Multi-Class Ovarian Cancer Subtype Classification in Histopathological Images Using Swin Transformer

Juliany P. Costa[1], **Igor R. B. Estrela**[1], **Antônio O. de Carvalho Filho**[2],
Anselmo C. Paiva[1], **Aristófanes C. Silva**[1], **João O. B. Diniz**[13]

[1]Núcleo de Computação Aplicada (NCA)
Universidade Federal do Maranhão (UFMA)
São Luís – MA – Brasil

[2]Programa de Pós-Graduação em Engenharia Elétrica
Universidade Federal do Piauí (UFPI)
Picos – Piauí – Brasil

[3]Fábrica de Inovação
Instituto Federal do Maranhão (IFMA)
Grajaú – MA – Brasil

`juliany@nca.ufma.br, igor@nca.ufma.br, antoniooseas@ufpi.edu.br,`

`paiva@nca.ufma.br, ari@nca.ufma.br, joao.bandeira@ifma.edu.br`

Abstract. *Ovarian cancer has a high mortality rate due to late diagnosis, which drives ongoing research to improve histopathological image analysis through artificial intelligence (AI). This study proposes a method for the automatic classification of four ovarian cancer subtypes using the Ovarian Cancer and Subtypes Histopathology Dataset, incorporating image processing techniques, data augmentation, and deep learning. The best performance was achieved with the Swin Transformer architecture, reaching an average accuracy of 98%. These results suggest that the proposed method is a promising approach for classifying ovarian cancer subtypes in histopathological images.*

1. Introduction

In recent decades, society has undergone a significant transformation driven by technological advances, especially in the field of AI. This evolution has expanded the range of possible applications across multiple domains, including medicine, which has been benefiting from automated solutions for more accurate and efficient diagnoses, prognoses, and treatments. [Werner 2019]

Among the most promising emerging applications in the healthcare field is the use of deep learning (DL) for interpreting medical images, such as X-rays, magnetic resonance imaging (MRI), and histopathological slides. These images are commonly used in the diagnosis of various diseases, including cancer. AI-based methods have demonstrated the ability to identify subtle patterns that assist healthcare professionals in the early and accurate detection of pathologies. [Vedana et al. 2024]

Ovarian cancer is a neoplasm that is difficult to detect at an early stage and has a high mortality rate. According to Instituto Nacional de Câncer (INCA), it is one of the most aggressive types of cancer and is often diagnosed only in advanced stages, when the chances of cure are significantly reduced. Among gynecological cancers, it ranks second in incidence, while globally, it is among the 7th to 8th most common cancers in women. [INCA 2023]

In addition to traditional convolutional neural networks (CNNs), more recent architectures such as Transformers have gained prominence due to their ability to capture complex patterns. The Swin Transformer, for instance, employs sliding windows and hierarchical attention mechanisms, enabling efficient modeling of spatial dependencies at multiple scales. [Liu et al. 2021]

Considering these challenges, the main objective of this study is to propose a computational method for the automatic classification of ovarian cancer subtypes in histopathological images, using image processing techniques and deep learning. The automatic classification of these subtypes is essential for early diagnosis, significantly increasing patients' chances of survival and cure. The main contributions of this work include:

- Enhancement of the intrinsic features of the images through image processing techniques;
- Use of the Swin Transformer architecture for cancer subtype classification, achieving results that surpass those reported in the literature.

The remainder of this paper is organized as follows: Section 2 presents the related work; Section 3 describes the materials and methods used; Section 4 discusses the results obtained; and finally, Section 5 presents the conclusions and suggestions for future work.

2. Related Work

Recent studies on the application of deep learning (DL) to the classification of epithelial ovarian cancer subtypes were analyzed. Most of the reviewed works employ CNN-based architectures, although some more recent studies also explore the use of Transformer-based models, as presented below:

Based on a modified version of the AlexNet architecture developed by the authors themselves, [Kasture et al. 2021] demonstrated that the application of data augmentation techniques improved accuracy from 70% to 83,93%. The study underscored the effectiveness of this approach in enhancing the generalization capability of deep learning models for multi-class classification involving four tumor subtypes and one non-cancerous class.

In the method proposed by [Kussaibi et al. 2024] An AI pipeline was developed to classify the main subtypes of epithelial ovarian cancer using slide images. ResNet-50 was employed to extract features, which were subsequently processed by two classifiers (CNN and LightGBM), both achieving a precision of 97%. The inclusion of a class-weighted loss function contributed to a better differentiation among the subtypes.

To evaluate the performance of different deep learning architectures in subtype classification, [Radhakrishnan et al. 2024] compared seven models, with InceptionV3 standing out by achieving an accuracy of 97,96%. In addition, they applied explainable artificial intelligence (XAI) techniques, such as Grad-CAM, saliency maps, Integrated Gradients, and DeepLift, to identify the image regions that influenced the model's decisions and to enhance its interpretability.

In turn, [Breen et al. 2024], compared 14 pretrained models, including CNNs and Transformers, for the classification of ovarian cancer subtypes. Transformer-based models stood out, achieving an accuracy of up to 97%. In a subsequent study, [Breen et al. 2025] focused on reducing the computational cost of these models. Despite the resource savings, using only 25% of the computational capacity, they managed to maintain the same performance, making the solution more feasible for clinical application.

The cited studies employed various techniques to automate the classification of epithelial ovarian cancer, including data augmentation and comparisons between CNN and Transformer architectures, approaches also explored in this work, as well as hyperparameter tuning, which was conducted in only a few of the analyzed studies.

It is worth noting that most of the reviewed studies did not apply specific filters or preprocessing techniques beyond basic color normalization. In addition, while many works relied on larger datasets, Kasture et al. (2021) used the same dataset adopted in the present research. None of the studies, however, explored preprocessing methods explicitly designed to enhance morphological structures, which could support a more accurate differentiation between tumor subtypes.

Relevant challenges therefore remain, such as distinguishing subtypes with overlapping characteristics and ensuring that models generalize well when trained on limited data. In this scenario, architectures like the Swin Transformer stand out for their ability to capture complex structural patterns and, when combined with tailored image processing techniques, hold strong potential to advance histopathological image analysis.

3. Materials and Method

The method developed in this study aims to improve precision in the classification of histopathological images of epithelial ovarian cancer subtypes, enabling the identification of complex tissue patterns and enhancing the quality of the results obtained.

The process began with image pre-processing to enhance texture-related features. Next, online data augmentation was applied to increase dataset diversity. The data were then split into folds for training, and classification was performed using a neural network model. Finally, model performance was evaluated using metrics such as Accuracy, Precision, F1-Score, and Recall. Figure 1 illustrates these steps, which are detailed below.

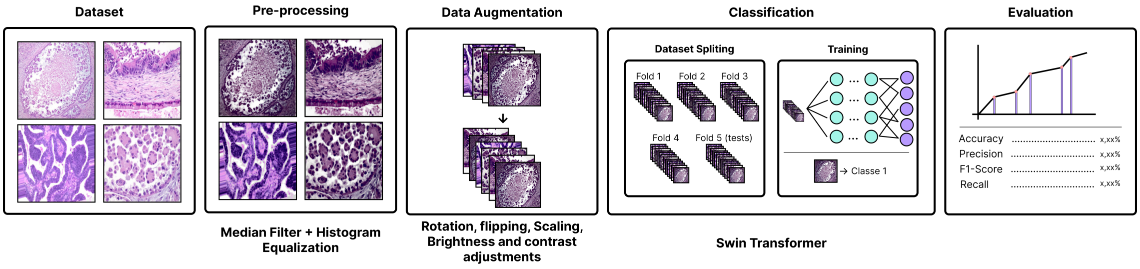

Figure 1. Main Steps of the Proposed Method.

3.1. Dataset

The dataset used in this study was the Ovarian Cancer and Subtypes Dataset Histopathology [Kasture et al. 2021], a publicly available dataset comprising 500 histopathological images of ovarian cancer, categorized into five subtypes: Clear Cell, Endometrioid, Mucinous, Serous, and Non-Cancerous. The classes are balanced, with 100 samples per category.

The images are available in various resolutions, ranging from 225 to 888 pixels in height and 225 to 550 pixels in width, and are presented in RGB color format. For method validation and input standardization for CNNs and Transformers, all images were resized to 224×224. Figure 2 shows examples of the classes included in the dataset.

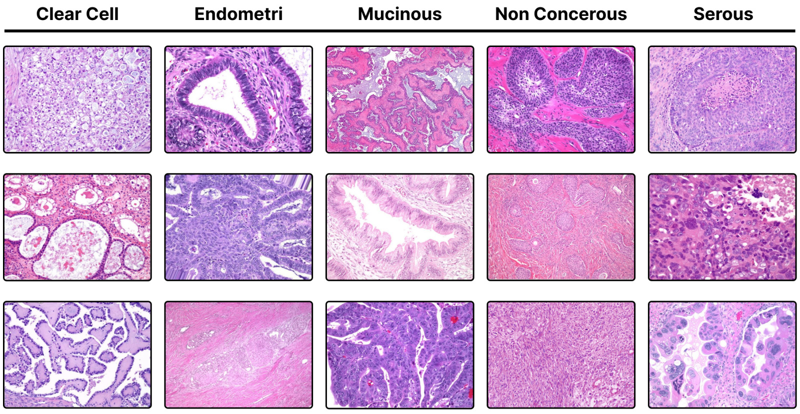

Figure 2. Sample Images from the Dataset.

3.2. Preprocessing

The preprocessing of histopathological images aimed to standardize intensity distribution and reduce interference that could compromise the extraction of relevant features by neural networks. To this end, median filtering and histogram equalization techniques were applied, respectively. Combined, these techniques aimed to adjust image contrast and highlight cancerous cellular structures, which often resemble the surrounding tissue. This procedure also sought to mitigate chromatic variations resulting from different staining protocols used in histological slides.

In the application of histogram equalization, the LAB color space was used, with the equalization applied specifically to the L (luminance) channel. This approach aimed to prevent abrupt color shifts and to facilitate the differentiation between regions and subtypes. According to [Gonzalez and Woods 2018], this technique redistributes gray levels, enhancing structural details of the cells.

However, when applied in isolation, it may intensify image noise, hindering morphological analysis. To mitigate this effect, median filtering was applied first, as it removes noise while preserving cellular edges and contours. As described by [Gonzalez and Woods 2018], this method avoids excessive smoothing of morphological structures, distinguishing itself from linear filters, which may compromise the distinction between cellular components.

Figure 3 presents the transformations applied to the images in the dataset.

3.3. Data Augmentation

As highlighted in Section 3.1, each class contains only 100 samples, which may lead to overfitting or ineffective training of deep learning architectures. To mitigate this issue,

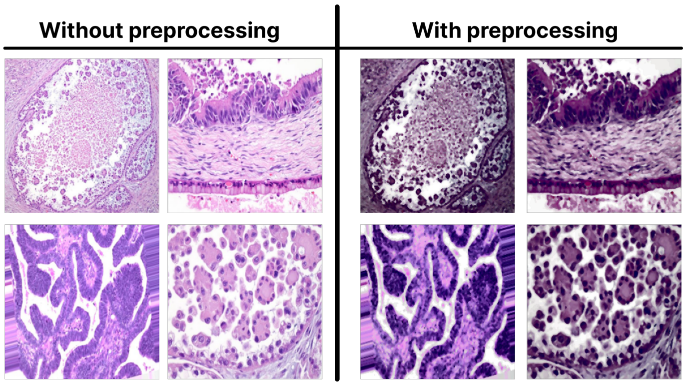

Figure 3. Example of the Applied Preprocessing.

data augmentation was proposed, applying subtle modifications to the images in order to increase both the variability and the number of samples in the dataset. The following transformations were employed:

- **Random rotation:** simulates different orientations of histopathological slides, making the model more robust to variations in tissue positioning.
- **Horizontal and vertical flipping:** introduces structural variations, enabling the model to learn symmetric patterns in different directions.
- **Scaling variations:** resizes images to expose the model to different magnification levels, facilitating the identification of textural features at varying resolutions.
- **Brightness and contrast adjustments:** modifies illumination levels to simulate different image acquisition conditions.

It is worth noting that data augmentation was used exclusively during the training phase, with the goal of increasing dataset diversity, reducing the risk of overfitting, and improving the model's generalization capability. This technique generated new synthetic samples by applying transformations to the original images, allowing the model to learn natural variations in histological patterns without the need to manually expand the dataset. However, during the testing phase, the images remained unaltered, ensuring that the model's performance was evaluated solely on real, unmodified data.

3.4. Classification

After data augmentation, different architectures were trained to classify the ovarian cancer subtypes. However, it is first necessary to define the dataset splitting strategy for training. In this work, cross-validation was used, as described below.

3.4.1. Dataset Splitting

To ensure reliable evaluation and reduce bias during the training phase, the k-fold cross-validation technique was adopted, with k=5. he data were divided as follows:

- Four folds were used for training the model with data augmentation, enabling it to learn relevant patterns from the available samples.
- One fold was reserved for testing, allowing an unbiased evaluation of the model's performance on unseen data.

In addition to reducing the risk of overfitting, this approach allowed the model to be trained and tested on different subdivisions of the dataset, ensuring better generalization and making it more robust for classifying new samples. As a result, the technique ensured that the outcomes were more accurate and representative, realistically reflecting the model's effectiveness in a clinical setting.

After the data split, different deep neural network architectures were trained.

3.4.2. Training

For the classification stage, several deep neural network architectures were tested, all widely used in medical image analysis and classification. Each model presents specific characteristics that influence its performance in extracting relevant patterns from histopathological images. The networks used are described below:

- **Swin Transformer:** an architecture based on the self-attention mechanism, designed for efficient image processing. Unlike traditional Transformers, the Swin Transformer employs hierarchical sliding windows to capture long-range spatial relationships more efficiently [Liu et al. 2021]. Its structure allows modeling of global dependencies while maintaining reduced computational cost.
- **VGG16:** a CNN composed of 16 layers, known for its sequential architecture and structural simplicity [Simonyan and Zisserman 2015]. It is characterized by the use of small filters (3×3) and a large number of trainable parameters, enabling the extraction of detailed features.
- **ResNet50:** a model based on residual networks, designed to mitigate the vanishing gradient problem in deep architectures [He et al. 2016]. Its main innovation lies in residual blocks, which use skip connections to facilitate gradient flow during training. This approach enhances the learning capacity of deep networks, making ResNet-50 one of the most widely used architectures in medical image classification tasks.
- **DenseNet121:** an architecture that connects all layers in a dense manner, where each layer receives as input the feature maps of all preceding layers [Huang et al. 2017]. This strategy reduces the number of trainable parameters and improves feature reuse throughout the network.

The selected networks were chosen based on their documented performance in the literature, demonstrating consistent and satisfactory results in similar tasks. At the end of the training process, the models were compared across different scenarios, with and without image preprocessing. This allowed for the assessment of preprocessing impact and the selection of the most effective approach for automated ovarian cancer detection in histopathological images.

3.5. Evaluation

The performance of the models was evaluated using the metrics accuracy, precision, recall, and F1-score, which are widely used in the literature to measure the effectiveness of machine learning algorithms. These metrics are essential for assessing the model's ability to correctly identify the classes, balancing false positives and false negatives, an aspect that is particularly relevant in medical applications. [Vilela et al. 2022]

4. Results and Discussion

This section presents the experiments conducted to validate the proposed method. First, the training environment is described, followed by an analysis of the experiments carried out to assess the influence of preprocessing on the performance of deep learning models. Finally, the results obtained are compared with those reported in the literature, highlighting the contributions of this study and the relevance of its findings.

4.1. Development Environment

The experiments were conducted in a computational environment configured on a computer with an Intel® Core™ i7-9700K processor, 16GB of RAM, and an NVIDIA GeForce GTX 1650 Super GPU. The model development was carried out using Python 3.8 and libraries such as TensorFlow, PyTorch, OpenCV, and Scikit-Learn. All networks were trained for 20 epochs with batches of 32 images, using the Categorical Cross-Entropy Loss function optimized by the Adam algorithm, with a learning rate of 0.0001.

4.2. Experiments

This section highlights the experiments conducted to validate the proposed method.

4.2.1. Experiment 1: Without Preprocessing

The first experiment evaluated the performance of the neural networks without applying preprocessing techniques, allowing for the analysis of each model's capability to classify ovarian cancer subtypes using raw images. The results obtained are presented in Table 1.

Table 1. Results of the Networks Without Preprocessing

Architecture	Accuracy	Precision	F1-Score	Recall
Swin Transformer	87,00%	87,93%	87,00%	86,66%
VGG16	78,00%	79,17%	78,00%	77,36%
ResNet50	65,00%	66,91%	65,00%	64,39%
DenseNet121	57,00%	71,20%	57,00%	52,80%

The results indicate that, without preprocessing, the Swin Transformer achieved the highest accuracy, reaching 87%, followed by VGG16 with 78%. ResNet50 and DenseNet121 showed lower performance, with accuracies of 65% and 57%, respectively. The variation in performance suggests that neural networks with different architectures have varying capabilities in extracting relevant features from raw images. These findings indicate that the Swin Transformer is an effective option for classifying ovarian cancer subtypes in a small-scale dataset such as the one used in this study.

4.2.2. Experiment 2: With Preprocessing

In the second experiment, a preprocessing step was applied with the aim of enhancing the morphological structures in the images prior to model training. The filters described in Section 3.2 were used. Table 2 presents the results obtained after image preprocessing.

Table 2. Results of the Networks with Median Filtering and Histogram Equalization

Architecture	Accuracy	Precision	F1-Score	Recall
Swin Transformer	98,00%	98,10%	98,00%	98,00%
VGGNet 16	83,00%	84,62%	83,00%	82,75%
ResNet 50	69,00%	71,41%	69,00%	64,74%
DenseNet 121	71,00%	76,41%	71,00%	71,56%

After applying preprocessing, all networks showed improvements in performance metrics. The Swin Transformer achieved an accuracy of 98%, indicating that the preprocessing step contributed to better extraction of relevant image patterns. VGG16 also improved, increasing from 78% to 83%. ResNet50 and DenseNet121, which had the lowest accuracies in the first experiment, reached 69% and 71%, respectively.

The comparative analysis between experiments suggests that preprocessing helped reduce the impact of noise and color variation in the images, enabling the neural networks to more accurately identify the morphological structures relevant to classification. The performance of DenseNet121, for example, improved significantly after the application of filters, surpassing ResNet50 in tumor subtype classification. This indicates that models which initially struggled to extract features can benefit from techniques that enhance image quality prior to the training phase.

4.3. Confusion Matrix Analysis

To analyze the results, a confusion matrix was generated for a single fold using the best-performing model with preprocessing, the Swin Transformer. Figure 4 presents the result.

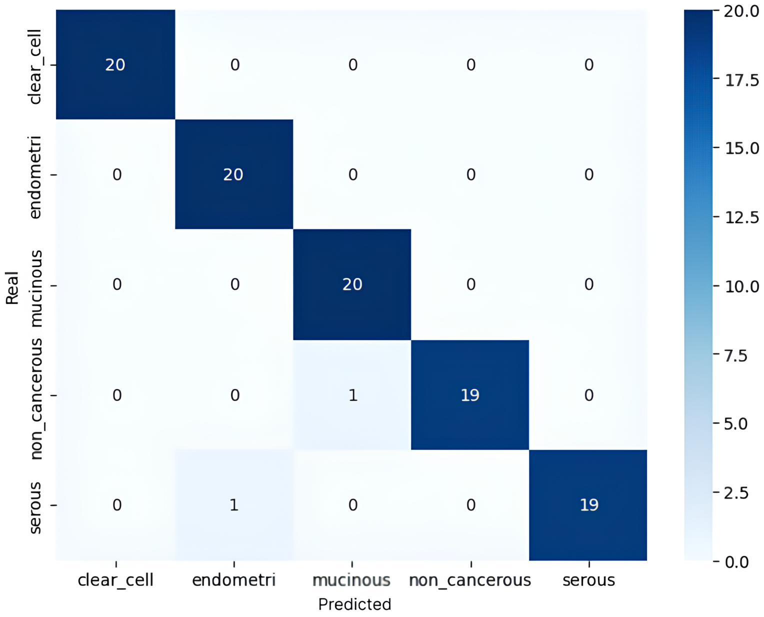

Figure 4. Confusion Matrix of the Swin Transformer

The confusion matrix of the Swin Transformer model demonstrates excellent performance in the classification of histopathological images, with high precision across all classes. The model achieved 100% accuracy in classifying the clear cell, endometrioid, and mucinous classes. However, minor misclassifications occurred in the non-cancerous and serous classes: one non-cancerous sample was incorrectly classified as mucinous, and one serous sample was classified as endometrioid. These errors suggest slight overlap in the features extracted from these classes, possibly due to morphological similarities in the images. Despite these small mistakes, the model showed highly satisfactory performance, with an overall accuracy close to 98%. This result suggests that the Swin Transformer is capable of capturing relevant patterns for class differentiation.

4.4. Comparison with Related Work

This section presents the results achieved in comparison with the studies discussed in Section 2. Table 3 provides a summary of this comparison.

Table 3. Comparison of Deep Learning Models for the Classification of Epithelial Ovarian Cancer Subtypes

Author	Deep Learning Model	Accuracy
[Kasture et al. 2021]	AlexNet	83.93%
[Kussaibi et al. 2024]	ResNet50	97%
[Radhakrishnan et al. 2024]	InceptionV3	97.96%
[Breen et al. 2024]	Transformer	97%
[Breen et al. 2025]	Transformer	97%
Proposed Method	Preprocessing + Swin Transformer	**98%**

The results of this study demonstrate superior performance compared to the approaches discussed in Section 2. The main distinguishing factor was the image preprocessing step, which had a significant impact on the accuracy of the tested models.

The experiments showed that preprocessing enhanced the accuracy of the models. Moreover, while some studies rely on large datasets to avoid overfitting, the results indicate that efficient preprocessing enables high performance even with limited datasets.

5. Conclusion and Future Work

This study presented a method for the automatic classification of ovarian cancer subtypes in histopathological images, combining image preprocessing with the Swin Transformer architecture. The proposed approach resulted in an efficient model, achieving 98% accuracy, 98% precision, 98% F1-score, and 98% sensitivity, demonstrating its effectiveness in distinguishing tumor subtypes. The integration of image processing techniques, such as median filtering and histogram equalization, significantly contributed to performance improvement, highlighting the importance of preprocessing in computational histopathology. These findings reinforce the potential of the model as a supporting tool for pathologists, enabling faster and more reliable diagnoses.

For future work, we plan to perform hyperparameter optimization along with the exploration of additional preprocessing techniques. Furthermore, the method will be tested on other datasets to assess its generalization ability, given that the current dataset is limited in size. We believe that the combination of deep learning and advanced image processing techniques could represent a significant advance in the analysis of histopathological images, contributing to the earlier and more accurate detection of ovarian cancer.

References

Breen, J., Allen, K., Zucker, K., Godson, L., Orsi, N., and Ravikumar, N. (2024). Histopathology foundation models enable accurate ovarian cancer subtype classification. *arXiv preprint arXiv:2405.09990.*

Breen, J., Allen, K., Zucker, K., Godson, L., Orsi, N. M., and Ravikumar, N. (2025). A comprehensive evaluation of histopathology foundation models for ovarian cancer subtype classification. *NPJ Precision Oncology*, 9(1):33.

Gonzalez, R. C. and Woods, R. E. (2018). *Digital Image Processing*. Pearson, 4th edition.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778.

Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4700–4708.

INCA (2023). Cancer incidence estimates in brazil – 2023 [in portuguese]. `https://www.inca.gov.br`.

Kasture, S., Mahajan, A., Patil, P., and Deshpande, S. (2021). A New Deep Learning Method for Automatic Ovarian Cancer Prediction. *International Journal of Advanced Computer Science and Applications*, 12(5):560–566.

Kussaibi, H., Alibrahim, E., Alamer, E., Alhaji, G., Alshehab, S., Shabib, Z., Alsafwani, N., and Menezes, R. G. (2024). Lightgbm-based classification of ovarian cancer subtypes from histological images using resnet50. *medRxiv*.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., and Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 10012–10022.

Radhakrishnan, M., Sampathila, N., Muralikrishna, H., and Swathi, K. S. (2024). Advancing ovarian cancer diagnosis through deep learning and explainable ai: A multiclassification approach. *IEEE Access*, 12:116968–116986.

Simonyan, K. and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In *Proceedings of the International Conference on Learning Representations (ICLR)*.

Vedana, A. B., Benevides, A. B., Farage, A. P. C., Moreira, G. A., Máximo, L. R. M. I., Brandão, R. S., Ubiali, I. R., Jardini, E. P., Bezerra, P. T. B., Da Silva, R. J. C., et al. (2024). Artificial intelligence in diagnostic medicine [in portuguese]. *Brazilian Journal of Implantology and Health Sciences*, 6(11):765–794.

Vilela, A. D., Silva, J. P., and Oliveira, M. C. (2022). Evaluation metrics for machine learning models in healthcare [in portuguese]. *Revista de Engenharia e Pesquisa Aplicada*, 7(2):34–45.

Werner, D. A. (2019). The fourth industrial revolution and artificial intelligence: A study on its concepts, impacts, and possible applications in law through legal text analysis for predictive classification of judicial decisions [in portuguese]. Master's thesis, Universidade do Vale do Rio dos Sinos (UNISINOS), São Leopoldo, RS, Brazil.