



# Development of an Artificial Intelligence Algorithm for Colposcopic Recognition of Cervical Features - Introduction of an Annotation Protocol

Georgi Danielov Prandzhev<sup>1,2</sup>, Grigor Angelov Gortchev<sup>1,2</sup>, Dobromir Dimitrov Dimitrov<sup>3,4</sup>,  
Radoslav Iliev Miltchev<sup>5</sup>, Slavcho Tomov Tomov<sup>1,2</sup>

<sup>1</sup>Department of Obstetrics and Gynecology, Medical University - Pleven, Pleven, Bulgaria

<sup>2</sup>Department of Obstetrics and Gynecology, University Hospital Saint Marina - Pleven, Pleven, Bulgaria

<sup>3</sup>Department of Surgical Propaedeutics, Medical University - Pleven, Pleven, Bulgaria

<sup>4</sup>Department of Surgical Oncology, Dr. Georgi Stranski University Hospital, Pleven, Bulgaria

<sup>5</sup>Faculty of Industrial Technology, Technical University of Sofia, Sofia, Bulgaria

Cervical cancer, which is largely preventable via effective screening and vaccination, remains the fourth most common cancer among women globally. In 2020, approximately 604,000 new cases and 342,000 cancer-related deaths were reported, with most occurring in low- and middle-income countries due to inadequate access to screening, human papillomavirus (HPV) vaccination, and timely treatment.<sup>1</sup> The Balkans exhibit some of the highest rates, with Romania leading at 20.44 cases per 100,000 women, followed by Bulgaria (16.78) and Serbia (15.67). In contrast, countries like Slovenia (7.63) and Greece (6.67) report much lower rates due to robust prevention strategies. Although North Macedonia (9.9) and Croatia (10.01) also demonstrate lower rates, better screening and public health education is required.<sup>2</sup>

A recent Cochrane review revealed that HPV testing detects high-grade cervical lesions with 90-95% sensitivity, outperforming cytology's sensitivity of 55-80%. However, the specificity of Pap smear, which often exceeds 90%, helps minimize false-positive results. HPV tests can be performed every 5 years, whereas Pap tests are performed every 3 years. Thus, Pap smears offer practical advantages despite their higher costs and false-positive rates. Combining HPV testing with the Pap smears ensures sensitivity of up to 95%, making it the most reliable method for detecting cervical cancer in women aged > 30 years. However, the cost of this combination limits feasibility in low-resource settings where cytology is more accessible. Colposcopy, a magnified visualization of the cervix, and biopsy remain the gold standard for diagnosing atypical patterns of the cervical epithelium. Its sensitivity ranges from 50% to 80% and specificity ranges from

60% to 90%, with variability attributable to observer experience and technical factors.<sup>3</sup>

Artificial intelligence (AI) is broadly defined as the development of computational systems capable of performing tasks, such as decision-making, learning, and problem-solving, that typically require human intelligence. In medical diagnostics, AI leverages machine learning (ML) and deep learning (DL) algorithms to analyze complex datasets, identify patterns, and provide insights to clinicians to assist in decision-making. AI applications in diagnostics include automated image analysis, predictive modeling, and risk stratification. Convolutional neural networks (CNNs) are often employed to process medical imaging data, which enables the identification of subtle abnormalities that may escape human detection. By integrating AI into diagnostic workflows, healthcare systems aim to enhance accuracy, reduce diagnostic errors, and provide scalable solutions for resource-limited settings without replacing the critical expertise of medical professionals.<sup>4</sup>

The annual percentage of publications on AI in cervical cancer screening has grown exponentially from 2009 to 2022. A noticeable rise was observed between 2016 and 2018, and it was driven by advancements in DL and its application in medical imaging. The sharpest growth in publications on AI was observed from 2019 to 2022, with 28% of the studies being published in 2020 and 16% being published in both 2021 and 2022. These data highlight AI's increasing integration into diagnostic tools and growing importance in research.<sup>5</sup> DL models, particularly CNNs, have consistently outperformed traditional ML approaches such as support vector



**Corresponding author:** Georgi Danielov Prandzhev, Department of Obstetrics and Gynecology, Medical University - Pleven; University Hospital Saint Marina - Pleven, Pleven, Bulgaria

**e-mail:** georgi.prandzhev@gmail.com

**Received:** December 09, 2024 **Accepted:** February 02, 2025 **Available Online Date:** 01.07.2025 • **DOI:** 10.4274/balkanmedj.galenos.2025.2024-12-35

Available at [www.balkanmedicaljournal.org](http://www.balkanmedicaljournal.org)

**ORCID iDs of the authors:** G.D.P. 0000-0001-9612-3838; G.A.G. 0000-0002-4710-9340; D.D.D. 0000-0003-3313-1093; R.I.M. 0000-0003-1914-1172; S.T.T. 0000-0001-7919-946X.

**Cite this article as:** Prandzhev GD, Gortchev GA, Dimitrov DD, Miltchev RI, Tomov ST. Development of an Artificial Intelligence Algorithm for Colposcopic Recognition of Cervical Features - Introduction of an Annotation Protocol. *Balkan Med J.*; 2025; 42(4):373-5.

Copyright@Author(s) - Available online at <http://balkanmedicaljournal.org/>

machines and random forest, especially in terms of accuracy and sensitivity.<sup>6</sup> Models that have demonstrated notable performances include MobileNetv2-YOLOv3 in combination with EfficientNetB0 (96.84% accuracy)<sup>7</sup> and gated recurrent CNN (96.87% accuracy),<sup>8</sup> which have demonstrated that DL models provide robust diagnostic support. These models exhibit sensitivities exceeding 95%, indicating their effectiveness in detecting true positives. Furthermore, specificities as high as 98.72% indicate the ability of DL models to reduce the incidence of false-positive results, which are critical for clinical reliability.<sup>6</sup>

Our protocol employs an advanced digital colposcope equipped with adjustable white LED illumination, with an intensity of 45,000-52,000 lux and a color temperature of 5,700-6,000 K. For uniform visualization, it offers up to 30x optical zoom, and a high-definition camera captures images at 1920 x 1080 resolution. These specifications ensure detailed imaging that is crucial for identifying cervical abnormalities. The process begins with appropriate patient positioning and optimal cervix visualization. A 3-5% acetic acid solution is applied to highlight the abnormal epithelium. Thereafter, 5% Lugol's iodine is applied to contrast normal glycogen-rich and abnormal areas. Images are systematically captured, starting with low magnification (7.5x) for general views and progressing to high magnification (15x and 30x) for detailed assessments. The images are subsequently uploaded to an annotation platform, where expert

colposcopists label key structures. The main areas of interest are the squamocolumnar junction and the three types of transformation zones (TZ). The expert colposcopists also tag atypical patterns such as mosaicism, punctations, acetowhite changes, vessels, and abnormal findings such as polyps, nabothian cysts, or leukoplakia. Thereafter, the annotated images are exported in JPG format along with the metadata, creating comprehensive datasets for AI training (Figure 1). All patient information is securely stored in compliance with medical and data privacy standards. Our algorithm leverages YOLOv8x, an advanced object detection framework that offers high accuracy, speed, and flexibility, making it ideal for medical imaging. YOLOv8 addresses challenges such as observer variability and subjectivity in colposcopic evaluations, ensuring greater precision and consistency in detecting and classifying cervical features. Our model, which was trained using 800 images, achieved 73% validation accuracy in identifying the TZ, which is the first step in colposcopic analysis. Written informed consent was obtained from the patients before being included in the study and for using the accompanying images. A confusion matrix illustrates the ability of the model to differentiate between fully visible (type 1), partially visible (type 2), and completely invisible (type 3) TZs (Table 1).<sup>9</sup> The analysis revealed that our algorithm exhibited the best classification accuracy for type 2 TZs (0.60), with relatively low misclassifications to type 1 TZs (0.20) and type 3 TZs (0.10). Conversely, the algorithm demonstrated the lowest classification accuracy for type 3 TZs



FIG. 1. A partial overview of the annotated database.

TABLE 1. Confusion Matrix - Demonstration of Correctly and Incorrectly Classified Examples for Each Class.

Predicted\actual	Type 1	Type 2	Type 3	Accuracy
Type 1	0.58	0.05	0.04	0.58
Type 2	0.20	0.60	0.10	0.60
Type 3	0.10	0.15	0.50	0.50
Total	0.88	0.80	0.64	

(0.50), with significant misclassifications to type 2 TZs (0.15) and some misclassifications to type 1 TZs (0.10). Based on the provided dataset, the accuracy rates for each TZ type were relatively low. This indicates that the current database may have been insufficient for consistently identifying the TZ types. Using larger and more diverse datasets could significantly enhance prediction accuracy, thereby improving diagnostic outcomes.

The integration of AI has demonstrated promising results in enhancing colposcopic sensitivity, specificity, and accuracy. Preliminary training and classification tests have demonstrated that a robust balance between accuracy and speed can be achieved, paving the way for practical clinical applications. To further enhance model performance, the classification of type 3 TZs should be improved, as their misclassification to type 2 TZ remains an issue. Thus, analyzing the features that cause confusion between type 1 and type 2 TZs is crucial. Model optimization and better feature analysis could help mitigate these challenges, reduce errors, and ultimately improve diagnostic accuracy in cervical screening. Although small datasets may be sufficient for initial evaluations, larger datasets are crucial for a more reliable and accurate diagnostic process. Continued studies and fine-tuning of processes are essential to achieve a reliable and efficient model.

**Informed Consent:** Written informed consent was obtained from the patients before being included in the study and for using the accompanying images.

**Authorship Contributions:** Concept- G.D.P., G.A.G., D.D.D., S.T.T.; Design- G.A.G., D.D.D., S.T.T.; Supervision- G.A.G., S.T.T.; Fundings- G.A.G.; Materials- R.I.M., S.T.T.; Data Collection or Processing- G.D.P., R.I.M., S.T.T.; Analysis or Interpretation- G.D.P., R.I.M.; Literature Search- G.D.P., R.I.M.; Writing- G.D.P.; Critical Review- G.A.G., D.D.D., S.T.T.

**Conflict of Interest:** No conflict of interest was declared by the authors.

**Funding:** This research was funded by the Bulgarian National Science Fund (BNSF), through "Competition for Financial Support for Basic Research Projects - 2022", project KII-06-H63/5 from 13.12.2022 - "Application of AI Methods for Colposcopic Recognition and Categorization of Cervical Features", with a contract no.: BG-175467353-2022-04-0257-C01.

## REFERENCES

1. Singh D, Vignat J, Lorenzoni V, et al. Global estimates of incidence and mortality of cervical cancer in 2020: a baseline analysis of the WHO Global Cervical Cancer Elimination Initiative. *Lancet Glob Health*. 2023;11:e197-e206. [\[CrossRef\]](#)
2. Todorovic J, Stamenkovic Z, Stevanovic A, et al.; COST Action 18218 participants Burden of Disease Collaborator Network. The burden of breast, cervical, and colon and rectum cancer in the Balkan countries, 1990-2019 and forecast to 2030. *Arch Public Health*. 2023;81:156. [\[CrossRef\]](#)
3. Koliopoulos G, Nyaga VN, Santesso N, et al. Cytology versus HPV testing for cervical cancer screening in the general population. *Cochrane Database Syst Rev*. 2017;8:CD008587. [\[CrossRef\]](#)
4. Hashimoto DA, Rosman G, Meireles OR. Artificial intelligence in surgery: understanding the role of AI in surgical practice. New York: McGraw Hill Medical Books; 2021. [\[CrossRef\]](#)
5. Vargas-Cardona HD, Rodriguez-Lopez M, Arrivillaga M, et al. Artificial intelligence for cervical cancer screening: scoping review, 2009-2022. *Int J Gynecol Obstet*. 2024;165:566-578. [\[CrossRef\]](#)
6. Park YR, Kim YJ, Ju W, Nam K, Kim S, Kim KG. Comparison of machine and deep learning for the classification of cervical cancer based on cervicography images. *Sci Rep*. 2021;11:16143. [\[CrossRef\]](#)
7. Habtemariam LW, Zewde ET, Simegn GL. Cervix type and cervical cancer classification system using deep learning techniques. *Med Devices (Auckl)*. 2022;15:163-176. [\[CrossRef\]](#)
8. Yu Y, Ma J, Zhao W, Li Z, Ding S. MSCl: a multistate dataset for colposcopy image classification of cervical cancer screening. *Int J Med Inform*. 2021;146:104352. [\[CrossRef\]](#)
9. Prendiville W, Sankaranarayanan R. Colposcopy and treatment of cervical precancer. Lyon (FR): International Agency for Research on Cancer; 2017. [\[CrossRef\]](#)