Integrating Autoencoder-Based Hybrid Models into Cervical Carcinoma Prediction from Liquid-Based Cytology

Ferdaous Idlahcen[1] [a], Ali Idri[1,2] [b] and Hasnae Zerouaoui[1]

[1]*Al Khwarizmi College of Computing, Mohammed VI Polytechnic University, 43150 Ben Guerir, Morocco*
[2]*Software Project Management Research Team, ENSIAS, Mohammed V University, 10000 Rabat, Morocco*

Keywords: Uterine Cervical Neoplasms, Liquid-Based Cervical Cytology (LBCC), Squamous Cell Carcinoma (SCC), Negative for Intraepithelial Lesion or Malignancy (NILM), AI-Assisted Screening, Digital and Computational Pathology (DCP).

Abstract: Artificial intelligence (AI)-assisted cervical cytology is poised to enhance sensitivity whilst lessening bias, labor, and time expenses. It typically involves image processing and deep learning to automatically recognize pre-cancerous lesions on a given whole-slide image (WSI) prior to lethal invasive cancer development. Here, we introduce autoencoder (AE)-based hybrid models for cervical carcinoma prediction on the Mendeley-liquid-based cytology dataset. This is built on fourteen combinations of AE, DenseNet-201, and six state-of-the-art classifiers: adaptive boosting (AdaBoost), support vector machine (SVM), multilayer perceptron (MLP), decision tree (DT), k-nearest neighbors (k-NN), and random forest (RF). As empirical evaluations, four performance metrics, Scott-Knott (SK), and Borda count voting scheme, were performed. The AE-based hybrid models integrating AdaBoost, MLP, and RF as classifiers are among the top-ranked architectures, with respective accuracy values of 99.30, 99.20, and 98.48%. Yet, DenseNet-201 remains a solid option when adopting an end-to-end training strategy.

1 INTRODUCTION

Cervical cancer (CxCa) is a prominently occurring gynecologic neoplasm (Dasari et al., 2015). It implies an unregulated cell cycle and invasiveness of the cervix uteri (Dasari et al., 2015) – the lower, narrow end of the uterus. Precancerous cervical lesions are strongly associated with human papillomavirus (HPV), a viral infection spread at an 80% rate via skin-to-skin or skin-to-mucosa contact (Hu et al., 2018; Petca et al., 2020). While 80%–90% of HPV infections are transient/latent and regress by host immunity within two years spontaneously, persistent or repeated infections with strains of high-risk HPV (HR-HPV) evolve into high-grade lesions or invasiveness (Huber et al., 2021). With such a well-known causal agent and a slower disease progression, cervical cancer is regarded as preventable and the best candidate for screening principles; its morbimortality appears thereof to be declining with the licensure of HPV- vaccines and mass-screening programs (Dasari

et al., 2015; Hu et al., 2018). Howbeit, CxCa persists to be a heavy global burden, largely encountered by women in low- and middle-income countries (LMICs) – accounting for 9 out of 10 deaths and an estimated 27% rise by 2030, while increasing by only 1% in high-income countries (HICs) according to the World Health Organization (WHO) (Ginsburg et al., 2017; Woo et al., 2021). Still, cervical cancer remains the second most prevalent malignancy in women under the age of 45 in HICs despite its disparity trend (Koliopoulos et al., 2017). Reckon with the status quo, by 2030, vaccine alone would have little effect on CxCa mortality with just a 0.1% decline, yet accelerated twice-lifetime screening in conjunction with treatment would lower mortality by 34.2%, sparing 300,000–400,000 lives lost (Canfell et al., 2020; Gangopadhyay, 2022).

Liquid-based cervical cytology (LBCC) has evolved as the gold standard of CxCa screening, owing to its superior sensitivity and specificity over traditional smear cytology (SC) (Sanyal et al., 2019).

[a] https://orcid.org/0000-0001-5888-6404
[b] https://orcid.org/0000-0002-4586-4158

LBCC procedure not only offers the advantage of lessening artifacts caused by low cellularity and blood contamination, yet it permits pathologists to conduct ancillary tissue assays previously restricted to histological material (Sanyal et al., 2019; Zhang et al., 2021). Nonetheless, cervical screening is generally labor-intensive. It highly demands skilled cytologists, with conflicting findings attributed to (i) population diversity, (ii) inter-examiner discrepancy in both sampling and preparation processes, and (iii) inter-observer variability in interpreting specimens (Bao et al., 2020; Sanyal et al., 2019; Thakur et al., 2022).

Modern pathology practice is shifting toward a digital scheme. Herein, computer displays are used to evaluate scanned cytology glass slides, enabling automated AI image-analysis on tissue sections (Bao et al., 2020). In contrast to shallow machine learning (ML), the strength of neural networks resides in their ability to extract highly representative features over several layered architectures, letting them suit high-dimensional data. A performance overview achieved by various deep convolutional neural networks (dCNNs) in both branches of cervical cancer pathology, i.e. histo-and cyto- pathology, can be found in (Idlahcen et al., 2022). As ML algorithms rely heavily on optimal feature extraction and selection schemes, a hybrid learning model (HLM) built on dCNN and ML remains more appealing than single learners (SLs) due to more robust features and classification lifting both performance and interpretability (Qaid et al., 2021). Still, the amount of training data continues to have a strong impact on models' performances (Fan et al., 2022).

In tumor pathology, gathering a large amount of noiseless data with correct labeling is quite tricky due to plenty of issues that impede automated WSI analysis, such as (Khened et al., 2021): stain variability, tissue artifacts, limited representative training samples, lack of labeling during acquisition, and extraction of clinically relevant patterns. The scarcity of expert-labeled and artifacts-free data poses barriers to the broadly adopted supervised learning approaches in computational pathology (Försch et al., 2021). Another less apparent challenge is the large dimensionality of WSIs compared to existing medical imaging modalities. Typically, a glass slide of 20 mm × 15 mm yields at least a 4.8 gigapixel image at an extremely high resolution equivalent to 40× magnification on a microscope, limiting end-to-end training (Khened et al., 2021).

To handle the above drawbacks, the present paper explores (i) the use of an unsupervised learning strategy, the autoencoder (AE), to overcome supervised feature learning limitations in digital cytology, and (ii) whether HLMs surpass SLs (end-to-end) in cervical LBCC smears classification. Herein, we built and assessed fourteen architectures for differentiating healthy controls from cervical carcinoma patients on the Mendeley- LBCC WSIs.

Recall that all the empirical evaluations were performed under Scott-Knott (SK) and Borda count voting schemes. Various domains, including software engineering (Idri et al., 2016; Ottoni et al., 2019), adopted the SK algorithm to compare clusters when scoring ML techniques for parameter tuning. Ergo, we applied the SK test since (i) it selects the top non-overlapping sets and (ii) surpasses past statistical methods. Likewise, the Borda count is used to score optimally the SK-selected techniques.

The present study addresses three key research questions (RQs):
- RQ1: Do dCNN-based HLMs outperform end-to-end dCNN architecture for classifying cervical cytology WSIs?
- RQ2: Do AE-based HLMs outperform end-to-end AE architecture for classifying cervical cytology WSIs?
- RQ3: Do dCNN-based HLMs outperform AE-based HLMs?

The major contributions of this study are three-fold:
- As far as we know, this work adopts for the first time autoencoders to (i) automatically extract robust features from cervical liquid-based cytology whole-slides and (ii) address supervised feature learning limitations.
- Analyze the effect on cervical cytology classification performance by modeling fourteen various combinations of AE, dCNN, and ML/DL classifiers on the same dataset.
- Assess the performances of the proposed architectures through four measures, SK clustering, and Borda count schemes.

This document is organized as follows. Data acquisition and pre-processing are described in Section 2. Section 3 reports the implemented empirical scheme. The experimental findings and discussion are provided in Section 4. Section 5 sums up this study.

2 DATASET

Data preparation is a key asset in an ML pipeline, consisting of (i) data acquisition, (ii) data pre-processing, and (iii) data augmentation, as depicted in Fig 1.

Figure 1: Data preparation scheme.

2.1 Data Acquisition

A total of 963 hematoxylin and eosin (H&E) -stained SurePath™ liquid-based cytology WSIs were retrieved from the Mendeley data repository (Hussain et al., 2020). The specimens were collected from 460 patients in the Gynaecology and Obstetrics Department of Gauhati Medical College and Hospital. All slides were captured at 400x magnification using a Leica ICC50 HD microscope and sampled into four sets as per The Bethesda System (TBS) standards: negative for intraepithelial lesion or malignancy (NILM, 613 slides), low-grade squamous intraepithelial lesion (LSIL, 163 slides), high-grade SIL (HSIL, 113 slides), and squamous cell carcinoma (SCC, 74 slides). A board-certified pathologist reviewed patient reports as ground truth.

As our purpose is to identify which patients are healthy and which are diagnosed with cervical carcinoma, we regard SCC as the "case or carcinoma" class, whereas NILM is labeled "control or healthy".

2.2 Stain Normalization

H&E-stained tissue sections are the main pillar of anatomic pathology (Idlahcen et al., 2020). It highlights the cellular structures, allowing for convenient differentiation of the nuclear, cytoplasmic, and extracellular matrix components (Chan, 2014). While hematoxylin binds to nucleic acid and stains it blue-purple, eosin grants the cytoplasm a bright pink hue that contrasts the nuclear color (Idlahcen et al., 2020). But uneven stains are ubiquitous in samples, posing one of the biggest hurdles to whole-slide image analysis (Khened et al., 2021). To avert such color variations, tissue stain normalization techniques are required. In this study, we implemented the (Macenko et al., 2009) stain normalization approach from the StainTools (Otálora et al., 2022) Python package on all the Mendeley-LBCC slides as a preprocessing step to avert color variation-driven biases.

2.3 Data Augmentation

In this study, we applied six augmentation techniques as follows: 90-degree rotation, horizontal flip, vertical flip, random scale, gaussian noise, and brightness.

A class imbalance in the Mendeley- LBCC dataset is perceived since 63% of WSIs pertain to a "control" class. As it stands, all the samples underwent data augmentation using the six aforesaid techniques for resampling to avert such limitations as well as a misleading classification. Accordingly, we generated new data from every single slide making an overall total of 2000 images for each class.

3 EMPIRICAL DESIGN

This section depicts the empirical design of the present study. The designed architectures were shortened using acronyms.

3.1 Performance Measures

We used four metrics: accuracy (Acc), precision (Pr), recall (Re), and F1-measure (F1). Accuracy is defined as the ability to correctly detect cases from controls. While precision denotes the proportion of the cases out of the total noted cases instances, recall indicates the number of cases successfully identified out of the instances of the total case; it reduces the total controls declared under cases. F-measure ranges from 0 as its worst value to 1 as its best one and refers to the harmonic mean of precision and recall.

As for evaluation, we adopted (k=5)-fold cross-validation (fCV). Recall that cross-validation schemes give better insights into complex and unseen data at every level, averting bias issues.

3.2 Scott-Knott & Borda Count Voting Schemes

Scott-Knott is proposed by Scott and Knott in 1974 as a hierarchical clustering algorithm (Ottoni et al., 2019). Its core use is variance analysis (ANOVA) although it is extensively used to achieve multiple comparisons of treatment means for distinct homogenous overlapping groups due to its simplicity yet robustness. Further, Borda count is adopted to pick the ideal architecture given four metrics with equal weight. Although other candidates or options could be picked instead of the bulk-favored option - the consensus-based voting process is the inverse of

the majority system. Recall that the Borda count voting system was performed to guarantee that no biases existed in the selection of any metric.

3.3 Experimental Scheme

The empirical scheme followed throughout this study is inspired by prior research in (Idri et al., 2016; Lahmar et al., 2022), involving three steps as follows:

- Assess the accuracy of each variant of the 14 architectures through Mendeley- LBCC dataset: one dCNN end-to-end architecture, six dCNN-based hybrid architectures, one AE end-to-end architecture, and six AE-based hybrid architectures.
- Cluster the designed architectures using the Scott-Knott algorithm, then select the SK top-cluster as per accuracy.
- Rank the designed architectures of the SK top-cluster using the Borda count voting system as per four performance measures, i.e. accuracy, precision, recall, and F1-measure. At last, select the top architecture.

3.4 Configuration

We built 14 architectures consisting of (i) end-to-end DenseNet-201. (Idlahcen et al., 2022) reports on preliminary work over the same dataset that led to the selection of DenseNet-201 as dCNN for this study.; (ii) six dCNN-based hybrid architectures involving DenseNet-201 as FE with respective six classifiers (AdaBoost, SVM, MLP, DT, k-NN, and RF); (iii) an end-to-end AE; and (iv) six AE-based hybrid architectures involving AE as FE with the same classifiers. All are designed to achieve a binary classification on the Mendeley- LBCC dataset. Herein, the following configurations were adopted:

- Since the default input size differs amongst dCNNs, we downsized all the images from an original size of 2048×536 px. into 224x224 px. to match the processed size when implementing a DenseNet-201 network.
- To avoid repetitions throughout the process, NumPy files (.npz) were used to store the resized images.
- We used Keras and TensorFlow frameworks as deep learning backends - particularly for end-to-end architectures. As per hybrid ones, we used the Scikit-learn library to implement the default configuration of the six classifiers.

All empirical schemes were performed using Google Colab's TPU.

3.5 Acronyms

For the convenience of the reader, we shorten the name of each variant as follows: DesNet for DenseNet-201; DERF for DenseNet-201 + RF; DEAda for DenseNet-201 + AdaBoost; DEMLP for DenseNet-201 + MLP; DETREE for DenseNet-201 + DT; DEKNN for DenseNet-201 + k-NN; DESVM for DenseNet-201 + SVM; AuEn for AutoEncoders; AuEnRF for AutoEncoders + RF; AuEnAda for AutoEncoders + AdaBoost; AuEnMLP for AutoEncoders + MLP; AuEnTREE for AutoEncoders + DT; AuEnKNN for AutoEncoders + k-NN; and AuEnSVM for AutoEncoders + SVM.

4 RESULTS & DISCUSSION

This section presents the empirical findings of the proposed designs. As stated, four performance metrics were used for assessment. Initially, the accuracy of DenseNet-201is compared against the hybrid architectures by a set of classifiers, each of which is conducted individually in conjunction with DenseNet-201 as a feature extractor. Likewise, end-to-end AE with hybrid architectures and dCNN-based hybrid architectures with those based on AE respective per each classifier. Then, the SK statistical test is performed to cluster the elected techniques. At last, the architectures of the SK top-cluster are ranked using the Borda count voting system.

4.1 Do dCNN-Based HLMs Outperform End-to-end dCNN Architecture for Classifying Cervical Cytology WSIs?

Table 1 displays the accuracy values of (i) end-to-end DenseNet-201 and (ii) dCNN-based HLMs, on augmented Mendeley- LBCC dataset. Through the results obtained:

- The end-to-end outperformed the others with an accuracy value of 99.66%.
- The HLM integrating SVM scored the worst, with an accuracy value of 83.04%.
- The remaining architectures, i.e. DenseNet-201 + AdaBoost, DenseNet-201 + MLP, DenseNet-201 + DT, DenseNet-201 + k-NN, and DenseNet-201 + RF, had an accuracy rating greater than 95%.

Table 1: Accuracy values of dCNN-based and AE-based end-to-end and hybrid architectures.

dCNN Archi.	Acc [%]	AE Archi.	Acc [%]
DesNet	99.66	AuEn	78.50
DEAda	97.88	AuEnMLP	99.20
DESVM	83.04	AuEnAda	98.48
DEMLP	97.80	AuEnSVM	51.06
DETREE	95.12	AuEnTREE	95.40
DEKNN	96.38	AuEnKNN	94.14
DERF	98.00	AuEnRF	99.30

Based on accuracy values, the seven architectures were clustered using the SK test as displayed in Fig 2. Through this figure, we notice that:

- Cluster 1 got just one architecture, i.e. end-to-end DenseNet-201, which performs the best out of all our models.
- The elements of the second cluster comprise three dCNN-based HLMs, (i) DenseNet-201 + RF, (ii) DenseNet-201 +AdaBoost, and (iii) DenseNet-201 + MLP, all of which have an accuracy greater than 96%.
- The third and fourth clusters each feature one architecture only, namely (i) DenseNet-201 + DT and (ii) DenseNet-201 + k-NN respectively. The two models' accuracy range from 95.12 to 96.38%.
- The last cluster is made up of the dCNN-based HLM integrating SVM, which performs the worst out of all our models.

Recall that Borda Count is not required to rank the models since the first cluster contains just one architecture.

4.2 Do AE-Based HLMs Outperform End-to-end AE Architecture for Classifying Cervical Cytology WSIs?

Table 1 displays the accuracy values of (i) end-to-end AE and (ii) AE-based HLMs, on augmented Mendeley- LBCC dataset. Through the results obtained:

- The HLM integrating RF scored the best, with an accuracy value of 99.30%.
- The HLM integrating SVM scored the worst, with an accuracy value of 51.06%.
- Except for the "AE + SVM", a significant variance in performance between end-to-end and hybrid architectures is perceived.
- The hybrid performance has improved in comparison to some of the prior dCNN-

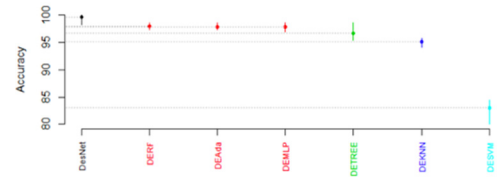based HLMs, notably for AdaBoost, MLP, and RF.

Figure 2: SK test results for the dCNN-based architectures.

In this sub-section, the SK test serves the same purpose as in RQ1. Through Fig 3, we notice four clusters including:

- The best cluster comprises three AE-based HLMs, (i) AE + RF, (ii) AE + MLP, and (iii) AE + AdaBoost, all of which have an accuracy greater than 98%.
- Both 'AE + DT' and 'AE + k-NN' come second given an accuracy range from 94.14 to 95.4%.
- The third cluster features the end-to-end AE given an accuracy value under the 80%.
- The last cluster features also one architecture only, i.e. the AE-based HLM integrating SVM, which performs the worst out of all our models.
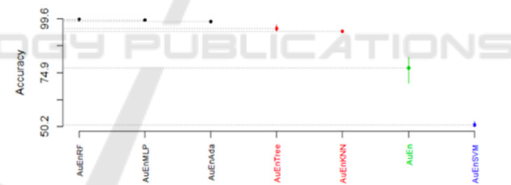- Except for the "AE + SVM", HLMs outperform the end-to-end architecture.

Figure 3: SK Test results for the AE-based end-to-end and hybrid architectures.

Table 2: Performance criteria values and Borda count ranking of the AE-based architectures belonging to the SK top-cluster.

Archi.	AE + RF	AE + MLP	AE + AdaBoost
Rank	1	2	3
Scores	11	9	4
Acc [%]	99.30	99.20	98.48
Pr [%]	99.65	99.10	98.35
Re [%]	98.96	99.30	98.60
F1 [%]	99.30	99.20	98.47

Next, the Borda count voting system was used to rank the proposed architectures belonging to the SK top-cluster. Herein, HLMs integrating AdaBoost, MLP, and RF as classifiers are statistically similar as

per cluster 1, indicating that all were used for the ranking according to the four performance measures. The related performance scores and Borda count ranking are depicted in Table 2. The findings are as follows:

- HLM with RF is ranked top.
- HLM with MLP comes second with a close score.
- HLM with AdaBoost is ranked last.

4.3 Do dCNN-Based HLMs Outperform AE-Based HLMs?

Table 1 summarizes the obtained accuracy values on the augmented Mendeley- LBCC dataset. Through the results, we notice that:

- Except for (i) end-to-end AE, (ii) AE + SVM, and (iii) dCNN + SVM, all the proposed architectures have an accuracy value superior to 95%.
- Among the 14 designs, DenseNet-201 performs the best with an accuracy value of 99.66%.
- AE-based HLMs integrating RF, AdaBoost, and MLP, perform as well favorably, with accuracy values ranging from 98.48 to 99.30%.
- All the HLMs incorporating SVM perform poorly, with the paring 'AE + SVM' yielding the worst accuracy value of 51.06%.
- Except for 'AE + k-NN', the remaining HLMs yielded accuracy values ranging from 95.12 to 98%.
- End-to-end AE performs poorly in contrast to DenseNet-201, with an accuracy rating of 78.5%.

The SK test fulfills the same purpose as in RQ1/RQ2. Through Fig 4, we notice five clusters including:

- The best cluster is made up of seven designs. All, apart from end-to-end DenseNet-201, are HLMs – particularly built with AdaBoost, MLP, and RF classifiers.
- The second cluster comprises four HLMs of k-NN and DT as classifiers only.
- The last clusters are made up of poorly performing architectures, namely end-to-end AE and HLMs built on SVM.

Next, the Borda count voting scheme was performed. Table 3 summarizes the performance scores and ranking of the SK top-cluster-related models. The findings are as follows:

- AE-based HLMs are highly ranked, with the RF classifier performing the best.
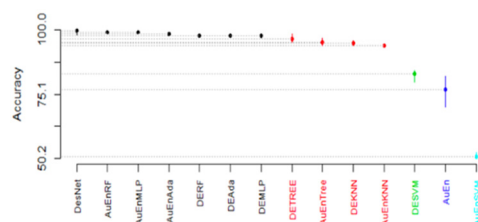

Figure 4: SK test results for all the proposed architectures.

- The 'AE + MLP' receives a similar score as DenseNet-201, ranking both seconds.
- AE-based HLM with AdaBoost is ranked third.
- The remaining are built on DenseNet-201 with RF, AdaBoost, and MLP.

Here, incorporating AE demonstrated its efficacy in classification tasks within cervical computational pathology. It is consistent with the fact that extracted features supplied as input to the classifiers are more informative and, ergo, cervical lesions are better distinguished. When paired with RF, the classification accuracy improves. One of the appealing benefits of RF is it searches for the relevant features among a random subset of pathological ones, in which complex nuclear elements (intended to identify abnormalities) could be wasted. Instead, DenseNet-201 remains a viable choice as an end-to-end strategy over whole-slide imaging for its structure adapted to prevent feature redundancy while employing fewer parameters.

5 CONCLUSION

The present paper proposed AE-based hybrid learning models for cervical cancer screening and investigated the impact of fourteen combinations on classification performance. All the architectures were evaluated under four key metrics, Scott-Knott, and Borda count schemes over Mendeley- LBCC WSIs. The main findings are as follows:

- RQ1: Do dCNN-based HLMs outperform end-to-end dCNN architecture for classifying cervical cytology WSIs? As per accuracy, the end-to-end dCNN outperforms the hybrid architectures. The SK test revealed the optimum cluster as having just such one architecture.
- RQ2: Do AE-based HLMs outperform end-to-end AE architecture for classifying cervical cytology WSIs? Except for the AE-based hybrid architecture integrating SVM as a classifier, the AE-based HLMs surpass

the end-to-end AE by a wide margin. Recall that AE with RF, MLP, and Adaboost classifiers come first, second, and third in the Borda count ranking, respectively.

- RQ3: Do dCNN-based HLMs outperform AE-based HLMs? As per Borda count, the AE-based HLMs are among the top 3 ranked architectures, whereas dCNN-based HLMs are all rated after the AE-based designs. Ergo, the feature extractions are more successful when the AE is implemented.

Table 3: Performance criteria values and Borda count ranking of the architectures belonging to the SK top-cluster.

Archi.	R.; Sc.	Acc [%]	Pr [%]	Re [%]	F1 [%]
AuEn RF	1; 25	99.30	99.65	98.96	99.30
Des Net	2; 23	99.66	99.89	98.53	99.10
AuEn MLP	3; 23	99.20	99.10	99.30	99.20
AuEn Ada	4; 17	98.48	98.35	98.60	98.47
DE RF	5; 10	98.00	98.35	97.67	98.01
DE Ada	6; 8	97.88	97.85	97.91	97.87
DE MLP	7; 6	97.80	97.90	97.74	97.80

The present study is limited by the cost of training DL models and the difficulty of interpreting their predictions. Further validation is required to ensure their reliability. Another weakness is the total number of images remains relatively small. Although the slides used were collected from three distinguished medical diagnostic centers, most of the study population was Indian, locally trained, so generalizability to other populations and settings is not known. To be objective, the usefulness of the proposed architectures will be concretely evaluated in future work on the Herlev dataset to confirm or refute this study's findings regarding conventional cytology. Extending it toward a multi-class problem mimicking pathologists for screening cervical intraepithelial neoplasia is also necessary.

ACKNOWLEDGEMENTS

REFERENCES

Bao, H., Bi, H., Zhang, X., Zhao, Y., Dong, Y., Luo, X., Zhou, D., You, Z., Wu, Y., Liu, Z., Zhang, Y., Liu, J., Fang, L., & Wang, L. (2020). Artificial intelligence-assisted cytology for detection of cervical intraepithelial neoplasia or invasive cancer: A multicenter, clinical-based, observational study. *Gynecologic Oncology*, *159*(1), 171–178. doi: 10.1016/J.YGYNO.2020.07.099

Canfell, K., Kim, J. J., Brisson, M., Keane, A., Simms, K. T., Caruana, M., Burger, E. A., Martin, D., Nguyen, D. T. N., Bénard, É., Sy, S., Regan, C., Drolet, M., Gingras, G., Laprise, J. F., Torode, J., Smith, M. A., Fidarova, E., Trapani, D., … Hutubessy, R. (2020). Mortality impact of achieving WHO cervical cancer elimination targets: a comparative modelling analysis in 78 low-income and lower-middle-income countries. *Lancet (London, England)*, *395*(10224), 591. doi: 10.1016/S0140-6736(20)30157-4

Chan, J. K. C. (2014). The Wonderful Colors of the Hematoxylin–Eosin Stain in Diagnostic Surgical Pathology. *Http://Dx.Doi.Org/10.1177/10668969135 17939*, *22*(1), 12–32. doi: 10.1177/1066896913517939

Dasari, S., Wudayagiri, R., & Valluru, L. (2015). Cervical cancer: Biomarkers for diagnosis and treatment. *Clinica Chimica Acta*, *445*, 7–11. doi: 10.1016/J.CCA. 2015.03.005

Fan, F. J., & Shi, Y. (2022). Effects of data quality and quantity on deep learning for protein-ligand binding affinity prediction. *Bioorganic & Medicinal Chemistry*, *72*, 117003. doi: 10.1016/J.BMC.2022.117003

Försch, S., Klauschen, F., Hufnagl, P., & Roth, W. (2021). Artificial Intelligence in Pathology. *Deutsches Ärzteblatt International*, *118*(12), 199. doi: 10.3238/ARZTEBL.M2021.0011

Gangopadhyay, A. (2022). Elimination of cervical cancer as a public health problem—how shorter brachytherapy could make a difference during COVID-19. *Ecancermedicalscience*, *16*. doi: 10.3332/ECANCER. 2022.1352

Ginsburg, O., Bray, F., Coleman, M. P., Vanderpuye, V., Eniu, A., Kotha, S. R., Sarker, M., Huong, T. T., Allemani, C., Dvaladze, A., Gralow, J., Yeates, K., Taylor, C., Oomman, N., Krishnan, S., Sullivan, R., Kombe, D., Blas, M. M., Parham, G., … Conteh, L. (2017). The global burden of women's cancers: an unmet grand challenge in global health. *Lancet (London, England)*, *389*(10071), 847. doi: 10.1016/S0140-6736(16)31392-7

Hu, Z., & Ma, D. (2018). The precision prevention and therapy of HPV-related cervical cancer: new concepts

and clinical implications. *Cancer Medicine*, *7*(10), 5217. doi: 10.1002/CAM4.1501

Huber, J., Mueller, A., Sailer, M., & Regidor, P. A. (2021). Human papillomavirus persistence or clearance after infection in reproductive age. What is the status? Review of the literature and new data of a vaginal gel containing silicate dioxide, citric acid, and selenite. *Women's Health*, *17*. doi: 10.1177/1745506521 1020702

Hussain, E., Mahanta, L. B., Borah, H., & Das, C. R. (2020). Liquid based-cytology Pap smear dataset for automated multi-class diagnosis of pre-cancerous and cervical cancer lesions. *Data in Brief*, *30*, 105589. doi: 10.1016/J.DIB.2020.105589

Idlahcen, F., Himmi, M. M., & Mahmoudi, A. (2020). *CNN-based Approach for Cervical Cancer Classification in Whole-Slide Histopathology Images*.

Idlahcen, F., Mboukou, P., Zerouaoui, H., & Idri, A. (2022). Whole-slide Classification of H&E-stained Cervix Uteri Tissue using Deep Neural Networks. *Proceedings of the 14th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, 322–329. doi: 10.5220/0011578700003335

Idri, A., Hosni, M., & Abran, A. (2016). Improved estimation of software development effort using Classical and Fuzzy Analogy ensembles. *Applied Soft Computing*, *49*, 990–1019. doi: 10.1016/J.ASOC.2016.08.012

Khened, M., Kori, A., Rajkumar, H., Krishnamurthi, G., & Srinivasan, B. (2021). A generalized deep learning framework for whole-slide image segmentation and analysis. *Scientific Reports 2021 11:1*, *11*(1), 1–14. doi: 10.1038/s41598-021-90444-8

Koliopoulos, G., Nyaga, V. N., Santesso, N., Bryant, A., Martin-Hirsch, P. P. L., Mustafa, R. A., Schünemann, H., Paraskevaidis, E., & Arbyn, M. (2017). Cytology versus HPV testing for cervical cancer screening in the general population. *The Cochrane Database of Systematic Reviews*, *2017*(8). doi: 10.1002/14651858.CD008587.PUB2

Lahmar, C., & Idri, A. (2022). Deep hybrid architectures for diabetic retinopathy classification. *Https://Doi.Org/10.1080/21681163.2022.2060864*. doi: 10.1080/21681163.2022.2060864

Macenko, M., Niethammer, M., Marron, J. S., Borland, D., Woosley, J. T., Guan, X., Schmitt, C., & Thomas, N. E. (2009). A method for normalizing histology slides for quantitative analysis. *Proceedings - 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, ISBI 2009*, 1107–1110. doi: 10.1109/ISBI.2009.5193250

Otálora, S., Marini, N., Podareanu, D., Hekster, R., Tellez, D., Laak, J. Van Der, Müller, H., & Atzori, M. (2022). stainlib: a python library for augmentation and normalization of histopathology H&E images. *BioRxiv*, 2022.05.17.492245. doi: 10.1101/2022.05.17.492245

Ottoni, A. L. C., Nepomuceno, E. G., de Oliveira, M. S., & de Oliveira, D. C. R. (2019). Tuning of reinforcement learning parameters applied to SOP using the Scott–Knott method. *Soft Computing*, *24*(6), 4441–4453. doi: 10.1007/S00500-019-04206-W

Petca, A., Borislavschi, A., Zvanca, M., Petca, R.-C., Sandru, F., & Dumitrascu, M. (2020). Non-sexual HPV transmission and role of vaccination for a better future (Review). *Experimental and Therapeutic Medicine*, *20*(6), 1–1. doi: 10.3892/ETM.2020.9316

Qaid, T. S., Mazaar, H., Al-Shamri, M. Y. H., Alqahtani, M. S., Raweh, A. A., & Alakwaa, W. (2021). Hybrid Deep-Learning and Machine-Learning Models for Predicting COVID-19. *Computational Intelligence and Neuroscience*, *2021*. doi: 10.1155/2021/9996737

Sanyal, P., Barui, S., Deb, P., & Sharma, H. C. (2019). Performance of A Convolutional Neural Network in Screening Liquid Based Cervical Cytology Smears. *Journal of Cytology*, *36*(3), 146. doi: 10.4103/JOC.JOC_201_18

Thakur, N., Alam, M. R., Abdul-Ghafar, J., & Chong, Y. (2022). Recent Application of Artificial Intelligence in Non-Gynecological Cancer Cytopathology: A Systematic Review. *Cancers*, *14*(14), 3529. doi: 10.3390/CANCERS14143529

Woo, Y. L., Gravitt, P., Khor, S. K., Ng, C. W., & Saville, M. (2021). Accelerating action on cervical screening in lower- and middle-income countries (LMICs) post COVID-19 era. *Preventive Medicine*, *144*, 106294. doi: 10.1016/J.YPMED.2020.106294

Zhang, X. H., Ma, S. Y., Liu, N., Wei, Z. C., Gao, X., Hao, Y. J., Liu, Y. X., Cai, Y. Q., & Wang, J. H. (2021). Comparison of smear cytology with liquid-based cytology in pancreatic lesions: A systematic review and meta-analysis. *World Journal of Clinical Cases*, *9*(14), 3308. doi: 10.12998/WJCC.V9.I14.3308.