

Doctoral dissertation review
Segmentation and Classification of Pap Smear Images to Detect Cervical Cancer
using Deep Learning

This review of the doctoral dissertation, authored by Ms. Betelhem Wubineh, M.Sc. Eng., was commissioned by the Chairman of the Disciplinary Council for Technical Information Technology and Telecommunications, Prof. Wojciech Bożejko (Letter dated 25.06.2206). The supervisor¹ of this dissertation is Prof. Andrzej Rusiecki, and the auxiliary supervisor is Dr. Krzysztof Halawa². The doctoral dissertation is written in English³.

1. The research problem and its significance

What is the most important issue considered in the dissertation?

The most important issue addressed in the dissertation is the development of an accurate, robust, and clinically useful deep learning framework for the automatic detection of cervical cancer from Pap smear images, combining cervical cell segmentation, classification, data augmentation, uncertainty estimation, and explainability methods. The dissertation investigates how advanced deep learning architectures, including modified U-Net models, DeepLabV3+, SegNet, Vision Transformers, attention mechanisms, ensemble learning, and generative models, can improve the identification of abnormal cervical cells and support computer-aided diagnosis systems.

Is it scientific in nature?

Yes, its nature is clearly scientific. It formulates research questions and hypotheses, proposes new methodological solutions, develops novel neural network architectures and learning strategies, and verifies them through extensive experimental studies on several datasets. The work contributes to knowledge in computer science, especially in the fields of medical image analysis and artificial intelligence. It also includes comparative studies, statistical analyses, cross-validation, cross-dataset validation, uncertainty estimation, and comparisons with the state of the art, all of which are characteristic elements of scientific research.

Does it have any practical significance?

Yes, the dissertation has substantial practical significance. Cervical cancer remains a major global health problem, while the number of qualified cytotechnologists and pathologists is limited compared with the large volume of cytological samples that must be examined in screening programs. Consequently, manual assessment is time-consuming, costly, and susceptible to inter-observer variability, which increases the importance of reliable computer-aided diagnostic systems for image analysis and decision-making support.

¹ to choose

² if an auxiliary supervisor was appointed

³ optional

2. Author's contribution

What is the author's most important contribution to obtaining the results described in the dissertation?

The Author's most important contribution can be summarized in the following points:

1. Development and validation of novel and advanced deep learning pipelines for cervical cell segmentation and classification, including modified U-Net, SPP-SegNet, SE-DeepLabV3+, ViT-U-Net, and uncertainty-aware multi-task approaches, evaluated across several datasets
2. Design and experimental verification of ensemble-learning and attention-based classification strategies, demonstrating that ensembles of pretrained CNNs, residual refinement blocks, attention mechanisms, and hybrid CNN-Transformer architectures provide better generalizing results than individual classification models.
3. Comprehensive investigation of the role of segmentation and synthetic data generation in cervical cancer detection, including the proposed RES-DCGAN augmentation strategy and the demonstration that BBox-based representations, which preserve limited contextual information around the cell, frequently outperform tightly cropped ROI-based inputs.

From a practical perspective, the proposed solutions address an important real-world problem. The number of qualified cytotechnologists and pathologists is limited compared with the large number of cytological samples that must be assessed, particularly within population-based cervical cancer screening programs. Thus, the development of reliable computer-aided diagnostic systems capable of supporting image analysis is of practical importance.

However, the practical applicability of the proposed methods should currently be regarded as potential rather than demonstrated. The obtained results indicate that the developed algorithms could, in the future, support cytotechnologists and pathologists by pre-screening samples reducing workload. Nevertheless, the dissertation does not include prospective clinical validation or evaluation in routine diagnostic practice. Therefore, the proposed methods cannot yet be considered clinically validated decision-support tools. Before deployment in real screening programs, clinical validation would be necessary to verify their diagnostic reliability, robustness, and usefulness in everyday pathological practice.

An additional argument supporting the significance of the obtained results is that a substantial part of the dissertation is based on research published in highly ranked journals, including *Scientific Reports*, *Cancers*, *IEEE Access*, and *Neural Computing and Applications*, as well as several international conference proceedings. This indicates that the work has undergone external scientific evaluation and that its results are well aligned with current international research trends in artificial intelligence and medical image analysis. The publication record presented in the dissertation demonstrates an above-average level of scientific activity for a doctoral candidate.

3. Correctness

Are the claims made in the thesis trustworthy? Are the justifications sound? Point out any weaknesses and errors you've noticed. Also, highlight the aspects of validity that are most valuable (elegance of evidence, experimental design, analysis of empirical data, quality of prototype software/hardware, etc.).

Overall, the claims made in the dissertation are largely trustworthy and adequately supported by the presented evidence. The Author formulates clear research questions and hypotheses, proposes specific methodological improvements, and validates them through extensive experiments performed on several independent datasets. The experimental methodology is generally rigorous and includes train-test splits, 5-fold cross-validation, statistical testing, cross-dataset validation, ablation studies, and comparisons with baseline architectures and previously published methods. Such a broad evaluation framework increases confidence in the reported conclusions.

One of the strongest aspects of the dissertation is the experimental design. The Author does not rely on a single dataset or a single evaluation protocol (which is typical for most Ph.D. theses) but systematically investigates the behaviour of the proposed methods across multiple datasets with different characteristics. The inclusion of external validation and cross-dataset experiments is particularly valuable, since many studies in this field report very high performance on individual datasets without examining generalization. Also, the use of statistical analyses to compare competing methods further strengthens the credibility of the conclusions. Another valuable aspect is the systematic comparison of alternative architectural choices. This makes the conclusions more convincing than studies that only present a newly proposed model without adequate baselines. The ablation studies also help identify which components contribute most strongly to the observed improvements.

At the same time, several limitations should be acknowledged.

1. The private Pomeranian dataset plays an important role throughout the dissertation, but the segmentation masks for this dataset were generated automatically using OpenCV-based procedures rather than manually verified expert annotations. Consequently, the quality of the ground truth used for segmentation training and evaluation may be imperfect, which can affect the reliability of some segmentation results. The Author openly acknowledges this limitation.
2. Although the reported classification results are often exceptionally high, occasionally approaching or reaching 99-100%, such performance should be interpreted cautiously because the datasets used are relatively small compared with those typically employed in large-scale clinical AI studies. Additional validation on larger multi-centre cohorts would be desirable before drawing stronger conclusions about deployment in real clinical settings.
3. Another limitation concerns already mentioned lack of clinical validation. While the proposed methods demonstrate excellent performance under experimental conditions, the dissertation does not include prospective clinical studies or direct evaluation involving practicing cytotechnologists and. Therefore, although the methods show strong potential for computer-aided diagnosis, their real-world clinical effectiveness remains to be demonstrated.
4. A particular concern is the lack of validation of synthetic images by clinical experts. Without such an assessment, it cannot be unequivocally ruled out that the classification models are learning, in part, artefacts characteristic of the image generation process, rather than solely the actual cytological features associated with pathology. For this reason, an expert assessment of the quality and diagnostic reliability of synthetic data should be included in the research methodology.
5. Finally, the explainability analysis using Grad-CAM and saliency maps is useful and improves transparency, but the highlighted image regions were not independently assessed by medical experts. As a result, the explanatory visualizations provide supportive evidence rather than fully validated clinical explanations.

4. Candidate knowledge

Which chapters (or sections within chapters) of the dissertation discuss the existing state of knowledge and thus confirm the candidate's general knowledge in Technical Information Technology and Telecommunications? What areas of these disciplines are covered in these chapters/sections? What is the reviewer's opinion of the quality of these chapters/sections? What is the reviewer's opinion of the bibliography? How complete is the bibliography? Please provide other arguments for or against stating that the candidate has general knowledge in Information and Communication Technology.

The existing state of knowledge is discussed primarily in Chapter 2 which review cervical cancer screening, deep learning in medical image analysis, cervical cell segmentation, cervical cell classification, data augmentation techniques, and related work. In addition, Chapter 1 contains a concise overview of the research background and motivation, while Chapter 3 demonstrates the candidate's ability to position the proposed methods within the broader context of existing solutions. These chapters confirm the candidate's broad knowledge in Information and communication technology, particularly in the areas of medical image analysis, computer vision, pattern recognition, machine and deep learning. The Candidate demonstrates familiarity with both methodological foundations and the current state of international research in AI-based medical imaging.

The quality of the literature review is high. The discussion is well organized, comprehensive, and logically structured. The candidate clearly motivates the research undertaken in the dissertation. The review demonstrates a good understanding of both the theoretical and practical aspects of the field.

The bibliography is extensive, relevant, and up to date. It is sufficiently complete for a doctoral dissertation and provides an appropriate representation of the current state of knowledge in the investigated research area.

5. Other notes

1. Minor weakness of the dissertation is the way in which the experimental results are presented. The study contains a very large number of experiments and performance evaluations, which clearly demonstrates the author's substantial effort. However, the results are distributed across numerous sections and tables, making it difficult for the reader to identify the overall best-performing configurations and the most important conclusions. The dissertation would benefit from a more concise and synthetic summary of the obtained results, supported by additional comparative tables highlighting the strongest methods across datasets and evaluation settings. I encourage the Candidate to prepare such a summary for the public defense.
2. The private Pomeranian dataset was obtained from the Pomeranian Medical University in Szczecin. The dissertation does not provide information regarding the conditions under which the dataset was made available to the doctoral candidate. This aspect is important because the dataset originates from clinical material and therefore involves sensitive medical data. Furthermore, the dissertation does not state whether approval from an ethics committee was obtained for the use of patient data in scientific research. Please clarify these issues.
3. Please explain the difference between Otsu thresholding and Binary mask shown in Fig. 2.
4. Figure 4.4 is unclear in illustrating the quality of the obtained segmentation results. The distinction between cytoplasm and nucleus segmentation is not explained, and the segmented regions are marked in a manner that is barely visible. Thus, the segmentation outputs appear

very similar to the original images, making it difficult for the reader to visually assess the segmentation quality.

5. The 6 pipelines defined in Section 3.6.1 were implemented and evaluated in Chapter 6; however, their comparative analysis is not presented in a sufficiently systematic manner. Although numerical results are reported for all pipelines, the dissertation lacks a discussion identifying the best-performing pipeline and explaining the relative contribution of segmentation, ROI extraction, and data augmentation strategies. As a result, the reader may find it difficult to follow the rationale behind the individual pipeline designs and their comparative effectiveness

Summary

Taking into account the opinions presented in the previous points and the requirements defined by Article 187 of the Act of 20 July 2018 - Law on Higher Education and Science (as amended)⁴ my assessment of the dissertation in terms of three basic criteria is as follows:

- A. Does the dissertation contain an original solution to a scientific problem? (select one option by marking X)

☒

Definitely YES

☐

Probably YES

☐

It's hard to say

☐

I don't think so

☐

Definitely NOT

- B. After reading the dissertation, do you agree that the candidate has general theoretical knowledge in the discipline of Technical Information Technology and Telecommunications?

☒

Definitely YES

☐

Rather YES

☐

It's hard to say

☐

I don't think so

☐

Definitely NOT

- C. Does the candidate have the ability to conduct research independently?

☒

Definitely YES

☐

Rather YES

☐

It's hard to say

☐

I don't think so

☐

Definitely NOT

Moreover, considering that the dissertation goes far beyond the development and evaluation of a single method and addresses the entire workflow of automated cervical cytology analysis, including segmentation, classification, data augmentation, uncertainty estimation, ensemble learning. In addition, the Author proposed and experimentally validated several original modifications of existing deep learning architectures, including Improved e.g. U-Net, SPP-SegNet, SE-DeepLabV3+. These solutions were extensively evaluated using multiple datasets, different validation protocols, and cross-dataset experiments. Therefore, I recommend awarding distinction to the doctoral dissertation ⁵.



Signature

⁴<http://isap.sejm.gov.pl/isap.nsf/DocDetails.xsp?id=WDU20190000276>

⁵ optional