

**Original Article – General Urology****AI-Assisted Cystoscopy Image Analysis for Detecting Urological Pathologies: A Novel Approach**

Ürolojik Patolojilerin Tespitinde Yapay Zeka Destekli Sistoskopi Görüntü Analizi: Yeni Bir Yaklaşım

Short Title: AI-Assisted Cystoscopy Image Analysis (Yapay Zeka Destekli Sistoskopi Görüntü Analizi)

Halil İbrahim İvelik, Okan Alkış, İbrahim Güven Kartal, Mehmet Sevim, Bekir Aras

Kutahya Health Science University, Department of Urology, Kutahya, Türkiye

Cite as: İvelik Hİ, Alkış O, Kartal İG, Sevim M, Aras B. AI-assisted cystoscopy image analysis for detecting urological pathologies: A novel approach. Grand J Urol 2025, DOI: 10.5505/GJU.2025.94824 [Epub Ahead of Print]

Submission date: 15 October 2024 **Acceptance date:** 18 February 2025 **Online first:** 06 March 2025

Publication date:

Corresponding Author: Halil İbrahim İvelik / Kutahya Health Science University, Department of Urology, Kutahya, Türkiye / halib_ive@hotmail.com / ORCID ID: 0000-0001-5298-0045

ORCID ID: O. Alkış 0000-0001-6116-9588 İ.G. Kartal 0000-0002-2313-3522 M. Sevim 0000-0002-7571-7669 B. Aras 0000-0002-7020-8830

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our authors, we are providing this early version of the manuscript. The version will undergo copyediting, typesetting and review before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



This is an Open Access article distributed under the terms of the Creative Commons Attribution NonCommercial License 4.0 (<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Objectives: This study aims to develop and evaluate an AI-assisted system for detecting urological pathologies using cystoscopy images.

Materials and Methods: A dataset comprising 500 pathological and 500 healthy cystoscopy images was collected from the urology clinic of training and research hospital. Images were obtained using three different endovision systems (Karl Storz [Germany], Stryker [USA], Richard Wolf [Germany]). The dataset was preprocessed, augmented, and used to train a Convolutional Neural Network (CNN) model to classify images as either normal or pathological. The model's performance was evaluated on a test set comprising 100 pathological and 100 healthy images, using metrics such as accuracy, sensitivity, specificity, and F1-score. Statistical analyses were performed using IBM SPSS version 25.0, with a p-value of <0.05 considered significant.

Results: The model achieved a sensitivity of 94% for detecting pathological cases and a specificity of 58% for correctly identifying healthy cases. For pathological images, precision, recall, and F1-score were 0.69, 0.94, and 0.80, respectively, while for healthy images, these metrics were 0.91, 0.60, and 0.72. The overall accuracy of the model was recorded as 76%.

Conclusion: The AI-assisted cystoscopy image analysis system demonstrates high sensitivity in detecting urological pathologies but requires further improvements to enhance specificity. Future studies should focus on increasing dataset diversity and improving the model's ability to distinguish between benign and malignant features. The integration of higher-quality images and advanced AI techniques holds great potential for enhancing the model's success and improving diagnostic accuracy.

Keywords: artificial intelligence, cystoscopy, bladder cancer, deep learning, image analysis

Özet

Amaç: Bu çalışma, sistoskopi görüntülerini kullanarak ürolojik patolojilerin tespitine yönelik AI destekli bir sistem geliştirmeyi ve değerlendirmeyi amaçlamaktadır.

Gereçler ve Yöntemler: Eğitim ve Araştırma Hastanesi Üroloji Kliniğinde; 500 patolojik ve 500 sağlıklı sistoskopi görüntüsünden oluşan bir veri seti toplanmıştır. Görüntüler, üç farklı endovizyon sistemi (Karl Storz [Almanya], Stryker [ABD], Richard Wolf [Almanya])

kullanılarak elde edilmiştir. Veri seti ön işleme tabi tutulmuş, artırılmış ve bir Konvolüsyonel Sinir Ağı (CNN) modeli, görüntüleri normal veya patolojik olarak sınıflandırmak üzere eğitilmiştir. Modelin performansı, doğruluk, hassasiyet, özgüllük ve F1 skoru gibi metriklerle, 100 patolojik ve 100 sağlıklı görüntüden oluşan bir test setinde değerlendirilmiştir. İstatistiksel analizler IBM SPSS versiyon 25.0 ile yapılmış, $p < 0.05$ anlamlı kabul edilmiştir.

Bulgular: Model, patolojik vakaların tespiti için %94 hassasiyet ve sağlıklı vakaların doğru sınıflandırılması için %58 özgüllük sağlamıştır. Patolojik görüntüler için kesinlik, geri çağırma ve F1 skoru sırasıyla 0.69, 0.94 ve 0.80 olarak bulunurken, sağlıklı görüntüler için bu değerler 0.91, 0.60 ve 0.72'dir. Modelin genel doğruluğu %76 olarak kaydedilmiştir.

Sonuç: AI destekli sistoskopi görüntü analiz sistemi, ürolojik patolojilerin tespitinde yüksek hassasiyet göstermektedir, ancak özgüllüğün artırılması için daha fazla iyileştirme gerekmektedir. Gelecek çalışmalarda, veri setinin çeşitliliğini artırmaya ve modelin benign ve malign özellikleri ayırt etme yeteneğini geliştirmeye odaklanılmalıdır. Daha kaliteli görüntülerin entegrasyonu ve ileri yapay zeka tekniklerinin kullanımı, modelin başarısını artırma açısından büyük bir potansiyel sunmaktadır.

Anahtar kelimeler: yapay zeka, sistoskopi, mesane kanseri, derin öğrenme, görüntü analizi

Introduction

Artificial Intelligence (AI) and machine learning (ML) are rapidly advancing fields with the potential to revolutionize medical practice. AI refers to the ability of computer programs to learn and solve problems autonomously. Within AI, ML involves building mathematical models from input data to make decisions without human intervention. A subset of ML, known as deep learning (DL), uses multi-layered neural networks that mimic brain neurons' structure and activity, significantly enhances image recognition through neural networks [1].

One specific class of DL algorithms, the Convolutional Neural Network (CNN), is particularly well-suited for image recognition and analysis due to its architecture, which resembles the visual cortex. CNNs have driven substantial breakthroughs in medical image recognition, enabling AI to classify medical images with high accuracy. In the past, ML models relied on hand-crafted features such as color, intensity, and texture, but DL has surpassed these by automatically learning these features from vast amounts of data [2].

AI's progress in medical imaging spans radiology, ophthalmology, dermatology, pathology, neurology, and gastroenterology, where systems like computer-aided diagnosis (CADx) and detection (CAdE) have addressed limitations in clinical practice [3,4]. Advances in computing power and big data analytics further facilitate AI integration into medical practice.

In urology, cystoscopy is a vital diagnostic tool for detecting urological pathologies. However, the interpretation of cystoscopy images relies heavily on the expertise and experience of clinicians, which can introduce variability and subjectivity into the diagnostic process. AI-supported systems can mitigate these issues by providing consistent and accurate image analysis, potentially enhancing diagnostic accuracy and efficiency [5].

This study develops and evaluates a CNN-based AI system for detecting urological pathologies from cystoscopy images. The system could be used both in clinical settings and at home, where patients might upload images captured using camera-equipped catheters for analysis, reducing the burden on healthcare professionals and offering a convenient monitoring tool for patients.

Developing such an AI system requires a multidisciplinary approach, combining expertise in urology, computer science, and data analytics. The involvement of clinical experts ensures that the system is clinically relevant and meets the practical needs of healthcare providers and patients. Additionally, the economic and societal benefits of such a system could be substantial, improving early detection rates and reducing healthcare costs through more efficient patient monitoring and follow-up.

AI-supported cystoscopy image analysis represents a promising advancement in urological diagnostics. This paper outlines the development of our AI system, details the methodology, and presents the results of our evaluations. By improving diagnostic accuracy and providing a scalable solution for patient monitoring, our system aims to enhance the overall quality of urological care.

Materials and Methods

Participants

This study was conducted at the Urology Clinic of Training and Research Hospital. The study included patients over 18 years of age who underwent cystoscopy between January 2018 and January 2024. Ethical approval for the study was obtained from the relevant institutional review board (08.07.2024-144011). All participants provided informed consent prior to

inclusion in the study. A total of 500 pathological and 500 healthy cystoscopy images were collected for analysis.

The pathological images in this study were specifically from patients diagnosed with bladder cancer, including images from papillary or solid tumor formations observed during follow-up. These images were taken from atypical tissue areas, and no other pathologies were included in the evaluation beyond bladder cancer. The pathological images did not focus on a single bladder region but were representative of various areas. The decision to design the study this way was to avoid the complexity of interpreting fibrotic and hyperemic areas in previously resected regions, which can be challenging even for expert urologists. The healthy images were from patients with intact bladder tissue, with no recurrence observed post-endoscopic resection.

Imaging Systems

Cystoscopy images were acquired using three different endovision imaging systems: Karl Storz (Germany), Striker (USA), and Richard Wolf (Germany). Each system was equipped with different quality telescopes, including two “Karl Storz 30° Hopkins Telescope” and one “R. Wolf 30° 4.0 mm Telescope”. This resulted in a varied dataset with differing image qualities and resolutions, which provided a comprehensive basis for training and evaluating the AI model.

Data Processing and Model Training

The collected cystoscopy images were classified into two categories: normal and pathological. Normal images were characterized by a smooth, cream-colored epithelial lining with non-prominent vasculature and minimal trabeculation. Pathological images were identified by the presence of raised, atypical structures such as tumors, which appeared distinct from the normal bladder lining.

To prepare the images for model training, they were resized to a consistent dimension of 224x224 pixels and normalized to a range of 0 to 1. Data augmentation techniques, including rotation, flipping, and brightness adjustments, were applied to increase the variability and robustness of the dataset.

A Convolutional Neural Network (CNN) was employed for image analysis and classification. The CNN architecture included multiple convolutional and pooling layers designed to extract relevant features from the images, followed by fully connected layers for

classification. The model was implemented using the TensorFlow and Keras libraries in Python [6].

The model was trained using a supervised learning approach. During training, the CNN learned to distinguish between normal and pathological images by optimizing the weights of the network to minimize the binary cross-entropy loss function. The Adam optimizer was used to update the model parameters, and the training process was monitored using validation data to prevent overfitting [7].

Training and Validation

The dataset was split into training and testing sets, with 80% of the images used for training and 20% reserved for testing. The training process involved iterating over the training data for multiple epochs, with each epoch consisting of a forward pass to compute the output and a backward pass to update the model parameters based on the loss gradient.

To enhance the model's generalization capabilities, k-fold cross-validation was employed. This technique involves partitioning the training data into k subsets and training the model k times, each time using a different subset as the validation data and the remaining subsets as the training data. The final model performance was averaged across the k folds to obtain a robust estimate of its accuracy, sensitivity, and specificity [8].

Performance Metrics

The performance of the trained model was evaluated using the test set. Key metrics included accuracy, precision, recall (sensitivity), and specificity. The confusion matrix was used to compute these metrics, providing a detailed understanding of the model's performance in distinguishing between normal and pathological images.

Precision was calculated as the ratio of true positive predictions to the sum of true positive and false positive predictions. Recall (sensitivity) was determined as the ratio of true positive predictions to the sum of true positive and false negative predictions. Specificity was computed as the ratio of true negative predictions to the sum of true negative and false positive predictions [9].

Technical Considerations

While different imaging systems and optics provided diverse data, they also introduced challenges related to image homogeneity and consistency. Variations in resolution, contrast,

and color profiles across the different systems potentially impacted the model's ability to generalize across all image types. This variability underscores the importance of incorporating a wide range of data augmentation techniques and rigorous cross-validation to ensure the robustness of the AI model.

Results

Model Performance

In our study, we developed an AI-assisted system to identify pathological and healthy bladder images from cystoscopy data. The model was trained on a dataset of 500 pathological and 500 healthy images and later tested on a separate set of 100 pathological and 100 healthy images. The initial testing within the controlled environment showed high accuracy, but real-world application yielded different results.

Confusion Matrix

The confusion matrix below illustrates the performance of our AI model on the test dataset. The matrix provides insights into true positive (TP), true negative (TN), false positive (FP), and false negative (FN) counts. While the AI model demonstrated high performance during initial testing, real-world application revealed significant challenges. The model achieved a sensitivity of 94%, indicating it could correctly identify 94 out of 100 pathological cases. However, the specificity was 58%, with 42 out of 100 healthy images incorrectly classified as pathological. This lower specificity suggests potential issues in distinguishing between certain benign structures (e.g., trabeculation, trigon area) and pathological ones (**Figure 1**).

The sensitivity and specificity of our model are key metrics that indicate its effectiveness: Sensitivity: 0.94 (94%); Specificity: 0.6 (60%)

Classification Report

The classification report provides additional metrics including precision, recall, and F1-score for both classes (healthy and pathological) (**Table 1**):

Precision: This metric indicates the accuracy of the model in predicting positive instances (i.e., how many of the instances predicted as pathological are actually pathological). The precision for healthy images is 0.91, meaning 91% of the images predicted as healthy are

indeed healthy. The precision for pathological images is 0.69, indicating that 69% of the images predicted as pathological are truly pathological.

Recall (Sensitivity): Recall measures the model's ability to identify all relevant instances. The recall for healthy images is 0.58, meaning the model correctly identifies 58% of the healthy images. The recall for pathological images is 0.94, indicating that the model correctly identifies 94% of the pathological images.

F1-Score: The F1-score is the harmonic mean of precision and recall, providing a single metric that balances both concerns. The F1-score for healthy images is 0.71, and for pathological images, it is 0.80. These scores indicate the overall effectiveness of the model in classifying each category.

Support: Support refers to the number of actual occurrences of each class in the dataset. Both healthy and pathological categories have 100 images in the test set.

Accuracy: Overall, the model has an accuracy of 75%, meaning it correctly classified 75% of the images in the test set.

Macro Average: This average calculates the mean performance across all classes without taking class imbalance into account. The macro average for precision, recall, and F1-score is around 0.80, 0.76, and 0.75, respectively.

Weighted Average: This average takes class imbalance into account, providing a more realistic measure of the model's performance. The weighted averages for precision, recall, and F1-score are approximately 0.80, 0.76, and 0.75, respectively.

Overall, while the model shows high sensitivity in detecting pathological images, its specificity in correctly identifying healthy images is lower. This indicates a tendency to incorrectly classify healthy images as pathological, which is an important consideration for further improvements and refinements in the model.

Discussion

The performance of our AI model, while promising in controlled test environments, exhibited lower specificity in real-world applications. This discrepancy can be attributed to several factors related to the variability and complexity of medical imaging, particularly in cystoscopy.

One significant challenge we faced was the variability in imaging systems and optics used for data collection. The images were sourced from three different endovision systems (Storz, Striker, R. Wolf) and varied in resolution, contrast, and color profiles due to different optical qualities (two Storz and one R. Wolf). These differences introduced inconsistencies in the data, making it harder for the model to generalize across all image types. As a result, the model's specificity was affected, leading to a higher rate of false positives (42 out of 100 healthy images were misclassified as pathological) [10].

Our model was primarily trained to identify pathological structures based on their elevation and texture compared to the smooth, flat surface of healthy bladder tissue. However, certain benign anatomical features, such as trabeculation and the trigon area, were sometimes misclassified as pathological due to their elevated appearance. Additionally, areas with increased angiogenesis were often flagged as pathological. This indicates that while the model is effective in detecting deviations from the norm, it requires further refinement to differentiate between benign and malignant variations more accurately [3].

Due to recent advancements in AI and machine learning, AI-assisted diagnostics has become an intriguing, yet not fully explored field. In our opinion, we should view neural network and deep learning-based models as a form of 'expert opinion' rather than an entirely objective diagnostic test. Notably, cystoscopy performed by a urologist is also, in essence, a form of 'expert opinion'. This similarity in approach makes AI-assisted diagnostic methods a potentially suitable application for urological procedures like cystoscopy. While AI can aid in identifying abnormalities and augment a clinician's ability to detect disease, human oversight remains crucial for interpretation, especially in complex cases where benign and malignant features might overlap. Therefore, AI should complement, rather than replace, the expertise of the clinician in these scenarios."

To enhance the model's performance, several strategies can be considered:

Larger and More Homogeneous Dataset: Increasing the size of the dataset with more diverse images from a single, high-quality imaging system can help reduce variability. This would allow the model to learn more consistent features and improve generalization [7].

Regional Mapping of the Bladder: Dividing the bladder into specific regions (e.g., trigon, dome, lateral walls) and training the model to recognize patterns within these regions can improve accuracy. This approach ensures that the model considers the anatomical context when making predictions [6].

Data Augmentation and Preprocessing: Implementing advanced data augmentation techniques, such as varying lighting conditions, rotations, and translations, can help the model become more robust to variations. Preprocessing steps like normalization and contrast adjustment can also standardize the input data, reducing discrepancies between images [11].

Advanced AI Techniques: Utilizing more sophisticated AI architectures, such as transfer learning with pre-trained models like ResNet or VGG, can enhance the model's ability to learn complex patterns. Ensemble learning, combining multiple models, can also provide more reliable predictions by mitigating the weaknesses of individual models [4].

The AI-assisted cystoscopy image analysis system developed in this study demonstrated high sensitivity in detecting urological pathologies. However, further work is needed to improve specificity. Our study employed a Weakly Supervised Learning approach, where not all images were manually labeled. To achieve more accurate results, more complex and data-intensive methods, such as Fully Supervised Learning, may be required. This approach could enhance the model's performance, particularly in distinguishing between benign and malignant structures more effectively."

Artificial intelligence, particularly deep learning, relies on large datasets and high computational power to learn and generalize effectively. The advancements in computing power and the availability of big data have facilitated the integration of AI into clinical practice. However, the success of AI models in medical imaging heavily depends on the quality and consistency of the training data [12].

In the future, AI models could benefit from more sophisticated learning mechanisms, such as continual learning, where the model can adapt to new data incrementally without forgetting previously learned information. This approach could be particularly useful in medical imaging, where new data continuously becomes available [13].

Our study contributes to the growing body of literature on AI-assisted medical imaging by highlighting the challenges and potential solutions for improving model performance in real-world applications. The successful implementation of AI in cystoscopy could significantly reduce the workload of urologists and improve patient outcomes by enabling earlier and more accurate detection of bladder pathologies.

Future research should focus on developing standardized imaging protocols and larger, more diverse datasets to train AI models. Additionally, integrating AI with other diagnostic

tools, such as MRI or CT scans, could provide a more comprehensive assessment of urological conditions, further enhancing diagnostic accuracy and patient care.

Conclusion

The developed AI model for cystoscopy image analysis shows promise but requires further refinement and testing with more diverse datasets to improve its specificity. Future work will focus on enhancing the model's ability to accurately classify benign anatomical variations and integrating higher-quality images from various endovision systems to improve overall performance.

Ethics Committee Approval: Ethical approval for this study was obtained from Kutahya Health Science University Clinical Research Ethics Committee (Approval number and date: 08.07.2024-144011).

Informed Consent: An informed consent was obtained from all the patients.

Publication: The results of the study were not published in full or in part in form of abstracts.

Peer-review: Externally peer-reviewed.

Authorship Contributions: Any contribution was not made by any individual not listed as an author. Concept – H.İ.İ., O.A.; Design – H.İ.İ., O.A.; Supervision – H.İ.İ., B.A.; Resources – İ.G.K., M.S.; Materials – İ.G.K., M.S.; Data Collection and/or Processing – İ.G.K., M.S.; Analysis and/or Interpretation – H.İ.İ., O.A.; Literature Search – İ.G.K., M.S.; Writing Manuscript – H.İ.İ., O.A.; Critical Review – H.İ.İ., B.A.

Conflict of Interest: The authors declare that they have no conflicts of interest.

Financial Disclosure: The authors declare that this study received no financial support.

References

- [1] Mishra C, Gupta D. Deep machine learning and neural networks: an overview. *IAES Int J Artif Intell* 2017;6(2):66-73. <https://doi.org/10.11591/ijai.v6.i2.pp66-73>
- [2] Yamashita R, Nishio M, Do RK, Togashi K. Convolutional neural networks: an overview and application in radiology. *Insights Imaging* 2018;9(4):611-29. <https://doi.org/10.1007/s13244-018-0639-9>
- [3] Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017;542(7639):115-8. <https://doi.org/10.1038/nature21056>
- [4] Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 2016;316(22):2402-10. <https://doi.org/10.1001/jama.2016.17216>

- [5] Ferro M, Falagario UG, Barone B, Maggi M, Crocetto F, Busetto GM, et al. Artificial intelligence in the advanced diagnosis of bladder cancer: comprehensive literature review and future advancement. *Diagnostics (Basel)* 2023;13(13):2308. <https://doi.org/10.3390/diagnostics13132308>
- [6] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*. 2012; p. 1097-105.
- [7] LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521(7553):436-44. <https://doi.org/10.1038/nature14539>
- [8] Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *Proceedings of the 14th international joint conference on artificial intelligence-Volume 2*. 1995; p. 1137-43.
- [9] Brownlee J. *Machine learning mastery with Python: understand your data, create accurate models, and work projects end-to-end*. Machine Learning Mastery; 2016.
- [10] Okagawa Y, Abe S, Yamada M, Oda I, Saito Y. Artificial intelligence in endoscopy. *Dig Dis Sci* 2022;67(5):1553-72. <https://doi.org/10.1007/s10620-021-07086-z>
- [11] Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. *J Big Data* 2019;6(1):60. <https://doi.org/10.1186/s40537-019-0197-0>
- [12] Bibault JE, Giraud P, Burgun A. Big data and machine learning in radiation oncology: state of the art and future prospects. *Cancer Lett* 2016;382(1):110-7. <https://doi.org/10.1016/j.canlet.2016.01.043>
- [13] Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med* 2019;25(1):44-56. <https://doi.org/10.1038/s41591-018-0300-7>

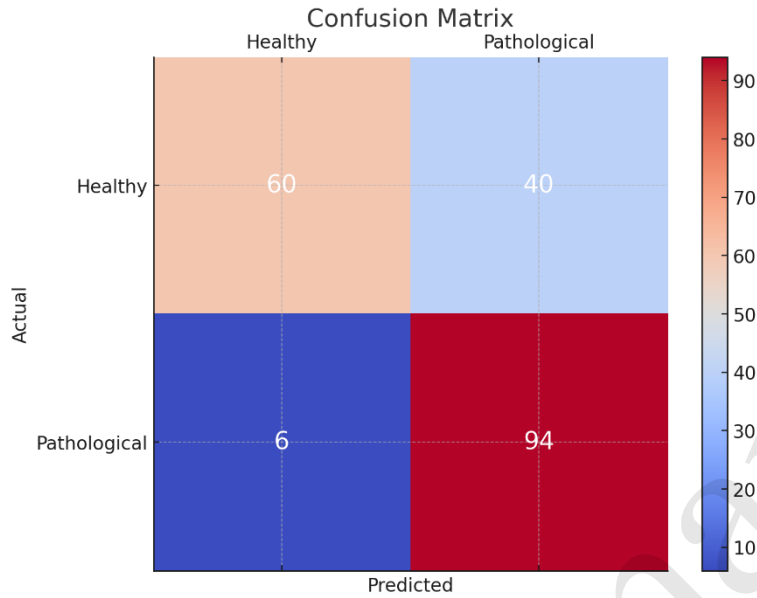


Figure 1. Illustrates the performance of our AI model on the test dataset. The matrix provides insights into true positive (TP), true negative (TN), false positive (FP), and false negative (FN) counts.

Table 1. The classification report provides additional metrics including precision, recall, and F1-score for both classes (healthy and pathological)

	Precision	Recall	F1-Score	Support
Healthy	0.91	0.60	0.72	100
Pathological	0.70	0.94	0.80	100
Accuracy			0.76	200
Macro Avg	0.80	0.77	0.76	200
Weighted Avg	0.80	0.77	0.76	200