

**A Non-Invasive Approach for Diagnosing Endometriosis Using Deep Learning**

Capstone Project Phase A 25-1-R-8

supervaisor: miri weiss cohen

daniel berkowitz

tal turjeman

https://github.com/DanielBerkowitzz/DeepLearningEndometriosisDetection

Table of Contents

[ABSTRACT 2](#_Toc201844797)

[1. INTRODUCTION 3](#_Toc201844798)

[2. LITRETURE OVERVIEW 6](#_Toc201844799)

[3. BACKGROUND 9](#_Toc201844800)

[4. RESEARCH PROCESS 13](#_Toc201844801)

[5. RESULTS 17](#_Toc201844802)

[6. DISCUSSION 27](#_Toc201844803)

[7. CONCLUSION 28](#_Toc201844804)

[8. USER GUIDE 29](#_Toc201844805)

[9. REFERENCES 36](#_Toc201844806)

ABSTRACT

This project presents a non invasive, deep learning approach for diagnosing endometriosis using magnetic resonance imaging. A DenseNet121 model, pretrained on ImageNet, was fine tuned to classify MRI slices as either showing signs of endometriosis or not. The model was trained and evaluated across different MRI modalities, including T1, T2, and a combination of both, with data augmentation to overcome the limited dataset size of 130 cases.

The experiments demonstrated that unfreezing the final dense block of the pretrained model significantly improved classification performance for all modalities. The model trained on combined modalities achieved the highest accuracy and generalization, suggesting that integrating multiple MRI types enhances diagnostic robustness.

The results demonstrate that deep learning can offer a promising approach for detecting endometriosis and may reduce the dependence on invasive diagnostic methods such as laparoscopy.

1. INTRODUCTION

Endometriosis is an inflammatory, chronic gynecologic disorder that affects approximately 10% of reproductive-age women worldwide [1].

Endometriosis is characterized by the growth of estrogen-dependent endometrial-like epithelial and stromal cells outside the uterine cavity. Endometriotic lesions can be found throughout the body, but they are more common in the pelvic cavity, where they can affect organs such as the ovaries, fallopian tubes, urinary bladders, intestines, or peritoneum [5].

The growth of these endometriotic lesions, driven by estrogenic hormonal stimulation, induces a chronic inflammatory state in the pelvic region. Unlike endometrial tissue, endometriotic tissue cannot be removed through menstruation at the end of the maturation process [2].

The condition commonly affects the ovary, usually on one side, as the disease progresses, endometrial cysts form on the ovary, resulting in increased bleeding and pressure within the cyst, especially near the ovarian surface. This pressure makes the cyst wall more prone to repeated ruptures, releasing its contents into the pelvic cavity and leading to significant adhesions.

Endometriosis' pathogenesis is complex and multifactorial, involving factors like sex hormones, immunity, inflammation, and genetics, but its exact cause isn't entirely understood.

The prevalent theory to suggest the development of endometriosis is Sampson's theory of retrograde menstruation, which describes the reflux of endometrial cells into the pelvic cavity, where they attach, invade surrounding tissues, and become vascularized, allowing them to implant, grow, and develop. Alternative theories include coelomic metaplasia, vascular and lymphatic transfer, and stem cell theory [3].

A diagram of internal organs

Description automatically generated

*Figure 1. Illustration of the typical localizations of endometriosis. [7]*

An illustration of the typical localizations of endometriosis is shown in [Figure 1.]

1. ovarian endometrioma

2. retro cervical endometriosis

3. deep bowel endometriosis

4. bladder endometriosis

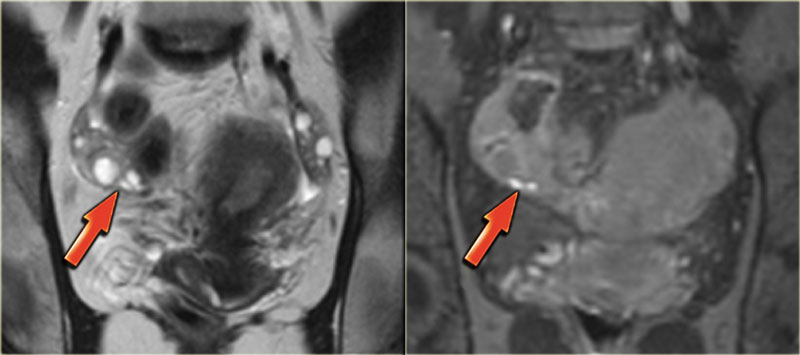
5. abdominal wall endometriosis

There are many manifestations of endometriosis, from incidentally discovered asymptomatic lesions to more severe cases.

It is common for symptoms to appear before the age of 20. Chronic pelvic pain, dysmenorrhea, dyspareunia, dysuria, painful defecation, abdominal bloating, and constipation are some of the primary symptoms of endometriosis. In addition, it has been associated with mental health issues such as anxiety and depression. Infertility is another common manifestation of endometriosis; approximately 40-50 percent of women with infertility are diagnosed with the disease. Endometriosis is primarily classified based on its localization and histopathological features, with three main subtypes: superficial peritoneal endometriosis, ovarian endometriotic cysts, and deep infiltrating endometriosis.

Superficial peritoneal endometriosis, which is located on the surface of pelvic cavity organs and often attaches to the peritoneum, rarely causes severe clinical symptoms.

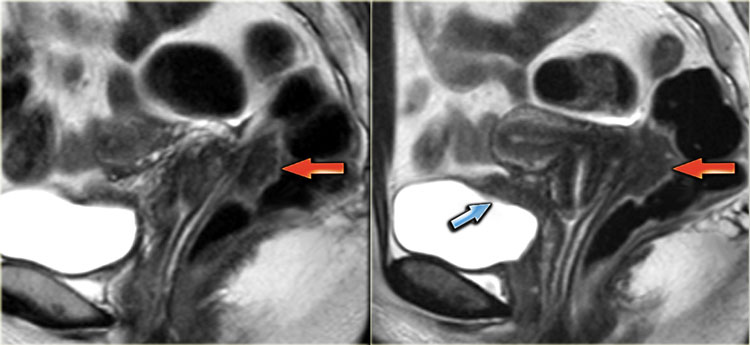
Ovarian endometriotic cysts, also known as endometriomas, develop on the ovaries and form cysts with fluid of varying sizes, this subtype is closely associated with infertility and an increased risk of ovarian cancer. Deep infiltrating endometriosis, the most severe form, invades visceral organs to a depth of 5 mm or more, either within or beyond the pelvic cavity, often distorting local anatomy. This type is known to cause severe symptoms and requires frequent surgical treatment [4].



*Figure 2. Coronal T2 and T1-Fatsat images: superficial serosal implants of endometriosis [7]*

A case of superficial endometriosis is illustrated in [Figure 2.]

Superficial endometriosis, also known as Sampson's syndrome, is characterized by the presence of superficial plaques distributed across the peritoneum, ovaries, and uterine ligaments. This form of the condition is often associated with minor symptoms and relatively fewer structural changes in the pelvic region. When examined through laparoscopy, these plaques typically appear as superficial powder-burn or gunshot lesions [7].



*Figure 3. Sagittal T2-weighted images demonstrating endometriosis infiltrating the rectum and endometriosis infiltrating the bladder [7]*

Figure 3. demonstrates a case of Deep pelvic endometriosis.

Deep pelvic endometriosis is characterized by sub peritoneal infiltration of endometrial deposits. This form of the disease tends to present more severe symptoms, which are related to the location and depth of the invasion.

Magnetic resonance imaging (MRI) contributes significantly to diagnosing deep infiltrating endometriotic lesions and assessing the extent of the disease.

Preoperative mapping of disease spread is essential for determining the necessity of surgical intervention and for planning a comprehensive surgical extraction if it is required [7].

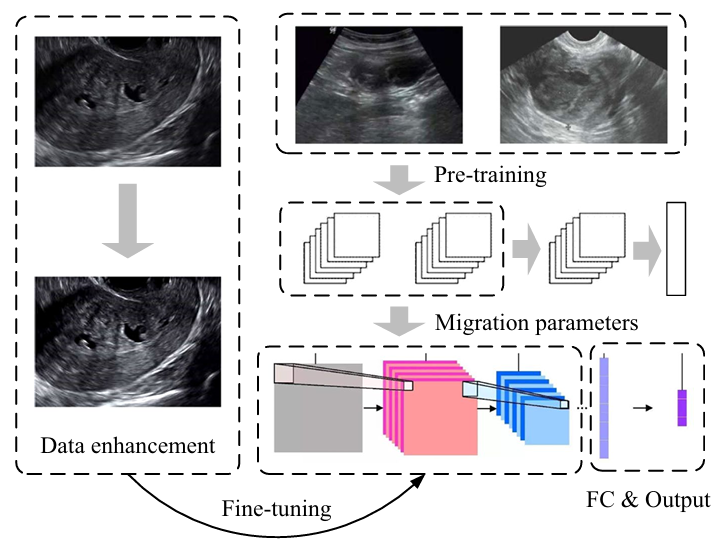
Endometriosis frequently demands extensive medical and surgical interventions, involving considerable costs and risks. Although the condition is widespread and leads to significant morbidity, it is often underdiagnosed and undertreated, with a diagnostic delay of 8 to 12 years from the beginning of symptoms. This delay is primarily due to the nonspecific nature of many symptoms and the lack of effective non-invasive diagnostic methods [2].

In order to diagnose endometriosis, tissue specimens are collected from the abdominal cavity with invasive procedures such as laparoscopy or transabdominal surgery and histologically examined. These methods are associated with risks, invasiveness, and high costs [3].

1. LITRETURE OVERVIEW

Minmin Yang, Min Liu, Yan Chen, Suhui He, Yan Lin [6] proposed a CNN based model of automatic classification of DPE using visual ultrasound (VU) images.

The model was based on the VGG-Net structure, which consists of multiple convolutional layers and fully connected (FC) layers. To improve efficiency and handle overfitting caused by the large number of parameters in the FC layers, they modified the architecture by replacing the final pooling layer with a Global Average Pooling (GAP) layer.



*Figure 4. Framework for automatic recognition of VU images of DPE [6]*

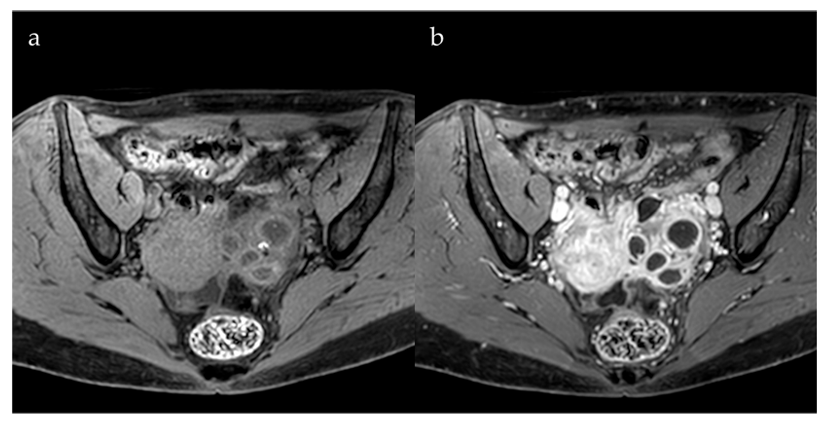
Figure 4. presents the overall framework for automatic recognition of VU images of DPE.

The process includes data enhancement to improve input quality, pre-training of convolutional neural networks on existing datasets, and integration of parameters from pre-trained networks. Then the model is fine-tuned using specific data for the research, followed by FC layers to produce the final output.

The proposed VGG-GAP model was tested for the classification of VU images of DPE using a dataset of 2,328 images collected from 140 patients with DPE confirmed by surgery or postoperative pathology and 206 patients with other gynecological diseases.

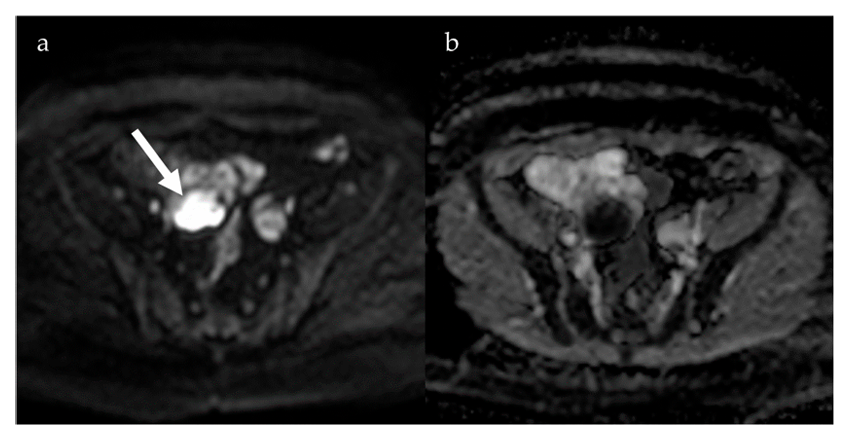
The model was compared with DSIFT, CNN, VGG-16, VGG-19, and AlexNet models and achieved the best results with an accuracy rate of 96.5%.

Laura Alonzo, Roberto Cannella, Giuseppe Gullo, Alessandra Lopez, Giulia Piombo, Giuseppe Cicero, Valentina Billone, Alessandra Andrisani, Gaspare Cucinella, Antonio Lo Casto and Giuseppe Lo Re [9] discuss several advanced MRI techniques used to identify endometriosis.



*Figure 6. Role of Contrast Agents [9]*

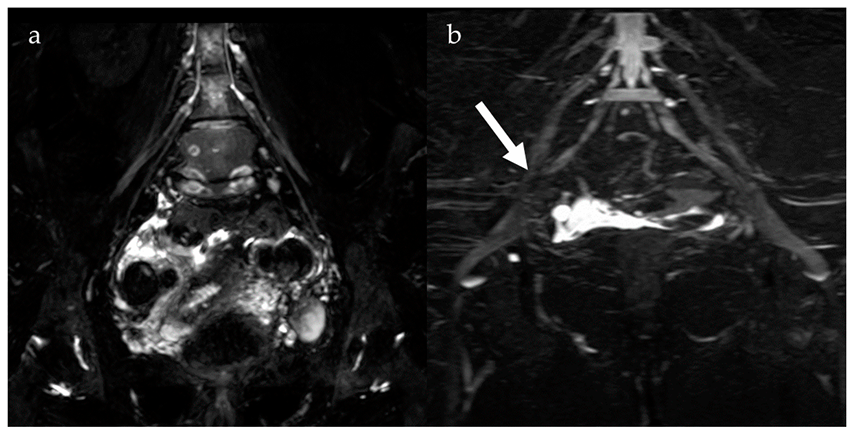
Gadolinium-based contrast agents in MRI scans help distinguish endometrial cysts from other pelvic lesion types. This method is beneficial when intense wall enhancement is observed, improving the ability to identify and evaluate such lesions.



*Figure 7. Diffusion-Weighted Imaging [9]*

Diffusion-weighted imaging (DWI) is an MRI method that measures the movement of water molecules in tissues. Areas with restricted water movement show up as darker regions on Apparent Diffusion Coefficient (ADC) maps. DWI helps in distinguishing between DPE and other gynecological conditions because endometrial cysts usually have a significantly lower ADC than functional ovarian cysts. Studies have also highlighted the effectiveness of DWI in detecting rare endometriotic lesions.

Susceptibility-weighted imaging (SWI) is sensitive to blood products, iron deposits, and microhemorrhages. It is useful for identifying deep endometriotic lesions, especially those that involve repeated bleeding but is less effective for detecting superficial endometriosis. The limitation of SWI is that it can create image distortions due to the presence of intestinal gas, which may reduce its accuracy.



*Figure 8. Magnetic Resonance Neurography and Diffusion Tensor Imaging [9]*

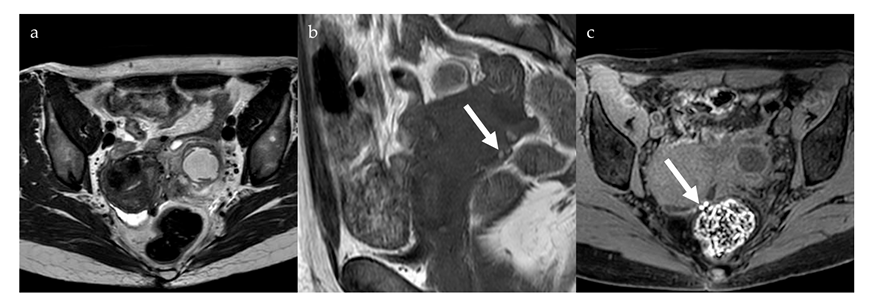
Magnetic resonance neurography (MRN) and diffusion tensor imaging (DTI) are useful for evaluating nerve involvement caused by endometriosis. MRN detects nerve abnormalities, such as thickening or misalignment. DTI measures water movement along nerve fibers, identifying disruptions caused by nerve compression or damage. Tractography, a visualization technique, links these abnormalities to clinical symptoms like pelvic pain and sciatica.

1. BACKGROUND

**3.1 MRI**

MRI is a non-invasive imaging technology that generates high-resolution, three-dimensional anatomical images without the use of damaging radiation. It is widely utilized for disease detection, diagnostic evaluation, and treatment monitoring. The technology works by stimulating and detecting changes in the direction of the rotational axis of protons found in the water molecules that form living tissues.

MRI technology uses powerful magnets to generate a strong magnetic field, causing protons in the body to align with that field. When a radiofrequency pulse is applied, the protons are stimulated and pushed out of equilibrium, resisting the magnetic field's pull. Once the radiofrequency pulse stops, the protons release energy as they return to alignment with the magnetic field, which is detected by MRI sensors. The time it takes for the protons to realign and the energy released can change depending on the surrounding environment and the chemical composition of the tissue. These differences allow physicians to distinguish between various types of tissues based on their magnetic properties.

To produce an MRI image, the patient is placed inside a large magnet and must remain completely still to prevent image blurring. In some cases, contrast agents containing gadolinium are administered intravenously before or during the procedure. These agents speed up the realignment of protons, resulting in brighter and clearer images [10].

*Figure 9. Hypointense nodules in the pouch of Douglas on anaxial T2-weighted image (a) with some small high-signal foci on a coronal T1-weighted image (b, arrow), more evident on anaxial fat saturated T1-weighted image (c, arrow), compatible with deep pelvic endometriosis with evidence of bleeding. [9]*

On MRI, DPE presents with non-specific signal patterns, including hypointense nodular lesions or soft tissue thickening with irregular, indistinct, or stellate margins on both T1- and T2-weighted images. In some cases, hyperintense lesions may appear on T1-weighted images, particularly on fat-saturated sequences, suggesting the presence of hemorrhagic foci. MRI demonstrates strong diagnostic performance for detecting DPE, with a sensitivity of 90% and specificity of 91% [9].

**3.2 DENSE NET**

The Dense Convolutional Network (DenseNet) is a convolutional network architecture where each layer is directly connected to every other layer in a feed forward fashion. Unlike traditional networks with layers, which have connections, one between each layer and its subsequent layer, DenseNet has direct connections. In this structure, each layer receives the feature maps of all preceding layers as inputs and passes its own feature maps to all subsequent layers. Instead of summing the feature maps like in other architectures such as ResNets, DenseNet concatenates the feature maps, preserving all the information from earlier layers. As a result, the network becomes very dense, with many connections.

DenseNet is designed to address key challenges in deep networks. It reduces the vanishing gradient problem, strengthens feature propagation, and encourages feature reuse, significantly reducing the number of parameters making the network more efficient.

DenseNet explicitly distinguishes between new information added to the network and information that is preserved. The layers in DenseNet are narrow, typically using a small number of filters, which allows them to add a limited set of new feature maps while keeping the existing ones. The final classifier then makes its decision based on all the feature maps present in the network.

A key advantage of DenseNet is its ability to improve the flow of information and gradients across the network. Each layer has direct access to the gradients from the loss function and the original input signal, enabling implicit deep supervision that simplifies the training process and by that supports the training of very deep networks. Additionally, the dense connections in DenseNet provide a regularizing effect, which helps reduce overfitting on tasks with smaller training set sizes.

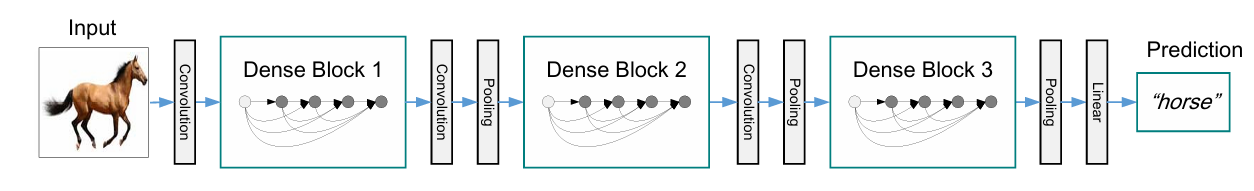


*Figure 10. A 5-layer dense block with a growth rate of k = 4. Each layer takes all preceding feature-maps as input [11]*

In DenseNet, the composite function includes three sequential operations: batch normalization (BN), followed by a rectified linear unit (ReLU), and finally a 3x3 convolution.

Direct concatenation of feature maps is not possible when they are in different sizes, so to handle changes in the size of feature maps, the network is divided into dense blocks and uses down-sampling layers. Between these dense blocks, there are transition layers.

The transition layers perform down-sampling with a combination of operations: first, a batch normalization layer, then a 1x1 convolution, and finally a 2x2 average pooling layer.

If some function of a layer ​ generates k feature maps, then the total number of input feature maps for the layer is , where is the number of channels in the input layer. The growth rate is represented by the hyper parameter k

*Figure 11. A deep DenseNet with three dense blocks. The layers between two adjacent blocks are referred to as transition layers and change feature-map sizes via convolution and pooling [11]*

Each layer in DenseNet can access all feature maps from previous layers within its block, forming what can be described as the network's “collective knowledge.” This cumulative set of feature maps acts as the global state of the network. Each layer adds its own k feature maps to this state. The growth rate regulates how much new information each layer adds to the global state. The global state can be accessed from everywhere in the network and unlike in traditional network architectures, there is no need to replicate it from layer to layer.

Although each layer in DenseNet generates k output feature maps, the number of inputs is often much larger. To address this, a 1×1 convolution is introduced as a bottleneck layer before each 3×3 convolution, and by this it improves the computational efficiency [11].

DenseNet is versatile and can be applied to different computer vision tasks. In image classification, its ability to extract rich feature representations makes it highly effective for identifying objects in images. For object detection, DenseNet serves as a strong backbone network, providing detailed feature maps that contribute to accurate detection of objects. Additionally, in semantic segmentation, its dense connections enable the model to capture fine details, making it effective for semantic segmentation tasks.

DenseNet-121 consists of 121 layers and provides a good balance between computational efficiency and accuracy, making it suitable for tasks with moderate resource requirements.

DenseNet-169 offers 169 layers, enabling deeper feature extraction, which is particularly useful for more complex datasets that require higher accuracy.

DenseNet-201 and DenseNet-264, These deeper architectures are designed for highly complex tasks that require extensive feature representation [12].

**3.4 Transfer learning**

Transfer Learning (TL) is a machine learning technique that addresses challenges like limited or costly data by reusing knowledge from previous tasks to solve new ones. It enhances learning by transferring insights from a source task to a target task. However, the success of TL depends on how well the source and target tasks align. If the knowledge transferred is not compatible, it can result in negative transfer, which harms performance. In contrast, a positive transfer leads to better results for the target task. TL eliminates the need to build models from scratch for every task, making model development faster and more efficient [13].

A diagram of a method

Description automatically generated

*Figure 13. Traditional ML vs. TL [13] [13]*

Transfer Learning works by using a pre-trained model. It starts with a model that has already been trained on a large dataset, allowing it to learn general features and patterns that can be useful for other tasks. This pre-trained model, called the base model, contains layers that capture both simple and complex features. The next step is to identify which layers of the base model hold information that can be reused for the new task. These layers usually contain broad, generic features that work well across similar tasks. After the reusable layers are selected, they are fine-tuned with data from the new task. Fine-tuning adjusts the parameters of the model so it can meet the specific needs of the new task while the useful knowledge from the original training remains. This process makes the model more accurate and adaptable to the new problem.

In Transfer Learning, models are adapted for new tasks using frozen layers and modifiable layers. Frozen Layers are layers from the pre-trained model that remain unchanged during fine-tuning. These layers retain the general features and patterns learned from the original task, which are often universal and applicable across related tasks. Modifiable Layers are the layers that are adjusted during fine-tuning. These layers learn task-specific features from the new dataset, allowing the model to adapt and meet the unique requirements of the new task [12].

TL strategies are categorized into three types based on the conditions between the source domain, target domain, and tasks. Inductive TL, in this strategy the target task is different from the source task, but both share the same domain. Traditional learning usually focuses on the target task or domain alone. However, in multi-task learning or multi-task settings, which are subsets of Inductive TL, the goal is to perform well on all available tasks. Transudative TL, here the tasks in both the source and target domains are identical, but the domains themselves are different. In this case, the target domain does not have labeled data, whereas the source domain contains a large amount of labeled data. Unsupervised TL is similar to Inductive TL in that the source and target tasks are different but related. However, this strategy focuses on unsupervised tasks such as clustering and dimensionality reduction. In this scenario, neither the source domain nor the target domain has labeled data [13].

## RESEARCH PROCESS

**4.1 MRI modalities**

Part of the research process included learning about the different MRI modalities.

T1W1 – T1-Weighted Imaging Produces high contrast between fat and water, making fat appear bright and water dark. It is useful for anatomical details and detecting hemorrhagic or fatty structures.

T2W1 – T2-Weighted Imaging Highlights fluid-filled structures, with water appearing bright and fat relatively darker. It is commonly used for identifying cysts.

DWI – Diffusion-Weighted Imaging Measures the movement of water molecules in tissues, with restricted diffusion often indicating pathology like tumors, stroke, or infections.

ADC – Apparent Diffusion Coefficient Mapping works with DWI to measure water movement in tissues, helping to tell the difference between harmless and harmful lesions.

SWI – Susceptibility-Weighted Imaging Enhances visualization of blood products, iron deposits, and microhemorrhages by using magnetic field variations.

CE-MRI – Contrast-Enhanced MRI Uses gadolinium-based contrast agents to highlight vascularized tissues, improving the detection of lesions.

Each of these MRI modalities contributes to different aspects of endometriosis detection, allowing for a more detailed analysis of lesion type, location, and severity.

**4.2 Hyper parameters optimization**

The researched hyperparameters in our model are:

* Learning rate –
* Batch size –
* Epochs –
* Dropout –
* Batch normalization – with / without

**4.3 Evaluation**

Evaluation metrics are used to measure the performance of a machine learning model by assessing how well it makes predictions.

Evaluation metrics, including accuracy, precision, and recall, will be adopted to compare the performances of our model and compare the performances of different methods.

TP = True positive, the model predicted that there is **Endometriosis**, and the MRI scan was correctly classified.

TN = True Negative, the model prediction was that there was no **Endometriosis**, and the MRI scan was correctly classified.

FP = False Positive, the model predicted there was **Endometriosis**, and the MRI scan was misclassified.

FN = False Negative, the model prediction was that there was no **Endometriosis**, and the MRI scan was misclassified.

Accuracy is the relative share of the positive answer in all our parameters. The result will be between zero and one, with one being the best level of accuracy and zero being the worst.

Accuracy = (TP + TN ) / (TP + TN + FP + FN)

Precision is the number of MRI scans correctly classified as **Endometriosis** out of all those classified as **Endometriosis**.

Precision = TP / (TP + FP)

Recall = Out of all those with **Endometriosis**, the number of MRI scans correctly classified as **Endometriosis**.

Recall = TP / (TP + FN)

**4.4 Research challenges and solutions**

**Challenges:**

**Variability in lesion appearance**  
Endometriotic lesions can look very different from one case to another. They vary in size, shape, intensity, and location within the pelvis. This wide variation can make it more difficult for the model to learn consistent features that apply across all variations.

**Mimicking conditions**  
Endometriosis often appears similar on MRI to other pelvic conditions like ovarian cysts, fibroids, or scar tissue. These overlaps can lead to confusion in diagnosis, not just for deep learning models, but also for experienced radiologists, and make accurate classification less achievable.

**Superficial lesions**  
Superficial peritoneal endometriosis is especially challenging to detect, as its imaging features are subtle and often blend with surrounding tissue. These lesions are often small, faint, and difficult to distinguish from normal tissue, making them harder for both radiologists and deep learning models to identify.

**Lack of labeled data**  
There is a lack of large, labeled MRI datasets specific to endometriosis. The datasets available are small, unbalanced, and vary widely in quality. This limits the ability of deep learning models to generalize and increases the risk of overfitting to specific patterns in the training data.

**Solutions:**

**Data augmentation**  
To address the limited number of labeled MRI cases, data augmentation was applied extensively. For each subject, multiple axial slices around the central pelvic region were extracted to ensure focus on the relevant anatomical area. Each slice was rotated by 90°, 180°, and 270°, increasing the diversity of appearances per case. This process expanded the dataset from 130 MRI scans to over 30,000 augmented images. The increased variety helped the model learn different specific details about each case and made it generalize better.

**Modality combination**  
To enhance the exposure of the model to diverse anatomical representations of endometriosis, several training experiments were conducted using different MRI modality configurations. These included separate runs on T1, T1FS weighted images, T2, T2FS weighted images, and both modalities combined. The combination set was constructed by merging T1, T1FS, T2, and T2FS slices into a unified dataset, with class balance maintained during preprocessing. Each configuration was trained independently using the same preprocessing, augmentation, and with different model architectures. By training on each modality individually as well as in combination, the experiments aimed to investigate how variations in MRI contrast and tissue characterization influence the learning process of the model.

**Transfer learning with fine tuning**

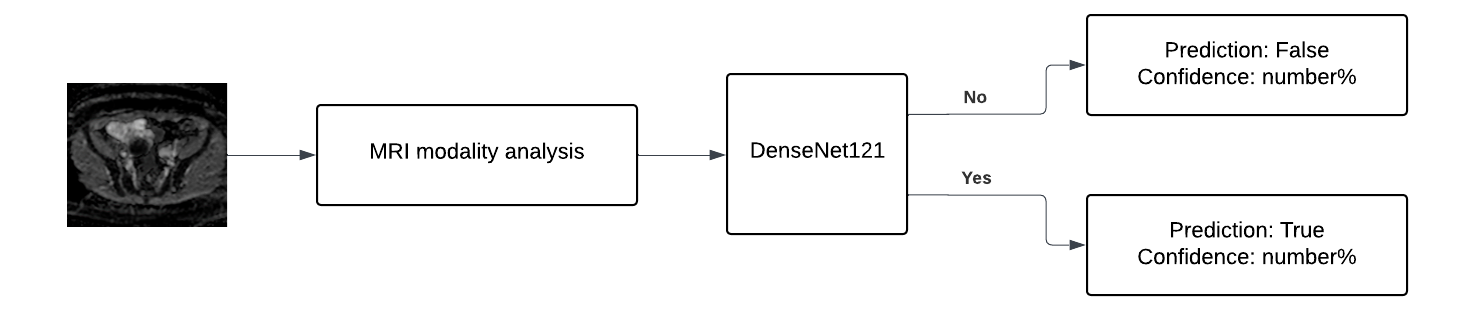
DenseNet121, pretrained on ImageNet, was used as the base model. Initially, all convolutional layers were frozen, meaning their weights remained unchanged, and only the final classifier layer was trained on the MRI data. However, since medical images differ substantially from regular images, keeping the pretrained feature extractor entirely frozen limited the ability of the model to adapt to the specific characteristics of endometriosis.

To overcome this limitation, the fourth dense block, which is part of the deeper layers of the DenseNet121 architecture, was unfrozen. This allowed the model to update the weights of these layers during training and learn more relevant features for identifying endometriotic lesions in MRI scans. By fine tuning only the deeper layers while keeping the earlier layers fixed, the model kept its general feature extraction ability while learning to adapt to the unique patterns in the medical images. The fine tuning significantly improved the performance for all modality combinations, suggesting that the specific features were better captured with this approach.

**Loss smoothing and regularization**

To improve generalization and reduce overfitting, label smoothing and dropout were implemented. Label smoothing is a method that slightly softens the ground truth labels during training. Instead of assigning a hard label of 1.0 to the correct class and 0.0 to the incorrect class, a small value was distributed such that the correct class was set to 0.9 and the incorrect class to 0.1. This reduces the confidence of the model in its predictions, encouraging it to remain flexible rather than becoming overly certain. Dropout is another regularization method that randomly drops out a group of neurons during each training step. A dropout rate of 0.5 means that half of the neurons in the classifier layer are turned off randomly. This forces the network to learn redundant representations and prevents any single neuron from becoming too important. In this project, dropout was applied in the final classifier layer of DenseNet121. These techniques were particularly effective in improving performance when dealing with subtle lesion boundaries and overlapping class features.

**4.5 Flow chart**



## RESULTS

In the training phase, we examined how various hyperparameters and input configurations affected model performance in classifying MRI scans for endometriosis detection. We focused on evaluating different learning rates, epoch counts, and different MRI modalities such as T1, T2 separated and combined to determine the optimal setup. The batch size was fixed at 32 or 64 depending on the configuration, balancing computational efficiency and model stability. Learning rates of e⁻⁴ and e⁻⁵ were tested, along with epoch counts of 50, 75, and 100, to explore the impact of training depth and convergence speed.

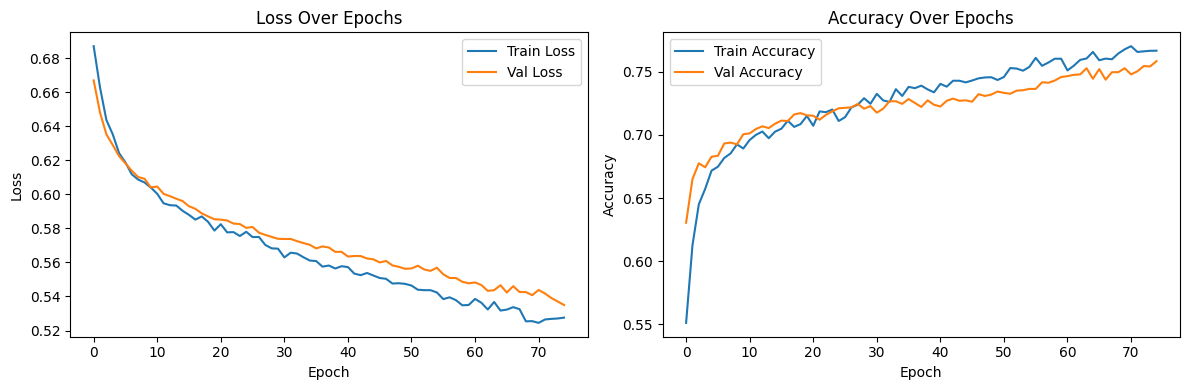
**Learning Rate**  
The learning rate is a key hyperparameter that controls the size of the steps taken during the optimization process. A lower learning rate provides greater training stability but increases training time, whereas a higher learning rate speeds up convergence but can prevent the model from reaching an optimal solution.

**Epochs**  
Epochs refer to the number of complete passes the model makes through the training dataset. To achieve a balance between underfitting and overfitting.

**Batch Size**  
The batch size determines how many samples are processed together in each training iteration. It directly affects the memory usage and the the stability of the training.

**5.1 Results of T1 and T2 separated**

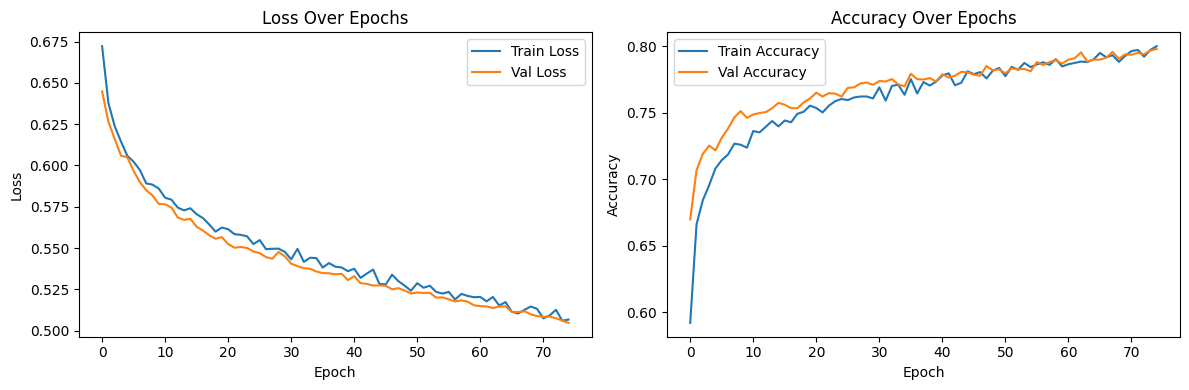
**Batch size = 32, dropout = 0.5, learning rate = , 75 epochs.**

*Figure 14. The model trained on T1 with Batch size 32, dropout 0.5, learning rate and 75 epochs.*

The model, trained on T1 weighted MRI slices, shows a steady improvement in both loss and accuracy over 75 epochs, with training and validation curves closely aligned indicating stable learning and good generalization. The model achieved a final test accuracy of **0.7615**, with a test loss of **0.5355**, showing strong alignment with training and validation results.

**Precision**: 83.74%, **Recall**: 63.64%, **F1 score**: 72.32%

The high precision suggests the model is good at avoiding false positives, but the lower recall means it misses a notable portion of true cases.

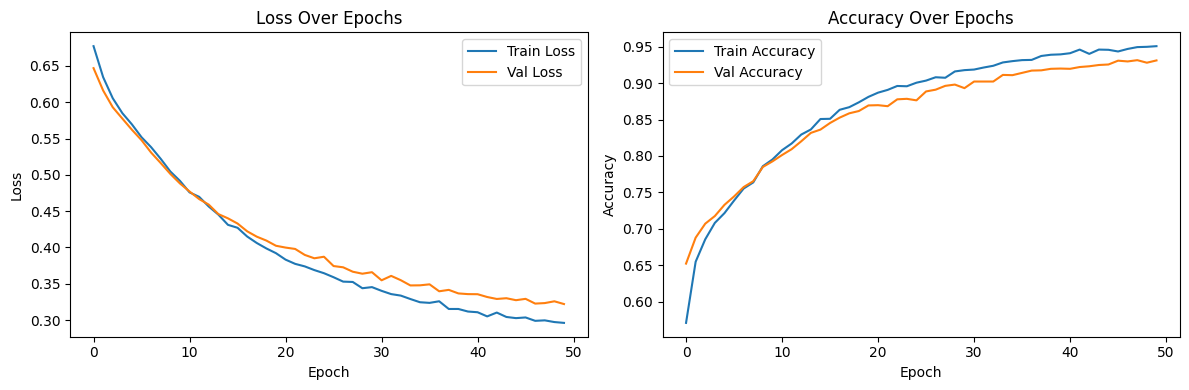
*Figure 15. The model trained on T2 with Batch size 32, dropout 0.5, learning rate and 75 epochs.*

The model, trained on T2 weighted MRI slices, reached a final test accuracy of **0.7966** with a test loss of **0.5031**, showing improved generalization compared to the T1 model. Throughout 75 epochs, both training and validation accuracy achieved nearly 80%, and validation loss consistently decreased to around 0.5048, indicating stable learning. These results suggest that T2 weighted images may have more discriminative features for endometriosis classification.

**Precision**: 80.51%, **Recall**: 76.51%, **F1 score**: 78.46%

The stronger recall and F1 score indicate that the model is more balanced and effective in identifying both positive and negative cases correctly.

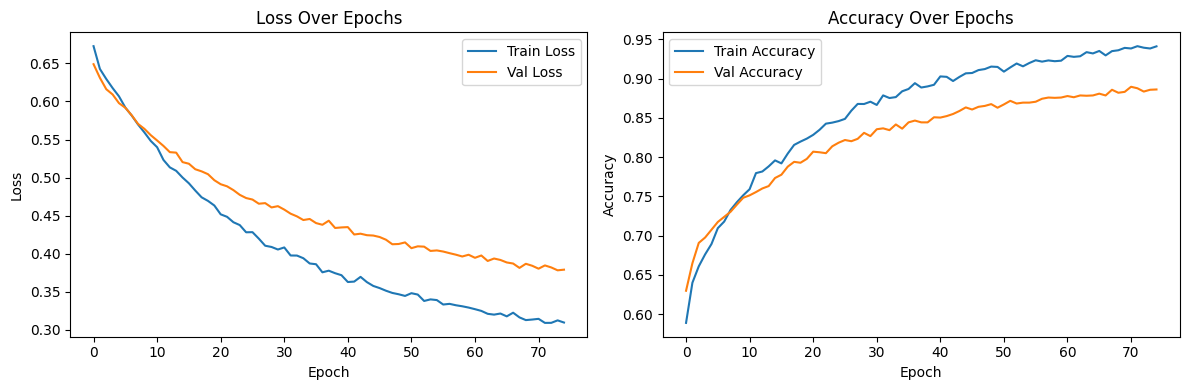
**Batch size = 64, dropout = 0.5, learning rate = ,** **denseblock4 unfreeze, 75 epochs**

*Figure 16. The model trained on T1 with Batch size 64, dropout 0.5, learning rate , denseblock4 unfreeze and 75 epochs.*

The model, trained on T1 weighted MRI slices, reached a final test accuracy of **0.9339** with a test loss of **0.3250**.The training and validation curves show stable and parallel improvement, with loss decreasing consistently and accuracy curves remaining closely aligned.

**The training stopped early after 50 epochs**, the early stopping was trigered because the model didn’t improve further over a number of epochs, likely due to convergence. This may result from the larger batch size and unfreezing of denseblock4, which allowed the model to learn more efficiently.

**Precision**: 95.52%, **Recall**: 90.77%, **F1 score**: 93.08%

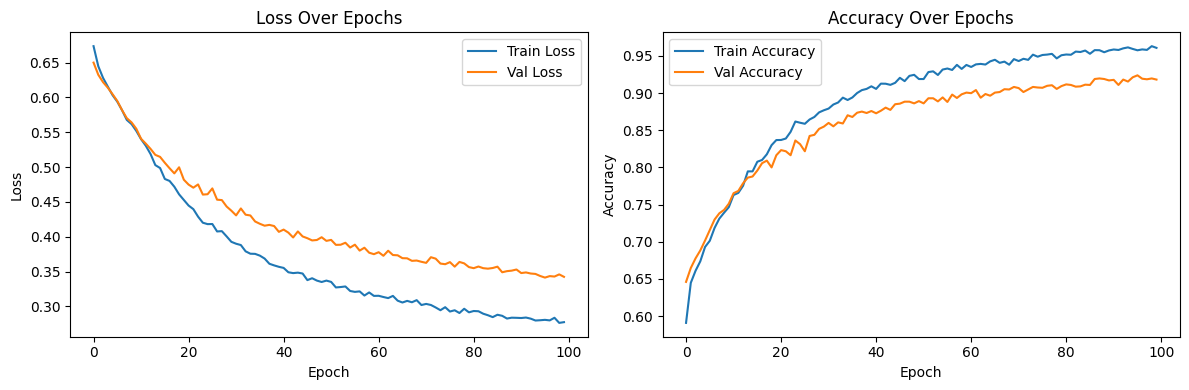
 *Figure 17. The model trained on T2 with Batch size 64, dropout 0.5, learning rate , denseblock4 unfreeze and 75 epochs.*

The model, trained on T2 weighted MRI slices, reached a final test accuracy of **0.9011** with a test loss of **0.3725**.The training and validation curves show consistent learning over 75 epochs, with both losses decreasing and accuracies rising steadily. The growing gap between the training and the validation accuracy and loss suggests slight overfitting, but overall the generalization remained strong.

**Precision:** 90.70%**, Recall:** 89.61%, **F1-score:** 90.15%

These results confirm that unfreezing deeper layers and using a larger batch size led to a well-optimized model with high performance across all metrics.

**Batch size = 64, dropout = 0.5, learning rate = , denseblock4 unfreeze, 100 epochs**

 *Figure 18. The model trained on T2 with Batch size 64, dropout 0.5, learning rate , denseblock4 unfreeze and 100 epochs.*

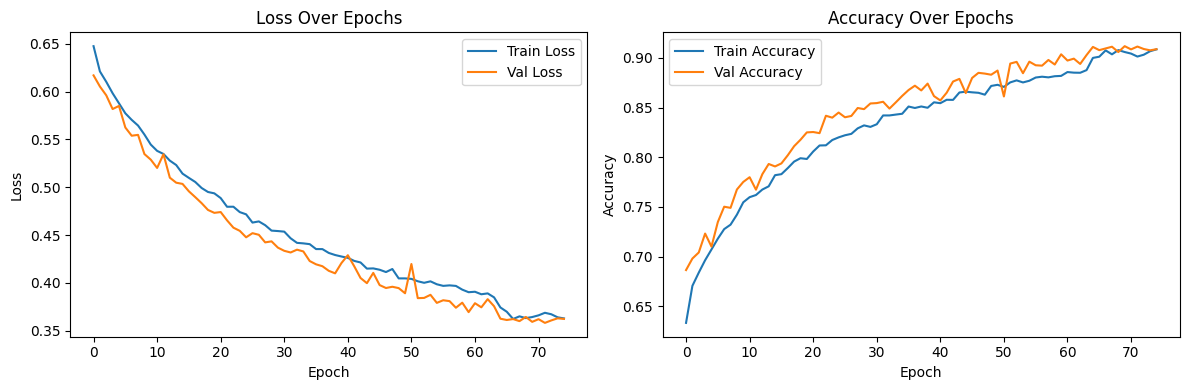
The model, trained on T2 weighted MRI slices, reached a final test accuracy of **0.9247** with a test loss of **0.3417**.The training and validation loss curves decrease steadily over 100 epochs, while the accuracy and loss curves continue to rise with a small gap between them, indicating slight overfitting.

**Precision:** 93.53%, **Recall:** 91.42%, **F1 score:** 92.46%

These results confirm its high performance and generalization capability after the more extended training.

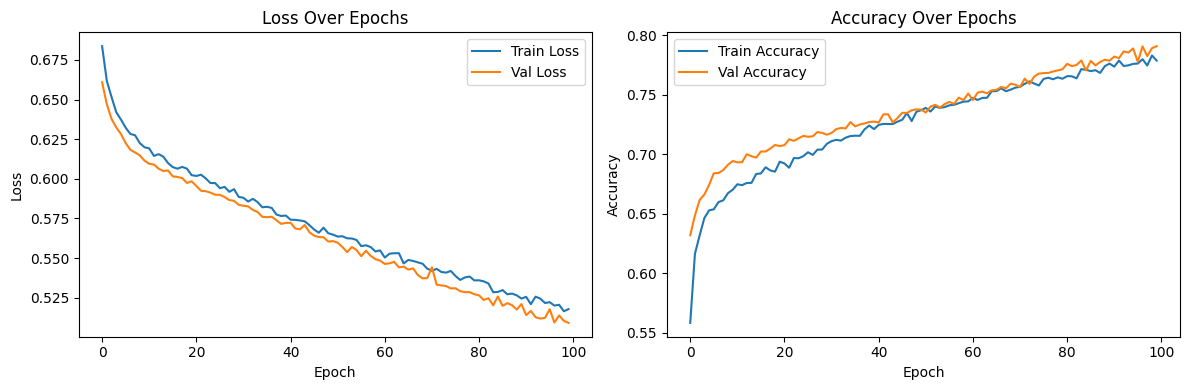
**5.2 Results of T1 and T2 combined**

**Batch size = 32, dropout = 0.3, learning rate = , 75 epochs**

*Figure 19. The model trained on T1 and T2 with Batch size 32, dropout 0.3, learning rate and 75 epochs.*

The model, trained on T1 and T2 weighted MRI slices with intermediate parameters, reached a final test accuracy of **0.9091** with a test loss of **0.3608**. The training and validation curves show convergence by the end of 75 epochs, with final accuracy and loss values closely aligned. However, both loss and accuracy curves display frequent spikes on the validation side, indicating some instability during the training, possibly because of the learning rate or the data variability. It is important to note that the validation was consistently better than the training across all epochs, which may indicate that there is a problem in the training process.

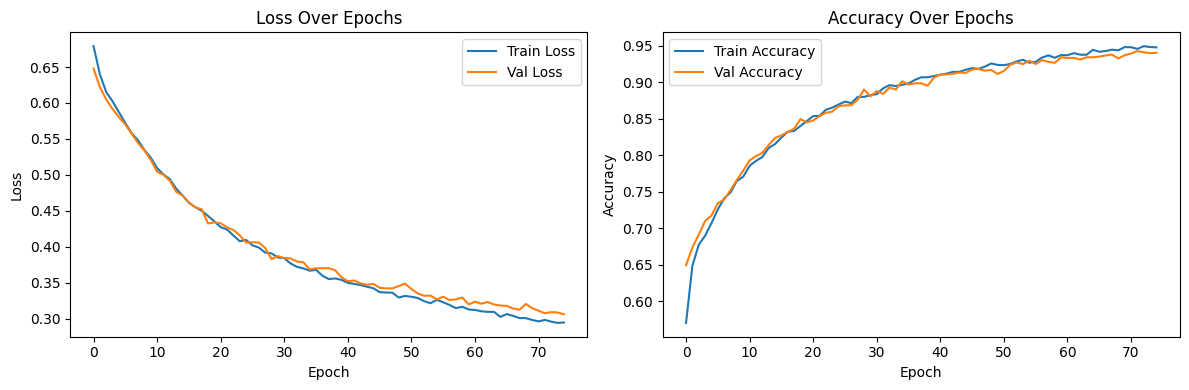
**Batch size = 32, dropout = 0.5, learning rate = , 100 epochs**

 *Figure 20. The model trained on T1 and T2 with Batch size 32, dropout 0.5, learning rate and 100 epochs.*

The model, trained on T1 and T2 weighted MRI slices, reached a final test accuracy of **0.7934** with a test loss of **0.5104**.The training and validation curves showed consistent and stable improvement, with validation loss decreasing smoothly and accuracy closely tracking training accuracy. **Precision: 78.30%**, **Recall: 77.61%**, **F1 score: 77.95%**

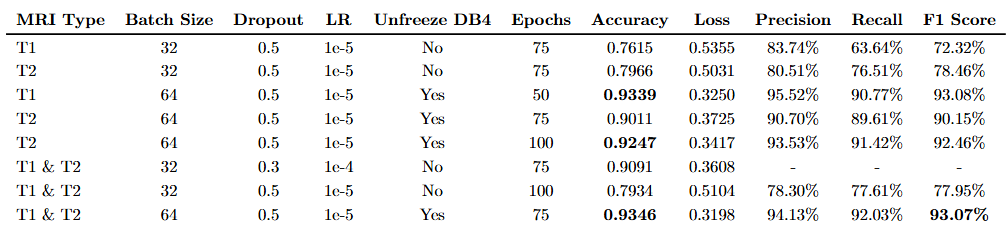
It is important to note that the validation was consistently better than the training across all the epochs, which may indicate that there is a problem in the training process.

**Batch size = 64, dropout = 0.5, learning rate = , denseblock4 unfreeze, 75 epochs**

 *Figure 21. The model trained on T1 and T2 with Batch size 64, dropout 0.5, learning rate , denseblock4 unfreeze and 75 epochs.*

The model, trained on T1 and T2 weighted MRI slices with batch size 64 and partial fine tuning by unfreezing fourth denseblock, allowing deeper features to adapt to the specific MRI data, reached a final test accuracy of **0.9346** with a test loss of **0.3198**.The training and validation curves show the strongest alignment, with stable convergence in both loss and accuracy by epoch 75. **Precision:** 94.13%, **Recall:** 92.03%, **F1 score:** 93.07%

These results indicate that it correctly identifies both endometriosis and non endometriosis cases with high reliability and small bias toward one class.

Table of results

Prediction Examples

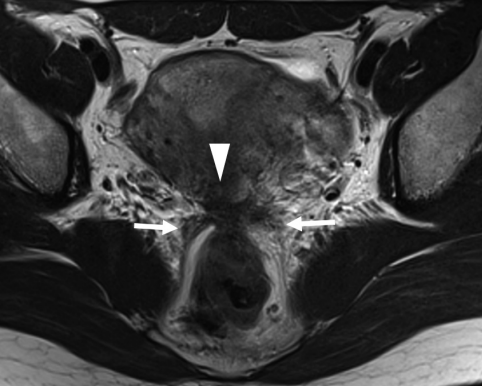
A close up of a scan

AI-generated content may be incorrect. Each example shows the prediction of the model from left and the labeled image from right.

*Figure 22. Pelvic endometriosis in a 42-year-old female with abdominal pain [16]*

Axial T2WI with fat saturation. Bilateral endometriotic cysts are observed posterior to the uterus, appearing as a ‘kissing ovary’. At the posterior surface of the uterus, fibrotic irregular thickening (arrows) is observed as a low signal intensity on T2WI. The bilateral ovaries and rectum converge at this point. This point corresponds to the insertion area of the uterosacral ligament, called the torus uterinus. During laparoscopic cystectomy, the pouch of Douglas is closed by adhesion.

**Model prediction: true with 97.82% confidence.**

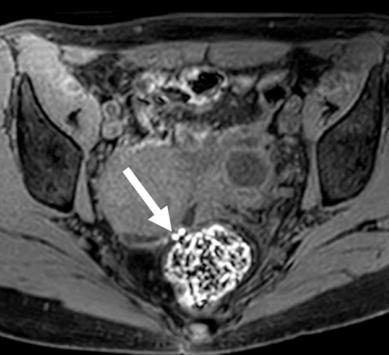
An x-ray of a pelvis

AI-generated content may be incorrect.

*Figure 23. Pelvic endometriosis in a 45-year-old female with severe dysmenorrhea and hypermenorrhea. [16]*

Axial T2 weighted image shows fibrous thickening of the torus uterinus and continuous irregular thickening of the bilateral uterosacral ligament . The rectal wall is stretched strongly to the torus uterinus, suggesting severe adhesions.

**Model prediction: true with 86.14% confidence.**

An x-ray of a person's pelvic body

AI-generated content may be incorrect.

*Figure 24. Axial T1FS [9]*

Hypointense nodules are seen in the pouch of Douglas. Small high-signal foci are visible on a coronal T1-weighted image, becoming more distinct on an axial fat-saturated T1-weighted image. These findings are consistent with deep pelvic endometriosis and suggest the presence of bleeding.

**Model prediction: true with 77.63% confidence.**

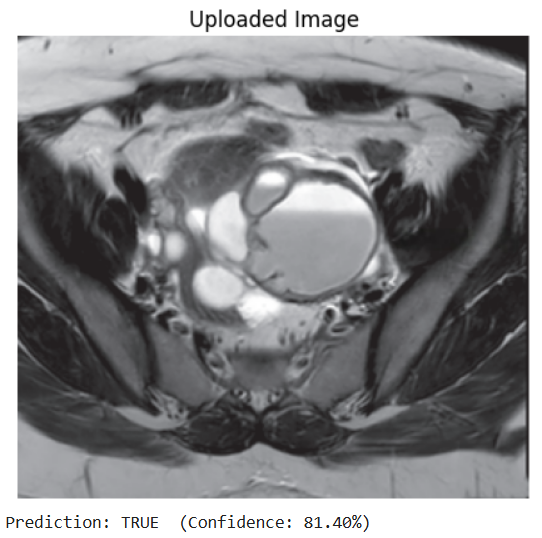
An ultrasound of a fetus

AI-generated content may be incorrect.A close-up of a person's uterus

AI-generated content may be incorrect.

*Figure 25. Endometrioma on axial T1FS [17]*

An x-ray of a person's chest

AI-generated content may be incorrect.

*Figure 26. Endometrioma on axial T2 [17]*

Endometrioma: showing well-defined, uniformly hyperintense cyst (white arrows) with no suppression of signal on T1 fat suppressed sequence, rather becoming more hyperintense giving a “light bulb appearance”. The lesion shows loss of the signal on T2W images (C) suggesting “T2 shading” and peripheral hypointense wall.

**Model prediction for T1FS: true with 96.15% confidence.**

**Model prediction for T2: true with 81.40% confidence.**

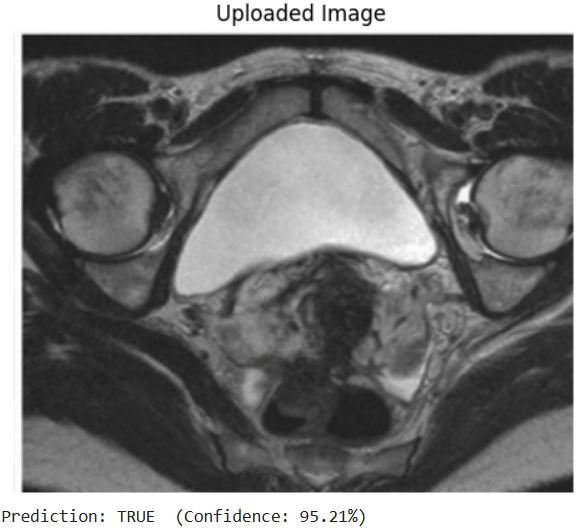
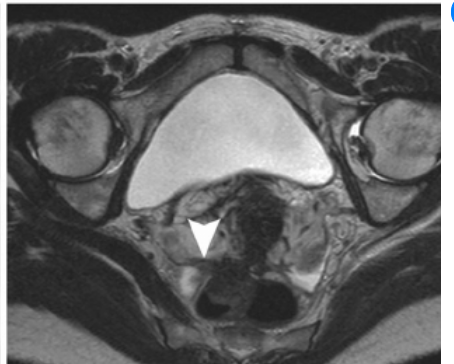
A close up of a mri

AI-generated content may be incorrect.

*Figure 27. Axial T2 [18]*

34-year-old woman with chronic pelvic pain. T2 weighted MR image in the axial plane clearly depicts a thick, low signal intensity band extending from the posterior uterine wall to the bowel, consistent with an adhesion (white arrow). Bilateral endometriomas and retraction of the ovaries (asterisks) to the midline due to co-existing adhesions also shown (kissing ovaries).

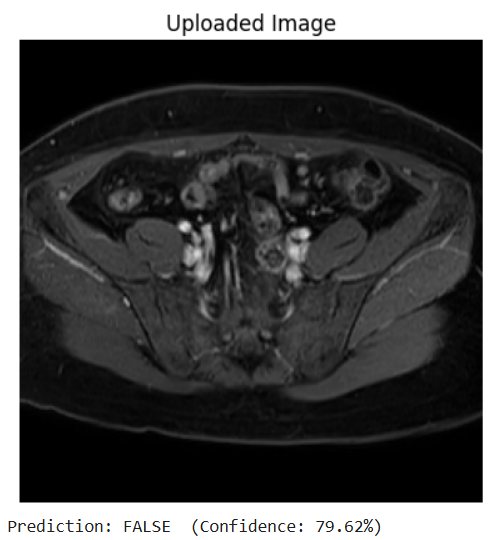
**Model prediction: true with 92.07% confidence.**



*Figure 28. Deep infiltrating endometriosis on axial TSE T2 [19]*

Deep infiltrating endometriosis: HR MRI (axial TSE T2 images). The image shows an endometriotic lesion infiltrating the rectum, creating adhesions with the right ovary and thickening of right USL (arrowhead).

**Model prediction: true with 95.21% confidence.**

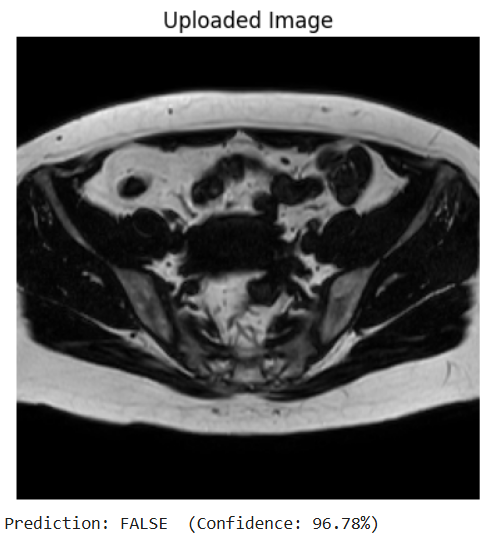
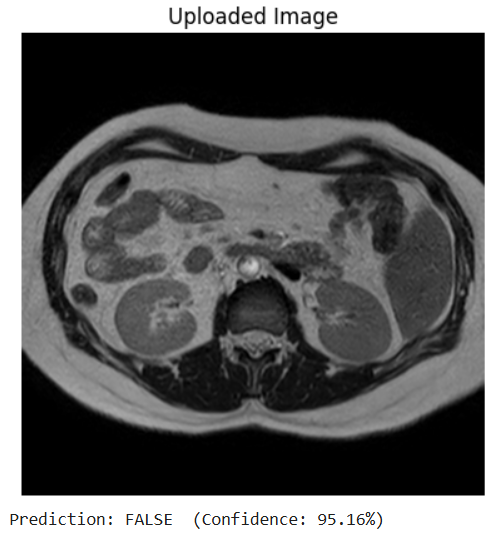
A close up of a body

AI-generated content may be incorrect.

*Figure 29. Axial T1FS with no endometriosis [20]*

Two different slices from the same T1FS MRI scan of a patient with no endometriosis.

**Model prediction for left slice: false with 84.59% confidence.**

**Model prediction for right slice: false with 79.62% confidence.**

*Figure 30. Axial T1 and T2 with no endometriosis [20]*

Two different modalities from the same MRI scan of a patient with no endometriosis.

**Model prediction for left T1: false with 96.78% confidence.**

**Model prediction for right T2: false with 95.16% confidence.**

### DISCUSSION

This project focused on developing a deep learning based approach for diagnosing endometriosis using MRI scans. By training and evaluating DenseNet121 models with transfer learning on T1 weighted, T2 weighted, and combined MRI data, the goal was to create a reliable, non invasive alternative to traditional laparoscopic diagnosis.

Dataset and Preprocessing

The MRI data set used in this study was obtained from the UT-EndoMRI dataset provided by UTHealth and published on Zenodo. The dataset included 130 MRI cases, each labeled as either diagnosed with endometriosis or healthy. Each case contained multiple MRI modalities, including T1, T1FS, T2, and T2FS. To overcome the relatively small dataset size, we applied extensive data augmentation, increasing the dataset to approximately 30,000 images. This was done by repeatedly slicing each MRI scan around its center, where pelvic organs are most visible, at small intervals. Each slice was then rotated by 90°, 180°, and 270° to generate four distinct images from a single slice. Despite this augmentation, we think that the dataset was not sufficient for the model to develop strong generalization. The MRI cases in the dataset differ in appearance from one another, and this variability likely made it difficult for the model to learn shared patterns across cases.

Transfer Learning and Model Behavior

Throughout the experiments, we used transfer learning with a densenet121 model pretrained on ImageNet. In the first training attempts, all feature layers were frozen and only the classifier was trained. However, since MRI images are different from the images in ImageNet, this limited the ability of the model to adjust to our task. To improve this, we later unfroze the final dense block, denseblock4 of the network, allowing deeper layers to be fine tuned. This change helped the model learn more specific features from the MRI scans while keeping the earlier, more general layers fixed. This approach led to noticeable improvements in all the results for each MRI modality used.

When comparing the MRI modalities, the model performed better on T2 images than on T1 when all layers were frozen. However, after unfreezing the fourth denseblock, the model trained on T1 images outperformed the model trained on T2 images in both accuracy and loss. One possible explanation is that T1 images contain condition specific features that are harder to learn when the model is restricted to general pretrained weights. And once the model is allowed to fine tune its deeper layers, it can better extract relevant features from T1 scans. This shift in performance might also be influenced by the specific distribution and characteristics of the T1 and T2 samples in the dataset.

To improve the robustness and generalization of the model, we trained it on a combined dataset that included both T1, T1FS and T2, T2FS MRI images. Each modality provides different contrast and highlights distinct anatomical features, which can be valuable in identifying subtle signs of endometriosis. By exposing the model to both types of images during training, we aimed to increase its ability to recognize relevant patterns across a wider variety of clinical imaging conditions.

Optimization Strategy

To optimize the model, we used the Adam optimizer with a learning rate of 1e-5 and a weight decay of 1e-4. Adam was chosen due to its adaptive learning rate properties, which help accelerate convergence and are particularly effective in deep learning tasks with noisy gradients, such as training on medical image data.

For the loss function, we used Cross Entropy Losswith label smoothing of 0.1, the label smoothing encourages the model to be less confident in its predictions, which can help reduce overfitting and improve generalization, this is useful when working with relatively small datasets.

To further improve the stability during the training, we implemented a Reduce LR On Plateau scheduler, this scheduler monitors the validation loss and reduces the learning rate by a factor of 10 if no improvement is observed after 3 epochs, this mechanism helps the model converge more smoothly and avoid getting stuck in local minima or plateaus.

We applied early stopping with a patience limit of 5 epochs to prevent overfitting. If the validation loss did not improve for five epochs, the training stopped. This approach helped ensure that the model did not continue learning unnecessarily once generalization performance plateaued, saving computational resources and reducing the risk of overfitting to the training set.

Future Work

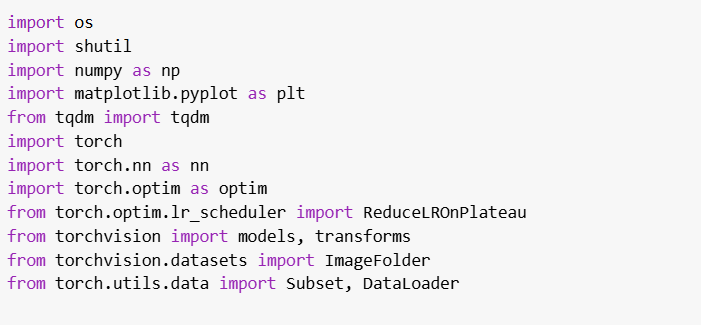
Future work can include adding a segmentation component following the classification part. Once the model identifies whether a case is positive for endometriosis, a segmentation model such as YOLO could be applied to localize and highlight specific regions associated with the condition. This approach could improve clinical interpretability by visually indicating the areas of interest within the MRI scan and can support more precise assessments and follow up decisions by medical professionals.

### CONCLUSION

This study demonstrated the potential of using deep learning techniques for non invasive diagnosis of endometriosis from MRI scans. By applying transfer learning with a DenseNet121 model, we were able to classify MRI cases with promising results. The experiments highlighted that the model was capable of learning meaningful patterns from the data. Overall, this work supports the idea that deep learning can assist in the detection of endometriosis, potentially reducing the need for invasive diagnostic procedures.

### USER GUIDE

Building the environment



This cell imports all the required packages used by the model. The imports can be grouped into the following categories:

**File handling:** os and shutil for copying datasets into the local runtime.

**Numerical and plotting libraries:** numpy for numerical operations, matplotlib.pyplot for plotting training metrics.

**Training utilities:** tqdm is used for progress bars.

**PyTorch:** Core machine learning framework used to define the models, optimization, and hardware control.

**TorchVision:** Provides the pretrained model, image transformations, and dataset handling.

The second cell is responsible for mounting google drive and copying the datasetand itperforms the following operations:

1. Mounts google drive using the colab built in drive.mount() method.
2. Defines the source paths for two datasets:
   * dataset\_t1\_new containing T1 weighted MRI images.
   * dataset\_t2\_new containing T2 weighted MRI images.
3. Copies each dataset folder from Google Drive into the local file system if not already present.
4. Counts and prints the number of image files in each dataset for validation.

For this model there were two different datasets T1 and T2 because we trained the model on these datasets separately. For future work it can be changed to work on one dataset that has true and false folders.

Preprocessing

A screenshot of a computer code

AI-generated content may be incorrect.

This cell defines two different image transformation pipelines:

**train\_transform:** Applies different filters to simulate imaging variation and improve generalization. This includes:

* RandomHorizontalFlip: Flips the image horizontally with a 50% chance.
* RandomRotation(15): Rotates the image within ±15 degrees.
* ColorJitter: Adjusts brightness and contrast slightly.

**val\_test\_transform:** Only includes resizing and normalization, without any augmentation, to ensure consistency during the validation and testing.

These transforms are later applied automatically when datasets are loaded.

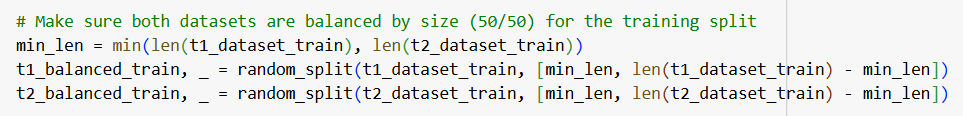
Preparing the datasets

This cell is responsible for loading the T1 and T2 MRI image datasets, balancing them, and splitting them into training, validation, and test sets. It then constructs PyTorch DataLoader objects for each set.

A screen shot of a computer code

AI-generated content may be incorrect.

1. Loads training, validation and test datasets separately for both T1 and T2 modalities using the defined transforms.



2. Randomly downsamples both datasets to ensure class balance and shuffles the images.

3. Concatenates the balanced datasets into one training set.

A computer screen shot of a code

AI-generated content may be incorrect.

4. Extracts class labels and applies a 70, 20, 10 split to train, validation and test, for each class.

5. Wraps the splits in PyTorch Subset objects so they can be used with DataLoader.

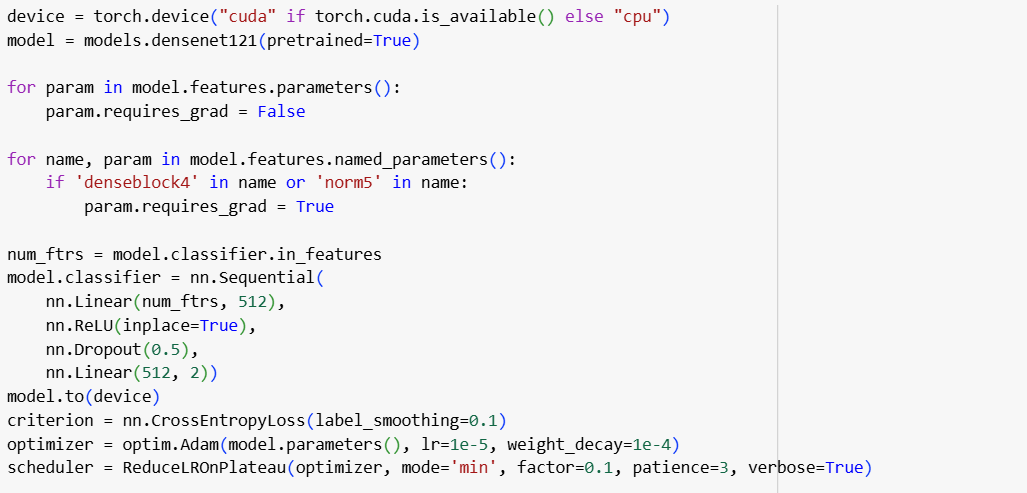
6. Initializes DataLoaders for efficient batch wise data access during training and evaluation.

A screen shot of a computer code

AI-generated content may be incorrect.

Each loader is configured to use a batch size of 64, with shuffling enabled only for training.

Model Architecture and Optimization Setup



1. Loads DenseNet121 with ImageNet pretrained weights.
2. Freezes all convolutional layers to preserve learned features.
3. Unfreezes the deepest layer denseblock4 and norm5 to allow domain-specific fine-tuning on medical data.
4. Replaces the original classifier with a new head for binary classification of two classes.
5. Defines training strategy:
   * CrossEntropyLoss with label smoothing to reduce overconfidence.
   * Adam optimizer with L2 regularization.
   * ReduceLROnPlateau dynamically lowers the learning rate if validation loss stagnates.

Training loop



A screenshot of a computer program

AI-generated content may be incorrect.

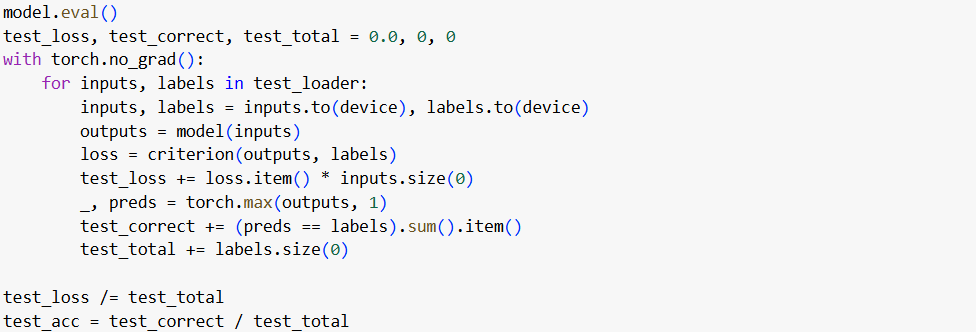
This loop trains the model for up to 75 epochs. During each epoch:

* The model is set to training mode.
* It performs forward passes, computes losses, updates weights, and tracks metrics.
* It then switches to evaluation mode for validation and stores those results.
* If validation loss does not improve for 5 epochs, training is stopped early.

All training and validation losses and accuracies are stored for later visualization.

Test evaluation

This cell evaluates the trained model on the test data.



model.eval(): Puts the model into evaluation mode.

with torch.no\_grad(): Disables gradient tracking to save memory and improve inference speed.

Loops over the test\_loader:

* Moves input images and labels to the correct device, CPU or GPU.
* Runs the model to get predictions.
* Calculates the loss using the same criterion as during the training.
* Tracks the number of correct predictions and total images.

At the end:

* Computes average test loss and test accuracy across all test batches.

Confusion Matrix and Evaluation Metrics

This cell evaluates the performance of the model on the test data.

A screenshot of a computer program

AI-generated content may be incorrect.

1. Uses predictions from the test set to compute:

* Confusion matrix
* Accuracy, Precision, Recall, F1 Score

2. Displays the confusion matrix as a plot with labeled axes.

Visualization of the training

Plots two charts:

* Train vs. Validation Loss over epochs.
* Train vs. Validation Accuracy over epochs.

These plots are essential to understand if the model was underfitting, overfitting, or converging well.

Upload and Classify New MRI Image

Allows a user to upload an image and run inference using the trained model.



* Preprocesses the image using validation transforms.
* Classifies the image as "true" or "false".
* Displays confidence of the prediction.

This step enables practical testing of the model on new MRI slices.

### REFERENCES

1. Horne, A. W., & Missmer, S. A. (2022). Pathophysiology, diagnosis, and management of endometriosis. *BMJ*, *379*, o70750. <https://doi.org/10.1136/bmj-2022-070750>

2. **Signorile, P. G., Viceconte, R., & Baldi, A.** (2022). New insights in pathogenesis of endometriosis. Frontiers in Medicine, 9, 879015. <https://doi.org/10.3389/fmed.2022.879015>

3. Zhao, N., Hao, T., Zhang, F., Ni, Q., Zhu, D., Wang, Y., Shi, Y., … Mi, X. (2024). Application of machine learning techniques in the diagnosis of endometriosis. *BMC Women’s Health*, *24*, 491. <https://doi.org/10.1186/s12905-024-03334-2>

4. Lamceva, J., Uljanovs, R., & Strumfa, I. (2023). The main theories on the pathogenesis of endometriosis. *International Journal of Molecular Sciences, 24*(5), 4254.

<https://doi.org/10.3390/ijms24054254>

5. Toré, U., Abilgazym, A., Asunsolo-del-Barco, A., Terzic, M., Yemenkhan, Y., Zollanvari, A., … & Sarria‑Santamera, A. (2023). Diagnosis of endometriosis based on comorbidities: A machine learning approach. *Biomedicines, 11*(11), 3015.

<https://doi.org/10.3390/biomedicines11113015>

6. Yang, M., Liu, M., Chen, Y., He, S., & Lin, Y. (2021). Diagnostic efficacy of ultrasound combined with magnetic resonance imaging in diagnosis of deep pelvic endometriosis under deep learning. *The Journal of Supercomputing*, *77*(7), 7598–7619.

<https://doi.org/10.1007/s11227-020-03535-0>

7. van Waesberghe, J. H., Hazewinkel, M., & Busard, M. P. (2011, November 1). *Endometriosis – MRI detection*. Radiology Assistant. VU University Medical Center Amsterdam, the Netherlands.

<https://radiologyassistant.nl/abdomen/unsorted/endometriosis-mri-detection>

8. Bendifallah, S., Puchar, A., Suisse, S., Delbos, L., Poilblanc, M., Descamps, P., Golfier, F., Touboul, C., Dabi, Y., & Daraï, É. (2022). Machine learning algorithms as new screening approach for patients with endometriosis.

*Scientific Reports, 12*(1), 639. <https://doi.org/10.1038/s41598-021-04637-2>

9. Alonzo, L., Cannella, R., Gullo, G., Piombo, G., Cicero, G., Lopez, A., Billone, V., Andrisani, A., Cucinella, G., Lo Casto, A., & Lo Re, G. (2024). Magnetic resonance imaging of endometriosis: The role of advanced techniques. *Journal of Clinical Medicine, 13*(19), 5783. <https://doi.org/10.3390/jcm13195783>

10. National Institute of Biomedical Imaging and Bioengineering. (n.d.). *Magnetic resonance imaging (MRI)*. National Institutes of Health.

<https://www.nibib.nih.gov/science-education/science-topics/magnetic-resonance-imaging-mri>

11. Huang, G., Liu, Z., van der Maaten, L., & Weinberger, K. Q. (2017).

*Densely connected convolutional networks* (arXiv:1608.06993v5) [Computer software]. arXiv. <https://doi.org/10.48550/arXiv.1608.06993>

12. GeeksforGeeks. (n.d.). *GeeksforGeeks: A computer science portal for geeks*.

<https://www.geeksforgeeks.org/>

13. Hosna, A., Merry, E., Gyalmo, J., Alom, Z., Aung, Z., & Azim, M. A. (2022). Transfer learning: A friendly introduction. *Journal of Big Data, 9*(1), Article 102.

<https://doi.org/10.1186/s40537-022-00652-w>

14. Timilsina, A. (2023, May 14). *YOLOv8 architecture explained*. Medium.

<https://abintimilsina.medium.com/yolov8-architecture-explained-a5e90a560ce5>

15. Ultralytics. (2024). *Ultralytics – Home of YOLOv5 and YOLOv8 models*.

<https://www.ultralytics.com/>

16. Kido, A., Himoto, Y., Moribata, Y., Kurata, Y., & Nakamoto, Y. (2022). MRI in the diagnosis of endometriosis and related diseases: A pictorial review.

*Korean Journal of Radiology, 23*(4), 426–445. <https://doi.org/10.3348/kjr.2021.0405>

17. Zuber, M., Shoaib, M., & Kumari, S. (2022). Magnetic resonance imaging of endometriosis: A common but often hidden, missed, and misdiagnosed entity. *Polish Journal of Radiology, 87*, e448–e461. <https://doi.org/10.5114/pjr.2022.119032>

18. Bourgioti, C., Preza, O., Panourgia, E., Chatoupis, K., Antoniou, A., Nikolaidou, M. E., & Moulopoulos, L. A. (2017). MR imaging of endometriosis: Spectrum of disease. *Diagnostic and Interventional Imaging, 98*(10), 751–767. <https://doi.org/10.1016/j.diii.2017.05.009>

19. Scardapane, A., Lorusso, F., Scioscia, M., Ferrante, A., Stabile Ianora, A. A., & Angelelli, G. (2014). Standard high-resolution pelvic MRI vs. low-resolution pelvic MRI in the evaluation of deep infiltrating endometriosis.

*European Radiology, 24*(10), 2567–2575. <https://doi.org/10.1007/s00330-014-3297-4>

20. X. Liang, L.A. Alpuing Radilla, K. Khalaj, H. Dawoodally, C. Mokashi, X. Guan, K.E. Roberts, S.A. Sheth, V.S. Tammisetti, L. Giancardo. "A Multi-Modal Pelvic MRI Dataset for Deep Learning-Based Pelvic Organ Segmentation in Endometriosis."

[UTHealth - Endometriosis MRI Dataset (UT-EndoMRI)](https://zenodo.org/records/13749613)