UNIVERSITY OF HULL

APPLICATION OF DEEP LEARNING TO DETECT ENDOMETRIOSIS IN LAPAROSCOPIC IMAGING

FINAL PROJECT

MSc Research Project (771764_B23_T1)

STUDENT ID

202306432

STUDENT FULL NAME

Happiness Chidinma Nwaigwe

SUPERVISOR

Dr. Ruben Valcarce-Dineiro

ACADEMIC YEAR

2022/2023

1.0 ABSTRACT

Endometriosis, a widespread and complex gynaecological ailment impacting roughly 190 million women globally, is identified by the presence of tissue resembling the endometrium outside the uterus. This condition gives rise to persistent and inflammatory symptoms, encompassing infertility, multisite pain, and additional health complications. The absence of distinct clinical indicators and a minimally invasive diagnostic approach contribute to a diagnostic delay spanning 4 to 11 years. This extended period results in the endurance and recurrence of symptoms, even following treatment. Effectively addressing the diverse facets of endometriosis calls for heightened research endeavours and enhanced medical strategies centred on early detection and management. This work aims to detect and localize endometriosis in laparoscopic images using state-of-the-art and region based deep learning models. This study addresses the challenge of detecting endometriosis in laparoscopic images through advanced deep learning models. Employing a variety of models including EfficientDet (D0, D1, D2), Faster R-CNN, SSD ResNet 152, MobileNet, and CenterNet, the research navigated through computational limitations to achieve significant diagnostic accuracy. Notably, EfficientDet D1 emerged as a standout, demonstrating high mean Average Precision (mAP) of 0.36 and mean Average Recall (mAR) of 0.34 across 50% - 95% Intersection over Union (IoU) thresholds. Additionally, CenterNet recorded the highest mAR of 40%. The qualitative analysis revealed the models' effectiveness in visualizing localized endometriosis lesions, underscoring the potential of AI in enhancing gynaecological healthcare. This study sets a new precedent in the application of deep learning for endometriosis detection, paving the way for more accurate and less invasive diagnostic techniques.

Keywords: Endometriosis detection, Laparoscopic Images, Deep Learning Models, EfficicientDet D0 – D2, Faster RCNN, MobileNet, SSD ResNet, CenterNet

2.0 INTRODUCTION

Laparoscopy, also known as keyhole surgery or minimally invasive surgery, is a surgical technique that provides a surgeon access to the abdomen and pelvis without the need for large incisions in the skin (MedlinePlus, 2023). This is achieved using a specialized instrument called a laparoscope, a small tube equipped with a light source and camera. The laparoscope transmits images of the internal structures to a television monitor, allowing the surgeon to perform procedures with greater precision. This versatile technique is employed for both diagnosis and treatment, addressing a range of conditions within the abdomen and pelvis, with common applications in gynaecology, gastroenterology, and urology (NHS, 2018). Laparoscopy is the only conclusive diagnostic test for endometriosis (Spire Healthcare, 2023).

Endometriosis is a painful medical condition characterized by the growth of tissue similar to the inner lining of the uterus outside the uterus itself. This abnormal tissue growth commonly occurs in the ovaries, fallopian tubes, and the pelvic lining. This condition which affects roughly 10% (190 million) of reproductive age women and girls globally, can lead to the formation of cysts on the ovaries, known as endometriomas (World Health Organization, 2023). This condition also causes irritation of nearby tissues and the development of scar tissue (adhesions), potentially resulting in the pelvic organs and tissues adhering to one another. Endometriosis often results in pain and contribute to fertility issues (Mayo Clinic, 2023). Endometriosis has a profound impact on women's lives, causing pain, infertility, reduced quality of life, and disruptions in daily activities and relationships. Diagnosing endometriosis is often delayed from 4 to 11 years due to various challenges, including the reliance on invasive laparoscopy and misdiagnosis (Agarwal et al., 2019). Endometriosis should be seen as a chronic, systemic condition primarily based on symptoms, and implementing a practical diagnostic approach can expedite diagnosis and improve management for affected women globally. A study conducted by Bontempo and Mikesell (2020), reveals a diagnostic delay of 8.6 years as many patients were mistakenly diagnosed with other physical and mental health problems. Specific symptoms and endometriosis locations increased the likelihood of misdiagnosis. These findings highlight the persistent complexity of diagnosing endometriosis and stress the importance of including intelligent machines/models in diagnostic research.

Deep learning (DL), a technique within artificial intelligence, uses convolutional neural networks (CNNs) to process complex data, such as medical images, in a manner akin to the human brain. This approach enhances diagnostic accuracy in healthcare, notably in detecting endometriosis through laparoscopic imaging, assisting clinicians in diagnosis and treatment planning (Sivajohan et al., 2022). DL's application extends beyond endometriosis detection, proving effective in analysing radiographs (Rehman et al., 2023), MRIs (Ebrahimi & Luo, 2021), and CT scans (Soni et al., 2022), thereby underscoring its significant role across various aspects of medical imaging and automated systems (Sarker, 2021). For example, Leibetseder et al. (2022) deployed region-based deep neural networks such as Faster R-CNN and Mask R-CNN, for the detection and localization of endometriosis in gynaecological laparoscopy. Similarly, Visalaxi and Sudalaimuthu (2022) implemented a cutting-edge technique called U-Net to segment the affected areas, while Zhang et al. (2021) used the VGGNet-16 model to classify the stages of endometriosis. Building on this foundation, my study aims to broaden the scope of deep learning applications by integrating a variety of architectures, including CenterNet-Hourglass (Iancu et al., 2023), Faster R-CNN (Xiao et al., 2020), EfficientDets D0 – D2 (AlDahoul et al., 2022), MobileNet (Muwardi et al., 2023), as well as Single Shot Detector (SSD) ResNet152 (Maktab Dar Oghaz et al., 2022), to enhance the detection of endometriosis in laparoscopic images.

3.0 LITERATURE REVIEW

Deep learning has emerged as a valuable tool in detecting endometriosis from laparoscopic images. The work by Leibetseder et al. (2022) is particularly noteworthy for its use of R-CNN algorithms and the creation of the ENdometrial Implants Dataset (ENID), which led to enhanced lesion segmentation precision. The study employed transfer learning and data augmentation techniques to train its models, achieving high precision at lower IoU thresholds. Despite these advancements, the study faced challenges such as reduced precision at higher IoU overlaps and limited improvements from traditional data augmentations, underscoring areas for future research in this field.

A recent study conducted by Visalaxi and Sudalaimuthu (2022) emphasized the effectiveness of deep learning in medical image segmentation, particularly for detecting endometriosis. It notes the successful use of transfer learning with ResNet50 and the U-Net architecture for accurate segmentation, validated by strong performance metrics. Techniques like semantic segmentation with CNNs and models such as CA-CNN and 3D U-Net advanced the detection of ovarian and uterine endometriosis. Despite these successes, the research identifies a significant gap in the availability of accurately annotated images for training, which limits the current deep learning approaches and underscores the need for improved image annotation and segmentation methods.

Zhang et al. (2021) utilized a VGGNet-16 convolutional neural network to analyse hysteroscopic images for endometrial lesion classification. The deep learning model, trained on a dataset that included a variety of endometrial conditions, outperformed the diagnostic accuracy of gynaecologists with an 80.8% accuracy in detailed classification and 90.8% in binary classification. Despite its success, the study's limitations were noted, including the use of a homogeneous image set and the exclusion of rarer lesion types. The findings suggest deep learning can be an effective diagnostic aid in gynaecology, yet the model requires further validation and dataset diversification for clinical application.

Furthermore, Tadepalli and Lakshmi (2021) explored the use of CNN and ensemble machine learning models to diagnose endometriosis, a condition contributing to infertility in a significant portion of women. The research highlights the complexity of diagnosing endometriosis due to its varied symptoms and the potential mental health impacts. Utilizing patient health claims data, the study employed models such as CA-CNN and DFKZ Net, inspired by U-Net architecture, for image segmentation. The performance evaluation showed that ensemble models, which combine multiple machine learning techniques, outperformed single models in the segmentation tasks.

The literature review identifies advancements in using deep learning for identifying endometriosis in medical images, acknowledging both progress and challenges. This project aims to further this work by using advanced deep learning models for a more nuanced analysis, aiming to fill the research voids and improve clinical treatment of endometriosis.

4.0 METHODOLOGY

4.1 Dataset Acquisition

The research utilizes the ITEC Gynecologic Laparoscopy Endometriosis Dataset (GLENDA, v1.5) (Leibetseder et al., 2020), comprising approximately 373 frames and 373 annotated images of endometriosis lesions derived from over 100 gynecologic laparoscopy surgeries. Both frames and images have consistent dimension of 640x360 pixels. GLENDA encompasses diverse classes such as Peritoneum, Ovary, Uterus, and Deep Infiltrating Endometriosis (DIE), each vital for a comprehensive model training and evaluation process focused on the automated detection of endometriosis.

4.2 Dataset Preprocessing

The dataset is rigorously pre-processed, the first step involves loading the COCO annotation file, which contains information about the images, annotations, and categories using Python's Json module. Afterwards, the images and their corresponding annotations are extracted into separate variables. The images are then partitioned into training, validation, and testing sets at ratios of 70%, 20%, and 10%, respectively. This partition results in 256 training, 84 validation, and 38 test images, which are then meticulously annotated and saved in JSON format. The processed data for each split is organized in a structure compatible with COCO format, which includes 'images', 'annotations', and 'categories' for each of the training, validation, and test splits. Subsequent data augmentation, applied via the `imgaug` library, enhances the dataset with transformations like flipping, scaling, translating, and rotating, effectively doubling the training images to 522. These augmentations aim to emulate diverse imaging conditions to bolster the model's generalization capabilities. A label map translates class names into numerical identifiers, and file paths are configured to ensure seamless model training and evaluation. The images and annotations are converted into TensorFlow's TFRecord format to expedite data handling efficiency. TFRecord is a binary file format that ensures faster data loading and is particularly useful when working with large diverse datasets in TensorFlow. The TFRecord files, inclusive of segmentation masks, are prepared for training with the TensorFlow Object Detection API.

It is important to note that this preprocessing involves careful file organization, with separate directories for different types of data, ensuring that the data pipeline is well-structured and executable during model training. As part of the preprocessing step, the annotations were unveiled by loading the coco.json file that contains each annotation image ID, its bounding box coordinates, and class labels, then formatted and saved back to the folder as the ground truth images. A sample of the unveiled ground truth image, the corresponding annotation and frame is shown in Figure 1 below.

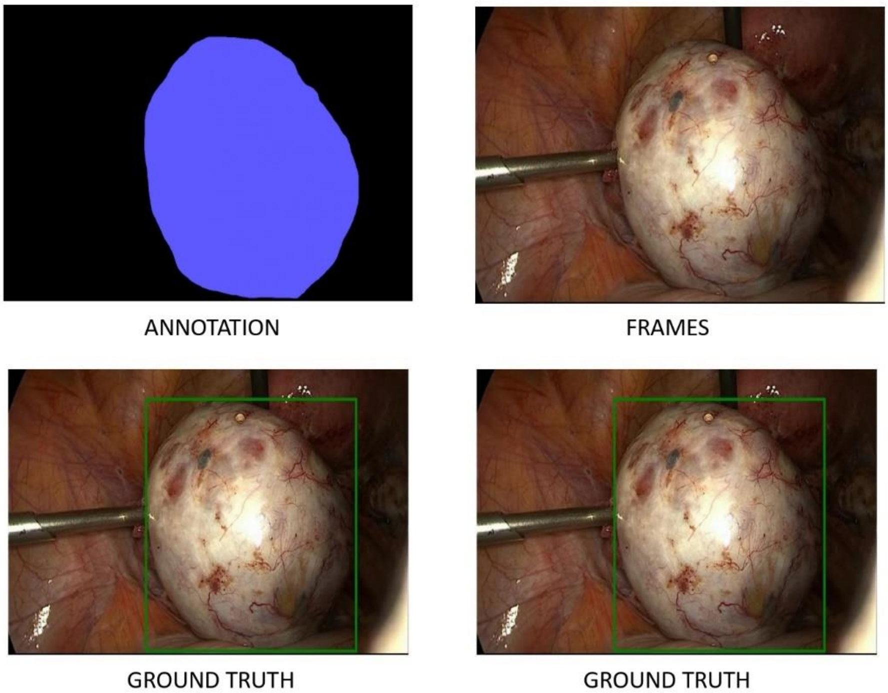


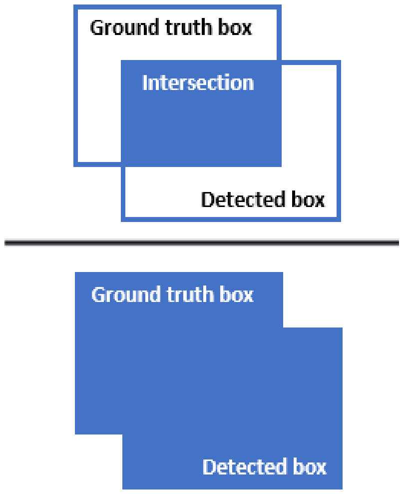
Figure 1: *Sample of an Image in Annotation format, Frame, and the Ground truth (after preprocessing)*

4.3 Model Training and Evaluation

The research evaluates three state-of-the-art object detection models: EfficientDet-D0, EfficientDet-D1, and Faster R-CNN. Each model's configuration comprises a unique model name, base pipeline file, pretrained checkpoint, and a batch size tailored to computational constraints. Training encompasses 80,000 steps for EfficientDet-D0 and 40,000 steps for both EfficientDet-D1 and Faster R-CNN, with intermittent evaluations.

Model performance is quantitatively assessed using mean Average Precision (mAP) and mean Average Recall (mAR) metrics over various IoU thresholds and object sizes. These metrics provide a multifaceted perspective on the models' detection accuracy and the ability to identify true positives. AP is evaluated for object sizes categorized as small, medium, and large, while AR is considered for up to 100 detections, offering an in-depth appraisal of each model's precision and recall capabilities.

$$\textbf{IoU} = \frac{\textbf{Area of Overlap}}{\textbf{Area of Union}} =$$



$$\textbf{Mean Average Precision} = \boldsymbol{mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i}$$

Where, $\boldsymbol{N}$ is the number of classes and $\boldsymbol{AP_i}$ is the average precision of class *i*.

$$\textbf{Mean Average Recall} = \boldsymbol{mAR = \frac{1}{N}\sum_{i=1}^{N} AR_i}$$

Where, $\boldsymbol{N}$ is the number of classes and $\boldsymbol{AR_i}$ is the average recall of class *i*.

5.0 EXPERIMENT AND RESULT

Seven deep learning models were successfully implemented:

Table 1: *The Models Experimental Setup and Result*

Model	Base Backbone	Batch Size	Training Steps	Total loss	IoU	mAP	mAR
EfficientDet-D0	EfficientNet-b0 + BiFPN	16	80000	2.34	0.50:0.95	0.27	0.29
EfficientDet-D1	EfficientNet-b1 + BiFPN	16	40000	2.14	0.50:0.95	0.36	0.34
EfficientDet-D2	EfficientNet-b2 + BiFPN	16	20000	1.77	0.50:0.95	0.34	0.24
Faster R-CNN	Inception Resnet v2	16	100000	0.65	0.50:0.95	0.27	0.27
SSD ResNet 152	Resnet 152 v1 FPN (RetinaNet)	8	20000	1.84	0.50:0.95	0.11	0.26
MobileNet	MobileNet v2 FPN-lite (mobile version of RetinaNet)	64	65000	3.71	0.50:0.95	0.23	0.22
CenterNet	Hourglass-104	8	10000	8.92	0.50:0.95	0.26	0.40

This study utilized a range of advanced deep learning models to detect endometriosis, including EfficientDet variants (D0, D1, D2), Faster R-CNN, SSD ResNet 152, MobileNet, and CenterNet. Each model was chosen for its specific strengths in medical image analysis. NVIDIA A40 GPUs were used for their robust processing power, which was crucial but not efficient in handling the computational demands of these sophisticated models. Training times varied from 18 to 52 hours, depending on each model's complexity and configuration, with batch sizes and steps determined by GPU availability.

In detail, as shown in Table 1 above, EfficientDet models combined SSD frameworks with EfficientNet backbones and BiFPN, optimized with a momentum optimizer and a cosine decay learning rate starting at 0.08. Faster R-CNN utilized Inception ResNet v2, trained with a cosine decay learning rate of 0.008. SSD with ResNet 152 v1 FPN (RetinaNet) and MobileNet v2 FPN-lite were both optimized for specific TPUs and trained on COCO, using momentum optimizers and cosine decay learning rates (0.04 and 0.08, respectively). The CenterNet model, with an Hourglass-104 backbone, was optimized for TPU-32 v3, using an Adam optimizer and a starting rate of 1e-3.

Regularizers like l2 and dropout were employed to prevent overfitting. Optimization techniques and evaluation metrics, including mean Average Precision (mAP) and mean Average Recall (mAR), were tailored to each model's architecture, ensuring comprehensive performance assessment in detecting endometriosis in laparoscopic images.

5.1 Quantitative Result

Employing the established evaluation metrics for object detection (Zenggyu, 2018), a comprehensive analysis was conducted on various deep learning models to assess their detection accuracy for endometriosis localisation. The assessment focused on a range of Mean Average Precision (mAP) and Mean Average Recall (mAR) values calculated across different Intersection over Union (IoU) thresholds. These thresholds are crucial as they determine the necessary overlap between the model's predicted area and the actual ground truth for a prediction to be deemed correct. The mAP values are reported at different IoU thresholds, denoted by subscript numbers mAP_{50-95} indicates the mAP averaged over 10 thresholds ranging from 50% to 95% IoU overlap at 5% intervals, while mAP_{50} and mAP_{75} represent the mAP at specific 50% and 75% IoU thresholds, respectively. This logic applies to mAR as well.

EfficientDet D0: Achieved mAP_{50-95} of 0.150, with a higher precision at the more lenient IoU threshold of 50%(mAP_{50} = 0.266). The Average Recall (AR) at maximum detections (maxDets) of 100 was 0.294, indicating moderate recall capabilities.

EfficientDet D1: Showed an improved mAP_{50-95} of 0.176, and a significant increase in precision at mAP_{50} of 0.356. The AR at 100 maxDets was notably higher at 0.335.

EfficientDet D2: Recorded a marginal decline in precision at mAP_{50-95} of 0.132 yet showed better performance at mAP_{50} of 0.344, with an AR at 100 maxDets of 0.235.

Faster R-CNN: Displayed mAP_{50-95} of 0.146 and an improved precision at 50% threshold of 0.269. The AR at 100 maxDets was 0.267, demonstrating a balanced recall rate.

SSD with ResNet152: This configuration scored lower in precision, with mAP_{50-95} of 0.044, and mAP_{50} of 0.111 as depicted in Figure 2. It also has one of the least effective recall rates, particularly for medium and large-sized objects.

MobileNet: This model achieved a mAP_{50-95} of 0.154, with an enhanced precision of 0.226 at the 50% IoU threshold. It exhibited an AR of 0.215 at 100 maxDets.

CenterNet: Achieved encouraging outcomes with mAP_{50-95} of 0.153 and mAP_{50} of 0.263, suggesting efficient detection across a range of object size. It also demonstrated a strong ability to recall objects at maxDets of 100 at 0.405 AR.

Each model demonstrated distinct capabilities and constraints in detecting endometriosis. EfficientDet D1 and CenterNet were notable for their high average precision and recall rates, particularly effective in identifying medium-sized objects. The MobileNet v2 FPN-lite model displayed a balanced performance across various object sizes, showing effectiveness in detecting large-sized objects. In contrast, the Faster R-CNN model excelled in the detection of medium-sized objects. However, the SSD model, while valuable in the study, identified areas that require further optimization.

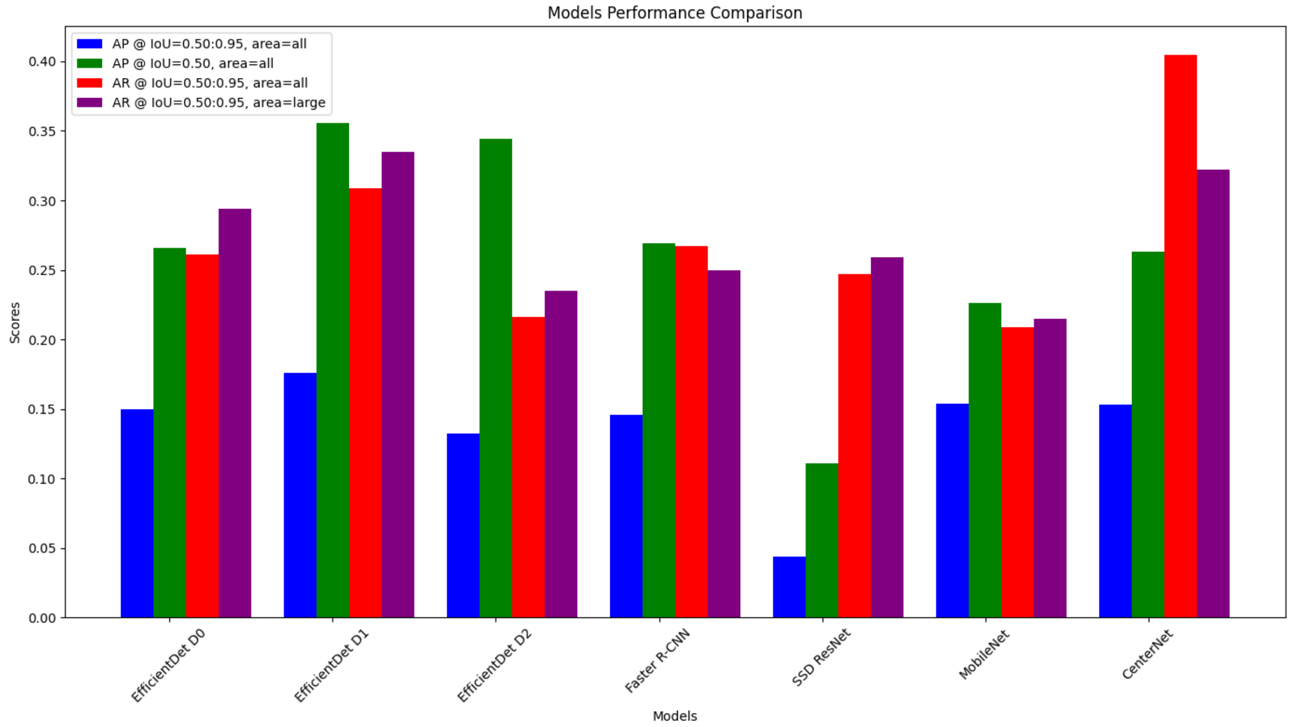


Figure 2: *The Models Comparative Analysis*

5.2 Qualitative Result

This qualitative analysis delves into the visual performance of the different deep learning models employed for detecting endometriosis. Their predictions are compared against ground truth images to assess their accuracy in real-world scenarios. CenterNet effectively identified primary lesions but was less reliable for peripheral or smaller ones. EfficientDet models, particularly D1 and D2, excelled in detecting a broader range of lesions, including smaller and more diffused types. Faster R-CNN and MobileNet demonstrated high accuracy in lesion detection and was adept at distinguishing between lesions and normal tissues. ResNet152, while effective in high-contrast situations, struggled with less defined lesions.

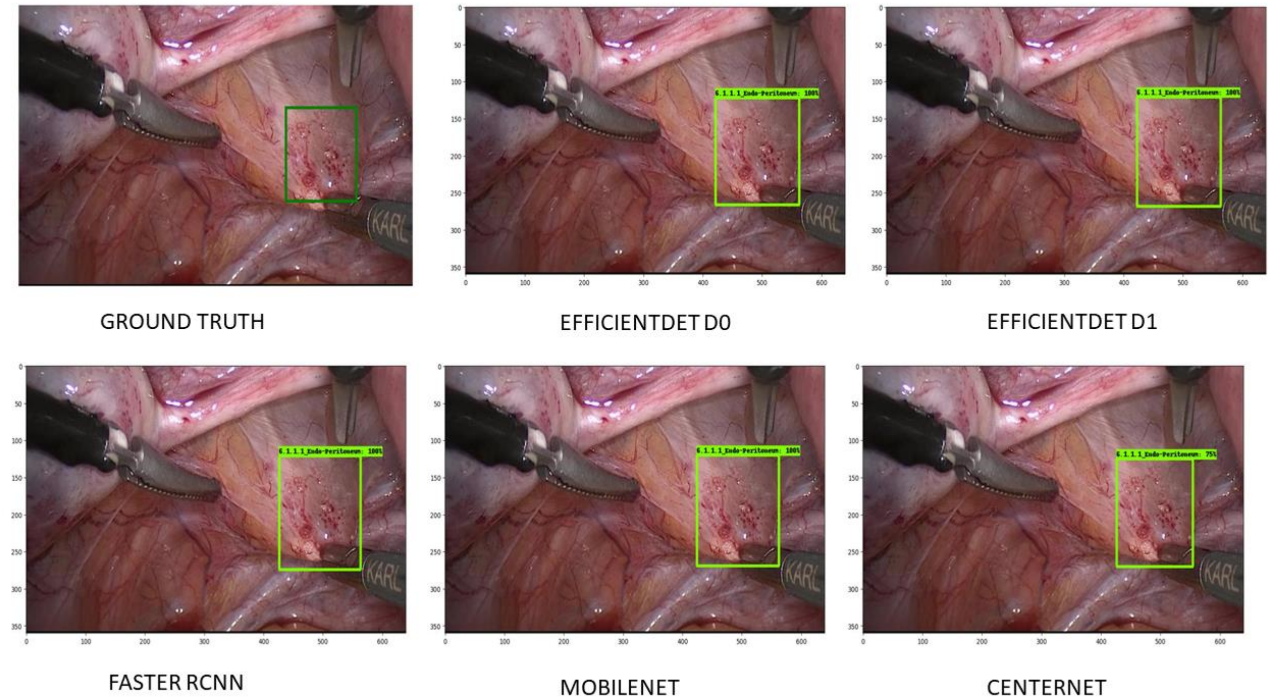


***Figure 3:** The Selected Qualitative Result Comparison for the Models against the Ground truth Image*

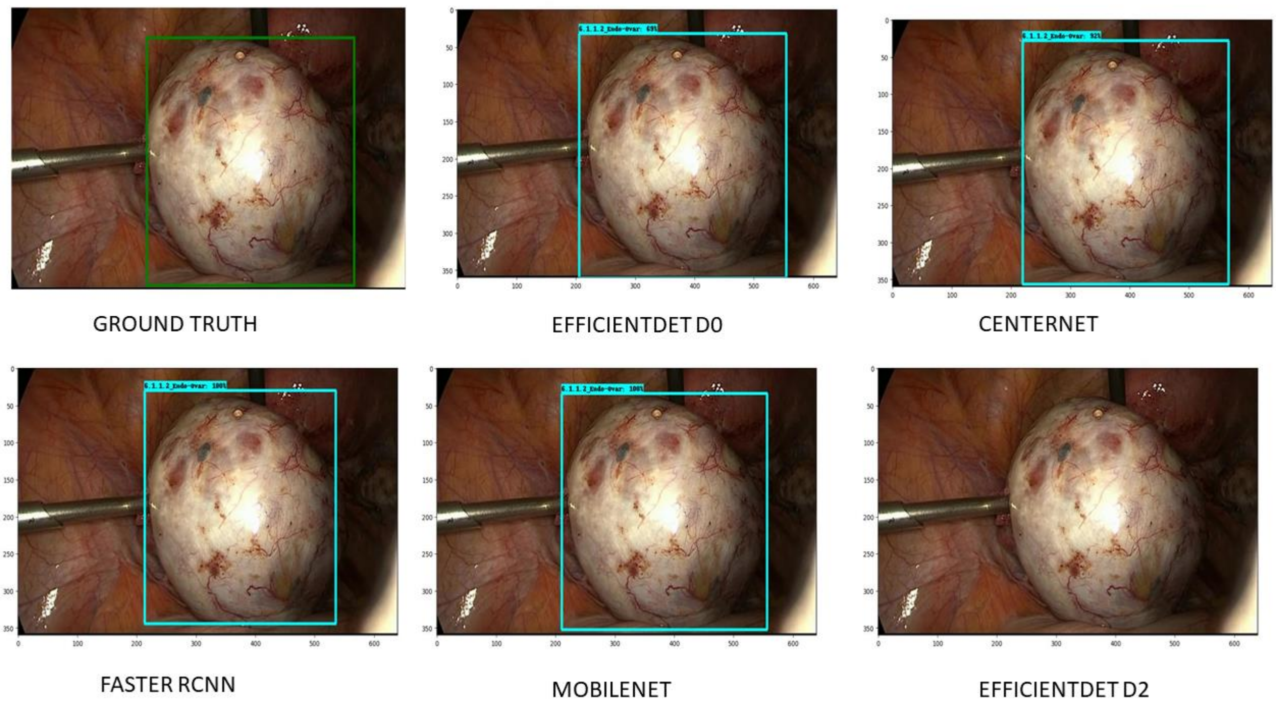


***Figure 4:** The Selected Qualitative Result Comparison for the Models against the Ground truth Image*

In Figures 3 and 4, Faster R-CNN and MobileNet demonstrate exceptional performance with confidence scores consistently at 100%, indicating a strong certainty in their lesion detection capabilities. EfficientDet D0 and D1 also exhibit high confidence and closely match the ground truth annotations. CenterNet, however, shows a lower confidence level of 75% for detecting one of the lesions in Figure 3, implying some uncertainty in its detection capability compared to the other models. Nonetheless, CenterNet shows improvement in Figure 4, with a confidence level of 92%. On the other hand, EfficientDet D2 did not detect any lesions in Figure 4, which suggests a limitation of this model in identifying larger lesions.

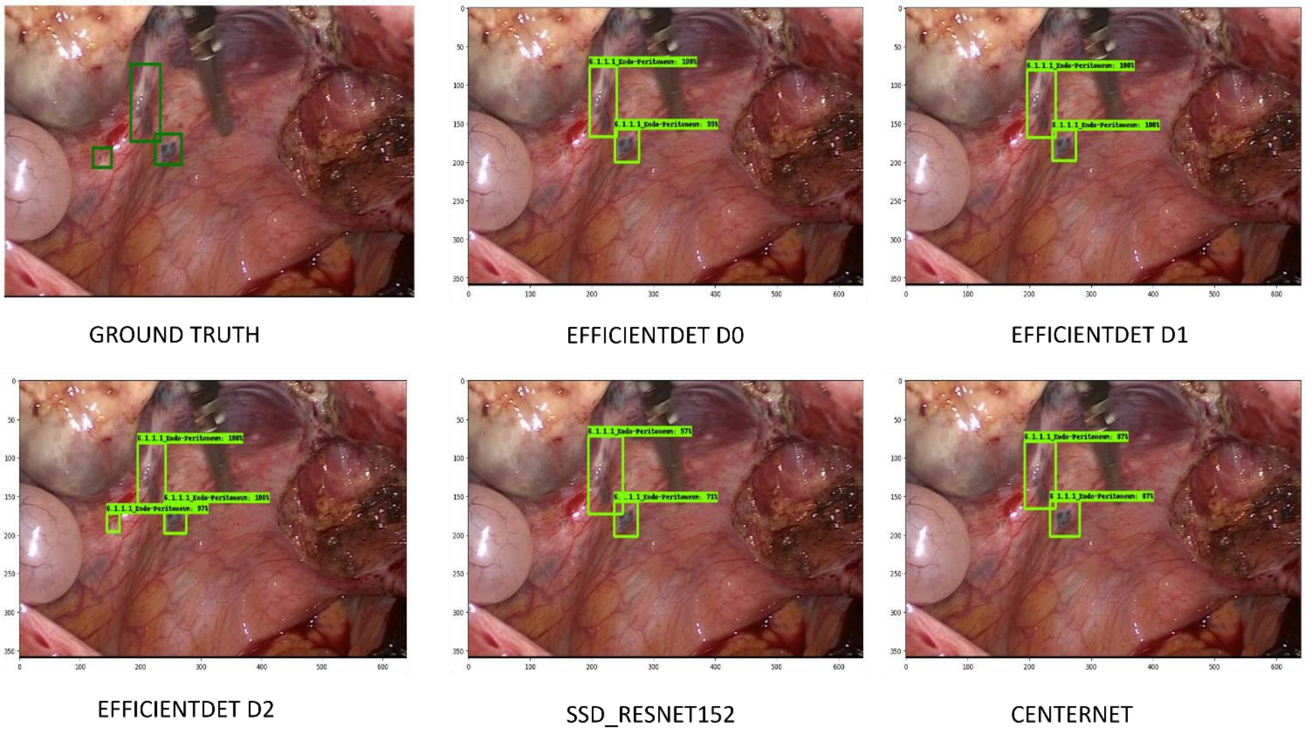


Figure 5: *The Selected Qualitative Result Comparison for the Models against the Ground truth Image*

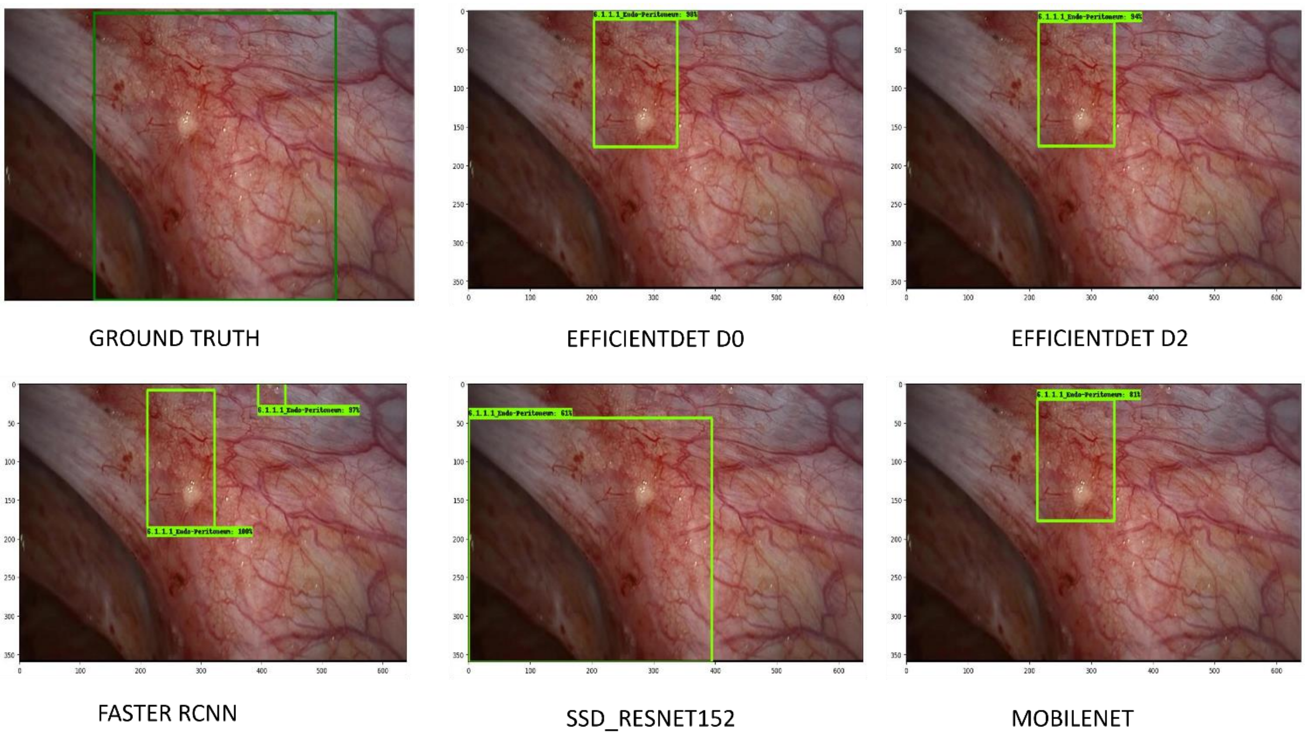


Figure 6: *The Selected Qualitative Result Comparison for the Models against the Ground truth Image*

The analysis of Figure 5 reveals variable confidence levels among EfficientDet D0, D1, and CenterNet, with scores ranging from 57% to 100%. EfficientDet D2 distinctly outperforms other models, with high confidence between 97% to 100%, effectively identifying areas with smaller and more diffused manifestations of the disease. In Figure 6, the models demonstrate confidence scores from 61% to 100%. Despite minor discrepancies in bounding box alignments, which resulted in slight overlaps and underlaps, the models generally identify the regions indicative of endometriosis. The detection of multiple potential sites of endometriosis by the models indicates a broad but not always precise recognition of the disease's extent.

5.3 Uniqueness and Mispredictions

Figure 7 highlights the distinct capabilities of certain models, such as Faster RCNN, EfficientDet D1, and D2, in accurately identifying lesions that were overlooked by other models. This distinction may stem from the unique architectural features or the diversity of their training data, which allows them to detect patterns that elude other models. In contrast, the RCNN models analysed by Leibetseder et al. (2022) showed a tendency to mispredict the few lesions they did detect.

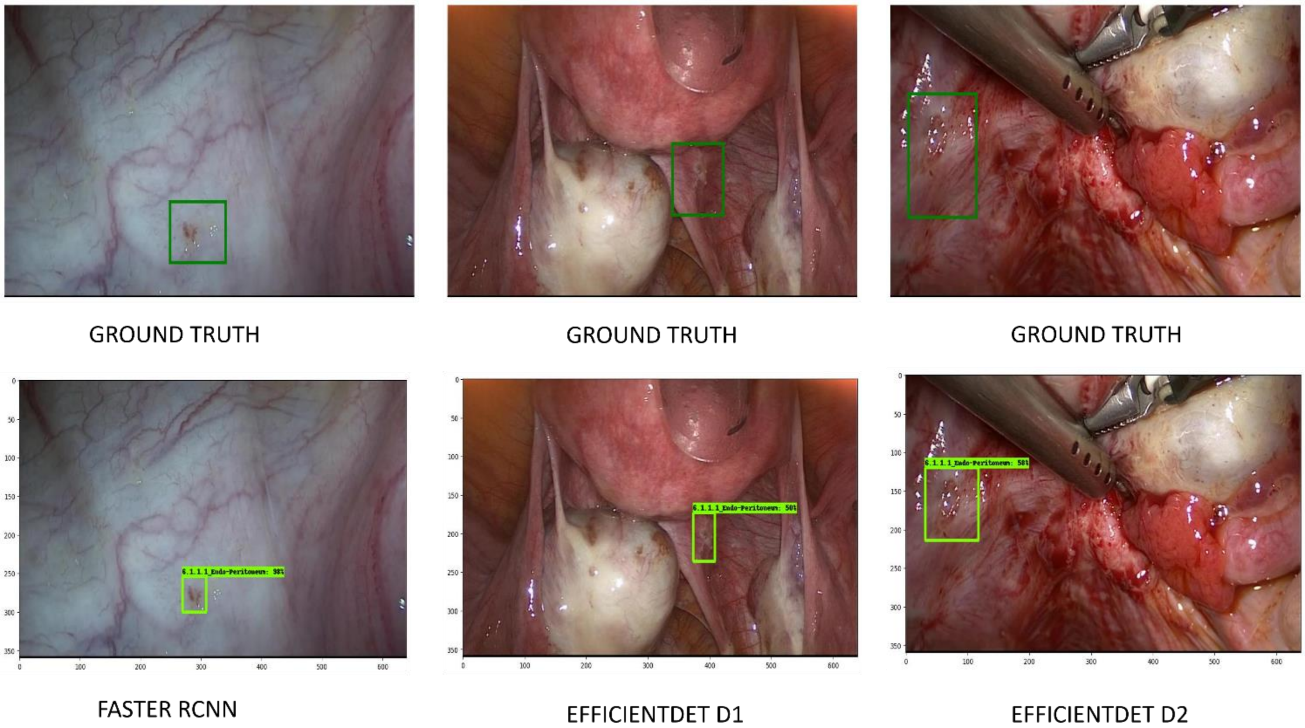


Figure 7: *Some of the Models Unique Detections*

However, mispredictions were not uncommon in this research (Figure 8). For example, some models extended the bounding boxes beyond the lesion area or misaligned them, possibly due to the models misinterpreting the texture and colour variations within the images as indicative of lesions. In other instances, objects, reflections, or shadows may have been mistaken for disease features, leading to false positives.

The mispredictions can be attributed to several factors:

Overfitting to Training Data: Where models have learned specific patterns too well, failing to generalize to new, unseen images.

Model Architecture Limitations: Certain architectures may not be conducive to capturing the full complexity of the medical imagery as they require longer training time.

Annotation Inconsistencies: Variability in the ground truth annotations could lead to confusion during model training.

Resource Constraints: Limited GPU memory and computational power may have forced models to train on down sampled images and ineffective hyperparameters, which affected performance gains by losing detail necessary for accurate lesion identification.

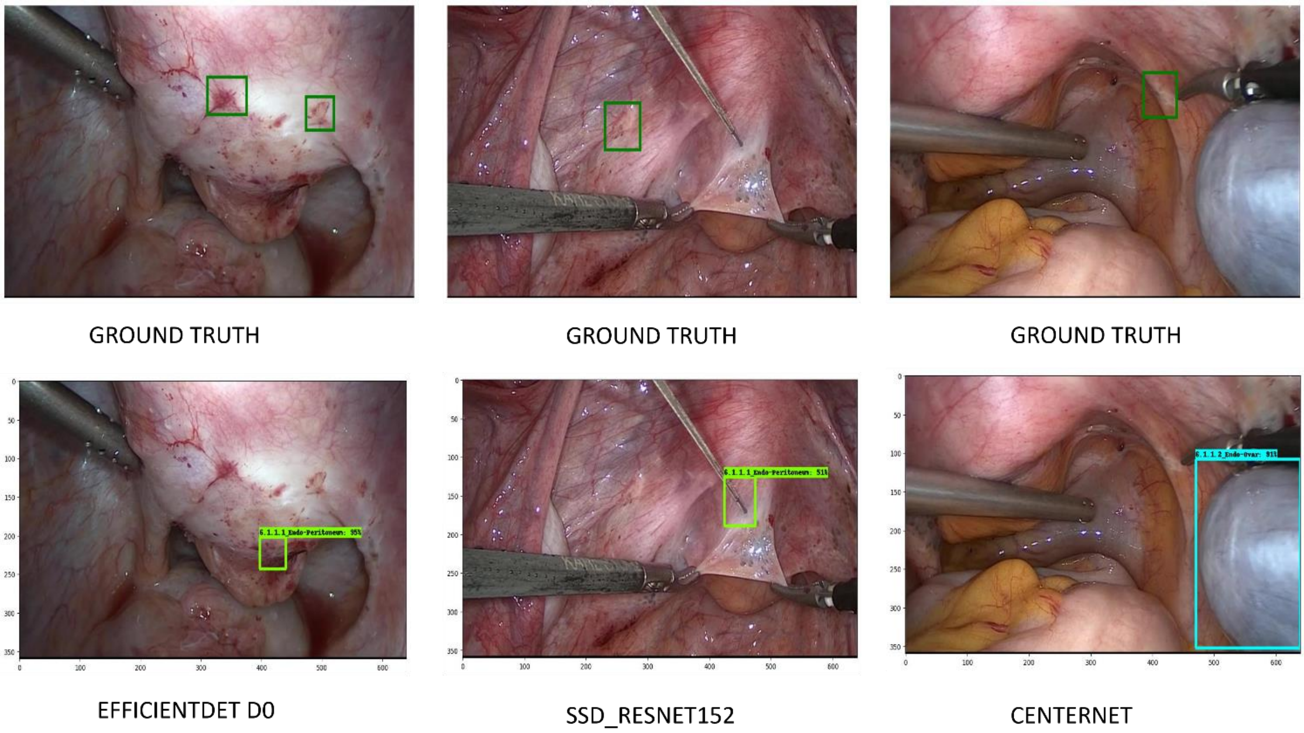


Figure 8: *Some of the Models Mispredictions*

6.0 DISCUSSION

The exploration of deep learning models for detecting endometriosis in laparoscopic images has provided profound insights into their capabilities, limitations, and suitability for clinical application. The models utilized in this study were rigorously trained and tested for enhanced detection and efficiency. Overall, EfficientDet D0 and D1 showcased high precision in detecting lesions, their bounding boxes closely matching the ground truth. This alignment shows their ability to pinpoint the lesions' precise boundaries. However, D0 was only reasonable in images with clear lesion boundaries, as the model occasionally missed smaller lesions or confused them with normal tissue, indicating limitations in detecting fine details and subtle textural differences. D1, on the other hand, exhibited improved detection accuracy over D0, with fewer false negatives and better delineation of lesions. It effectively identified lesions of varying sizes and shapes, demonstrating its robustness in complex laparoscopic scenes. EfficientDet D2, while showing a lower total loss, indicating effective learning as seen in its qualitative results, exhibited a slight decline in mAR, reflecting the complexity of balancing loss reduction with precise lesion detection. Faster R-CNN, despite its extensive training requirements, demonstrated a balanced performance with the lowest total loss (Figure 9) and satisfactory mAP and mAR scores. The model's strength lies in its ability to maintain precision and recall, even under the constraints of heavy computational demands. It was particularly adept at distinguishing lesions in challenging scenarios, such as those with low contrast or surrounded by similar textures. However, the model sometimes generated false positives in areas of high textural complexity. ResNet 152 and MobileNet faced challenges in achieving high mAP and mAR scores. MobileNet's high total loss, despite its high accurate localisation of endometriosis, points to difficulties in dealing with the high-resolution, complex imagery of endometriosis lesions, due to its limited architectural design favouring computational resources over accuracy. CenterNet, despite a high total loss as illustrated in Figure 9, achieved a notable mAR score. It reliably identified a range of lesion sizes and was less prone to false positives compared to some of the other models. However, its ability to delineate the exact boundaries of the lesions varied, with some inaccuracies observed in complex regions.

Incorporating these findings with previous studies resulted to a unique contribution to the field. Prior research by Visalaxi and Sudalaimuthu (2022), Zhang et al. (2021), and Tadepalli and Lakshmi (2021) primarily focused on classification and ensemble machine learning in diagnosing endometriosis. This study diverges by concentrating on nuanced lesion detection and localization, employing metrics such as mAP and mAR over F1 scores for a more comprehensive evaluation of model performance in accurately identifying lesions.

This research addresses the gap left by Zhang et al. (2021), who focused on classification accuracy using VGGNet-16 but did not extensively explore different models and lesion localization. In comparison, this study emphasizes the detailed detection and positioning of endometriosis lesions, advancing beyond mere classification. Similarly, while Tadepalli and Lakshmi's research was innovative in diagnosing endometriosis using CNN and ensemble models based on patient data, this work applies deep learning directly to medical imagery, offering a more direct and visual detection approach.

The study slightly aligns with Leibetseder et al. (2022) in methodological approach (localisation) but goes beyond the scope of their research by incorporating advanced deep learning models and evaluating the performance with not only mAP but mAR, which ensures a comprehensive detection of positive samples, crucial in medical imaging for complete lesion coverage. This project, despite the challenges faced, particularly regarding resource constraints, outperformed prior research, including Leibetseder et al. (2022) which was limited to segmenting only a few lesions. While this study, by contrast, identified a broader range of lesions with unprecedented accuracy and confidence. This capability to detect diverse lesion patterns with high precision significantly enhances diagnostic accuracy, offering new insights into endometriosis's varied manifestations and potentially informing more effective treatment strategies.

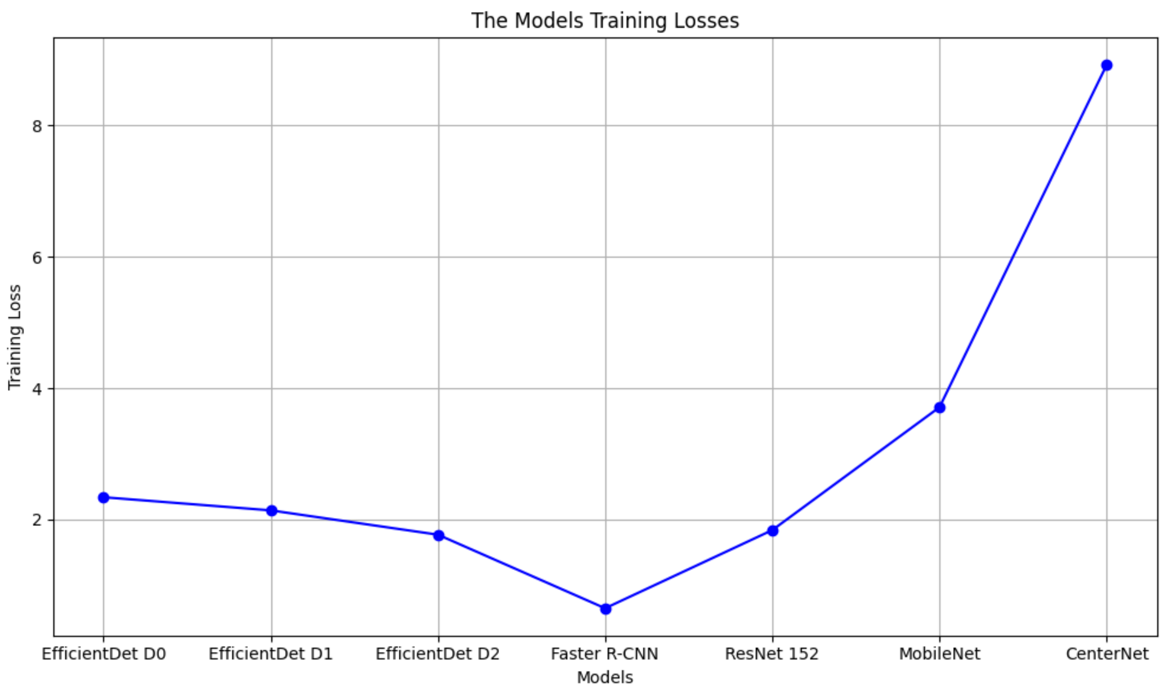


Figure 9: *The Models Training Losses*

7.0 CONCLUSION AND FUTURE WORK

This research presents a significant advancement in endometriosis detection using deep learning techniques in laparoscopic imaging, surpassing previous methodologies like those by Leibetseder et al. (2022). Advanced models, particularly EfficientDet D1 and CenterNet, showed superior diagnostic accuracy, as evidenced by their mean Average Precision (mAP) and mean Average Recall (mAR) scores. Meanwhile, Faster R-CNN and MobileNetV2 excelled in qualitative analysis, effectively visualizing localized lesions. This underscores AI's potential in transforming gynaecological healthcare.

However, the project encountered substantial computational challenges, including GPU allocation and memory issues, due to the complex model architectures and high-resolution image processing demands. These led to frequent out-of-memory errors and execution failures, and efforts to adjust batch sizes and training steps were only partially effective. This situation highlights the urgent need for greater GPU resources in research and academic environments to handle complex medical imaging tasks and enhance diagnostic accuracy.

Future work should aim at refining model architectures, exploring more efficient algorithms, and utilizing advanced computational resources like NVIDIA A100. The A100's capabilities are especially suitable for managing GLENDA image analysis complexities. In conclusion, this study not only sets a new benchmark in deep learning for endometriosis detection but also opens avenues for future research. Addressing computational challenges and maximizing AI's potential will pave the way for more accurate, faster, and less invasive diagnostic methods in gynaecological health, ultimately improving patient outcomes and care.

8.0 REFERENCES

Agarwal, S.K., Chapron, C., Giudice, L.C., Laufer, M.R., Leyland, N., Missmer, S.A., Singh, S.S. & Taylor, H.S. (2019) Clinical diagnosis of endometriosis: A call to action. *American Journal of Obstetrics and Gynecology*, 220(4).

AlDahoul, N., Karim, H. A., De Castro, A. & Tan, M. J. T. (2022) Localization and classification of space objects using EfficientDet detector for space situational awareness. *Scientific Reports*, 12(1), 21896.

Bontempo, A.C. & Mikesell, L. (2020) Patient perceptions of misdiagnosis of endometriosis: Results from an online national survey. *Diagnosis*, 7(2), 97–106.

Ebrahimi, A. & Luo, S. (2021) Convolutional neural networks for Alzheimer's disease detection on MRI images. *J Med Imaging (Bellingham)*, 8(2), 024503.

Iancu, B., Winsten, J., Soloviev, V. & Lilius, J. (2023) A Benchmark for Maritime Object Detection with Centernet on an Improved Dataset, ABOships-PLUS. *Journal of Marine Science and Engineering*, 11(9), 1638.

Leibetseder, A., Schoeffmann, K., Keckstein, J. & Keckstein, S. (2022) Endometriosis detection and localization in laparoscopic gynecology. *Multimedia Tools and Applications*, 81(5), 6191–6215.

Leibetseder, A., Kletz, S., Schoeffmann, K., Keckstein, S. & Keckstein, J. (2020) *ITEC datasets*. Glenda: The ITEC Gynecologic Laparoscopy Endometriosis Dataset (v1.5). Available online: https://ftp.itec.aau.at/datasets/GLENDA/ [Accessed 8/8/2023].

Maktab Dar Oghaz, M., Razaak, M. & Remagnino, P. (2022) Enhanced Single Shot Small Object Detector for Aerial Imagery Using Super-Resolution, Feature Fusion and Deconvolution. Sensors, 22(12), 4339.

Mayo Clinic. (2023) *Endometriosis*. Available online: https://www.mayoclinic.org/diseases-conditions/endometriosis/symptoms-causes/syc-20354656 [Accessed 6/11/2023].

MedlinePlus (2023) *Laparoscopy*. Available online: https://medlineplus.gov/lab-tests/laparoscopy/#:~:text=A%20laparoscopy%20is%20a%20type,in%20your%20bell. [Accessed 02/11/2023].

Muwardi, R., Permana, J., Gao, H. & Yunita, M. (2023) Human Object Detection for Real-Time Camera using Mobilenet-SSD. *Journal of Integrated and Advanced Engineering (JIAE)*, 3, 141-150.

NHS. (2018) *Laparoscopy (keyhole surgery)*. Available online: https://www.nhs.uk/conditions/laparoscopy/ [Accessed 2/11/2023].

Rehman, A., Khan, A., Fatima, G., Naz, S. & Razzak, I. (2023) Review on chest pathologies detection systems using deep learning techniques. *Artif Intell Rev*, 1-47.

Sarker, I.H. (2021) Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions. *SN Computer Science*, 2(6), 420.

Sivajohan, B., Elgendi, M., Menon, C., Allaire, C., Yong, P. & Bedaiwy, M.A. (2022) Clinical use of artificial intelligence in endometriosis: A scoping review. npj Digital Medicine, 5(1), 109.

Soni, M., Singh, A. K., Babu, K. S., Kumar, S., Kumar, A. & Singh, S. (2022) Convolutional neural network based CT scan classification method for COVID-19 test validation. *Smart Health (Amst)*, 25, 100296.

Spire Healthcare. (2023) *How is endometriosis diagnosed?* Available online: https://www.spirehealthcare.com/health-hub/specialties/womens-health/how-is-endometriosis-diagnosed/ [Accessed 6/11/2023].

Tadepalli, K. & Lakshmi, P. V. (2021) 3D Convolution Neural Network-Based Ensemble Model to Detect Endometrium Issues at Early Stages and Enhance Fertility Chances in Women. *Design Engineering*, 32, 1032-1044.

Visalaxi, S. & Sudalaimuthu, T. (2022) Automated segmentation of endometriosis using transfer learning technique. *F1000Research*, 11, 360.

World Health Organization. (2023) *Endometriosis*. Available online: https://www.who.int/news-room/fact-sheets/detail/endometriosis [Accessed 2/11/2023].

Xiao, Y., Wang, X., Zhang, P., Meng, F. & Shao, F. (2020) Object Detection Based on Faster R-CNN Algorithm with Skip Pooling and Fusion of Contextual Information. *Sensors (Basel)*, 20(19).

Zenggyu (2018*) An introduction to evaluation metrics for object detection*. Zenggyu的博客 - An Introduction to Evaluation Metrics for Object Detection. Available online: https://blog.zenggyu.com/posts/en/2018-12-16-an-introduction-to-evaluation-metrics-for-object-detection/index.html [Accessed 21/12/2023].

Zhang, Y., Wang, Z., Zhang, J., Wang, C., Wang, Y., Chen, H., Shan, L., Huo, J., Gu, J. & Ma, X. (2021) Deep learning model for classifying endometrial lesions. *Journal of Translational Medicine*, 19(1), 10.