International Journal of Computer Applications (0975 - 8887)
Volume 187 - No.132, August 2026

Risk Assessment System for Endometriosis in Medical Ultrasound Exams using Machine Learning

Luciano S. de Souza
Futuro Tech
Recife, PE - Brasil

Maria Clara F.G. Barretto
Universidade Federal de Pernambuco
Recife, PE - Brasil
Futuro Tech
Recife, PE - Brasil

Elliot Q.C. Garcia
Universidade Federal Rural de Pernambuco
Recife, PE - Brasil
Futuro Tech
Recife, PE - Brasil

Adriana F.G. Barretto
Futuro Tech
Recife, PE - Brasil

João B.C. Garcia
Futuro Tech
Recife, PE - Brasil

Tiago A.E. Ferreira
Universidade Federal Rural de Pernambuco
Recife, PE - Brasil

ABSTRACT

Endometriosis is a prevalent gynecological disease affecting millions of women worldwide, frequently causing chronic pelvic pain and infertility. Despite the widespread availability of pelvic ultrasound as a non-invasive diagnostic tool, its efficacy remains heavily dependent on examiner expertise, leading to notorious diagnostic delays. To address these limitations, this paper proposes a novel clinical decision support system based on serially organized machine learning models to automate and enhance endometriosis screening. Utilizing a real-world dataset of 7,020 medical exams (comprising 294,983 ultrasound images), it trained two complementary Multi-Layer Perceptron (MLP) expert models—one specialized in the positive class and the other in the negative class—integrated via an OR logic framework to strictly mitigate Type II errors (false negatives). Evaluated on a completely independent, out-of-sample test set of 6,080 exams, the proposed methodology achieved an overall Accuracy of 0.815 and an F1-score of 0.844. Crucially, the system reached a perfect Recall of 1.00, ensuring a complete absence of false negatives, which significantly outclasses existing deep learning baselines and traditional classifiers in the literature. These findings demonstrate that the proposed framework serves as a highly robust, low- and high-risk alert system that can optimize clinical monitoring strategies and enable timely therapeutic interventions.

General Terms

Pattern Recognition, Machine Learning, Medical diagnosis support.

Keywords

Endometriosis Diagnosis, Machine Learning, Medical Ultrasound, Model Combination, Decision Support System

1. INTRODUCTION

Endometriosis is a gynecological disease characterized by the presence of functional endometrial tissue outside the uterine cavity, primarily affecting the pelvic peritoneum and ovaries. Its prevalence reaches approximately 20% of women of reproductive age and up to 50% of infertile women, representing one of the main causes of chronic pelvic pain and female infertility~[14]. This illness significantly impacts the healthcare system and society, reducing women's quality of life and increasing economic costs [12, 5, 10].
There are several different medical exams to detect endometriosis, where the ultrasound is a cheap method that can provide a non-expensive clinical diagnosis. Although ultrasound is a widely used method for the clinical diagnosis of endometriosis, it remains challenging. This is due to the absence of specific biomarkers and the limitations of conventional imaging, the accuracy of which heavily depends on the examiner's expertise and the anatomical visibility of the lesions [3, 9]. In this context, machine learning has shown promise in supporting diagnostic imaging, especially in ultrasound. These approaches have improved tasks such as segmentation, classification, lesion detection, and assistance in clinical diagnosis [4, 15, 11].
In recent years, machine learning techniques have been applied to assist in medical diagnosis, including the analysis of ultrasound images. Hu, Ping, et al. in [6] propose a deep learning model applied to gynecological ultrasound images to differentiate endometrioma cysts from other pathologies, such as tubo-ovarian abscesses. The approach uses supervised learning with automatic feature extraction, reducing the dependence on the specialist's subjective interpretation.
Balica, Adrian, et al. [2] investigate the use of deep learning for the diagnosis of endometriosis from ultrasound images, and multiple architectures (such as ResNet, DenseNet, and EfficientNet) are trained for automatic classification of the disease. The study demonstrates that CNN-based models can serve as clinical decision

support systems, reducing diagnostic variability and potentially anticipating diagnosis, which often takes years using traditional methods.

Therefore, machine learning techniques have high potential to improve the diagnosis of endometriosis through imaging. This work proposes a tool based on serially organized machine learning models to assist in the diagnosis of endometriosis through the analysis of ultrasound images. The approach uses multiple models with complementary functions, allowing for improved disease detection capabilities and reduced classification errors. Thus, it seeks to offer a clinical decision support system, contributing to a more precise and earlier identification of endometriosis.

This article is organized as follows: Section 2 describes the methodology adopted; Section 3 shows and discusses the results obtained, and finally, Section 4 presents the conclusions of the study and perspectives for future work.

2. METHODS AND MATERIALS

This section describes the methodology proposed for the recognition problem. It presents the data preprocessing, the model architecture used, the training procedure, and the metrics adopted to evaluate performance.

2.1 Data collection

The dataset consists of a collection of real-world ultrasound medical exams of women. There are 7020 ultrasound exams from different regions of the female pelvis, where 3505 exams have a positive diagnosis of endometriosis and 3515 have a negative diagnosis. In this collection of ultrasound exams, there are $294,983$ ultrasound images, which averages out to approximately 42 images per exam. All diagnoses are reported by exam, not by image. Thus, it is assumed that the label of an image is exactly the same as the label given to the exam. This assumption is not entirely true, because the endometriosis signs may be present in just one image of the ultrasound exam. Given this assumption, an image can be mislabeled as positive (endometriosis), but the labels cannot be noisy if the diagnosis is negative (no endometriosis).

This dataset is divided into 150,996 positive images (diagnosed with endometriosis) and 143,897 negative images (no endometriosis diagnosis). The images are received in DICOM (Digital Imaging and Communications in Medicine [1]) data format, which is the international standard for handling, storing, and transmitting medical images (X-ray, CT scan, MRI).

2.2 Data Pre-processing

Figure 1 shows the image preprocessing workflow. The image preprocessing is divided into three parts. The first part is an initial image transformation. An initial verification is performed to adjust the format of the medical images. The raw images are in DICOM format. A conversion from 16-bit to 8-bit and an export to a PNG image with dimensions of 1232×854 pixels are performed. The images are then converted to grayscale with a single channel.

The second preprocessing phase is employed to safeguard the identity of patients. All personal information of patients contained in the image is deleted, and the image becomes completely anonymous. In addition, a 10% crop is applied to the borders of the images to remove any technical information. After that, the final step of the preprocessing is reached: image standardization. A new image resize is employed, reducing the images to dimensions of 992×688. With this new size, the images are normalized so that each pixel has an intensity between 0 and 1.

[1] https://dicom.nema.org/medical/dicom/current/output/chtml/part01/chapter_1.html#sect_1.1

Table 1. : Dataset distribution to train the model expert for the negative class.

Datasets	**Positive Images** (Endometriosis)	**Negative Images** (no Endometriosis)	**Subtotal**
Train	–	7.041	7.041
Validation	7317	–	7317
Test	2891	2488	5379
Total	10208	9529	19737

Table 2. : Dataset distribution to train the model expert for the positive class.

Datasets	**Positive Images** (Endometriosis)	**Negative Images** (no Endometriosis)	**Subtotal**
Train	9.671	–	9.171
Validation	–	9.049	9.549
Test	537	480	1.017
Total	10.208	9.529	19737

2.3 Dataset Division

Because it does not know how to accurately label images, it uses the ultrasound medical exam report to label all images from that exam, regardless of whether they contain endometriosis. To address this limitation, we are training two MLP models. The first model focuses on the positive class (endometriosis), and the other on the negative class. We create a pipeline with these two models to minimize the possibility of false negatives.

In this way, to train machine learning models to recognize endometriosis signs, the ultrasound image dataset was divided into two collections of three sets: training set, validation set, and test set. All datasets are totally disjoint; i.e., in addition to ensuring that no images are repeated in more than one set, it was also guaranteed that the images of a given patient are contained in only one set, preventing data leakage between sets.

Thus, it creates two distinct collections of datasets: one for the first model, which acts as an expert for the negative class, and another for the second model, an expert for the positive class. Table 1 presents the data distribution used to train the negative class expert model, while Table 2 shows the datasets for the positive class expert model. Notably, the remaining images form an out-of-sample dataset, which will also be used to evaluate the final performance of the proposed methodology.

Each of these image datasets created to train the models came from 470 ultrasound exams, totaling 940 exams. Since the complete dataset contains $7,020$ ultrasound exams, there are $6,080$ remaining exams. These form an out-of-sample dataset consisting of $3,036$ positive and $3,044$ negative cases, which will serve as a completely independent test set to evaluate the performance of the proposed classification system.

2.4 Deep Learning Method and Metrics

In this work, the models are based on deep neural networks, particularly Multi-Layer Perceptrons (MLPs). The MLP is used to estimate the presence or absence of endometriosis for a given ultrasound image. During the training process, the target for each image is set to 0 or 1, where 0 represents the negative class (absence of

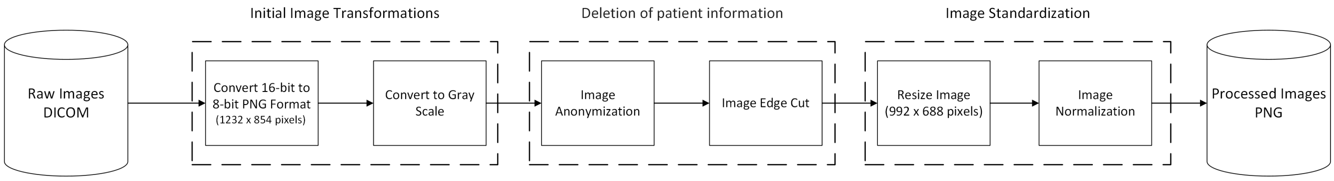


Fig. 1: Fluxogram for preprocessing ultrasound images.

endometriosis) and 1 represents the positive class (presence of endometriosis). However, after training, the MLP output can range between 0 and 1. The closer the output is to 0 or 1, the lower or higher the probability of endometriosis, respectively. Thus, the MLP output functions as a risk score, where 0 indicates the lowest risk and 1 the highest.

To model the relationship between the input features (ultrasound images) and the classes of interest (0 for non-endometriosis and 1 for endometriosis), these scores are subsequently used in a separability analysis. From the MLP output, a score distribution for each class can be constructed. Ideally, class 0 will exhibit a score distribution with an expected value near zero, whereas class 1 will have an expected value near one. In this ideal scenario, the distributions do not overlap. In practice, however, these distributions may show considerable dispersion, resulting in an overlap. To quantify this separability, the overlap metric, Equation 1, was used.

$$overlap = \sum_{j} min(f_E(x_j), f_N(x_j)) \tag{1}$$

which is calculated from the empirical distributions of the scores, estimated using histograms, and quantifies the area of intersection between these distributions, where $f_E(x_j)$ and $f_N(x_j)$ are the normalized frequency in bin j of the classes endometriosis and no endometriosis, respectively.

Based on the separability analysis, a decision threshold τ is calculated to define the classification boundary in the score space. If the MLP output is below τ, the image is classified as negative (no endometriosis); if it is above the threshold, the image is classified as positive. This decision threshold is empirically adjusted by evaluating the model's performance on an in-sample dataset. Specifically, the threshold τ is selected to maximize the classification accuracy:

$$\tau_i = \underset{\tau \in \tau'}{\text{arg max}} \ \text{Acc}(\tau)$$

where $Acc(\tau)$ represents the accuracy obtained by applying the threshold τ_i of model i to the classification of the ultrasound images on the validation set. Accuracy is defined as the ratio between the number of images correctly classified and the total number of images checked in the experiment. Its effectiveness depends directly on the degree of separation between the score distributions of classes 0 and 1. In addition to separability analysis and accuracy measures, traditional classifier evaluation metrics, such as precision, recall, and F1-score, were used to quantify the performance of MLP models in terms of predictive capacity.

Although the classification analysis is based on individual ultrasound images, a clinical diagnosis in practice relies on a complete set of images (i.e., a medical exam). Therefore, this work aims to develop a decision support system at the examination level. The system processes a group of images from a single exam. If at least one image is classified as positive, the patient is predicted to have endometriosis. Conversely, if no positive images are detected, the patient is classified as negative.

It is worth noting that the system is not flawless, and misclassifications can occur. In practice, the system is subject to two types of errors: Type I error (false positive) and Type II error (false negative). A false positive occurs when the system predicts that a healthy patient has endometriosis. Conversely, a false negative occurs when a patient who actually has the disease is predicted to be healthy.

In real life, the occurrence of a type II error is far more damaging than observing a type I error. If the system has a Type I error, the patient will be referred to a medical team for treatment and may be given a new analysis. However, if a false negative occurs, the patient will be discharged, and no treatment will be initiated, allowing the disease to progress.

The proposed system combines the two MLP expert models described to mitigate Type II errors (false negatives). Thus, given an ultrasound exam, both expert MLP models generate their respective outputs. The patient is classified as healthy only if both models indicate the negative class (i.e., they are below their respective thresholds). Otherwise, the patient is flagged as having endometriosis. In computational terms, the system applies an OR logic function to the MLP models' outputs. If the OR logic function is 0, then the patient is healthy. Otherwise, the patient has endometriosis. Figure 2 shows a scheme os the proposed process.

3. EXPERIMENTS AND RESULTS

This section presents and analyzes the results obtained from applying the proposed methodology. The performance of the models is analyzed in terms of separability between the score distributions, as well as traditional classification metrics such as accuracy, precision, recall, and F1-score. Additionally, the effect of the choice of decision threshold and serial classification strategy on the overall system performance is investigated.

Two classification models were built. One model was specialized with data from the negative class, and the other model with data from the positive class. Many initial experiments were executed to define a good architecture for the MLP models. In these experiments, the number of layers, the number of neurons per layer, the activation function, and the training algorithm were varied. After these initial experiments, the architecture for both models was defined as follows:

—9 Hidden Layers with 17062 neurons (an array of 4496×344 neurons), and an output layer with one neuron.

—Input and Hidden layers' Activation Function: $ReLu$

—Output Layer's Activation Function: $linear$

—Optimization Algorithm: Adam

—Learning rate: 0.0001

For both models, the number of epochs was set to 2700, intentionally reaching overfitting. Given that each model was trained exclu-

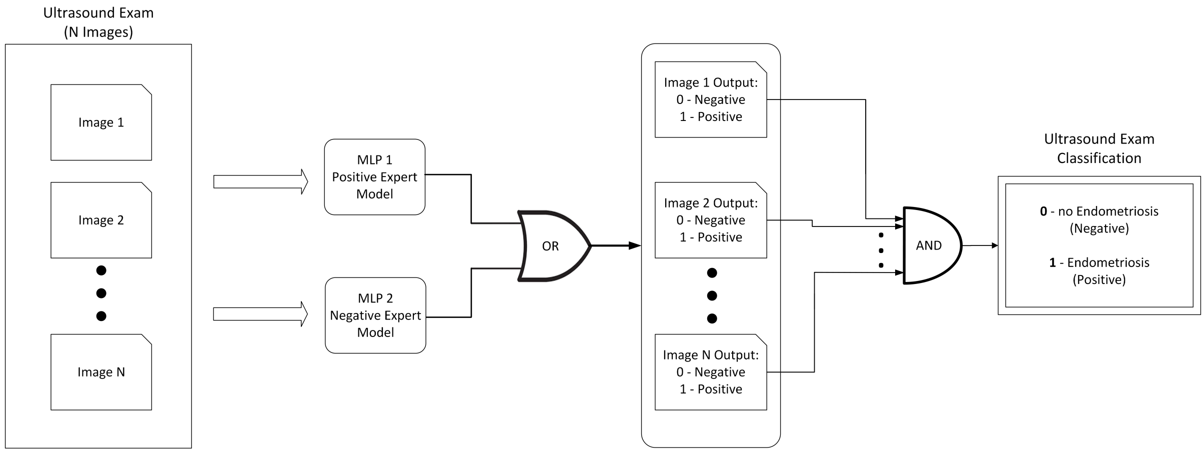
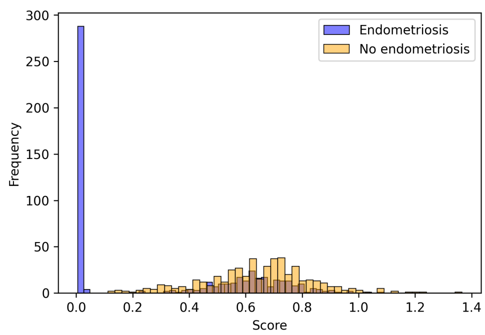


Fig. 2: Classification Ultrasound Exam Scheme.

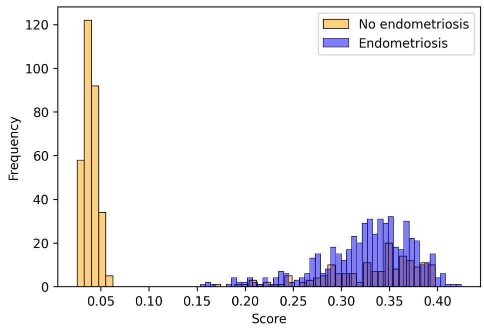
sively on data from a single class, they function as selectors that verify whether a given data point was observed during training. In other words, if data seen during the training process is presented to the model again, its output will tend toward the expected score of the overfitted class. Conversely, for unseen data, the generated score will deviate from that specialized class – a behavior that is further accentuated by the overfitting. For both models, in the training stage, the target class has a desired output equal to 0. Thus, for the Positive Expert Model (MLP1), the endometriosis class has an expected score of 0. For the Negative Expert Model (MLP2), the expected score of 0 is attributed to the healthy class.

Initially, the models are analyzed individually. Figure 3 shows the distribution of ultrasound image scores for each class, considering the two models used. The blue distribution corresponds to the target class, the endometriosis class to the MLP 1 model, and the no endometriosis class to the MLP 2 model. Figure 3a shows the distribution of scores generated by the first model (MLP 1). It can be observed that this model creates two score distributions, one for each possible class. The distribution of the target class (endometriosis) has a concentration close to zero and a spread of some class members with scores greater than zero. Concurrently, the distribution of the non-endometriosis class features high dispersion, with an expected score value between 0.6 and 0.7. Basically, the entire no endometriosis score distribution overlaps with the spread part of the endometriosis class. This behavior reflects the data mislabeling described in Section 2.1. However, as shown in Figure 3a, it is possible to notice a gap separating the concentration of scores close to zero in the endometriosis class from the lowest score in the no endometriosis class. In this way, it is possible to define the threshold $\tau_p = 0.0296$ (the threshold for the positive expert model). Therefore, any image that has a model score output less than τ_p will be classified as endometriosis. Otherwise, the image will be classified as no endometriosis.

Figure 3b shows a similar behavior for the MLP 2 model (the negative expert). Here, the no endometriosis class is also concentrated near a zero score but presents a greater dispersion when compared to the target class of the MLP 1 model. The endometriosis score distribution also exhibits a large dispersion, falling within the score range of $[0.15, 0.45]$. Once again, it is possible to note a gap between the two classes. By defining the threshold $\tau_n = 0.0658$ (the threshold for the negative expert model) and applying a logic con-



(a) MLP 1 – Positive Expert Model



(b) MLP 2 – Negative Expert Model

Fig. 3: Distribution of scores generated by the models used.

trary to that of the first model, any image with a score less than τ_n is classified as no endometriosis. Otherwise, it is classified as endometriosis.

Next, the models are analyzed jointly, and their classification outputs are standardized. Regardless of the model, if an image is classified as endometriosis, the information propagated to the OR logic gate is 1. If it is classified as no endometriosis, the propagated value is 0. Table 3 presents these classification rules (truth table). Notably, this classification logic is designed to minimize Type II errors

Table 3. : Standardization logic for image classification. MLP 1 and MLP 2 models propagate a 1 for positive (endometriosis) and a 0 for negative (no endometriosis) cases. The final decision is the logical OR of both model outputs.

MLP 1	MLP 2	OR Gate Output
0	0	0
0	1	1
1	0	1
1	1	1

(false negatives). In this context, classifying a healthy patient as ill is far less critical than classifying an ill patient as healthy. While a healthy patient misclassified as ill will be subsequently reviewed by a medical board, an ill patient misclassified as healthy would be sent home without treatment.

However, in a practical context, the objective is to classify the complete ultrasound exam rather than individual images. Therefore, the entire image set is presented to both models. Under this rule, an exam is classified as negative (no endometriosis) if and only if all its constituent images are classified as no endometriosis. Otherwise, if at least one image within the set is classified as positive, the exam is concluded to be positive (endometriosis). This rule is applied to minimize the type II error, and it is represented in Figure 2 with the AND logic gate.

To test the exam classification performance of the proposed system, the remaining dataset of $6,080$ ultrasound exams was employed. Figure 4 presents the confusion matrix of the proposed system for this out-of-sample test, where the values are normalized to sum to 1 (or 100%). In this test, the dataset comprises approximately 50% positive and 50% negative classes. The classification output achieved a Type II error of zero, meaning that all ultrasound exams labeled as endometriosis were correctly predicted. For the negative class, 62% of the non-endometriosis exams were predicted correctly (representing 31% of the full dataset), while 38% were incorrectly predicted as positive, resulting in a Type I error rate of 38% (or 19% misclassifications of the full dataset).

It is also possible to calculate other classification performance metrics, such as Accuracy, Precision, Recall, and F1-score, where values closer to 1 indicate better performance. Accuracy, the proportion of correct predictions, reached 0.815 (or 81.5%). Precision, which measures the accuracy of positive predictions, was 0.731 (or 73.1%). Recall measures the model's ability to detect all positive cases within the dataset. Here, the system achieved a Recall of 1.00 (or 100%), reinforcing that the Type II error is zero. Finally, the F1-score, which evaluates the classification performance by calculating the harmonic mean of precision and recall, reached 0.844, reflecting a high balance.

An important objective is to answer the question: how can it verify the performance of the proposed classification system in a practical sense? To achieve this, the results obtained must be compared with other works in the literature under similar conditions. Unfortunately, a direct performance comparison is not possible, since the database used here differs from those in previous studies. However, an indirect comparison can be made to provide a qualitative perspective on how the proposed methodology performs relative to others. In this context, the study by Balica et al. [2] presented the performance of various models in classifying ultrasound images with or without endometriosis. Table 4 presents a comparison between the proposed classification system and various deep architectures analyzed by Balica et al. [2]. It can be observed that the proposed method obtained the highest efficiency overall for all

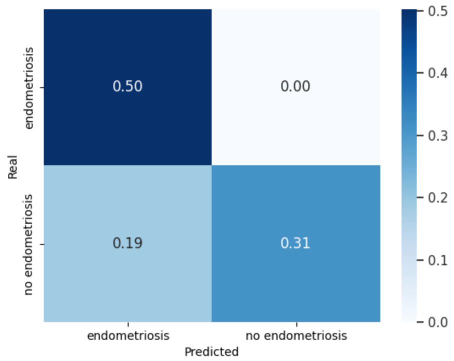


Fig. 4: Confusion Matrix of the proposed classification system. The information is given in a proportional rate, where the line information is the truth and the column information is the prediction of the proposed classification system.

Table 4. : Accuracy, Precision, Recall and F1-score to compare between the proposed method and various deep architectures [2] to classify ultrasound images in a qualitative sense.

Model	Acuracy	Precision	Recall	F1-score
Xception	0.78	0.78	0.78	0.78
InceptionResnet	0.79	0.79	0.79	0.79
ResNet50	0.78	0.80	0.78	0.76
EfficientnetB2	0.79	0.81	0.79	0.78
Densenet121	0.80	0.81	0.80	0.79
Proposed	**0.815**	**0.731**	**1.00**	**0.844**

measures of accuracy, precision (0.815), recall (1.00), and the F1-score (0.844), surpassing models with recognized applicability in the literature for health applications and disease recognition, such as DenseNet121 [1, 13] and EfficientNetB2 [7, 8]

In a broader comparison, the work of Santos et al. [14] used data from 298 women with severe dysmenorrhea and persistent acyclic pelvic pain, whose imaging examinations showed no significant abnormal findings. A systematic analysis with various machine learning methods was employed to classify images with and without endometriosis. Table 5 presents the results achieved by [14] and the currently proposed system. Once again, the qualitative comparison demonstrates the superior performance of the proposed method across all evaluation metrics.

With the proposed method, a robust decision support system for endometriosis diagnosis can be established. The proposed system achieved a Type II error equal to zero (or a Recall of 100%), indicating a very low risk for clinical applications. In this initial version, given an ultrasound exam, the system provides a binary output indicating a positive (endometriosis) or a negative (no endometriosis) exam, the latter with very high confidence. Thus, the proposed method can effectively stratify exams into low-risk and high-risk indications. Therefore, the proposed method functions as an alert system, highlighting ultrasound cases where human analysis needs to be more thorough.

4. CONCLUSIONS

The approach presented in this work is particularly relevant in medical applications, where the failure to detect positive cases can lead

Table 5. : Accuracy, Precision, Recall, and F1-score to compare the proposed method and various Machine Learning Methods [14] to classify images before the patient underwent laparoscopy, in a qualitative sense.

Classifiers	Performance Measurements (%)			
	Accuracy	Precision	Recall	F1-score
Soft Voting Ensemble[1]	59.0	54.8	73.9	64.8
Support Vector Machine	66.4	68.0	58.6	63.4
Naive Bayes	49.3	47.9	79.3	59.7
Multilayer Perceptron	61.6	60.7	58.6	59.6
Extremely Randomised Trees	58.2	57.1	55.2	56.1
Logistic Regression	59.7	60.0	51.7	55.6
Stacking Ensemble[2]	58.1	57.7	51.17	54.5
Categorical Boosting	58.1	57.7	51.7	54.5
Adaptive Boosting	55.0	53.3	55.2	54.2
Extreme Gradient Boosting	53.4	51.6	55.2	53.3
Light Gradient Boosting	53.4	51.6	55.2	53.3
Random Forest	54.7	54.2	44.8	49.1
Proposed System	**81.5**	**73.1**	**1.00**	**84.4**

[1] Extreme Gradient Boosting plus Naive Bayes.
[2] Categorical Boosting plus Naive Bayes with Logistic Regression as the final estimator.

to more serious consequences than false alarms. The results point to the potential of the technique in supporting the early diagnosis of endometriosis, reducing subjectivity in medical interpretation, and enabling automated screening systems based on machine learning.
The findings demonstrate that the proposed method achieved competitive performance compared to the baseline deep architectures in terms of accuracy (0.815) and F1-score (0.844), highlighting its effectiveness in diagnosing endometriosis from ultrasound images. The main contribution is achieving a recall of 1.00, implying a complete absence of false negatives, which contrasts with comparative models in the literature. This outcome is particularly vital in the clinical context, since misidentifying the disease can lead to significant treatment delays.
The obtained results demonstrate that the proposed approach is capable of identifying relevant patterns associated with endometriosis, showing consistent performance in both separability analysis and classification metrics. This strategy proved highly effective in exploiting the complementarity between the expert models. Furthermore, the risk stratification framework demonstrated significant potential for accurately identifying disease-free cases. Consequently, the model not only contributes to automated diagnosis but also offers valuable support for the long-term monitoring and early detection of the disease.
Given the definitions of the MLP output scores and the capability to calibrate the classification process by adjusting the thresholds (τ_p and τ_n), the distance between the model's output and these thresholds could serve as a metric to quantify anomalies detected by the neural network. This interesting research direction requires further investigation and will be addressed in future work.
The proposed method achieves a Type II error equal to zero, whereas the Type I error rate remains greater than zero. Conventionally, a Type I error (false positive) indicates that the model misclassifies negative cases as positive. However, preliminary follow-up observations of these misclassified cases revealed that several patients were subsequently diagnosed with endometriosis in follow-up exams one to two years later. These findings raise the hypothesis that the model may identify subtle patterns that precede the clinical onset of the disease. While this hypothesis requires further long-term study and validation, these initial results suggest that the method may offer a capacity for the early recognition of endometriosis. Such diagnostic anticipation is particularly relevant given the notorious diagnostic delays associated with endometriosis, as it could enable more effective monitoring strategies and early therapeutic interventions.
As future work, the intention is also to enhance the proposed architecture by including new models, with the goal of improving the treatment of false positives.

5. REFERENCES

[1] T. S. Arulananth, S. Wilson Prakash, Ramesh Kumar Ayyasamy, V. P. Kavitha, P. G. Kuppusamy, and P. Chinnasamy. Classification of paediatric pneumonia using modified densenet-121 deep-learning model. *IEEE Access*, 12:35716–35727, 2024.

[2] Adrian Balica, Jennifer Dai, Kayla Piiwaa, Xiao Qi, Ashlee N Green, Nancy Phillips, Susan Egan, and Ilker Hacihaliloglu. Augmenting endometriosis analysis from ultrasound data using deep learning. In *Medical Imaging 2023: Ultrasonic Imaging and Tomography*, volume 12470, pages 118–123. SPIE, 2023.

[3] Marc Bazot, Clarisse Lafont, Roman Rouzier, Gilles Roseau, Isabelle Thomassin-Naggara, and Emile Daraï. Diagnostic accuracy of physical examination, transvaginal sonography, rectal endoscopic sonography, and magnetic resonance imaging to diagnose deep infiltrating endometriosis. *Fertility and sterility*, 92(6):1825–1833, 2009.

[4] Elizabeth V Craig, LeAnn M Shannon, and Rochelle F Andreotti. The complementary role of ultrasound and magnetic resonance imaging in the evaluation of endometriosis: a review. *Ultrasound Quarterly*, 36(2):123–132, 2020.

[5] L. Della Corte, C. Di Filippo, O. Gabrielli, S. Reppuccia, V. L. La Rosa, R. Ragusa, M. Fichera, E. Commodari, G. Bifulco, and P. Giampaolino. The burden of endometriosis on women's lifespan: A narrative overview on quality of life and psychosocial wellbeing. *Int J Environ Res Public Health*, 17(13):4683, 2020.

[6] Ping Hu, Yanjuan Gao, Yiqian Zhang, and Kui Sun. Ultrasound image-based deep learning to differentiate tubal-ovarian abscess from ovarian endometriosis cyst. *Frontiers in physiology*, 14:1101810, 2023.

[7] Shikha Khullar, Vignesh Ramamoorthy. H, Sreedhar Kollem, Selvendran Ganesan, S. Kaliappan, and Indla Obulesu. Deepthyroscan:efficientnetb2 architecture for automated thyroid diagnosis. In *2025 IEEE 3rd Global Conference on Wireless Computing and Networking (GCWCN)*, pages 1–6, 2025.

[8] Jayashri Kolekar, Chhaya Pawar, Amol Pande, and Chandrashekhar Raut. Deep learning based ovarian cancer classification using efficientnetb2 with attention mechanism. *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, 8(1):32–52, Nov. 2025.

[9] Leandro A Mattos, Manoel O Goncalves, Marina P Andres, Scott W Young, Myra Feldman, Mauricio S Abrão, and Rosanne M Kho. Structured ultrasound and magnetic resonance imaging reports for patients with suspected endometriosis: guide for imagers and clinicians, 2019.

[10] S. A. Missmer, F. F. Tu, S. K. Agarwal, C. Chapron, A. M. Soliman, S. Chiuve, S. Eichner, I. Flores-Caldera, A. W. Horne, A. B. Kimball, M. R. Laufer, N. Leyland, S. S. Singh, H. S. Taylor, and S. As-Sanie. Impact of endometriosis on life-course potential: A narrative review. *Int J Gen Med*, 14:9–25, 2021.

[11] Sneha Mittal, Angela Tong, Scott Young, and Priyanka Jha. Artificial intelligence applications in endometriosis imaging. *Abdominal Radiology*, 50(10):4901–4913, 2025.

[12] Kelechi E. Nnoaham, Lone Hummelshoj, Premila Webster, Thomas d'Hooghe, Fiorenzo de Cicco Nardone, Carlo de Cicco Nardone, Crispin Jenkinson, Stephen H. Kennedy, and Krina T. Zondervan. Impact of endometriosis on quality of life and work productivity: a multicenter study across ten countries. *Fertility and Sterility*, 96(2):366–373.e8, 2011.

[13] Asif Raza, Mohammed S. Alshehri, Sultan Almakdi, Ali Akbar Siddique, Mohammad Alsulami, and Majed Alhaisoni. Enhancing brain tumor classification with transfer learning: Leveraging densenet121 for accurate and efficient detection. *International Journal of Imaging Systems and Technology*, 34(1):e22957, 2024.

[14] Letícia Santos, Mateus de Azevedo, Lia Shimamura, Antonio Nogueira, Francisco Candido dos Reis, Eduardo Schor, Julio Rosa e Silva, Daniel Tiezzi, and Omero Poli-Neto. Machine learning as a clinical decision support tool for diagnosing superficial peritoneal endometriosis in women with dysmenorrhea and acyclic pelvic pain. *Medical Research Archives*, 12(12), 2024.

[15] Niranjan Yadav, Rajeshwar Dass, and Jitendra Virmani. A systematic review of machine learning based thyroid tumor characterisation using ultrasonographic images. *Journal of Ultrasound*, 27(2):209–224, 2024.