Endometriosis Lesion Classification Using Deep Transfer Learning Techniques

Shujaat Ali Zaidi[1], Varin Chouvatut[2]*, Chailert Phongnarisorn[3]
Department of Computer Science-Faculty of Science, Chiang Mai University, Mueang, Chiang Mai 50200, Thailand[1,2]
Department of Obstetrics and Gynecology-Faculty of Medicine, Chiang Mai University,
Mueang, Chiang Mai 50200, Thailand[3]

***Abstract*—In resource-limited settings, assisting physicians with disease identification can significantly improve patient outcomes. Early diagnosis is crucial, as many patients could remain healthy with timely intervention. Recent advancements in deep learning models for medical image processing have enabled algorithms to achieve diagnostic accuracy comparable to that of healthcare professionals. This research aims to develop a comprehensive system for the rapid and precise detection of endometriosis lesions. We explore the several deep transfer learning architectures, specifically MobileNetV2, VGG19, and InceptionV3, on the Gynecologic Laparoscopy Endometriosis Dataset (GLENDA). Through extensive literature review and parameter optimization, we find that MobileNetV2 outperforms the other models in terms of accuracy. However, challenges remain, as healthcare imaging datasets often suffer from limited sample sizes and uneven class distributions. Collecting additional samples can be costly and time-consuming, which is a prevalent issue in medical imaging. To address this, we employ Deep Convolutional Generative Adversarial Networks (DCGAN) to enhance the dataset by generating synthetic images, thus improving class balance. This image augmentation strategy not only boosts model performance but also reduces the manual effort required for image labeling. We evaluate our proposed model using metrics such as accuracy, precision, recall, and F1-score. Initially, our model achieves an accuracy of 95%. The introduction of synthetic samples results in an increased accuracy of 99%, reflecting a 4% improvement and enhancing the model's overall efficacy.**

Keywords*—*Endometriosis classification; lesion detection; medical image classification; deep learning; transfer learning; DCGAN

I. Introduction

Endometriosis is a medical condition in which endometrial cells, naturally found within the uterus, begin to grow outside the uterine cavity [1], [2]. The preliminary hypothesis describing this is retrograde menstruation, where menstrual tissue pursues a distinctive path. Rather than being expelled during the menstrual cycle, the tissue flows in reversal, traveling up through the fallopian tubes and potentially implanting in the ovary or within the abdominal cavity, resulting in lesion construction. These lesions are differ in size, with classification based on dimensions critical to diagnosis and treatment. Mainly, lesions measuring five millimeters or more are typically classified as deep-seated endometriosis [3], [4].

Endometriosis is frequently associated with significant diagnostic delays [5], [6]. Endometriosis disorder has a broad spectrum of symptoms, such as menstrual pain, acute pelvic pain, infertility, and pain during intercourse. Non-gynecological symptoms can also appear, including pain while urinating, discomfort during bowel movements, flank pain, fatigue, and blood in the urine, among others. Furthermore, the biological exams are complicated, and results can differ widely, even in examinations conducted by professionals in the field [7], [8].

*Corresponding authors.

Endometriosis is a severe gynecological problem impacting an estimated 190 million women internationally [9], [10]. The disease is prevalent among women of different age class, from infant women to those who are post-menopause. Studies estimate that approximately 10 percent of women of reproductive age are affected by endometriosis [11], with a prevalence of 2 percent to 4 percent among postmenopausal women [12]. Additionally, The disease's influence on young women is significant, with nearly 50 percent of those undergoing persistent pelvic pain occurring before the age of 20 being diagnosed with endometriosis [4].

Endometriosis can be diagnosed through various procedures, including ultrasound, magnetic resonance imaging (MRI), and laparoscopy [13], [14]. Laparoscopy is the gold benchmark, as it permits explicit examination of the abdominal and pelvic areas through a camera inserted via a small incision. MRI, a non-invasive imaging method, uses magnetic areas to create detailed images of the body's internal structures, making it beneficial in recognizing deep lesions or more extensive cysts associated with endometriosis. Ultrasound, generally performed transvaginally, is usually utilized for initial screening and assists in detecting ovarian cysts or other abnormalities [15], [16], [17], [18].

In recent years, computer-aided methodologies have become essential in medical fields for disease diagnosis, including identifying conditions such as heart disease [19], endometriosis lesions detection [20], and classifying breast cancer [21]. However, Healthcare imaging datasets frequently suffer from insufficiency and uneven class distributions. Moreover, acquiring additional samples is both expensive and time-intensive. This is a common problem in the medical image domain. Researchers endeavor to conquer this problem by utilizing data augmentation. Image augmentation methods typically fall into two main categories: conventional approaches and deep learning techniques. Conventional image augmentation methods apply simple manipulations like translation, cropping, flipping, or resizing. Deep learning-based augmentation utilizes state-of-the-art neural networks to create more adaptive, data-rich transformations. These two techniques enhance data diversity, helping models to generalize more effectively [22],

[23].

A promising deep-learning technique for image synthesis inspired by the game hypothesis is the Generative Adversarial Network (GAN) [24], [25]. In GANs, two networks are learned in an adversarial manner: one network, the generator, constructs artificial images, while the other, the discriminator, learns to distinguish between real and synthetic images. By repeating this adversarial approach, GANs enhance both networks' abilities, allowing the generation of high-quality, realistic images. The computer vision community has adopted GANs, directing to various variations developed for photorealistic image generation [26], [27].

Deep learning has attracted significant attention in many fields (for more detail, see [28], [29], [30], [31], [32]) such as the medical domain because of its efficacy in image analysis. It is employed for classification, enhancement of image quality, and segmentation of medical images. In recent years, deep learning techniques have rapidly evolved, with MobileNetV2 attaining popularity due to its compact architecture, which makes it perfect for various applications [33]. The MobileNetV2 architecture contains an internal structure with a linear bottleneck, an element that minimizes memory requirements for better processing. Therefore, to provide an accurate and efficient classification of endometriosis lesions, in this study, we propose an endometriosis image classification founded on the MobileNetV2 architecture. The proposed method uses MobileNetV2 as the base model for the transfer learning process. We add a global pooling layer and two fully connected layers to enhance the model's performance and refine classification outcomes [34]. The primary contributions of this research are as follows.

- We propose a synthesis of high-quality endometriosis lesions from laparoscopy images employing deep convolutional generative adversarial networks (DCGANs).
- Several state-of-the-art deep learning architectures, such as VGG19, InceptionV3, and MobileNetV2, are analyzed for this research. These architectures are fine-tuned on the laparoscopy images dataset to obtain efficient and accurate endometriosis lesion detection.
- We improve the training set of our deep learning model by including synthetic data, a technique that guides to enhanced classification results.

The remainder of the article is organized as follows: Section 2 explains the literature review on utilizing deep learning to analyze endometriosis lesions. Sections 3 and 4 present the proposed methodology and experimental result analysis. Consequently, Sections 5 and 6 provide the discussion and conclusion of the study, respectively.

II. Literature Review

Detecting endometriosis through laparoscopy imaging is challenging due to the disease's complex and varied presentation. Its lesions usually have subtle or indistinct characteristics, making them challenging to identify accurately, even by specialists. Machine learning applications aimed at diagnosing endometriosis through laparoscopy imaging still need to be developed, partially due to the limited access to labeled, representative datasets. This section summarizes recent research in machine learning for endometriosis diagnosis, focusing on techniques, methodologies, and the insights they provide.

Zaidi [5] employed a deep-learning based approach to detect endometriosis lesions from laparoscopy images.The team achieved a accuracy of 0.93% by applying Inception V3 model with 5-fold cross-validation techniques. Deep convolutional networks, exemplified as GoogLeNet [35], have been effectively employed across various applications, showcasing their versatility and robust feature extraction capabilities. These networks serve as robust backbones for many state-of-the-art deep-learning architectures used in image and video research analysis. By leveraging hierarchical layers of convolutional filters, they can capture complicated patterns and attributes, making them ideal for object recognition, classification, and segmentation tasks. Visalaxi [36] employed a deep-learning approach to categorize laparoscopy images associated with endometriosis. They employed the dataset for gynecologic laparoscopy related to endometriosis [37], comprising about 6,000 images from laparoscopy videos. Sixty percent of this dataset was used for training, and several architectures were tested to determine the model with the best performance. The ResNet50 architecture [38], loaded with ImageNet pre-trained weights [39], achieved the highest accuracy at 90 percent, with sensitivity and precision scores of 82 percent and 83 percent, respectively.Leibetseder's study used transfer learning and the Faster R-CNN [40] and Mask R-CNN [41] models to accurately detect endometriosis in laparoscopic images, getting a 32.4 percent precision. In addition, the research investigated different data augmentation methods, revealing that a blend of cropping and rotation yielded optimal results. The GLENDA dataset was employed as the primary data source in this research.

In 2021, Yun [42] and colleagues introduced a neural network model specifically designed to aid in classifying endometriosis. This study employed a convolutional neural network (CNN) architecture called VGGNet-16. Using a dataset of 6,478 histopathology images for training. The researchers aimed to develop a highly accurate system that could support and potentially enhance the diagnostic work performed by radiologists. In their research, Takahashi and colleagues [43] explored how computer vision techniques can support detecting endometriosis cancer in laparoscopic images. The team achieved a prediction accuracy of 90.29% by applying advanced neural network techniques. Sudalaimuthu [44] proposed an innovative approach recognized as Structural Similarity Analysis of Endometriosis (SSAE). This approach evaluates endometriosis progression in areas such as the ovaries, uterus, rectum, and peritoneum founded on laparoscopy images from the GLENDA image dataset. To enhance the model's robustness, images were subjected to data augmentation techniques, including horizontal and vertical shifts, rotation, shear, and zoom. The data was split, with 70 percent allocated for training purpose and the 30 percent for testing purpose. The U-Net architecture was utilized to explore factors like filter sizes and optimization methods. The highest results recorded were dice coefficient is 0.74 and an IoU is 0.72, though it is unclear if these metrics were specific to the test set.The literature on automated identification and segmentation of endometriosis remains limited, with few works addressing these areas. Among the existing studies, most rely on images from laparoscopy procedures, which, while invasive, are the gold benchmark for diagnosis. Despite using these high-quality images, automated approaches are still being developed with the aim of improving accuracy and performance.

Our research differentiates itself from the abovementioned research by suggesting a data augmentation approach to conventional and deep learning approaches that encompass detection and classification tasks for analyzing deep or deep-seated endometriosis. Additionally, it focauses on laparoscopy images, seeking at premature examination beyond the essential for invasive procedures.

III. Materials and Methods

This section provides a detailed description of the materials used and the approach taken for classifying endometriosis lesions, including an overview of the image collection process, a summary of the overall methodology, and a discussion of the proposed methodology. Figure(1) provides a flowchart of the methodology followed in this research.

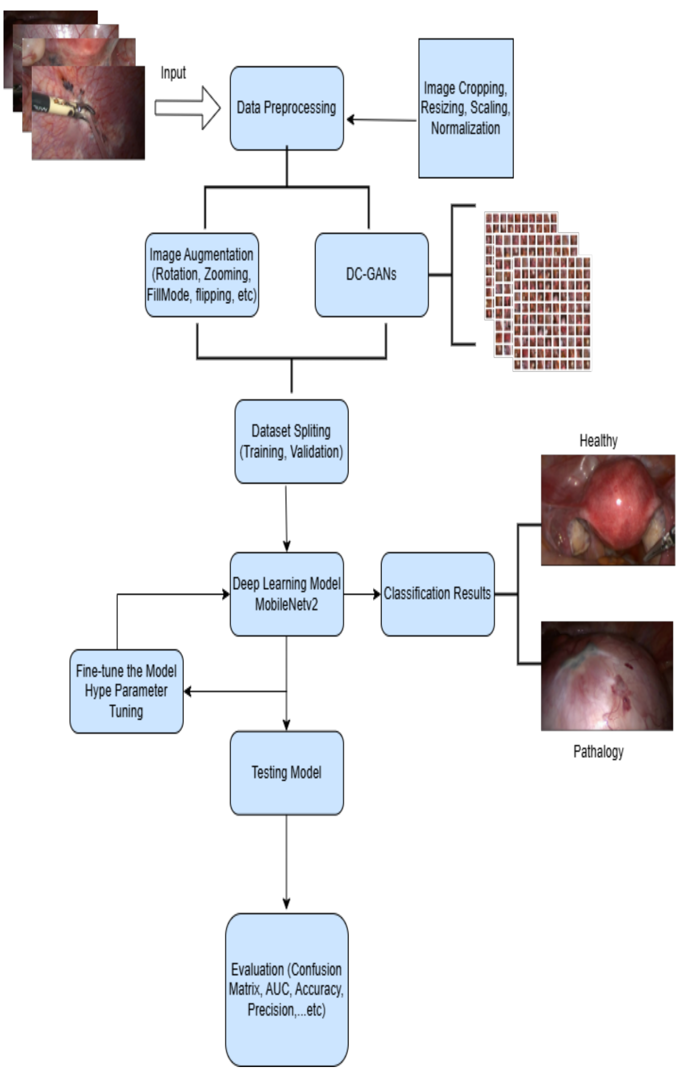

Fig. 1. A Flowchart of proposed methodology.

A. Data Collection

The Gynecologic Laparoscopy Endometriosis Dataset (GLENDA) is a comprehensive dataset derived from more than 400 gynecologic laparoscopy video recordings, a significant number of which depict instances of endometriosis with varying phases of severity. GLENDA comprises over 25,000 images, including more than 12,000 positive pathological photographs related to endometriosis and over 13,000 negative non-pathological images devoid of prominent endometriosis. The dataset is intentionally designed for various artificial content analysis tasks related to endometriosis recognition. The dimensions of the images are 640 by 360 pixels.

B. Image Pre-processing

Image preprocessing is a collection of procedures used on raw images to prepare them for additional examination or processing. The primary aim is to improve the image by identifying pertinent information while reducing artifacts, noise, and extraneous features that might interfere with further analysis. Prevalent preprocessing tasks include resizing, filtering, cropping, noise attenuation, color modification, and image enhancement. The phases are chosen according to the image properties and the particular application's requirements. In this research work, All images were resized to 224×224 pixels. We also applied image sharpening to the entire dataset to improve the quality of the images. Furthermore, the images are processed in RGB format. We implemented many rescaling methods, including multiplying each pixel by 1/255. This generalizes the input, conserves memory, and reduces the computational expense of applicable procedures. Moreover, it also facilitates the understanding of the ideal function.

C. Data Augmentation Using Image Data Generator Methods

This approach is used to generate data examples from the current samples. It is beneficial when the dataset contains a limited number of instances or exhibits class imbalance. We may use many geometric techniques to generate augmented data. This approach facilitates the learning of a class-imbalanced dataset and enhances the model's generalization, reducing overfitting. The model acquires the ability to handle unexplored versions of training samples. We used the Image Data Generator package from Kera's to do image augmentation. The following are the approaches that were executed for this objective:

1) Random zoom: It dynamically zoomed into the provided image. In this work, the zoomed value is fixed to 0.1 percent.

2) Random rotation: It dynamically rotates the provided image based on the specified value. The rotated value is fixed at 30 degrees.

3) Horizontal flip: This option randomly inverts the image horizontally to alter the positioning of its sides.

4) Width shift: It horizontally alters the image's width, with a value established at 0.2 percent of its overall width.

5) Height shift: It vertically alters the image's height, with a value established at 0.2 percent of its overall height.

D. Data Augmentation Using DCGAN Method

Deep Convolutional Generative Adversarial Networks (DCGANs) are a class of generative models that utilize deep convolutional layers to generate realistic images from random noise. They are built upon the framework of traditional GANs, which consist of two neural networks: a generator and a discriminator. Figure (2 and 3) illustrates its architecture. The generator learns to produce synthetic images, while the discriminator evaluates their authenticity. Both networks are trained simultaneously in an adversarial process, where the generator aims to create images that can deceive the discriminator, and the discriminator strives to distinguish between real and generated images.The generator creates progressively more naturalistic pictures throughout the training to deceive the discriminator. This adversarial strategy incrementally enhances the quality of the produced pictures. The generator learns explicitly to translate random noise vectors into picture spaces that mimic the distribution of authentic endometriosis images. At the same time, the discriminator acquires the ability to differentiate between actual and faked images. Ultimately, this leads to the generator creating high-fidelity synthetic endometriosis pictures that closely mimic authentic ones.

DCGANs introduce specific architectural enhancements, including the use of convolutional layers instead of fully connected layers, which improves spatial structure preservation in the generated images. Additionally, techniques like batch normalization and the use of Leaky ReLU activation functions enhance training stability and convergence. These improvements enable DCGANs to generate high-quality and coherent images across various domains, such as medical picture classification, and objects.

The structured use of convolutional layers helps the model learn hierarchical representations, which are crucial for producing realistic outputs. Furthermore, the adversarial training approach encourages the generator to refine its outputs iteratively, leading to visually convincing results. Due to their effectiveness and simplicity, DCGANs have become a foundational model in the field of generative image synthesis and serve as a basis for more advanced architectures in related applications.

DCGANs can generate data from existing samples. In situations like medical picture classification, gathering sufficient data to build a big deep learning model is crucial, which might be arduous. Utilizing presently accessible data, we may produce new samples with DCGANs; these synthetic samplings do not directly correspond to actual patients. This technology has expedited the use of deep learning in medical imaging; nonetheless, it requires significant improvement in accurately collecting delicate tissues from images. In this study employs a deep convolutional GAN (DCGAN) [46] to generate synthetic endometriosis images featuring pathologies diseases in the used dataset. A deep generative model of around five layers was used with a relatively slight discriminator network. The complex architecture assists in detecting subtle tissues within the endometriosis images. DCGAN was used for this analysis because of its stability in terms of simplicity and efficacy in producing high-quality pictures. In contrast to more intricate GAN variations like CGAN, BEGAN, WGAN, and DCGAN, it offers a simple architecture that is easy to build and tune, rendering it appropriate for our application in medical image creation. Moreover, DCGAN has been extensively evaluated across many image synthesis tasks, showing its reliability and effectiveness in creating realistic images, which coincides with our objective of enhancing medical datasets.

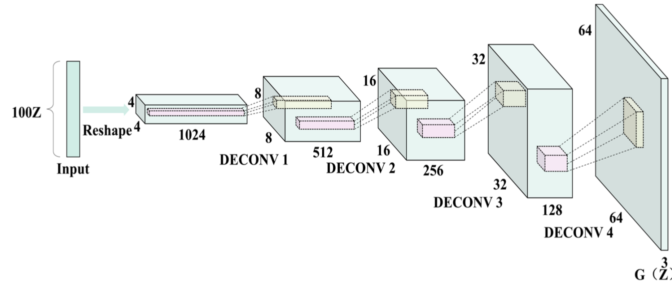

Fig. 2. Generator architecture [49].

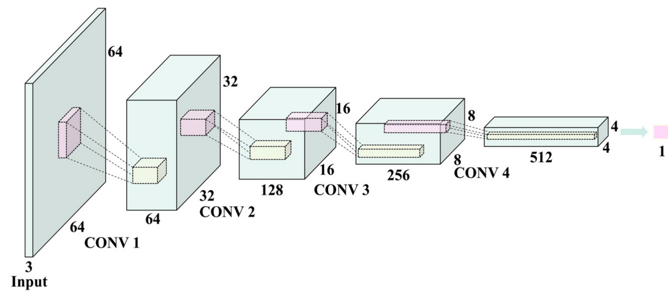

Fig. 3. Discriminator architecture [49].

We trained DCGANs to synthesize images of endometriosis lesions, with separate models developed for each lesion category. The training process involved alternating updates to the generator and discriminator networks in an iterative manner, allowing both networks to improve their performance progressively. A learning rate of 0.0002 was employed, and training was conducted for 200 epochs for each lesion category to ensure adequate learning of the underlying data distribution.

The input dataset consisted of original images with a resolution of 224 × 224 pixels, and a batch size of 128 was used during training to optimize computational efficiency. Upon completion of the training process, we generated 300 synthetic images for each lesion category. Each class-specific training session required approximately 20 hours to complete on a local PC equipped with an Intel i7-7820HQ CPU. These synthesized images were then merged with the original dataset to augment the data available for subsequent analysis or model training. This approach ensured an expanded and balanced dataset that could potentially improve downstream performance in tasks such as endometriosis lesions classification or segmentation.

To evaluate the visual quality of the synthesized images, a radiologist with expertise in endometriosis assessed the synthetic lesions for their resemblance to real clinical cases. The radiologist's analysis focused on key morphological features, such as texture, contrast, and spatial distribution, to determine whether the synthetic lesions accurately mirrored the appearance of genuine endometriosis implants. By conducting training independently for each lesion category and ensuring consistency in the training parameters, we aimed to achieve high-quality image synthesis across all categories. This method

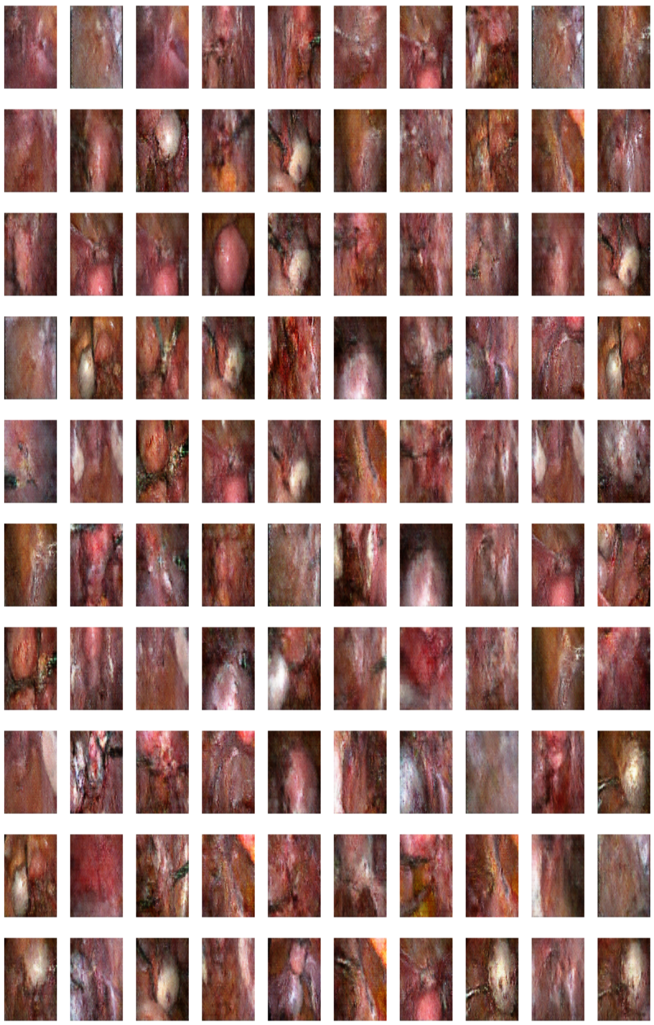

Fig. 4. DCGANs synthetic images.

highlights the potential of DCGANs in generating synthetic data that closely resembles real-world medical images, contributing to the development of robust datasets in medical imaging research.Figure (4) displays examples created for the pathology class after 200 epochs.

E. Training Phase

This section emphasizes some of the primary details applied in the training stage, such as kinds of models, hyperparameters, and classifier architectures.

1) Dataset split: In supervised machine learning and deep learning workflows, datasets are first divided into distinct sets for training, validation, and testing. Image augmentation, commonly performed on the training and potentially validation sets, aids in model generalization during the training and optimization stages. To ensure unbiased results, test data usually remains unaugmented to evade data leakage. Nevertheless, recent research highlights that augmenting test data can also be valuable in examining model robustness.In this work, the dataset is divided into 70% for training, 20% for validation, and 10% for testing.

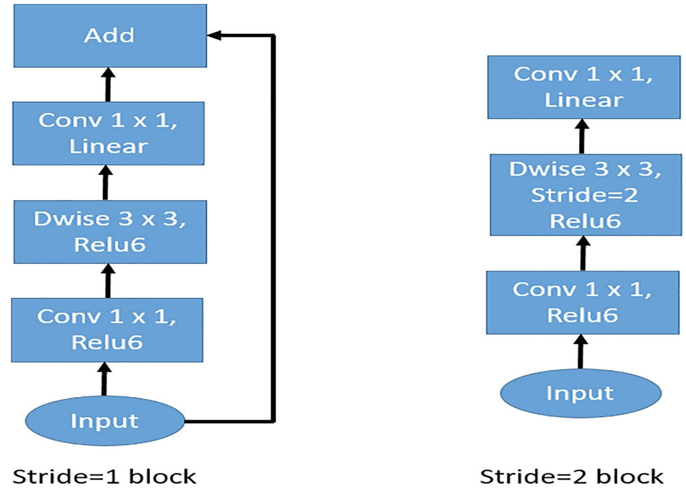

Fig. 5. MobileNetV2 architecture [47].

- Training Images: 17977
- Validation Images: 5136
- Testing Images: 2569

2) Model architectures: In this study, several deep learning models were evaluated using the Glenda dataset to assess their performance across different architectures. Three models (VGG19, InceptionV3, MobileNetV2)were implemented utilizing the Keras library in Python. Transfer learning was utilized for the convolutional layers to initialize these models, while random initialization was applied to the fully connected layers. The classifier network was built with a custom design that included two hidden layers and an output layer containing two neuron units, each assigned to a specific class. This minimalist design, featuring fewer layers and neurons, was strategically utilized to avoid overfitting.

3) MobileNetV2: MobileNetV2 [48] is recognized as a lightweight convolutional neural network (CNN) that is generally utilized in miscellaneous applications. It improves the MobileNetV1 model by introducing new modules, especially inverted residuals and linear tie-ups. The essence design of MobileNet is established on depthwise divisible convolution. This approach differs from standard 2-dimensional convolution, which treats all input channels uniformly to produce one output channel. Instead, depthwise convolution applies filters independently to each input channel, resulting in separate output channels that are subsequently combined. The separable depth-wise convolution process involves a subsequent 1×1 point-wise convolution, which merges these output channels into one final channel. This method provides the same output as traditional convolution while being more efficient due to a reduced number of parameters.Figure 5 presents the graphical representation of the MobileNetV2 framework.

4) Hyper parameters: In this study, we kept hyper parameters uniform across all models, allowing for a precise and fair comparison of each model's performance. This consistent approach allowed for a clear assessment of each model's relative performance.

a) Optimizer: In our optimization process, we rigorously tested two methods: Stochastic Gradient Descent (SGD) and Adaptive Momentum Estimation (Adam). After multiple iterations, we found that both optimizers performed similarly. However, we chose the Adam optimizer for training, as it is a more advanced technique for learning model weights and has been proven to enhance convergence and stability in deep learning models.

b) Learning rate: Since we utilized pre-trained weights through transfer learning for the convolutional networks in our models, we opted for a low learning rate, set explicitly at 0.0001, to ensure stable training.

c) Activation functions: ReLU was implemented across all layers, excluding the output layer, adding non-linearity to enhance model performance. In the output layer, we applied the softmax function, which is well-suited for labeled classification, as it allows each neuron to generate separate outputs.

d) Loss function: We opted for Binary Cross-Entropy (BCE) to calculate loss, which allows us to consider each label independently.

IV. Results

This section presents a detailed account of the results obtained from our experimental research and describes the programming environments employed to achieve these results. For clarity, each aspect is discussed in separate subsections.

A. Experimental Setup

The large dataset influenced our choice of programming environments to conduct deep learning experiments. We carried out model implementation and training on our Dell laptop, assembled with a 1.80 GHz Intel Core i5 CPU and 16 gigabytes of RAM. Our approach leveraged Python libraries such as TensorFlow, Keras, and sci-kit-learn for effective attribute classification. Jupyter Notebook enabled efficient program development and data analysis, and we observed satisfactory results across several models.

B. Evaluation Procedure

In order to make a comprehensive comparison of model classification performance, this paper leverages various metrics associated with the confusion matrix, including precision, recall, F1-score, and accuracy.Accuracy calculates the model's prevalent correctness, while precision focuses on the reliability of optimistic predictions. Recall quantifies the model's sensitivity by measuring the proportion of true positives, and specificity assesses the ability to identify negative cases correctly. We clearly understand each model's strengths and weaknesses by calculating and analyzing these metrics. The formulas for calculating these performance metrics are provided below.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

$$\text{Precision} = \frac{TP}{TP+FP} \tag{2}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{3}$$

$$F1\text{-score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

In this equation, TP, or True Positive, refers to the count of positive cases accurately identified as positive. FN, or False Negative, represents the number of positive cases mistakenly classified as negative. FP, or False Positive, denotes the count of negative cases incorrectly labeled as positive, while TN, or True Negative, captures the number of negative cases correctly classified as negative.

C. Area Under the Curve (AUC)

The ROC curve is commonly used to evaluate how well a model can distinguish between different classes. It is based on plotting the True Positive Rate (TPR) and False Positive Rate (FPR) across different thresholds. Since TPR and FPR are directly proportional, they increase together. The ROC curve visually represents this relationship and enables calculation of the Area Under the Curve (AUC), where a larger AUC value implies a more robust classification capability. Thus, a higher AUC indicates better model performance.

D. VGG19

In analyzing endometriosis lesion images, spatial data is crucial for effective disease classification. Deep learning-based models are frequently employed to capture these essential features, and VGG19, a highly layered convolutional model, was chosen for this study. We applied VGG19 with pre-trained weights for feature extraction and added a custom three-layer classifier. The model was trained across three configurations with varying epochs (3, 7, 14, and 20) and a batch size of (32, 64, and 128), resulting in above-average performance metrics with an accuracy of up to 88%. Table I compiles the results of tests conducted without data augmentation, providing insights into the baseline performance of our endometriosis lesion detection algorithm. Table II, on the other hand, presents the outcomes when data augmentation techniques are applied, showcasing the improvements in detection accuracy and robustness achieved through augmentation. Figures 6 and 7 illustrate the loss trajectories and accuracy trends, respectively, for the VGG19 model trained on the baseline dataset (without augmentation). In contrast, the effects of incorporating data augmentation are evident in Figures 8 and 9, which depict the model's loss dynamics and classification performance under enhanced training conditions. Figure 10 shows the confusion matrices illustrating the endometriosis lesion classification outcomes from testing the VGG19 model. The AUC-ROC curve, which can be seen in Figure 11, also explains the result of the VGG19 model.

E. InceptionV3

The InceptionV3 model offers an optimized approach to achieve high classification accuracy without relying on an extensive network architecture. With fewer learnable parameters than the VGG19 model, InceptionV3 provides an efficient alternative that still delivers excellent classification performance.

TABLE I. VGG19 MODEL RESULTS WITHOUT DATA AUGMENTATION

Epochs	Batch-Size	Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)
3	32	56	54	55	56
	64	57	57	56	57
	128	59	59	59	60
7	32	63	63	63	63
	64	65	64	64	65
	128	66	65	66	66
14	32	70	70	69	70
	64	74	73	73	74
	128	77	77	77	78
20	32	80	80	79	80
	64	80	79	80	80
	128	81	81	81	82

TABLE II. VGG19 MODEL RESULTS WITH DATA AUGMENTATION

Epochs	Batch-Size	Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)
3	32	60	60	61	61
	64	62	61	62	63
	128	64	63	64	65
7	32	68	69	69	70
	64	72	72	73	73
	128	74	74	75	75
14	32	77	76	77	78
	64	79	79	79	80
	128	82	83	83	84
20	32	84	84	85	85
	64	87	87	87	88
	128	88	87	88	88

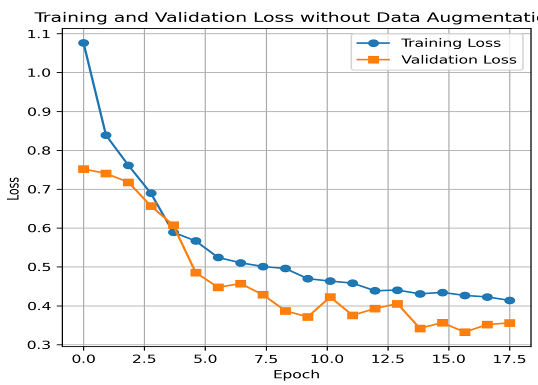

Fig. 6. VGG19 loss display without data augmentation.

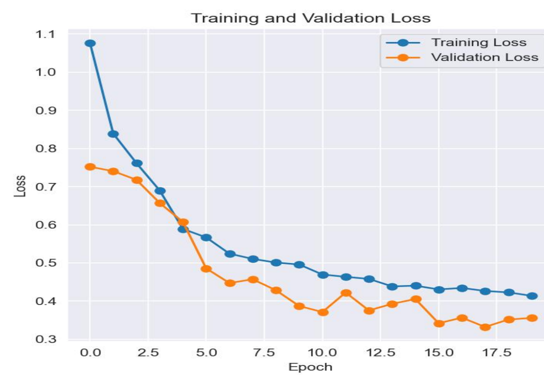

Fig. 8. VGG19 loss display.

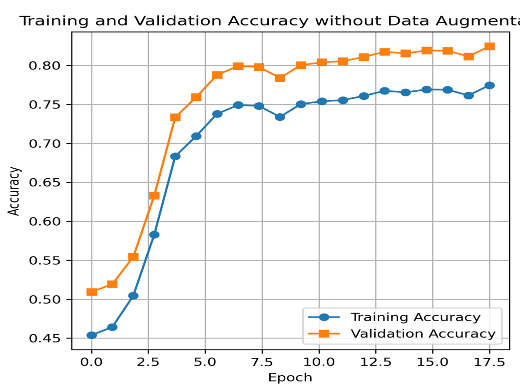

Fig. 7. VGG19 accuracy display without data augmentation.

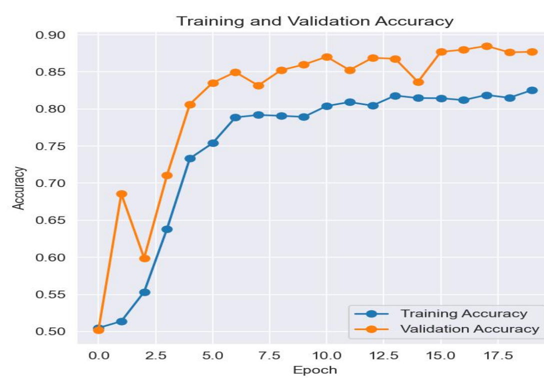

Fig. 9. VGG19 accuracy display.

Its design prioritizes streamlined architecture, balancing effectiveness with computational efficiency.InceptionV3 model was trained across three configurations with varying epochs (3, 7, 14, and 20) and a batch size of (32, 64, and 128), resulting in above-average performance metrics with an accuracy of up to 97%, that are higher than VGG19. Table III compiles the results of tests conducted without data augmentation, providing insights into the baseline performance of our endometriosis lesion detection algorithm. Table IV, on the other hand, presents the outcomes when data augmentation techniques are applied, showcasing the improvements in detection accuracy and robustness achieved through augmentation. Figures 12 and 13 illustrate the loss trajectories and accuracy trends, respectively, for the InceptionV3 model trained on the baseline dataset (without augmentation). In contrast, the effects of incorporating data augmentation are evident in Figures 14 and 15, which depict the model's loss dynamics and classification performance under enhanced training conditions. Figure 16 shows the confusion matrices illustrating the endometriosis lesion classification outcomes from testing the InceptionV3

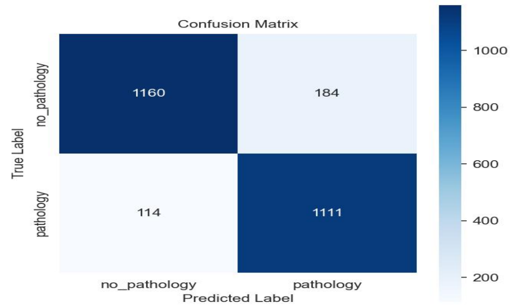

Fig. 10. VGG19 confusion matrix.

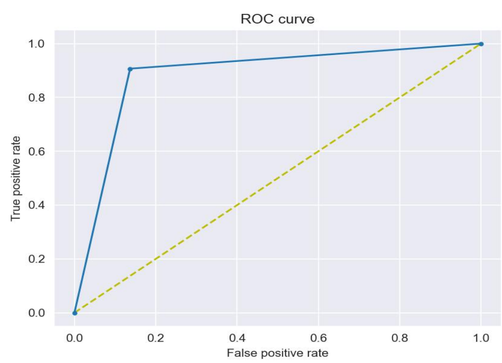

Fig. 11. VGG19 AUC-ROC curve.

model. The AUC-ROC curve, which can be seen in Figure 17, also explains the result of the InceptionV3 model.

TABLE III. INCEPTIONV3 MODEL RESULTS WITHOUT DATA AUGMENTATION

Epochs	Batch-Size	Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)
3	32	65	63	64	65
	64	67	67	67	67
	128	70	69	69	70
7	32	73	72	73	73
	64	75	75	75	76
	128	80	79	79	80
14	32	83	83	82	83
	64	87	85	86	87
	128	88	88	89	89
20	32	90	89	89	90
	64	90	91	90	91
	128	92	91	92	92

F. *MobileNetV2*

MobileNetV2 builds on the foundation of MobileNetV1 with an updated architecture featuring fewer learnable parameters for greater efficiency.MobileNetV2 model was trained across three configurations with varying epochs (3, 7, 14,

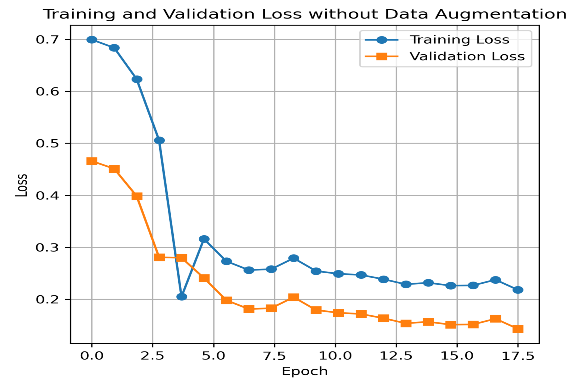

Fig. 12. InceptionV3 loss display without data augmentation.

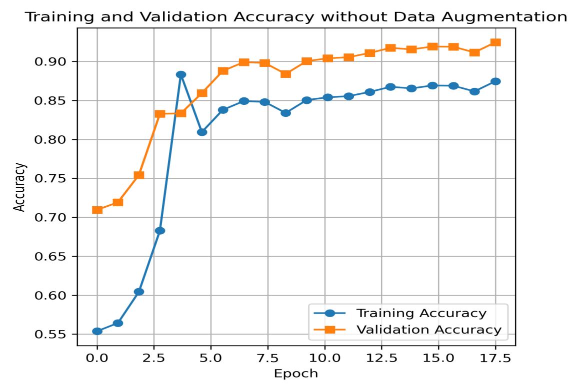

Fig. 13. InceptionV3 accuracy display without data augmentation.

TABLE IV. INCEPTIONV3 MODEL RESULTS WITH DATA AUGMENTATION

Epochs	Batch-Size	Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)
3	32	52	53	53	54
	64	56	56	57	58
	128	59	58	58	59
7	32	62	63	61	63
	64	68	69	71	72
	128	80	80	82	83
14	32	89	88	89	90
	64	90	91	89	92
	128	92	92	92	93
20	32	92	93	93	94
	64	95	94	96	96
	128	96	97	97	97

and 20) and a batch size of (32, 64, and 128), resulting in above-average performance metrics with an accuracy of up to 99%, which are higher than InceptionV3. Table V compiles the results of tests conducted without data augmentation, providing insights into the baseline performance of our endometriosis lesion detection algorithm. Table VI, on the other hand, presents the outcomes when data augmentation techniques are applied, showcasing the improvements in detection accuracy and robustness achieved through augmentation. Figures 18 and 19 illustrate the loss trajectories and accuracy trends,

Fig. 14. InceptionV3 loss display.

Fig. 15. InceptionV3 accuracy display.

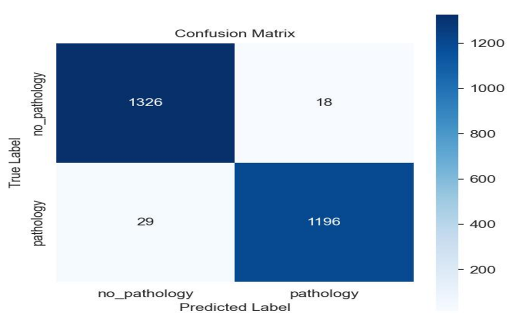

Fig. 16. InceptionV3 confusion matrix.

respectively, for the MobileNetV2 model trained on the baseline dataset (without augmentation). In contrast, the effects of incorporating data augmentation are evident in Figures 20 and 21, which depict the model's loss dynamics and classification performance under enhanced training conditions. Figure 22 shows the confusion matrices illustrating the endometriosis lesion classification outcomes from testing the MobileNetV2 model. The AUC-ROC curve, which can be seen in Figure 23, also explains the result of the MobileNetV2 model.

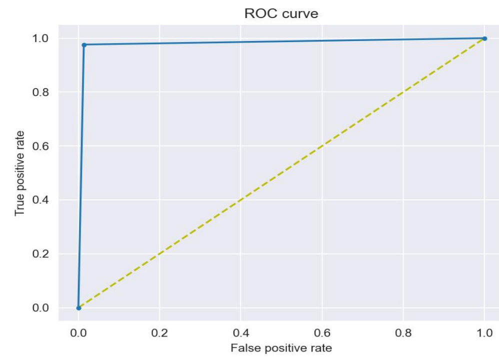

Fig. 17. InceptionV3 AUC-ROC curve.

TABLE V. MOBILENETV2 MODEL RESULTS WITHOUT DATA AUGMENTATION

Epochs	Batch-Size	Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)
3	32	70	70	69	70
	64	70	70	71	71
	128	72	72	71	73
7	32	73	74	74	75
	64	76	77	77	78
	128	79	80	79	80
14	32	83	84	83	84
	64	83	84	84	85
	128	85	86	86	87
20	32	90	90	89	90
	64	92	91	92	92
	128	93	94	94	95

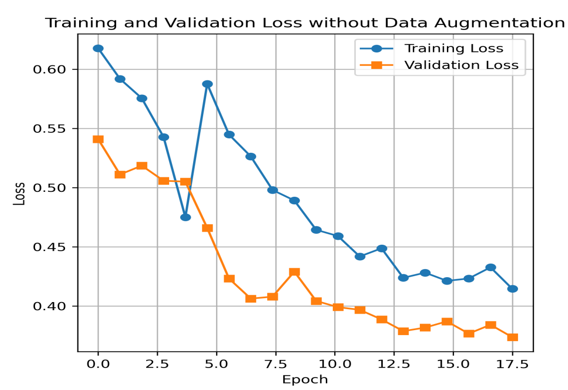

Fig. 18. MobileNetV2 loss display without data augmentation.

V. DISCUSSION

Endometriosis is one of the leading gynecological issues facing women worldwide. Diagnosing and treating this condition remains complex, particularly in settings with limited medical resources. Deep learning techniques applied to medical imaging on large datasets have enabled computer algorithms to achieve diagnostic accuracy similar to that of healthcare professionals. In this research, we propose a deep learning-based solution for classifying endometriosis lesions

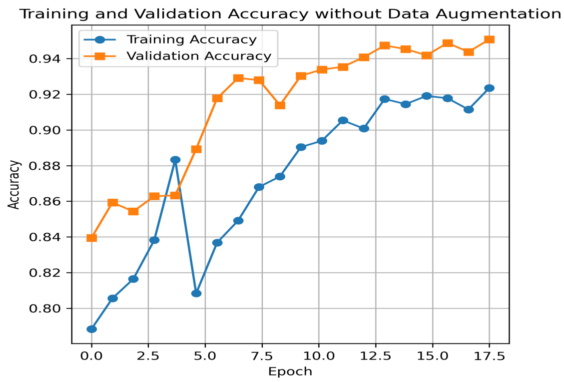

Fig. 19. MobileNetV2 accuracy display without data augmentation.

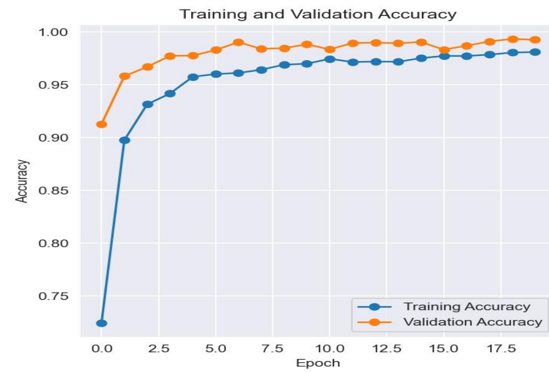

Fig. 21. MobileNetV2 accuracy display.

TABLE VI. MOBILENETV2 MODEL RESULTS WITH DATA AUGMENTATION

Epochs	Batch-Size	Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)
3	32	61	62	61	62
	64	61	61	62	63
	128	65	65	65	65
7	32	67	66	67	68
	64	70	70	71	73
	128	79	80	79	80
14	32	83	84	82	84
	64	86	87	87	87
	128	91	91	91	92
20	32	94	93	94	94
	64	95	96	96	97
	128	99	99	99	99

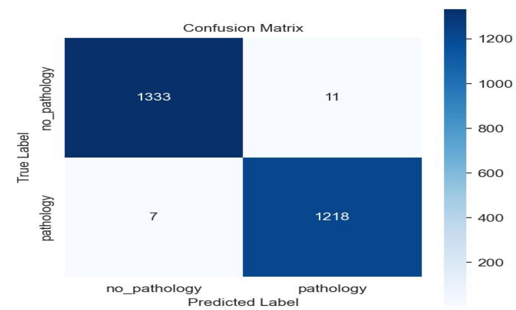

Fig. 22. MobileNetV2 confusion matrix.

Fig. 20. MobileNetV2 loss display.

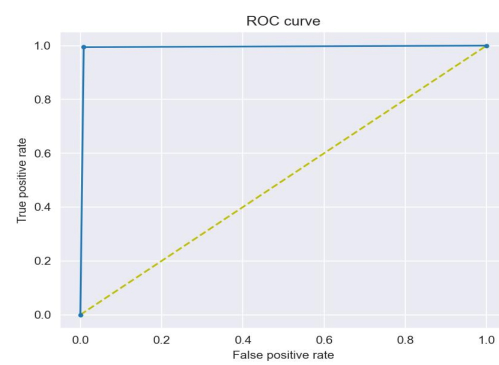

Fig. 23. MobileNetV2 AUC-ROC curve.

TABLE VII. OVERALL RESULTS

#	Deep Learning Models	Accuracy without Data Augmentation (%)	Accuracy with Data Augmentation (%)
1	VGG19	82	88
2	InceptionV3	92	97
3	MobileNetV2	95	99

in laparoscopy images, utilizing the MobileNetV2 architecture alongside a neural network classifier. We incorporate both conventional methods and deep learning-based image augmentation techniques. Experimental findings indicate that deep learning models are well-suited to accurately classify endometriosis lesions, facilitating a robust diagnostic tool for endometriosis.

The incorporation of synthetic images into the training process offers a substantial benefit to the performance of deep learning models, particularly when dealing with small or imbalanced datasets. In many real-world applications, such as

TABLE VIII. COMPARED WITH DIFFERENT TECHNIQUES EMPLOYED BY OTHER RESEARCHERS

#	Citation	Classifier	Accuracy (%)
1	Visalaxi et al. (2021) [36]	ResNet50	91
2	Yun et al. (2021) [43]	VGGNet-16	90.80
3	Takahashi al. (2021) [44]	DNN	90.29
4	Leibetseder al. (2022) [40]	Faster-RCNN	32.4 (precision)
5	Sudalaimuthu al. (2022) [45]	U-Net	74 (F1-score)
6	Figueredo al. (2024) [46]	Ensemble of Networks	96.67
7	Zaidi al. (2025) [5]	Inception V3	93
8	Our proposed	MobileNetV2	99

medical imaging, obtaining a large and representative dataset can be a significant challenge. Synthetic image generation addresses this limitation by augmenting the dataset, providing the model with a broader spectrum of examples, including rare or underrepresented cases. This allows the model to learn more robust and generalized features, which directly contributes to improvements in key performance metrics such as accuracy, precision, and recall.When compared to models trained solely on real images, those trained with synthetic images exhibit superior handling of rare classes. This is particularly important in fields like medical imaging, where certain conditions or abnormalities may be infrequent but critical for diagnosis. By increasing the diversity of the training data, synthetic images help to reduce bias toward more common classes, ensuring that the model does not develop a skewed understanding of the data. As a result, models trained with synthetic images are better equipped to identify and classify rare conditions, leading to a decrease in false negatives and an overall improvement in recall.

Moreover, synthetic images enhance the generalization ability of deep learning models. With more varied and comprehensive data, models are better prepared to perform well on unseen data, a crucial aspect for deployment in real-world scenarios. This greater ability to generalize ensures that the model's performance remains consistent when faced with new, previously unobserved examples. Thus, synthetic images play a vital role in improving not only the accuracy and reliability of deep learning models but also their ability to adapt to diverse and unpredictable data, particularly in specialized fields such as medical imaging.

In our comparative evaluation of convolutional neural network architectures, MobileNetV2 demonstrated superior performance relative to models such as VGG19 and InceptionV3, mainly due to MobileNetV2's lightweight and efficient structure, making it ideal for mobile and embedded deployment. With fewer parameters, it offers faster and simpler training and deployment. Experimenting with 32, 64, and 128 batch sizes, we achieved a favorable trade-off between accuracy and computational efficiency at a resolution of 224x224. Scaling pixel values from 0 to 1 was the most effective for normalization. Additionally, we observed that while fine-tuning the fully connected layer and freezing convolutional layers maintained stable performance, it did extend convergence time.

Table VII presents a comparative analysis of various deep learning algorithms tested on the GLENDA dataset, highlighting the performance of each model in detecting endometriosis. This comparison underscores the strengths and limitations of different algorithms within the same dataset, allowing for an evaluation of each model's effectiveness in classifying endometriosis lesions. Table VIII provides a summary of these findings and compares the performance of our models to existing state-of-the-art approaches, offering insights into how our methods advance current standards in accuracy, precision, and overall reliability for clinical decision support.

VI. CONCLUSION

Our approach, combining the MobileNetV2 model with both conventional and deep learning-based augmentation techniques, demonstrated high performance, achieving 99% in recall, binary accuracy, and F1-score. The integration of synthetic samples generated by DCGAN significantly improved training data diversity and addressed class imbalance issues within the dataset. Specifically, by generating synthetic images of endometriosis using DCGAN, we were able to enhance the MobileNetV2 model's accuracy, likely due to DCGAN's capacity to capture diverse manifestations of endometriosis present in the GLENDA dataset. These findings underscore the potential of deep learning models to accurately classify endometriosis lesions from laparoscopy images, supporting their use as clinical decision support tools for timely diagnosis. Additionally, synthetic data augmentation shows promise for addressing similar challenges in other areas of medical imaging, offering a pathway to more effective and reliable diagnostic support systems. Future work will aim to further improve system performance by expanding the dataset to include a wider variety of laparoscopy videos, enhancing both model accuracy and generalization across clinical scenarios.

AUTHOR'S CONTRIBUTIONS

All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

DATA AVAILABILITY

Data availability is not applicable to this article as no new data were created or analysed in this study.

CONFLICT OF INTEREST

The authors state that they do not have any conflicts of interest.

ACKNOWLEDGMENTS

The first author is a PhD degree student in the Computer Science Program at the Faculty of Science, Chiang Mai University (CMU), under the CMU Presidential Scholarship.

REFERENCES

[1] J. N. Samreen et al., 'MRI of endometriosis: a comprehensive review', Appl Radiol, vol. 48, no. 5, pp. 6–12, 2019.

[2] A. L. Shafrir et al., 'Risk for and consequences of endometriosis: a critical epidemiologic review', Best practice & research Clinical obstetrics & gynaecology, vol. 51, pp. 1–15, 2018.

[3] N. Agarwal and A. Subramanian, 'Endometriosis-morphology, clinical presentations and molecular pathology', Journal of laboratory physicians, vol. 2, no. 01, pp. 001–009, 2010.

[4] H. Kwok et al., 'Lesion distribution characteristics of deep infiltrating endometriosis with ovarian endometrioma: an observational clinical study', BMC women's health, vol. 20, pp. 1–7, 2020.

[5] S. A. Zaidi, V. Chouvatut, C. Phongnarisorn, and D. Praserttitipong, 'Deep learning based detection of endometriosis lesions in laparoscopic images with 5-fold cross-validation', Intelligence-Based Medicine, vol. 11, p. 100230, 2025.

[6] C. M. Becker et al., 'ESHRE guideline: endometriosis', Human reproduction open, vol. 2022, no. 2, p. hoac009, 2022.

[7] Y. Wang, K. Nicholes, and I. M. Shih, 'The origin and pathogenesis of endometriosis', Annual Review of Pathology: Mechanisms of Disease, vol. 15, no. 1, pp. 71–95, 2020.

[8] E. Pascoal et al., 'Strengths and limitations of diagnostic tools for endometriosis and relevance in diagnostic test accuracy research', Ultrasound in Obstetrics & Gynecology, vol. 60, no. 3, pp. 309–327, 2022.

[9] K. E. Nnoaham et al., 'Impact of endometriosis on quality of life and work productivity: a multicenter study across ten countries', Fertility and sterility, vol. 96, no. 2, pp. 366–373, 2011.

[10] C. M. Becker, S. A. Missmer, and K. T. Zondervan, 'Endometriosis: the authors reply', New England Journal of Medicine, vol. 383, no. 2, 2020.

[11] B. Smolarz, K. Szyłło, and H. Romanowicz, 'Endometriosis: epidemiology, classification, pathogenesis, treatment and genetics (review of literature)', International journal of molecular sciences, vol. 22, no. 19, p. 10554, 2021.

[12] A. G. Cope, W. M. Vanburen, and S. P. Sheedy, 'Endometriosis in the postmenopausal female: clinical presentation, imaging features, and management', Abdominal Radiology, vol. 45, no. 6, pp. 1790–1799, 2020.

[13] W. K. R. Figueredo, A. C. Silva, A. C. De Paiva, J. O. B. Diniz, A. Brandão, and M. A.P. Oliveira, 'Automatic segmentation of deep endometriosis in the rectosigmoid using deep learning', Image and Vision Computing, 2024.

[14] S. L. Vlek et al., 'Laparoscopic imaging techniques in endometriosis therapy: a systematic review', Journal of Minimally Invasive Gynecology, vol. 23, no. 6, pp. 886–892, 2016.

[15] M. Al-Taher et al., 'Intraoperative enhanced imaging for detection of endometriosis: A systematic review of the literature', European Journal of Obstetrics & Gynecology and Reproductive Biology, vol. 224, pp. 108–116, 2018.

[16] C. Mosbrucker, A. Somani, and J. Dulemba, 'Visualization of endometriosis: comparative study of 3-dimensional robotic and 2-dimensional laparoscopic endoscopes', Journal of Robotic Surgery, vol. 12, no. 1, pp. 59–66, 2018.

[17] G. Vizzielli et al., 'Real three-dimensional approach vs two-dimensional camera with and without real-time near-infrared imaging with indocyanine green for detection of endometriosis: A case-control study', Acta Obstetricia et Gynecologica Scandinavica, vol. 99, no. 10, pp. 1330–1338, 2020.

[18] J. Quesada et al., 'Endometriosis: a multimodal imaging review', European journal of radiology, vol. 158, p. 110610, 2023.

[19] B. Cherradi, O. Terrada, A. Ouhmida, S. Hamida, A. Raihani, and O. Bouattane, 'Computer-aided diagnosis system for early prediction of atherosclerosis using machine learning and K-fold cross-validation', in 2021 international congress of advanced technology and engineering (ICOTEN), 2021, pp. 1–9.

[20] C. Nifora, L. Chasapi, M. K. Chasapi, and C. Koutsojannis, Deep Learning Improves Accuracy of Laparoscopic Imaging Classification for Endometriosis Diagnosis. 2023.

[21] M. M. Rahman, Y. Ghasemi, E. Suley, Y. Zhou, S. Wang, and J. Rogers, 'Machine learning based computer aided diagnosis of breast cancer utilizing anthropometric and clinical features', Irbm, vol. 42, no. 4, pp. 215–226, 2021.

[22] H. Ding, N. Huang, and X. Cui, 'Leveraging GANs data augmentation for imbalanced medical image classification', Applied Soft Computing, vol. 165, 2024.

[23] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, 'GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification', Neurocomputing, vol. 321, pp. 321–331, 2018.

[24] I. Goodfellow et al., 'Generative adversarial nets', in Advances in neural information processing systems, 2014.

[25] H. Ding, Y. Sun, N. Huang, Z. Shen, and X. Cui, 'TMG-GAN: Generative adversarial networks-based imbalanced learning for network intrusion detection', IEEE Transactions on Information Forensics and Security, vol. 19, pp. 1156–1167, 2023.

[26] J. Gui, Z. Sun, Y. Wen, D. Tao, and J. Ye, 'A review on generative adversarial networks: Algorithms, theory, and applications', IEEE transactions on knowledge and data engineering, vol. 35, no. 4, pp. 3313–3332, 2021.

[27] W. Li et al., 'EID-GAN: Generative adversarial nets for extremely imbalanced data augmentation', IEEE Transactions on Industrial Informatics, vol. 19, no. 3, pp. 3208–3218, 2022.

[28] A. Turab, F. Ullah, and S. A. Zaidi, 'A Federated Learning Framework for Intermittent Clients in Collaborative Smart Healthcare Systems', in AI and Blockchain Applications for Privacy and Security in Smart Medical Systems, IGI Global Scientific Publishing, 2025, pp. 1–30.

[29] S. A. Zaidi and V. Chouvatut, 'Mae Mai Muay Thai Layered Classification Using CNN and LSTM Models', in 2022 26th International Computer Science and Engineering Conference (ICSEC), 2022, pp. 351–356.

[30] A. Turab, R. Shafqat, S. Muhammad, M. Shuaib, M. F. Khan, and M. Kamal, 'Predictive modeling of hepatitis B viral dynamics: a caputo derivative-based approach using artificial neural networks', Scientific Reports, vol. 14, no. 1, p. 21853, 2024.

[31] A. Turab, F. Ullah, and S. A. Zaidi, 'Explainable AI and Computational Modeling for Real-Time COVID-19 Detection Using Medical Imaging', in AI and Blockchain Applications for Privacy and Security in Smart Medical Systems, IGI Global Scientific Publishing, 2025, pp. 161–190.

[32] S. A. Zaidi and V. Chouvatut, 'Mae Mai Muay Thai Style Classification in Movement Appling Long-Term Recurrent Convolution Networks', Journal of Internet Services and Information Security, vol. 13, no. 1, pp. 95–112, 2023.

[33] K. Dong, C. Zhou, Y. Ruan, and Y. Li, 'MobileNetV2 model for image classification', in 2020 2nd International Conference on Information Technology and Computer Application (ITCA), IEEE, 2020, pp. 476–480.

[34] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, 'Mobilenetv2: Inverted residuals and linear bottlenecks', in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 4510–4520.

[35] P. Ballester and R. Araujo, 'On the performance of GoogLeNet and AlexNet applied to sketches', Proceedings of the AAAI conference on artificial intelligence, vol. 30, 2016.

[36] S. Visalaxi and T. S. Muthu, 'Automated prediction of endometriosis using deep learning', International Journal of Nonlinear Analysis and Applications, vol. 12, no. 2, pp. 2403–2416, 2021.

[37] A. Leibetseder, S. Kletz, K. Schoeffmann, S. Keckstein, and J. Keckstein, 'GLENDA: gynecologic laparoscopy endometriosis dataset', in International Conference on Multimedia Modeling, Cham: Springer International Publishing, 2019, pp. 439–450.

[38] K. He, X. Zhang, S. Ren, and J. Sun, 'Deep residual learning for image recognition', in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[39] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, 'Imagenet: A large-scale hierarchical image database', in 2009 IEEE conference on computer vision and pattern recognition, 2009, pp. 248–255.

[40] A. Leibetseder, K. Schoeffmann, J. Keckstein, and S. Keckstein, 'Endometriosis detection and localization in laparoscopic gynecology', Multimedia Tools and Applications, vol. 81, no. 5, pp. 6191–6215, 2022.

[41] S. Ren, K. He, R. Girshick, and J. Sun, 'Faster R-CNN: Towards real-time object detection with region proposal networks', IEEE transactions on pattern analysis and machine intelligence, vol. 39, pp. 1137–1149, 2016.

[42] K. He, G. Gkioxari, P. Dollár, and R. Girshick, 'Mask r-cnn', in Proceedings of the IEEE international conference on computer vision, 2017, pp. 2961–2969.

[43] Y. Zhang et al., 'Deep learning model for classifying endometrial lesions', Journal of Translational Medicine, vol. 19, pp. 1–13, 2021.

[44] Y. Takahashi et al., 'Automated system for diagnosing endometrial cancer by adopting deep-learning technology in hysteroscopy', PLoS One, vol. 16, no. 3, p. e0248526, 2021.

[45] S. Visalaxi and T. Sudalaimuthu, 'Automated segmentation of endometriosis using transfer learning technique', F1000Research, vol. 11, p. 360, 2022.

[46] W. K. R. Figueredo, A. C. Silva, A. C. de Paiva, J. O. B. Diniz, A. Brandão, and M. A. P. Oliveira, 'Automatic segmentation of deep endometriosis in the rectosigmoid using deep learning', Image and Vision Computing, vol. 151, p. 105261, 2024.

[47] A. Radford, L. Metz, and S. Chintala, 'Unsupervised representation learning with deep convolutional generative adversarial networks', arXiv preprint arXiv:1511. 06434, 2015.

[48] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, 'Mobilenetv2: Inverted residuals and linear bottlenecks', in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 4510–4520.

[49] N. Zhao, Y. Song, A. Yang, K. Lv, H. Jiang, and C. Dong, 'Accurate Classification of Tunnel Lining Cracks Using Lightweight ShuffleNetV2-1.0-SE Model with DCGAN-Based Data Augmentation and Transfer Learning', Applied Sciences, vol. 14, no. 10, p. 4142, 2024.