

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

Capstone Project Phase A 25-1-R-8

miri weiss cohen

Table of Contents

[1. INTRODUCTION 2](#_Toc189080455)

[2. LITRETURE OVERVIEW 5](#_Toc189080456)

[3. BACKGROUND 9](#_Toc189080457)

[4. RESEARCH PROCESS 19](#_Toc189080458)

[5. EXPECTED ACHIEVEMENTS 21](#_Toc189080459)

[6. EVALUATION AND VERIFICATION PLAN 22](#_Toc189080460)

[7. PRODUCT 23](#_Toc189080461)

[8. REFERENCES 25](#_Toc189080462)

ABSTRACT

Endometriosis is a chronic, inflammatory gynecologic disease that affects approximately two hundred million women worldwide. Individuals with endometriosis often experience profound symptoms, including severe pain during menstruation or sexual intercourse, infertility, persistent pelvic pain, exhaustion, and mental health challenges.

Endometriosis is characterized by the growth of endometrial-like tissue outside the uterus.

The current preferred method of diagnosis is laparoscopy, an invasive surgical procedure performed to visually inspect and confirm the presence of endometrial tissue within the abdominal cavity.

Since these tissues are visually observable and can be seen using magnetic resonance imaging (MRI), we intend to use a DenseNet model for initial classification of MRI scans, followed by a YOLO model for detecting and marking endometriotic lesions. This approach aims to replace the invasive and expensive diagnostic procedure of 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

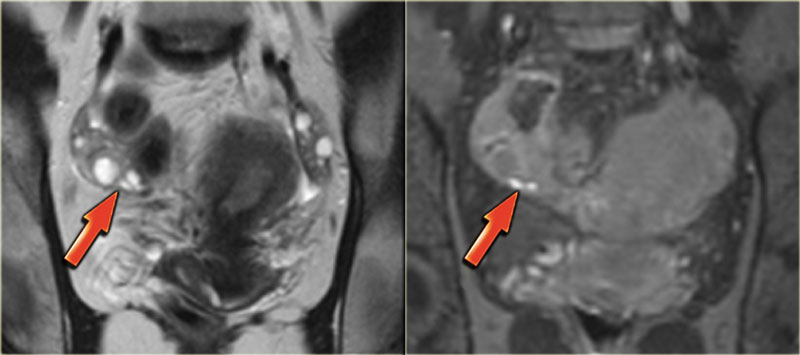
There are many manifestations of endometriosis, from incidentally discovered asymptomatic lesions to more severe cases.

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.

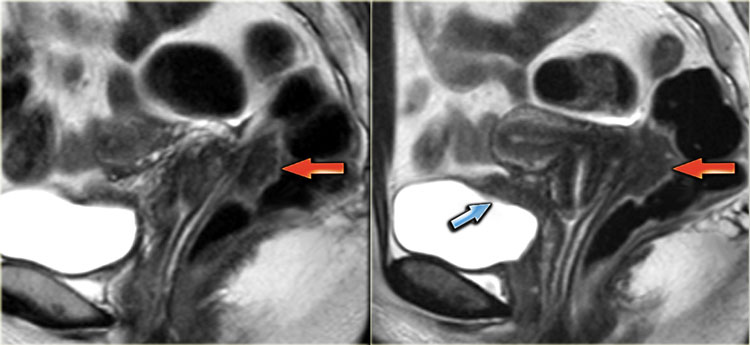
Ovarian endometriotic cysts, also known as endometriomas, develop on the ovaries and form cysts with fluid of varying sizes, this subtype is closely associated with infertility and an increased risk of ovarian cancer. Deep infiltrating endometriosis, the most severe form, invades visceral organs to a 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].

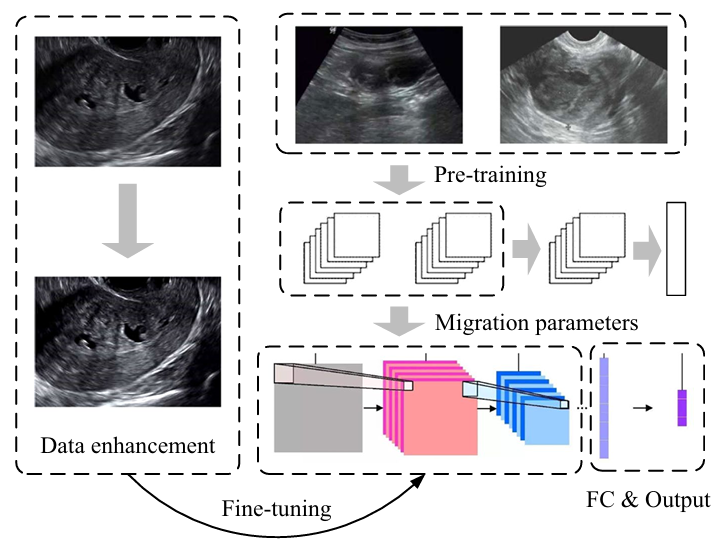
Endometriosis frequently demands extensive medical and surgical interventions, involving considerable costs and risks. Although the condition is widespread and leads to significant morbidity, it is often underdiagnosed and undertreated, with a diagnostic delay of 8 to 12 years from the beginning of symptoms. 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%.

Sofiane Bendifallah, Anne Puchar, Stéphane Suisse, Léa Delbos, Mathieu Poilblanc, Philippe Descamps, Francois Golfier, Cyril Touboul, Yohann Dabi, Emile Darai [8] developed a diagnostic model based on machine learning to predict endometriosis using clinical and symptom-based data provided by patients from the Ziwig Health platform. The data included features such as reported symptoms, medical history, demographic characteristics, and quality of life indicators.

Their research began with a dataset of 8,000 records, which was filtered down to 1,126 patients with endometriosis and 608 patients without the condition. They selected 16 significant features identified with recommendations from experts and statistical analysis.

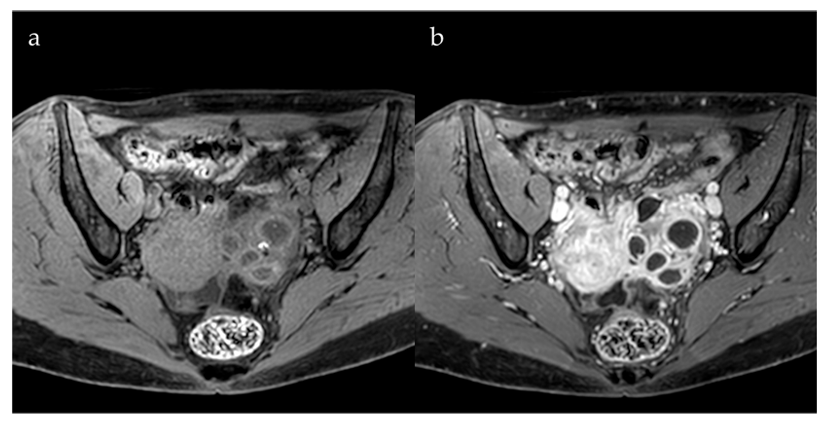
A flowchart of a patient

Description automatically generated

*Figure 5. Flow chart of population for model development and validation [8]*

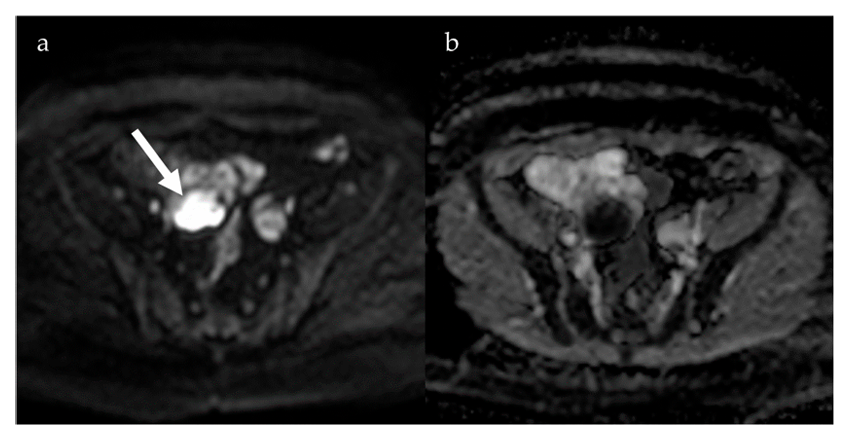
The models used included Logistic Regression, Decision Tree, Random Forest, extreme Gradient Boosting, and Voting Classifiers. The models were trained and validated on the dataset, with performance evaluated using sensitivity, specificity, F1 score, and AUC ROC metrics. The results demonstrated that Random Forest, XGB, and Voting Classifiers achieved the best performance, with sensitivity values reaching 95% and specificity ranging between 66% and 92%.

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]*

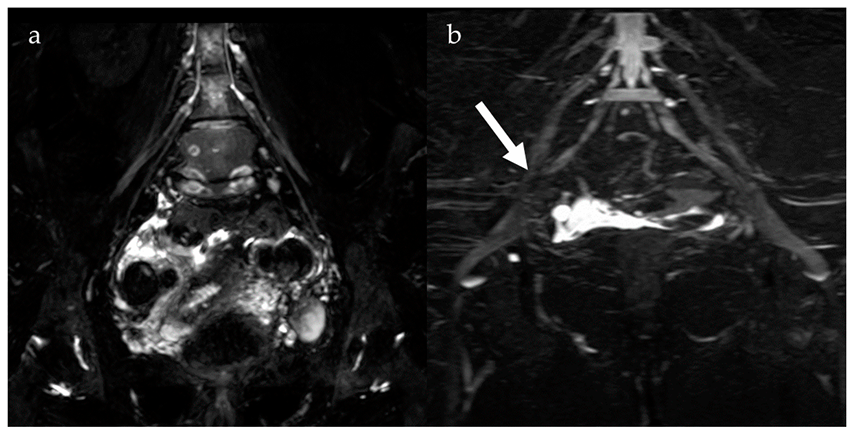
Gadolinium-based contrast agents in MRI scans help distinguish endometrial cysts from other pelvic lesion types. This method is beneficial when intense wall enhancement is observed, improving the ability to identify and evaluate such lesions.



*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]*

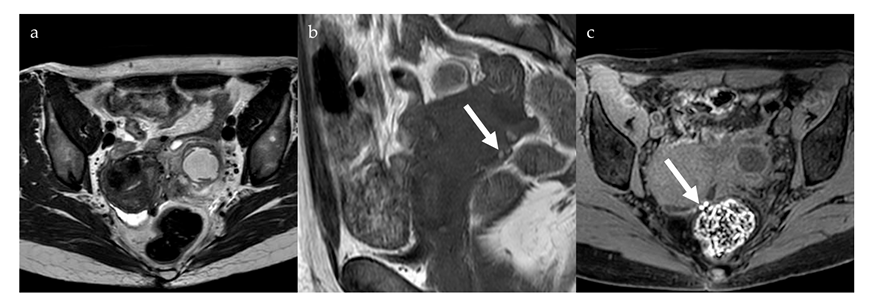
Magnetic resonance neurography (MRN) and diffusion tensor imaging (DTI) are useful for evaluating nerve involvement caused by endometriosis. MRN detects nerve abnormalities, such as thickening or misalignment. DTI measures water movement along nerve fibers, identifying disruptions caused by nerve compression 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].

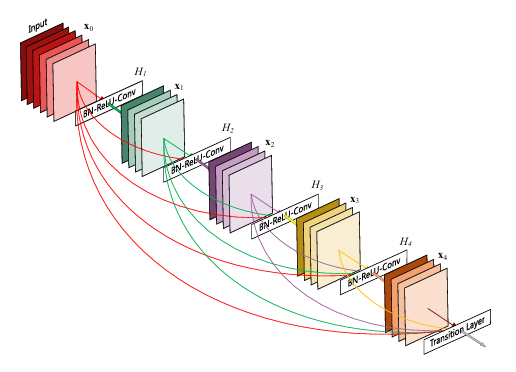
**3.2 DENSE NET**

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.

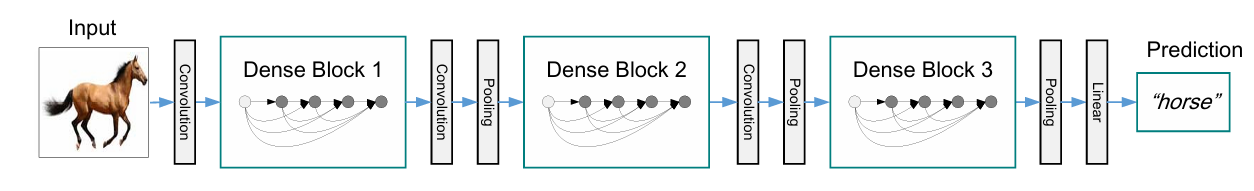


*Figure 10. A 5-layer dense block with a growth rate of k = 4. Each layer takes all preceding feature-maps as input [11]*

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 a 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].

**3.3 YOLO**

YOLO (You Only Look Once) is a high-speed object detection algorithm designed to classify and locate objects within an image in a single pass. By dividing the image into grid cells, YOLO assigns each cell the responsibility of detecting objects within its boundaries, enabling real-time performance while maintaining accuracy.

YOLOv8, developed by Ultralytics [15], introduces key advancements over its predecessor, YOLOv5, most notably the transition from Anchor-Based Detection to Anchor-Free Detection. Unlike YOLOv5, which relied on predefined anchor boxes and calculated offsets for predictions, YOLOv8 directly predicts the center, width, and height of objects. This reduces computational complexity, improves inference speed, and enhances accuracy across object scales. The adoption of Anchor-Free Detection allows YOLOv8 to generalize better to new datasets, eliminate the need for manual anchor tuning, and deliver a flexible, efficient, and highly precise solution for modern object detection tasks.

The YOLOv8 family consists of several model variants, each offering a trade-off between accuracy, speed, and size to suit diverse application needs. These variants are defined by parameters such as depth\_multiple (d), which adjusts the network’s depth by scaling the number of Bottleneck Blocks, and width\_multiple (w), which scales the number of channels in convolutional layers. Smaller values for d and w result in faster, compact models with reduced accuracy, while larger values increase accuracy at the cost of more computational resources. The max\_channels (mc) parameter limits the number of channels to control model size and prevent overfitting.

YOLOv8 offers five variants: n (nano) for maximum speed with lower accuracy, s (small) for a balance of speed and accuracy, m (medium) for higher accuracy with moderate speed, l (large) for precision in demanding tasks, and xl (extra-large) for the highest accuracy in resource-intensive applications. These configurations enable users to choose the optimal model based on their specific requirements, ensuring efficient and effective object detection.

Blocks in YOLOv8

In YOLOv8, each block processes feature maps defined by dimensions H×W×min (num, mc)×w. Here, H and W represent the spatial dimensions, min (num, mc) specifies the number of channels (chosen as the smaller value between num and mc), and w is the width multiplier that adjusts the number of channels. This structure enables flexibility and ensures efficient management of feature map complexity across the network.

Conv Block: The Conv Block is a key building block in YOLOv8, responsible for extracting features, improving data representation, and adjusting spatial dimensions when necessary. It consists of three layers: Conv2D, BatchNorm2D, and the SiLU activation function. The Conv2D layer performs convolution operations using filters to generate feature maps, with parameters such as k (kernel size), s (stride, which reduces spatial dimensions), p (padding, which preserves dimensions), and c (number of filters, determining the output channels). The BatchNorm2D layer normalizes feature map values for stability and faster training, while the SiLU activation function introduces non-linearity, ensuring smooth gradients and enhancing the network's ability to learn complex patterns. Together, these components allow the Conv Block to transform the input feature map into a refined representation, ready for further processing in the subsequent layers of the architecture.

C2f (Cross Stage Partial Block): The C2f block in YOLOv8 plays a key role in enhancing feature processing while retaining critical information. It extracts complex features, preserves raw data through shortcuts, and combines processed and unprocessed information into a unified output. The input feature map, sized 𝐻×𝑊×𝐶, is divided into two paths: the internal path, which processes data deeply using Bottleneck blocks, and the shortcut path, which bypasses processing to retain the original data. If shortcut=true, the Bottleneck blocks within the internal path further split the input, allowing part of the data to bypass convolutional layers and combine with processed data via addition. If shortcut=false, all the data is fully processed by the Bottlenecks.

Within the Bottleneck blocks, the Conv layers perform key transformations:

1×1: Reduces channels to minimize computation.

3×3: Extracts spatial features while maintaining input dimensions.

1×1: Restores or adjusts the number of channels for compatibility.

The shortcut path directly transfers its input to the output without modification. After both paths are processed, their outputs are concatenated along the channel dimension. If needed, a final 1×1 Conv layer adjusts the number of channels in the combined feature map.

By efficiently combining feature extraction with information retention, the C2f block ensures a balance of accuracy and efficiency, making it integral to YOLOv8's robust object detection capabilities.

SPPF Block: The SPPF Block (Spatial Pyramid Pooling - Fast) in YOLOv8 is designed to efficiently process feature maps while reducing computational complexity. It allows the network to learn multi-scale features by pooling information at various spatial resolutions, enabling adaptability to images of different sizes. Unlike standard Spatial Pyramid Pooling, the SPPF block employs a simplified, faster pooling scheme to balance speed and accuracy.

The block starts with a 1×1 convolutional layer that reduces the number of channels without changing the spatial dimensions, minimizing computational load. This is followed by multiple max-pooling operations using a fixed kernel size (e.g., 5×5). These operations retain the most prominent features by selecting maximum values within each kernel region. Padding ensures that the spatial dimensions remain consistent throughout the process. The resulting feature maps from the max-pooling layers are concatenated along the channel dimension, creating a unified feature map rich in multi-scale information. To prepare this output for subsequent stages, a final 1×1 convolution reduces the number of channels to the desired count while maintaining the spatial dimensions.

By combining multi-scale pooling and feature fusion, the SPPF block helps YOLOv8 capture both fine and coarse details effectively, enhancing its ability to detect objects of varying sizes with greater efficiency.

Detect Block: The Detect Block is the final stage of YOLOv8, responsible for predicting bounding boxes and object classes. Unlike earlier YOLO versions, YOLOv8 adopts an anchor-free detection approach, which directly predicts the center, width, and height of an object within a grid cell. This simplification reduces computational complexity and improves detection speed while maintaining accuracy. The Detect Block processes feature maps from previous layers, such as C2f and Concatenate, and predicts objects across multiple scales. High-resolution maps (e.g., 80×80) detect small objects, while low-resolution maps

(e.g., 20×20) capture larger objects. Each grid cell in the feature map produces a vector containing: x, y: Coordinates of the object center relative to the grid cell’s top-left corner.

w, h: Width and height of the bounding box. Confidence score: Probability of the object’s presence in the grid cell. Class probabilities: Likelihood of the object belonging to a specific category (e.g., dog, car). The block operates through two tracks: the bounding box prediction track, responsible for x, y, w, h and confidence scores, and the class prediction track, which outputs probabilities for each class. Both tracks use convolutional layers, concluding with a 1×1 convolution to produce predictions. After predictions, normalization functions are applied: Sigmoid for bounding box coordinates and confidence scores, and Softmax for class probabilities. A Non-Maximum Suppression (NMS) step follows, filtering overlapping boxes and retaining only the most confident detections. By combining anchor-free detection with multi-scale processing and efficient post-processing, the Detect Block ensures fast, accurate, and streamlined object detection.

YOLOv8 Architecture:

A diagram of a computer

Description automatically generated

*Figure 12. An illustrates of the YOLOv8 architecture, highlighting the Backbone, Neck, and Head components, which process feature maps for object detection [14]*

Backbone: The Backbone in YOLOv8 is the feature extraction module that transforms the input image into meaningful feature maps for object detection. Starting with an input size of 640×640, the Backbone uses a combination of Conv and C2f blocks to gradually reduce the spatial resolution while increasing the number of channels, enabling the network to capture patterns at multiple levels of granularity. The process begins with Conv blocks, where the image resolution is halved (e.g., 640×640→320×320) while channels are increased based on depth and width multipliers. Between these Conv layers, C2f blocks refine and enrich the feature maps using bottleneck structures with shortcut connections. These shortcuts allow part of the data to bypass processing, preserving essential information while the rest undergoes convolution for extracting detailed features. As the data alternates between Conv and C2f blocks, the resolution continues to shrink (e.g., 320×320→160×160→80×80) while the feature maps become deeper and more abstract. Intermediate outputs from the Backbone are also sent to the Neck for multi-scale feature fusion, ensuring that critical spatial and semantic information is retained. The final stage of the Backbone is the SPPF block, which consolidates multi-scale features into a fixed-size representation. By pooling and concatenating information, the SPPF captures global context while preserving spatial details, producing feature maps that are rich in semantic content and optimized for detection tasks.

Neck: The Neck in YOLOv8 processes the multi-scale feature maps from the Backbone and prepares them for the Head, where object detection is finalized. Its main purpose is to fuse spatial and semantic information across different resolutions, enhancing the network’s ability to detect objects of varying sizes. Feature maps from the Backbone, such as 20×20, 40×40, and 80×80, are refined using a combination of Upsample, Concat, and C2f blocks. The Upsample block doubles the spatial dimensions of the input map (e.g.,40×40→80×80) without altering the number of channels, ensuring alignment for concatenation. The Concat block merges feature maps from different layers along the channel dimension, preserving fine-grained details and high-level features. The C2f block in the Neck differs from the Backbone, as its shortcut parameter is set to False, ensuring that all data passes through the bottleneck blocks for deeper processing. This enhances the complexity and richness of the feature maps before moving to the next layers. By alternating between Upsample, Concat, and C2f blocks, the Neck fuses multi-scale features effectively, enabling the network to capture details across various object sizes. The final refined feature maps are then passed to the Head for bounding box prediction and classification.

Head: The Head in YOLOv8 finalizes the object detection process by predicting bounding boxes and object classes. It consists of three Detect blocks, each specialized for detecting objects of different sizes, using multi-scale feature maps from the Backbone and Neck. The first Detect block processes the high-resolution feature map (80×80) from the Neck to focus on detecting small objects. The second Detect block uses the intermediate-resolution feature map (40×40) to identify medium-sized objects. The third Detect block operates on the low-resolution feature map (20×20) to detect large objects, leveraging the global context retained in these maps. Each Detect block employs YOLOv8’s Anchor-Free Detection mechanism, directly predicting the center, width, and height of objects within a grid cell, along with class probabilities. This approach eliminates the need for predefined anchor boxes, reducing complexity and improving detection efficiency across a range of object sizes. By specializing in multi-scale detection, the Head enhances YOLOv8's ability to identify objects with precision and efficiency, making it adaptable to diverse real-world scenarios.

YOLOv8 is a cutting-edge object detection model that combines speed, accuracy, and efficiency. Its architecture, comprising the Backbone, Neck, and Head, processes images efficiently, extracts multi-scale features, and detects objects across various sizes. The adoption of Anchor-Free Detection simplifies predictions and enhances performance, making YOLOv8 ideal for modern computer vision challenges [14].

**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]*

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

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)

**4.3 Research challenges**

* Endometriotic lesions can be different in size, shape, and location, making them difficult to detect and classify in MRI scans.
* Conditions like ovarian cysts and other medical conditions often mimic the imaging features of endometriosis, leading to potential misdiagnosis.
* Superficial peritoneal endometriosis is difficult to detect on MRI due to its subtle and indistinct imaging features.
* The lack of large, labeled MRI datasets specific to endometriosis.

## EXPECTED ACHIEVEMENTS

We expect to build a system capable of detecting and marking evidence of endometriosis in MRI scans. The goal of this project is to achieve high accuracy and, by doing so, eliminate the need for diagnostic surgery.

This system will use a DenseNet structure concatenated with YOLO for final classification.

The system will detect signs of endometriosis and outline the affected areas using boundary boxes.

To train the network, we will utilize a dataset of MRI scans from various modalities.

After training, the network will be able to analyze MRI scans and generate diagnostic predictions.

This research aims to identify the optimal combination of parameters by training the network with varying learning rates, batch sizes, and epochs.

This research will use transfer learning to train the model and achieve better results.

Another goal is to determine the most relevant MRI modality and agent, considering their resolution and accuracy for detecting endometriosis.

We aim for this system to achieve an accuracy rate above 0.8 in correctly diagnosing endometriosis. Diagnosing endometriosis through MRI scans is challenging due to the diverse types of endometriosis and the variations in size, shape, and location of endometriotic lesions.

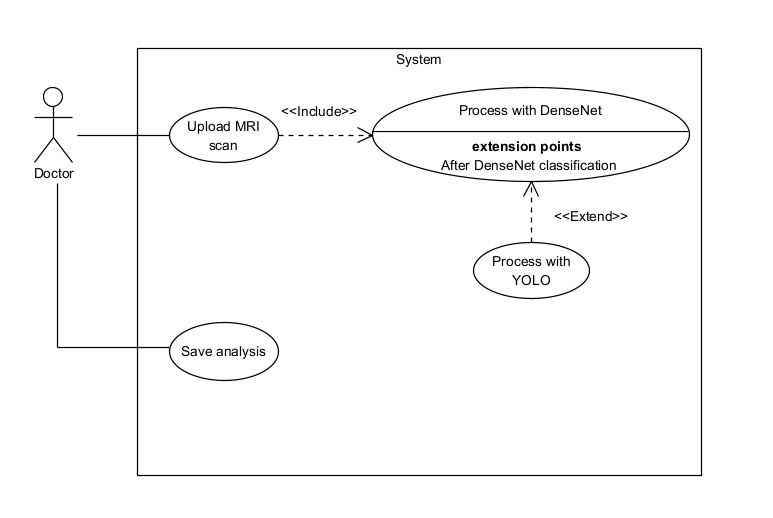
This system is designed to assist doctors in diagnosing endometriosis and tracking endometriotic structures. Currently, despite the advancements in deep learning in the medical field, there is no established model specifically for analyzing MRI scans related to endometriosis.

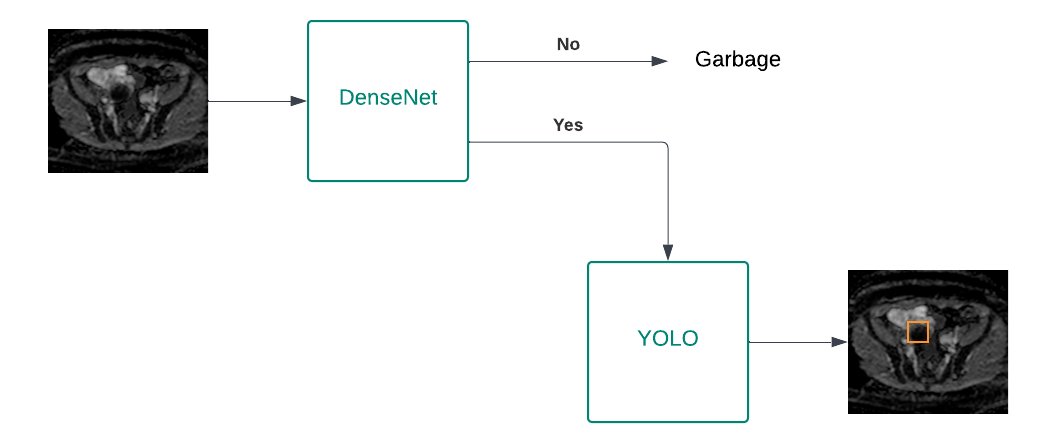
# EVALUATION AND VERIFICATION PLAN

|  |  |  |
| --- | --- | --- |
| **Case** | **Test description** | **Expected result** |
| 1 | Upload a file with an unsupported format | Throw error: "Unsupported file type. Please upload a .dcm file." |
| 2 | Upload the correct file type and press "Analyze" | The system identifies and segments endometriotic lesions if present. |
| 3 | Press "Save Analysis" after a successful analysis | Save the analysis results to a chosen directory and confirm saving. |
| 4 | Analyze an MRI scan with no endometriotic lesions | Display: "No evidence of endometriosis detected." |
| 5 | Analyze an MRI scan with endometriotic lesions | Highlight each lesion with a boundary box and display the total count. |
| 6 | Press back button after analysis | The system displays the main page |

1. PRODUCT

**Use case**

****

**Flow chart**

**GUI**

A person looking at a screen

Description automatically generated

A person looking at a screen

Description automatically generated

A screenshot of a medical scan

Description automatically generated

### REFERENCES

1. [Pathophysiology, diagnosis, and management of endometriosis](https://www.bmj.com/content/bmj/379/bmj-2022-070750.full.pdf)
2. [Frontiers | New Insights in Pathogenesis of Endometriosis](https://www.frontiersin.org/journals/medicine/articles/10.3389/fmed.2022.879015/full)
3. [Application of machine learning techniques in the diagnosis of endometriosis | BMC Women's Health](https://link.springer.com/article/10.1186/s12905-024-03334-2)
4. [The Main Theories on the Pathogenesis of Endometriosis - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC10001466/)
5. [Diagnosis of Endometriosis Based on Comorbidities: A Machine Learning Approach - PubMed](https://pubmed.ncbi.nlm.nih.gov/38002015/)
6. [Diagnostic efficacy of ultrasound combined with magnetic resonance imaging in diagnosis of deep pelvic endometriosis under deep learning | The Journal of Supercomputing](https://link.springer.com/article/10.1007/s11227-020-03535-0)
7. [The Radiology Assistant : Endometriosis - MRI detection](https://radiologyassistant.nl/abdomen/unsorted/endometriosis-mri-detection)
8. <https://rdcu.be/d3ion>
9. <https://doi.org/10.3390/jcm13195783>
10. [Magnetic Resonance Imaging (MRI)](https://www.nibib.nih.gov/science-education/science-topics/magnetic-resonance-imaging-mri)
11. [1608.06993](https://arxiv.org/pdf/1608.06993)
12. [GeeksforGeeks | A computer science portal for geeks](https://www.geeksforgeeks.org/)
13. [Transfer learning: a friendly introduction | Journal of Big Data | Full Text](https://journalofbigdata.springeropen.com/articles/10.1186/s40537-022-00652-w)
14. <https://abintimilsina.medium.com/yolov8-architecture-explained-a5e90a560ce5>
15. https://www.ultralytics.com/