently in the DR, wherein the explicit data is embedded in the shared data by affine conversion. The DR comprises four ResUnits, all have a design analogous to that of the ResUnits in SER excluding the IN layer is substituted by the FIN layer for intra-layer concatenation of shared and explicit data. Besides, the affine conversion variables γ and β in the standard IN layer are nearly associated with the scan pattern and must be trained via network learning, as follows:

$$IN(x) = \gamma \left(\frac{x - \mu(x)}{\sigma(x)} \right) + \beta \tag{1}$$

In Eq. (1), γ is the scale variable, β is the offset variable, x refers to the input example, $\mu(x)$ and $\sigma(x)$ are the average and the standard variance, correspondingly. Because the desired modality style is previously estimated, the constant variables Δ' and Ξ' captured from the explicit data are deliberated to substitute the affine variables in the IN layer for reducing the computation complexity. Therefore, the FIN layer is defined by

$$FIN(x) = \Delta' \left(\frac{x - \mu(x)}{\sigma(x)} \right) + \Xi' \tag{2}$$

In Eq. (2), x is the concatenated feature after the fusion first and then convolution scheme, Δ' is the standard variance of the explicit data from EER, and Ξ' is the respective average value. So, the shared and explicit data can be concatenated efficiently without complex coding at the FIN layer. After that, the feature after concatenation via 4 ResUnits is up-sampled with 3 transposed Conv layers to recreate the high-dimensional Fmap to a scan with an equal dimension as the source modality. At last, the synthesized target modality scan is obtained.

3.2.2. Discriminator network

In this study, $\mathcal{D}$ has a design analogous to the design of twofold categorization CNN. It comprises five Conv layers and obtains a set of scans such as the source MRI scans and the respective artificial desired scan as input. After the initial Conv layer, the amount of channels of Fmaps rises to 64 and is folded in all successive convolutions. The result value of the final tanh activation is applied to estimate whether the scan is actual or artificial, the nearer it is to 1, a highly expected the desired modality is accurate.

3.2.3. Objective functions

The objective function of this FE-GAN comprises three loss functions: (a) L_1 error between the artificial and the actual desired modality, which is applied to improve the excellence of the artificial desired modality; (b) L_1 error between the pseudo-target modality and the actual desired modality, which facilitates the pseudo-target modality to preserve many explicit data; (c) the adversarial error of the GANs that creates the artificial scan highly accurate to deceive $\mathcal{D}$. Consider X_1, X_2 are the scans of the two input modalities, y denotes the scan of the actual target modality, and F is the LFC unit. Then, $\bar{Y} = \mathcal{G}(X_1, X_2)$ denotes the artificial desired modality, and the pseudo-target modality is denoted by $\bar{\bar{Y}} = F(X_1, X_2)$. The total loss function is defined by

$$L_{tot} = \underset{\mathcal{G}}{\operatorname{argmin}}\, \underset{\mathcal{D}}{\max}\, L_{FE-GAN}(\mathcal{G}, \mathcal{D}) + L_1(\mathcal{G}, F) \quad (3)$$

In Eq. (3), $L_{FE-GAN}(\mathcal{G}, \mathcal{D})$ is defined as:

$$L_{FE-GAN}(\mathcal{G}, \mathcal{D}) = \mathbb{E}_Y\left[\log\left(\mathcal{D}(Y)\right)\right] + \mathbb{E}_{X_1, X_2}\left[1 - \log\left(\mathcal{D}\left(\mathcal{G}(X_1, X_2)\right)\right)\right] \quad (4)$$

Also, $L_1(\mathcal{G}, F)$ is computed by

$$\begin{aligned} L_1(\mathcal{G}, F) &= \lambda_1 \mathbb{E}_{X_1, X_2, Y}[\|Y - \bar{Y}\|_1] + \lambda_2 \mathbb{E}_{X_1, X_2, Y}\left[\left\|Y - \bar{\bar{Y}}\right\|_1\right] \\ &= \lambda_1 \mathbb{E}_{X_1, X_2, Y}[\|Y - \mathcal{G}(X_1, X_2)\|_1] + \lambda_2 \mathbb{E}_{X_1, X_2, Y}[\|Y - F(X_1, X_2)\|_1] \quad (5) \end{aligned}$$

Based on Eq. (4), $\mathcal{G}$ is projected to create a high-quality scan that deceives $\mathcal{D}$ into providing an estimation as near to 1 as promising, which is stable with the improvisation of $\mathcal{D}\left(\mathcal{G}(X_1, X_2)\right)$. As for $\mathcal{D}$, it tries to increase $\mathcal{D}(Y)$ when reduce $\mathcal{D}\left(\mathcal{G}(X_1, X_2)\right)$, therefore the actual modality is differentiated from the created scan. To create the artificial scan highly accurate, the pixel-wise variances between the actual and the artificial desired scans, and the pseudo-target scan, are computed in Eq. (5) and reduced in Eq. (3). The variables λ_1 and λ_2 are used to equilibrium 2 elements of L_1 error, both of which are assigned to 0.1 in the test.

***FE-GAN Training*:** The FE-GAN is learned by the Adam optimizer with an epoch number of 300 and a batch dimension of 5. In the initial 100 epochs, the training rates for $\mathcal{G}$ and $\mathcal{D}$ are set to 0.0001, then linearly degenerate to 0 in the consecutive 100 epochs. The IN is used in the SER and the FIN is used in the DR. The batch regularization is applied to differentiation stages. Also, the Leaky ReLU with a gradient of 0.2 is adopted for activation in all layers. For all epochs, $\mathcal{G}$ and $\mathcal{D}$ are trained alternatively.

3.3 EC staging system using multi-modal fusion of CNN-LSTM

After obtaining more MRI scans from the FE-GAN, a multi-modal classification system is proposed to improve EC detection by concatenating multiple modality sequences. It uses multiple CNN-LSTM branches to capture features from each modality sequence. The feature vectors are fused as input for the modifier ER with vector splicing, preserving individualities. An attention strategy using FC and softmax layers is applied to classify different stages of EC in MRI scans.

3.3.1. CNN-LSTM

The DeepECP model uses three CNN-LSTM branches, each with four layers, to capture spatiotemporal characteristics of MRI scans. Three Conv layers capture hierarchical features (e.g., region, structure and temporal) from multi-modality MRI sequences, followed by an LSTM network layer to learn relevant sequential features.

Extraction of region features: In the initial Conv layer, the kernel dimension of $S_1 \times N \times 1$ and the stride along 3 sizes, i.e., 1 temporal and 2 spatial sizes of $(1,1,1)$ are set to capture spatial characteristics from the entire MRI scan. The Conv with 2 spatial sizes is a feature representation for all region-of-interests (ROIs) when the Conv with the temporal size signifies various feature representations of a similar ROI. This assists in describing the time-based possessions of the respective ROI. Characteristics obtained from this layer define high-order temporal dynamics of endometrium areas.

Extraction of structural characteristics: In the 2nd Conv layer, the kernel dimension of $S_2 \times 1 \times N$ and the stride in 3 sizes $(1,1,1)$ are assigned to capture

Received: May 31, 2023. Revised: June 29, 2023.

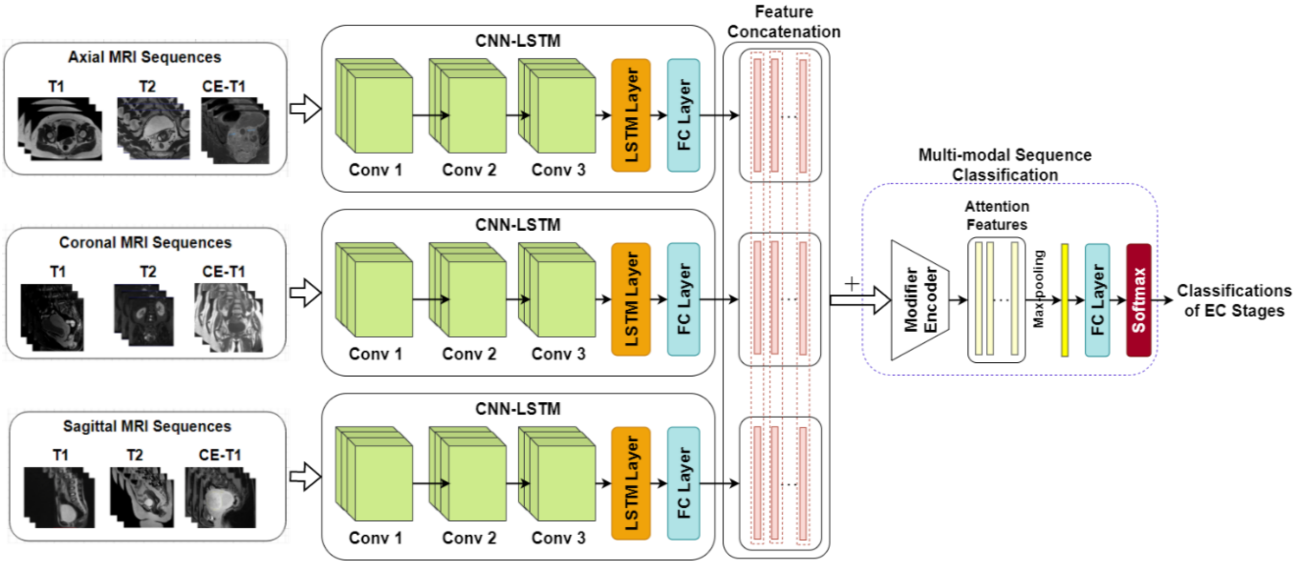

Figure. 3 Architecture of multi-modal sequence classification using CNN-LSTM for EC staging

structural features from the learned region features. The Conv with 2 spatial sizes is a feature representation for the entire MRI scan, revealing morphological modifications in the endometrium.

Extraction of temporal features: The 3rd Conv layer captures the temporal characteristics of the entire MRI scan. The kernel dimension of $S_3 \times 1 \times 1$ and the stride in three dimensions (2,1,1) are assigned. So, the characteristics captured in this layer with a kernel are called representations of the temporal dynamics of the multi-modality MRI sequences. It is observed that such 3 Conv layers are utilized to get high-level and high-order temporal characteristics from multi-modality MRI sequences.

Extraction of sequential features: The LSTM network layer is adopted to model the sequential data of EC, which learns the temporal dynamics of endometrium in multi-modality MRI sequences. To fit the input dimension needed by the LSTM, the temporal characteristics obtained in the earlier Conv layer are flattened and given to the LSTM layer, which is described as:

$$f^t = \sigma\left(U^f x^t + W^f h^{t-1} + b^f\right) \quad (6)$$

$$g^t = \tanh(U^g x^t + W^g h^{t-1} + b^g) \quad (7)$$

$$o^t = \sigma(U^o x^t + W^o h^{t-1} + b^o) \quad (8)$$

$$s^t = f^t s^{t-1} + g^t \sigma\left(U^i x^t + W^i h^{t-1} + b^i\right) \quad (9)$$

$$h^t = \tanh(s^t)\, o^t \quad (10)$$

In Eqs. (6) to (10), f^t, g^t, and o^t are the forget, external input, and the output gates at period t, correspondingly. x^t, h^t, and s^t are the input, hidden, and state vectors, correspondingly. U^k, W^k, and b^k with $k \in f, g, o, i$ are the weights and biases, correspondingly. Also, σ and tanh are sigmoid and hyperbolic tangent activation functions. The temporal characteristics of the LSTM layer are remodelled and given to the FC layer.

The CNN-LSTM branch features are concatenated and fed to the sequence classification layer, which uses a modifier network to learn cancer characteristics. The layer uses a pre-learned image categorization system and self-attention features, with 1 ER layers assigned. The characteristics are max-pooled with temporal sizes to achieve the cancer characteristic of the entire input sequence. The FC layer and softmax output layer classify EC stages.

4. Experimental results

The effectiveness of the DeepECP is assessed by implementing it using Python code. The multi-modal MRI image datasets considered in this study are discussed in section 3.1. Table 2 provides the statistics about the training and testing images in the CPTAC-UCEC and TCGA-UCEC datasets. Also, Fig. 4 presents a few examples of multi-modality MRI sequences. A comparative assessment is conducted between the DeepECP and some previous models (such as CNN [16], VGGNet-16 [21], SVM [22], and InceptionResNet [23]). To do this, these existing models are also implemented by Python 3.8 and tested for the CPTAC-UCEC and TCGA-UCEC datasets, which helps to understand the success rate of the DeepECP model.

4.1 Parameter settings

In the training phase, for proposed DeepECP, the

Table 2. Statistics of training and testing dataset

CPTAC-UCEC Dataset									
EC Stages	Modality	Training Set **(No. of images)**				Testing Set **(No. of images)**			
		Axial	*Coronal*	*Sagittal*	*Total*	*Axial*	*Coronal*	*Sagittal*	*Total*
Stage 1A	T1	26	21	68	115	15	10	40	65
	T2	32	36	31	99	22	20	19	61
	CE-T1	94	68	97	259	57	40	42	139
Total		152	125	196	***473***	94	70	101	***265***
Stage 1B	T1	100	16	46	162	50	8	22	80
	T2	61	28	21	110	30	16	10	56
	CE-T1	113	20	98	231	63	10	47	120
Total		274	64	165	***503***	143	34	79	***256***
Stage 2	T1	154	138	25	317	75	70	12	157
	T2	35	70	35	140	19	39	20	78
	CE-T1	26	21	43	90	14	10	20	44
Total		215	229	103	***547***	108	119	52	***279***
Stage 3A	T1	88	208	25	321	42	102	16	160
	T2	32	76	56	164	21	40	30	91
	CE-T1	21	21	89	131	10	10	45	65
Total		141	305	170	***616***	73	152	91	***316***
Stage 3B	T1	44	33	19	96	20	19	11	50
	T2	40	28	36	104	20	21	20	61
	CE-T1	101	21	156	278	50	10	70	130
Total		185	82	211	***478***	90	50	101	***241***
Stage 3C	T1	118	240	25	383	55	115	13	183
	T2	63	70	20	153	30	29	10	69
	CE-T1	88	56	29	173	42	28	11	81
Total		269	366	74	***709***	127	172	34	***333***
Total training images					***3326***	***Total test images***			***1690***
TCGA-UCEC Dataset									
Stage 1A	T1	56	31	31	118	26	24	19	69
	T2	31	23	13	67	20	15	50	85
	CE-T1	22	18	32	72	18	12	20	50
Total		109	72	76	***257***	64	51	89	***204***
Stage 1B	T1	22	26	18	66	10	13	10	33
	T2	54	30	29	113	20	19	15	54
	CE-T1	22	18	32	72	13	10	21	44
Total		98	74	79	***251***	43	42	46	***131***
Stage 2	T1	75	34	33	142	38	25	20	83
	T2	80	72	33	185	47	40	29	116
	CE-T1	66	24	33	123	43	18	20	81
Total		221	130	99	***450***	128	83	69	***280***
Stage 3A	T1	36	30	26	92	22	16	15	53
	T2	18	16	24	58	10	12	20	42
	CE-T1	60	22	30	112	47	18	18	83
Total		114	68	80	***262***	79	46	53	***178***
Stage 3B	T1	46	29	24	99	30	21	19	70
	T2	32	25	22	79	25	17	13	55
	CE-T1	23	40	31	94	14	20	21	55
Total		101	94	77	***272***	69	58	53	***180***
Stage 3C	T1	43	30	25	98	29	18	15	62
	T2	43	43	33	119	26	26	18	70
	CE-T1	47	28	27	102	28	20	16	64
Total		133	101	85	***319***	83	64	49	***196***
Total training images					***1811***	***Total test images***			***1169***

Received: May 31, 2023. Revised: June 29, 2023.

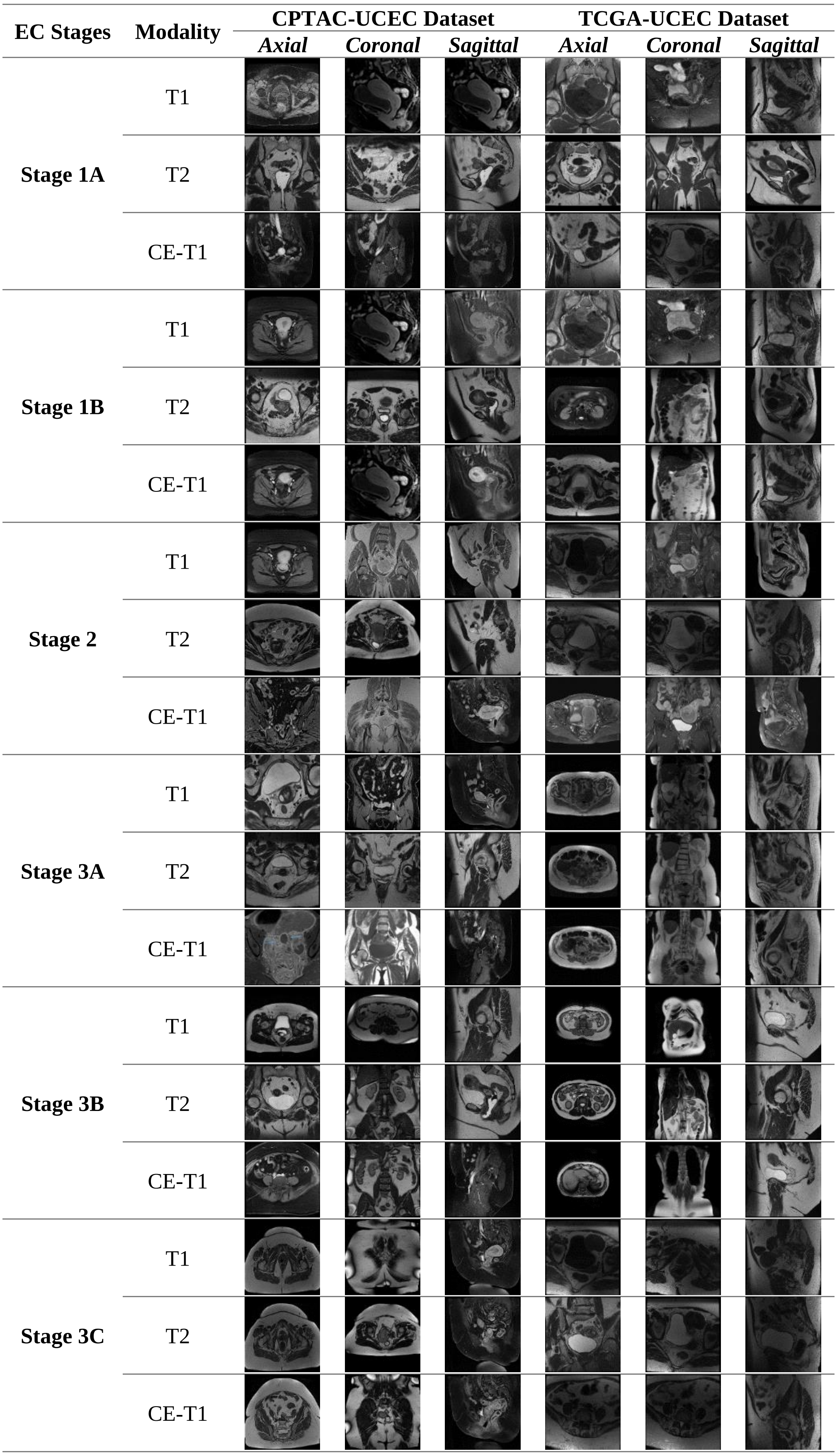

Figure. 4 Few samples of MRI sequences from CPTAC-UCEC and TCGA-UCEC datasets for different EC stages

Table 3. Parameter settings for existing models

Algorithms	Parameters	Range
CNN [16]	Learning rate	0.0001
	Epochs	150
	Optimizer	Adam
	Momentum	0.99
	Batch size	24
VGGNet-16 [21]	Learning rate	0.00001
	Number of epochs	200
	Batch size	64
	Momentum	0.9
	Optimizer	Stochastic gradient descent
	Dropout rate	0.5
SVM [22]	Kernel	Linear
	Penalty	0.1
InceptionResNet [23]	Learning rate	0.0001
	Batch size	64
	Dropout rate	0.5
	Optimizer	Adam

convolution kernel dimension in the 3 Conv layers is assigned as: $S_1 = 2, S_2 = 2$, and $S_3 = 8$. The respective channel numbers of 3 Conv layers are assigned as: $K_1 = 8, K_2 = 16$, and $K_3 = 32$. The Adam optimizer is utilized for model training with a primary training rate of 0.0001, an epoch number of 250, and a batch dimension of 20. As well, a dropout rate of 0.5 is applied on the FC layer to prevent overfitting. Table 3 lists the parameters used for the existing models.

The performance metrics considered in this study are described below.

- Precision: It is calculated in Eq. (11).

$$Precision = \frac{TP}{TP+FP} \tag{11}$$

- Recall: It is calculated in Eq. (12).

$$Recall = \frac{TP}{TP+FN} \tag{12}$$

- F-score: It is measured in Eq. (13).

$$F = \frac{2 \times Precision \times Recall}{Precision+Recall} \tag{13}$$

- Accuracy: It defines the fraction of accurate detection of EC stages over total images tested.

$$Accuracy = \frac{True\ Positive\ (TP)+True\ Negative\ (TN)}{TP+TN+False\ Positive\ (FP)+False\ Negative\ (FN)} \tag{14}$$

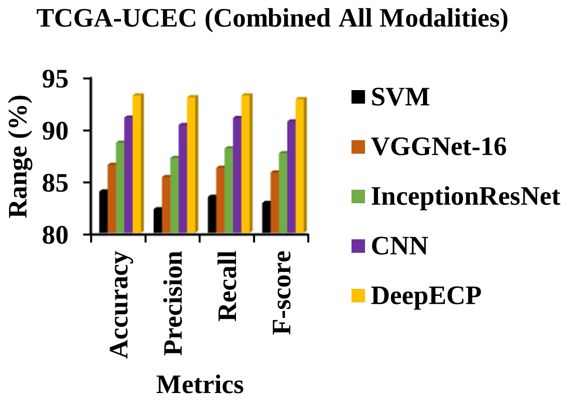

Figure. 5 Comparison of various EC detection models on TCGA-UCEC dataset

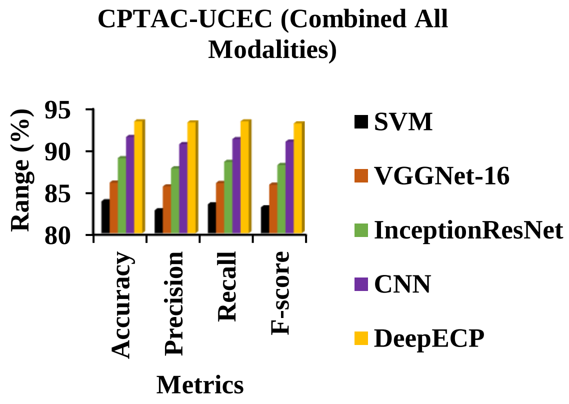

Figure. 6 Comparison of various EC detection models on CPTAC-UCEC dataset

In Eq. (11), TP is the quantity of positive images detected as positive, TN is the quantity of negative images detected as negative, FP is the quantity of negative images detected as positive and FN is the quantity of positive images detected as negative.

In Fig. 5, the performance of various existing and proposed EC detection models tested on the TCGA-UCEC dataset is plotted. It is reflected that the DeepECP model attains the greatest effectiveness regarding accuracy, precision, recall and F-score, contrasted to the other models. On average, for the concatenated MRI modalities from the TCGA-UCEC dataset, the accuracy of the DeepECP is improved by 10.96%, 7.69%, 5.11% and 2.33% compared to the SVM, VGGNet-16, InceptionResNet and CNN, respectively. The precision of the DeepECP model is increased by 13.04%, 8.96%, 6.68% and 2.94% compared to the SVM, VGGNet-16, InceptionResNet and CNN, respectively. The recall of the DeepECP is improved by 5.11%, 6.68%, 5.76% and 5.91% compared to the SVM, VGGNet-16, InceptionResNet and CNN models, respectively.

Also, the f-score of the DeepECP is increased by 12.01%, 8.2%, 5.91% and 2.36% compared to the SVM, VGGNet-16, InceptionResNet and CNN models, respectively.

Fig. 6 illustrates the different metrics resulted by the different EC detection models on the CPTAC-UCEC dataset. It is noted that the DeepECP model achieves maximum performance in detecting different EC stages automatically by learning various features from the multi-modality MRI sequences compared to the others.On average, the accuracy of the DeepECP on the combined MRI modalities from the CPTAC-UCEC dataset is increased by 11.37%, 8.5%, 4.94%, and 2.04% compared to the SVM, VGGNet-16, InceptionResNet, and CNN models, respectively. The precision of the DeepECP model is improved by 12.67%, 8.92%, 6.23%, and 2.86% compared to the SVM, VGGNet-16, InceptionResNet, and CNN models, respectively. The recall of the DeepECP model is better than 11.86%, 8.55%, 5.46%, and 2.32% compared to the SVM, VGGNet-16, InceptionResNet, and CNN models, respectively. Also, the f-score of the DeepECP model is improved by 12.05%, 8.52%, 5.64%, and 2.4% compared to the SVM, VGGNet-16, InceptionResNet, and CNN models, respectively.

5. Conclusion

This study developed the DeepECP model for EC detection. An extensive experiment was conducted to assess the efficiency of DeepECP on the two distinct MRI datasets. The results proved that the DeepECP model on the TCGA-UCEC dataset has 93.2% accuracy, which is 10.96%, 7.69%, 5.11% and 2.33% higher than the SVM, VGGNet-16, InceptionResNet and CNN models, respectively. Similarly, the DeepECP model on the CPTAC-UCEC dataset has 93.3% accuracy, which is 11.37%, 8.5%, 4.94% and 2.04% higher than the SVM, VGGNet-16, InceptionResNet and CNN models, respectively, for EC staging in contrast with the other CNN models. On the other hand, localizing cancer regions in the MRI scans are essential for effective medical diagnosis. To deal with this task, UNet-based segmentation models are broadly used, but it has a limitation because of the intrinsic locality of convolution operations. So future work will focus on solving this problem by adopting a new segmentation model for EC localization.

Conflict of interest

The authors declare no conflict of interest.

Author contributions

Conceptualization, methodology, software, validation, Karthick; formal analysis, investigation, Kamatchi; resources, data curation, writing—original draft preparation, Karthick; writing—review and editing, Karthick; visualization, supervision, Kamatchi;

References

[1] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, “Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries”, *CA: A Cancer Journal for Clinicians*, Vol. 71, No. 3, pp. 209-249, 2021.

[2] Cancer of the Endometrium - Cancer Stat Facts. SEER. (n.d.). Retrieved March 31, 2023, from https://seer.cancer.gov/statfacts/html/corp.html

[3] P. Huang, X. Fan, H. Yu, K. Zhang, H. Li, Y. Wang, and F. Xue, “Glucose Metabolic Reprogramming and its Therapeutic Potential in Obesity-Associated Endometrial Cancer”, *Journal of Translational Medicine*, Vol. 21, No. 1, pp. 1-14, 2023.

[4] K. Njoku, C. E. Barr, and E. J. Crosbie, “Current and Emerging Prognostic Biomarkers in Endometrial Cancer”, *Frontiers in Oncology*, Vol. 12, pp. 1-16, 2022.

[5] G. Trojano, C. Olivieri, R. Tinelli, G. R. Damiani, A. Pellegrino, and E. Cicinelli, “Conservative Treatment in Early Stage Endometrial Cancer: A Review”, *Acta Bio Medica: Atenei Parmensis*, Vol. 90, No. 4, pp. 405-410, 2019.

[6] R. V. D. Helder, B. M. Wever, N. E. van Trommel, A. P. V. Splunter, C. H. Mom, J. C. Kasius, and R. D. Steenbergen, “Non-invasive Detection of Endometrial Cancer by DNA Methylation Analysis in Urine”, *Clinical Epigenetics*, Vol. 12, No. 1, pp. 1-7, 2020.

[7] V. Bhardwaj, A. Sharma, S. V. Parambath, I. Gul, X. Zhang, P. E. Lobie, and V. Pandey, “Machine Learning for Endometrial Cancer Prediction and Prognostication”, *Frontiers in Oncology*, Vol. 12, pp. 1-16, 2022.

[8] Y. X. Gan, Q. H. Du, J. Li, Y. P. Wei, X. W. Jiang, X. M. Xu, and W. Q. Liu, “Adjuvant Radiotherapy after Minimally Invasive Surgery in Patients with Stage IA1-IIA1 Cervical Cancer”, *Frontiers in Oncology*, Vol. 11, pp. 1-8, 2021.

[9] W. C. Chen, L. T. Hsu, Y. T. Huang, Y. B. Pan, S. H. Ueng, H. H. Chou, and T. C. Chang,

"Prediction of Myometrial Invasion in Stage I Endometrial Cancer by MRI: The Influence of Surgical Diagnostic Procedure", *Cancers*, Vol. 13, No. 13, pp. 1-11, 2021.

[10] P. Gul, K. Gul, M. O. Altaf, A. Javaid, and J. Ashraf, "The Accuracy of MRI in the Local Staging of Endometrial Cancer: An Experience from a Tertiary Care Oncology Institute in Pakistan", *Cureus*, Vol. 14, No. 11, pp. 1-10, 2022.

[11] N. Concin, X. M. Guiu, I. Vergote, D. Cibula, M. R. Mirza, S. Marnitz, and C. L. Creutzberg, "ESGO/ESTRO/ESP Guidelines for the Management of Patients with Endometrial Carcinoma", *International Journal of Gynecologic Cancer*, Vol. 31, No. 1, pp. 12-39, 2021.

[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] K. A. Tran, O. Kondrashova, A. Bradley, E. D. Williams, J. V. Pearson, and N. Waddell, "Deep Learning in Cancer Diagnosis, Prognosis and Treatment Selection", *Genome Medicine*, Vol. 13, No. 1, pp. 1-17, 2021.

[14] A. Esteva, K. Chou, S. Yeung, N. Naik, A. Madani, A. Mottaghi, and R. Socher, "Deep Learning-Enabled Medical Computer Vision", *NPJ Digital Medicine*, Vol. 4, No. 1, pp. 1-9, 2021.

[15] A. Echle, N. T. Rindtorff, T. J. Brinker, T. Luedde, A. T. Pearson, and J. N. Kather, "Deep Learning in Cancer Pathology: A New Generation of Clinical Biomarkers", *British Journal of Cancer*, Vol. 124, No. 4, pp. 686-696, 2021.

[16] A. 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 Medical Imaging*, Vol. 22, No. 1, pp. 1-14, 2022.

[17] A. A. Mukhlif, B. A. Khateeb, and M. A. Mohammed, "An Extensive Review of State-of-the-art Transfer Learning Techniques Used in Medical Imaging: Open Issues and Challenges", *Journal of Intelligent Systems*, Vol. 31, No. 1, pp. 1085-1111, 2022.

[18] S. Candemir, X. V. Nguyen, L. R. Folio, and L. M. Prevedello, "Training Strategies for Radiology Deep Learning Models in Data-Limited Scenarios", *Radiology: Artificial Intelligence*, Vol. 3, No. 6, pp. 1-10, 2021.

[19] A. M. Praiss, Y. Huang, C. M. S. Clair, A. I. Tergas, A. Melamed, F. Khoury-Collado, and J. D. Wright, "Using Machine Learning to Create Prognostic Systems for Endometrial Cancer", *Gynecologic Oncology*, Vol. 159, No. 3, pp. 744-750, 2020.

[20] H. C. Dong, H. K. Dong, M. H. Yu, Y. H. Lin, and C. C. Chang, "Using Deep Learning with Convolutional Neural Network Approach to Identify the Invasion Depth of Endometrial Cancer in Myometrium Using MR Images: A Pilot Study", *International Journal of Environmental Research and Public Health*, Vol. 17, No. 16, pp. 1-18, 2020.

[21] Y. Zhang, Z. Wang, J. Zhang, C. Wang, Y. Wang, H. Chen, and X. Ma, "Deep Learning Model for Classifying Endometrial Lesions", *Journal of Translational Medicine*, Vol. 19, pp. 1-13, 2021.

[22] M. Akazawa, K. Hashimoto, K. Noda, and K. Yoshida, "The Application of Machine Learning for Predicting Recurrence in Patients with Early-Stage Endometrial Cancer: A Pilot Study", *Obstetrics & Gynecology Science*, Vol. 64, No. 3, pp. 266-273, 2021.

[23] R. Hong, W. Liu, D. D. Lair, N. Razavian, and D. Fenyö, "Predicting Endometrial Cancer Subtypes and Molecular Features from Histopathology Images Using Multi-Resolution Deep Learning Models", *Cell Reports Medicine*, Vol. 2, No. 9, pp. 1-17, 2021.

[24] X. Chen, Y. Wang, M. Shen, B. Yang, Q. Zhou, Y. Yi, and H. Zhang, "Deep Learning for the Determination of Myometrial Invasion Depth and Automatic Lesion Identification in Endometrial Cancer MR Imaging: A Preliminary Study in a Single Institution", *European Radiology*, Vol.30, pp. 4985-4994, 2020.

[25] 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", *Frontiers in Physiology*, Vol. 13, pp. 1-12, 2022.

[26] The Cancer Genome Atlas Uterine Corpus Endometrial Carcinoma Collection (TCGA-UCEC) - The Cancer Imaging Archive (TCIA) Public Access - Cancer Imaging Archive Wiki. (n.d.). Retrieved April 1, 2023, from https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=19039602.

[27] The Clinical Proteomic Tumor Analysis Consortium Uterine Corpus Endometrial Carcinoma Collection (CPTAC-UCEC) - The Cancer Imaging Archive (TCIA) Public Access

- Cancer Imaging Archive Wiki. (n.d.). Retrieved April 1, 2023, from https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=33948263