

Joint Classification and Supervised Contrastive Representation Learning for Cervical Cancer Detection in Histopathological Images

Nebras Sobahi

nsobahi@kau.edu.sa

King Abdulaziz University

Osman Fatih Koparir

Firat University

Abdulkadir Sengur

Firat University

Berrin Tarakci Gencer

Firat University

Research Article

Keywords: Cervical cancer classification, histopathological images, supervised contrastive learning, deep learning, balanced classification accuracy, explainable artificial intelligence

Posted Date: October 13th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-8610859/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Joint Classification and Supervised Contrastive Representation Learning for Cervical Cancer Detection in Histopathological Images

Nebras Sobahi¹, Osman Fatih Koparir², Abdulkadir Sengur³, Berrin Tarakci Gencer²

¹King Abdulaziz University, Faculty of Engineering, Department of Electrical and Electronics Engineering, P.O. Box 80204, Jeddah 21589, Saudi Arabia (nsobahi@kau.edu.sa)

²Firat University, Faculty of Veterinary Medicine, Department of Histology and Embryology, Elazig, Turkey (241304208@firat.edu.tr, btarakci@firat.edu.tr)

³Firat University, Technology Faculty, Electrical and Electronics Engineering Department, Elazig, Turkey (ksengur@firat.edu.tr)

Abstract

Correct categorization of histopathological images of the cervix is very crucial if one is to identify and treat any emerging cases of cervical cancer on time. However, this is not an easy task, given that there is so much variability within one class while there are only minute distinctions between normal and pathological tissues. In our study, we therefore develop a deep learning approach that combines contrast learning with the traditional approach of using the cross-entropy loss function to correctly categorize histopathological images of the cervix into normal or pathological tissues. Our approach is based on ResNet-50 and seeks to maximize both learning of appropriate features and correct classification of images into one of two classes: normal or pathological cervical tissues. Performance is evaluated using stratified five-fold cross-validation on a variety of metrics, including balanced accuracy, precision, recall, F1, ROC-AUC, as well as average precision. For the histopathological dataset, CAISHI, this approach in all cases produces very high balanced accuracy of 96.90 ± 0.23 on average. High precision, recall, and F1 values indicate a high accuracy of abnormal region detection. In comparison to existing machine learning techniques, CNN models, transformer networks, and a few-shot learning strategy, the proposed approach demonstrates improved robustness and generalization capabilities. From the results, it can be inferred that the integration of supervised contrastive learning with deep convolutional feature extraction leads to a dramatic improvement in the representation ability and classification accuracy in the field of cervical histopathology images. The proposed framework can indeed be used as a robust interpretation solution with high potential for medical support in cervical cancer diagnosis.

Keywords: Cervical cancer classification, histopathological images, supervised contrastive learning, deep learning, balanced classification accuracy, explainable artificial intelligence.

1. Introduction

Cervical cancer is a malignant neoplasm that originates in the cervical cells, primarily induced by chronic infection with high-risk human papillomavirus (HPV) [1]. It can be mostly avoided with HPV vaccination and regular screening methods like HPV testing and Pap smears [2]. Cervical cancer is one of the most common types of cancer that women get around the world. Around 660,000 new cases were found around the world in 2022, and about 350,000 women died from the disease [3]. More than 90% of these deaths happened in low- and middle-income countries, which shows that there are big differences in who can get vaccinated, screened, diagnosed early, and treated effectively [4]. The World Health Organization has reported that cervical cancer can be stopped from being a public health problem with the right prevention methods. Its global elimination goals include high rates of HPV vaccination,

widespread screening of women with high-performance tests, and prompt treatment of precancerous lesions and invasive cancer [5].

Deep learning-based methods have been used more and more in the last few years to find cervical cancer in histopathology. These methods aim to find cancerous and precancerous tissue patterns by looking at digitized histopathology slides automatically. Whole-slide images (WSI) are so big that models usually work with smaller image patches and combine their predictions at the slide level using methods like Convolutional Neural Networks (CNN) and multiple-instance learning. Histopathological systems that use deep learning have shown promise in helping pathologists by making diagnoses more accurate, repeatable, and efficient. Hemalatha et al. [6] proposed a self-supervised learning framework utilizing the SimCLR model to extract representations from unlabeled cell data, followed by a Hybrid Krill Herd Optimization (HKHO) technique for feature selection and clustering. The study utilized the SIPaKMeD dataset for cervical cancer cells, along with various unlabelled datasets like CytoImageNet for pre-training, and achieved an accuracy of 97.63% on the SIPaKMeD dataset. The limitations of the proposed study include the fact that features learned via self-supervised learning may not always be directly discriminative for the specific classification task without refinement, and the HKHO algorithm introduces high computational complexity.

Fei et al. [7] designed a two stage solution. In the first stage, the authors applied a Local Graph Attention Network (GAT) for the classification of cervical cells in tiles of images, followed by the application of a Global GAT driven by supervised contrastive learning for WSI classification. They had 10,701 field-of-view images of the Shanghai Medical College Hospital along with other institutions. The authors obtained 85.26% accuracy and 92.76% AUC for WSI classification. The method has limitations in terms of the handling of fully connected graphs, over-smoothing of nodes, required computation, and possible need for adaptation when applied to other datasets. Pacal et al. [8] presented a hybrid deep learning model that combines InceptionNeXt blocks for extracting local features, improved Swin Transformer blocks for capturing global context, and a Residual MLP (ResMLP). They validated their model on the NCT-CRC-HE-100K and Kather-5K histopathological image datasets and reported an accuracy of 99.96% and 99.06%, respectively. These models also reported the following limitations: the high computational requirements of the Swin Transformer blocks, a lack of clinical variability in publicly available data, and a tendency to generate false positives in tumor classification based on morphological ambiguity in benign lesions. AbdulAzeem et al. [9] proposed a computer-aided diagnosis system that utilizes Vision Transformers as classifiers, along with Majority Fusion for result integration, coupled with SHAP for explanation. The model was evaluated on the DTU/Herlev and SIPaKMeD datasets, reporting accuracies of 99.22% and 99.80%, respectively. The limitations of the current state of the art are said to be large computational requirements, along with GPUs for the training of Vision Transformers, which may not be feasible for the integration of these models into current clinical practices. AlZobi et al. [10] introduced the PSO-SV system that adaptively combined predictions of the MobileNetV2, VGG19, DenseNet121, Swin Transformer, and Vision Transformer using particle swarm-optimized weights. For validation, it was tested on an endometrial cancer dataset obtained from TCGA-UCEC that was comprised of 11,977 images and a second dataset that contained four classes. The obtained results were 99.67% and 99.00%, respectively. Limitations include the high training time attributed to the iterative PSO optimization process and the observation that standalone Transformer-based models performed significantly worse than CNNs for this specific task.

Mohanta et al. [11] performed a detailed survey on automatic cervical irregularity diagnosis methods using various microscopic imaging techniques like Pap smear, histopathology, and fluorescence imaging techniques. Several public and private databases like Herlev, SIPaKMeD, and ISBI challenges have been explored in this survey, although it neither reported any specific accuracy in experiments nor provided these databases specifically for experiments. Several key challenges have been pointed out in this survey, such as the few-shot learning technique being less explored, mostly

unbalanced databases, and further exploration in multimodality imaging techniques being required. A hybrid deep learning approach has been proposed by Sharma et al. [12], which uses combination techniques of fine-tuning and feature fusion of pre-trained models like DenseNet201 and InceptionV3 models. They experimented on a dataset consisting of Pap smear images belonging to five classes according to SIPaKMeD classes and reported an accuracy of 96.54%, although limitations such as high processing overhead due to the complexity of this proposed technique and further exploration in larger and diverse databases are required to make it robust and generalizable. Ain et al. [13] proposed a feature learning technique using a genetic programming-based approach called MD-FLGP, which uses multiple descriptors like LBP, HOG, and SIFT for malignancy level diagnosis in breast cancer. They experimented on BreakHis histopathological databases and reported accuracies of 83-84% in magnifications with less intensity compared to deep learning techniques, and improved over conventional techniques although its limitations are its reduced performances at greater magnifications compared to deep learning techniques and its high average training time of three days, although it requires only milliseconds to predict malignancy levels in histopathologies. Jeleń et al. [14] investigated a technique that utilizes morphological features and feature selection techniques such as step-wise and significant feature selections, and use Support vector machines for classifications. They experimented on a dataset consisting of LBC images acquired from Pomeranian Medical University and acquired 263 samples, and compared its performances to CNN and reported its performances well above other techniques and acquired accuracies of 93%, although limitations such as its smaller dataset sizes and biases in acquisition from single institutions have been pointed out in this paper. Vazquez et al. [15] performed a scoping review to map how machine learning and deep learning are being used for diagnosing, predicting, and treating cervical cancer. They pulled together 153 studies and, rather than presenting new experimental results themselves, they summarize what the literature shows. They point out several limitations in the field: scarce code and data sharing hurts reproducibility; few studies have external validation; challenges are thrown up by weaving AI into real clinical workflows while also safeguarding patient privacy. Yang et al. [16] proposed the CAISHI dataset, a benchmark histopathology image dataset that focuses on cervical AIS. They also conducted some baseline evaluations using models like ResNet50 and Vision Transformer. The dataset consists of 2,240 images; the highest accuracy achieved was 96.88% by ResNet50. However, this dataset has its own caveats: it contains only two classes, which are normal and AIS; the images were in PNG format, possibly with compression artifacts; and the data were from a single hospital, so it may not be very representative. Rasheed et al. [17] provided the most comprehensive review of nuclei segmentation in cervical smear images, from classic to deep learning approaches. They discussed datasets such as Herlev and ISBI, but the paper did not propose a new experiment with fresh accuracy metrics. The review pointed out several method-based limitations, including issues with clumped nuclei, noise, and diverging contrast, but also generally lacking high-quality annotated data and often imbalance. Gangrade et al. [18] developed the route of ensemble learning, blending Classic CNN, AlexNet, and SqueezeNet by averaging to classify squamous cells. On the SIPaKMeD dataset, their ensemble reached an accuracy of 94%, higher than the respective accuracies given by individual models. They further pointed out that there were limits due to the achievable gain given the limited number of available datasets, plus high computational and hardware demands linked to deploying a deep learning model. Chauhan et al. [19] implemented a hybrid learning network using progressive resizing for training ResNet-152 and VGG-16 models, followed by PCA for feature reduction and a majority voting classifier with SVM and Random Forest. The method achieved 99.29% accuracy for binary classification on the SIPaKMeD dataset and 100% accuracy on an LBC dataset. The study noted that training deep models is computationally expensive and suggested that future work is needed to integrate more advanced feature selection approaches to further improve efficiency. Raza et al. [20] proposed a framework integrating a Neural Feature Extractor based on VGG16 and an AutoInt model for feature interaction learning, classified by machine learning algorithms like KNN. Using a publicly available cervical cancer dataset of 25000 images, the KNN classifier achieved an accuracy of 99.96%. Limitations included the reliance on pre-trained weights from non-

medical domains which might limit adaptability to specialized tasks, and the potential that the dataset variability may not fully reflect diverse real-world clinical settings.

Histopathological images have made it much easier to find cervical cancer using both traditional machine learning and deep learning methods. CNN, in particular, have shown to be very good at automatically learning how to tell the difference between different visual patterns based on tissue morphology. However, most current methods mainly focus on improving the accuracy of classification and don't pay enough attention to the quality and structure of the learned feature representations. Histopathological images exhibit significant intra-class variability, nuanced inter-class distinctions, and appearance variations induced by staining, which may constrain the robustness and generalization capacity of models solely reliant on classification. Furthermore, while previous research has investigated patch-based learning, whole-slide image analysis, and transfer learning utilizing deep convolutional backbones, there exists a scarcity of methodologies that explicitly mandate discriminative representation learning at the embedding level. A lot of current models only use cross-entropy loss, which pushes for correct label prediction but doesn't directly limit the feature space to make sure that classes are close together and separate from each other. This can slow things down in difficult cases, especially when the visual cues separating healthy from malignant cervical tissue are subtle. In this respect, the current study proposes a deep learning framework for detecting cervical cancer in histopathology slides that improves both the model classification and representation learning. The proposed approach integrates a convolutional feature extractor together with supervised contrastive learning and is capable of chiseling out more structured yet discriminative embeddings while accurately predicting the class. Using a supervised contrastive loss, samples of the same class will be drawn closer together with respect to embedding space, while those of different classes move apart and yield more robust, interpretable representations. Our overall approach remains quite straightforward: image processing, deep feature extraction, simultaneous optimization regarding both classification and contrastive objectives, and extensive performance evaluation. A pre-trained ResNet-50 backbone uses high-level features from full-resolution histopathology images. For its classification procedure, the cross-entropy loss function was adopted, while the supervised contrastive loss is implemented to leverage the features at this level for representation learning.

The contributions of this work are listed below:

1. We propose a cervical cancer detection model for histopathological images by integrating classification with supervised contrastive learning to jointly improve both accuracy as well as representation quality.
2. We provide a structured way of learning embeddings where compactness within classes and separability among classes are ensured, filling the gaps of the conventional cross-entropy based histopathological prediction models.
3. In addition to the above-mentioned techniques for ensuring robustness to variations in image shape and staining characteristics in cervical tissue imaging, we make use of a convolutional backbone pretrained model, feature normalization, and supervised contrastive loss.
4. A thorough performance and embedding discriminate ability assessment strategy based on retrieval-based metrics is suggested, and this would provide a clearer understanding about the learning feature space.

2. Materials and Methods

This study introduces an automated framework for cervical cancer detection that integrates deep CNN with sophisticated representation learning techniques to conduct objective analyses of histopathological images. Fig. 1 shows the illustration of the proposed framework.

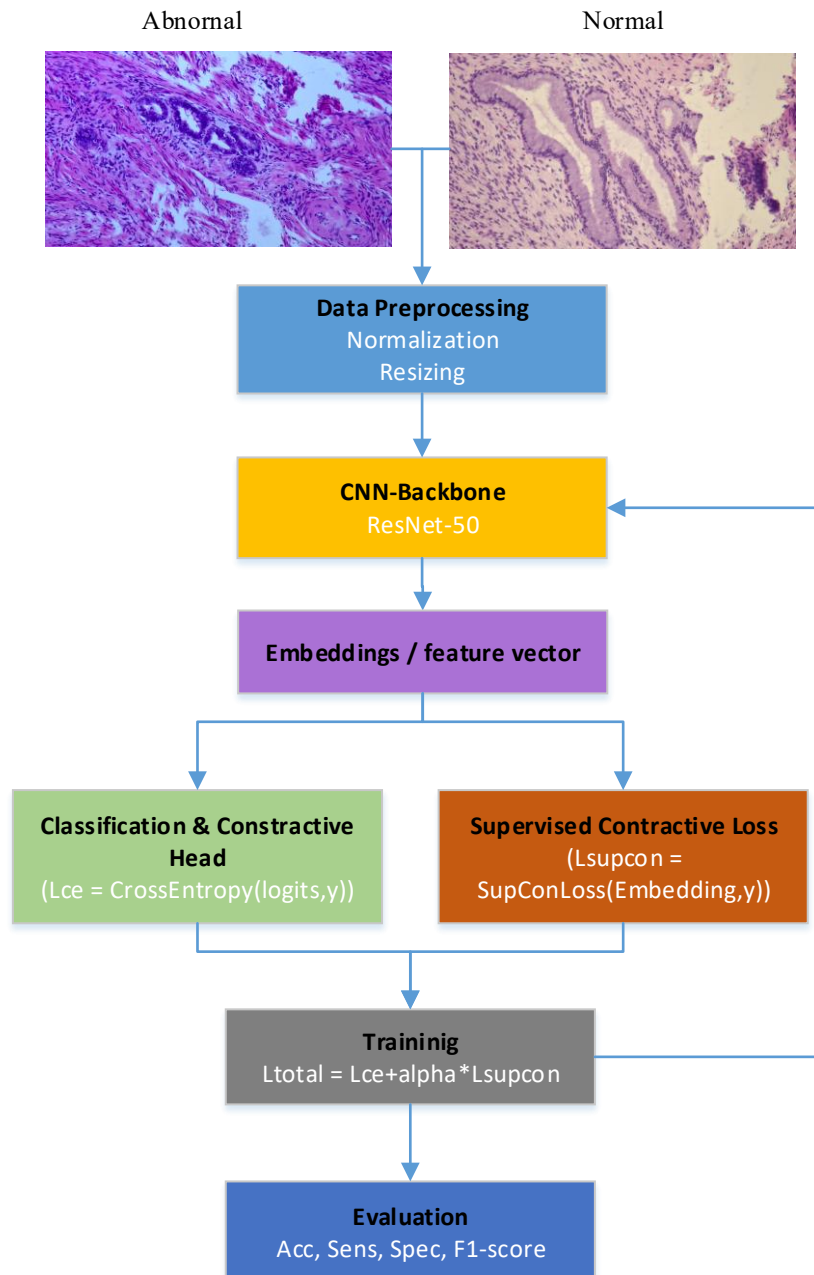


Fig. 1 The illustration of the proposed method

It applies directly to whole slide images of cervical tissue, benefiting from a deep convolutional backbone that is pretrained to extract high-level morphological information from images without requiring handcrafted descriptors. The deep feature describes well the intricate morphological details, such as how cells are arranged, irregularities in topography, and textural information, that indicate malignancy. The approach improves representation quality from training for both supervised classification and supervised contrastive learning simultaneously. The cross-entropy loss forces the model to be proficient at cancer classification, whereas the supervised contrastive loss further molds the embedding space, trying to bring closer points from the same category while pushing apart points from different categories. It requires a direct pipeline involving image preprocessing, deep feature extraction,

optimization of both losses, and model validation. It utilizes a data splitting approach that is stratified, employing cross-validation for robust results for a wide range of situations. Standard measures used for validation for a classifier for diagnosis—for instance, accuracy, precision, recall, specificity, and F1 score—are used, along with metrics that involve retrieval, like mean Average Precision, for understanding how effectively points lie in their respective categories as represented in their feature space.

2.1. Preprocessing

Let the histopathological dataset be defined as

$$D = \{(x_i, y_i)\}_{i=1}^N$$

where x_i denotes a full-resolution cervical histopathology image and $y_i \in \{0,1\}$ represents the corresponding class label indicating non-cancerous or cancerous tissue [21]. All images are resized and normalized according to the input requirements of the convolutional backbone. Pixel intensities are scaled to a standardized range to reduce illumination and staining variability commonly observed in histopathological slides. To ensure reliable performance estimation and to preserve class distribution, stratified sampling is employed.

2.2 Deep Feature Extraction with Convolutional Backbone

Feature extraction is performed using a ResNet-50 architecture pretrained on ImageNet [22]. Let $f_\theta(\cdot)$ denote the convolutional backbone parameterized by θ . The final fully connected layer is removed, and each input image x_i is mapped to a deep feature vector:

$$h_i = f_\theta(x_i), \quad h_i \in \mathbb{R}^d$$

To stabilize similarity computations and facilitate contrastive learning, the extracted features are L2-normalized [23]:

$$z_i = \frac{h_i}{\|h_i\|_2}$$

This normalization constrains embeddings to lie on a unit hypersphere, enabling cosine similarity to be computed via inner products and preventing feature magnitude dominance during training.

2.3 Classification Head and Cross-Entropy Loss

A linear classification head is attached to the normalized embeddings to perform binary classification [24]. The logits are computed as:

$$\ell_i = W z_i + b$$

where $W \in \mathbb{R}^{2 \times d}$ and $b \in \mathbb{R}^2$. Class probabilities are obtained using the softmax function:

$$p_i(c) = \frac{\exp(\ell_{i,c})}{\sum_{c'} \exp(\ell_{i,c'})}$$

The classification objective is defined using the cross-entropy loss:

$$L_{CE} = -\frac{1}{N} \sum_{i=1}^N \log p_i(y_i)$$

which encourages correct discrimination between cancerous and non-cancerous tissue samples.

2.4 Supervised Contrastive Representation Learning

To improve the discriminative quality of the learned feature space, supervised contrastive learning is incorporated [25]. For a mini-batch of size N , let $P(i)$ denote the set of indices corresponding to samples that share the same class label as sample i . The supervised contrastive loss is defined as:

$$L_{SupCon} = -\frac{1}{N} \left(\sum_{i=1}^N -\frac{1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp\left(\frac{z_i^T z_p}{\tau}\right)}{\sum_{a \in I/\{i\}} \exp\left(\frac{z_i^T z_a}{\tau}\right)} \right)$$

where τ is a temperature parameter controlling the sharpness of similarity distributions. This loss explicitly pulls embeddings of the same class closer while pushing embeddings of different classes apart, which is particularly beneficial for histopathological images where inter-class visual differences can be subtle.

2.5 Joint Optimization, Training, and Evaluation

The proposed framework is trained using a joint objective that combines standard supervised classification with supervised contrastive representation learning [26]. For each mini-batch, the overall loss is defined as a weighted sum of cross-entropy loss and supervised contrastive loss:

$$L = L_{CE} + \beta L_{SupCon}$$

where β controls the contribution of the contrastive term. This formulation encourages the model to learn both an accurate decision boundary for cervical cancer detection and a discriminative embedding space where samples from the same class are clustered while samples from different classes are separated.

2.6. Dataset

The CAISHI histopathological image dataset was applied in the experiments involving the classification of cervical adenocarcinoma in situ and normal cervical tissues [16]. CAISHI consists of a total of 2,240 H&E stained histopathological images obtained from the Shengjing Hospital of China Medical University. Among the images, 1,230 images are of cervical adenocarcinoma in situ, and 1,010 images represent normal cervical glands. The images are obtained from endoscopic biopsy methods using a microscope under 100 $\times$ magnification. Every image provided is of high resolution, measuring 3,840 pixels by 2,160 pixels and stored as a PNG file. The histopathological images come from 112 patients, 52 of whom had cervical adenocarcinoma in situ and 60 of whom were normal. Each patient contributed between 1 and 69 slices, depending upon tissue size. The annotations are accurate and from expert pathologists, and the study has been approved by an Institutional Review Board. This particular data set is useful for assessing deep learning algorithms because it has well-balanced resolution and relevance, as well as obvious applications for binary classification between normal and abnormal tissue. Fig. 2 shows samples of the normal and abnormal histopathological images. While the first row of Fig. 2 shows the normal sample images, the second row shows the abnormal histopathological images.

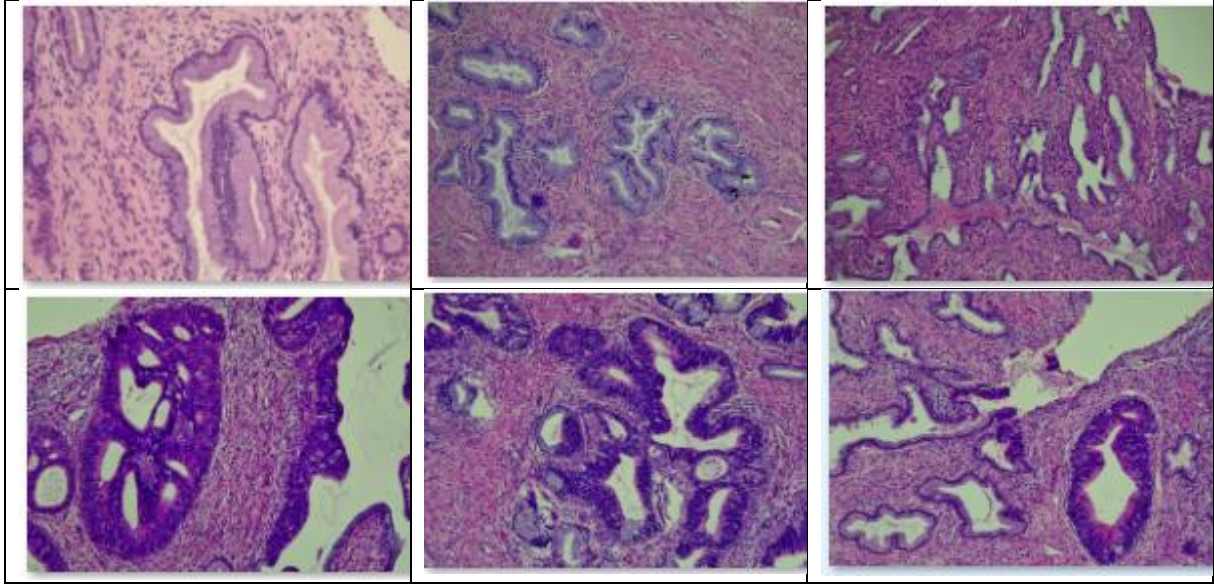


Fig. 2 Samples of normal and abnormal histopathological images

3. Experimental results

The proposed methodology was implemented in Python using PyCharm and run on a system equipped with an NVIDIA GeForce RTX 3090 GPU (24 GB RAM). A k-fold cross-validation was adopted, where k was chosen as 5 for testing of the proposed methodology. As mentioned earlier, in the experimental works, a pretrained ResNet-50 network is used as the convolutional backbone to extract features from full-resolution histopathological images. The final fully connected layer is removed, and the extracted feature vectors are L2-normalized to stabilize similarity computations. A linear classifier with two output units is applied for binary classification. Training is performed for 50 epochs with a batch size of 32 using the AdamW optimizer, a learning rate of $1e^{-4}$, and a weight decay of $1e^{-4}$. The loss function consists of cross-entropy loss and supervised contrastive loss, where the contrastive loss is scaled by $\beta = 0.2$ and uses a temperature value of 0.07. The obtained achievements were evaluated using five standard evaluation metrics namely, Accuracy, Precision, Recall, Specificity, and F1-score.

Fig 3 represents confusion matrices for each of the five folds in the cross-validation process. Beginning with Fold 1, there were 198 correct negatives and 209 correct positives, with only a few errors made, including 2 false positives and 4 false negatives. Continuing with Fold 2, there were 198 correct negatives and 210 correct positives, with only a few errors made, including 2 false positives and 3 false negatives. Turning to Fold 3, there were 196 correct negatives and 208 correct positives, with only a few errors made, including 5 false positives but only 4 false negatives. Moving on to Fold 4, there were 198 correct negatives and 209 correct positives, with only a few errors made, including 3 false positives and 3 false negatives. Finally, looking at Fold 5, there were 196 correct negatives and 209 correct positives, with only a few errors made, including 4 false positives and 3 false negatives.

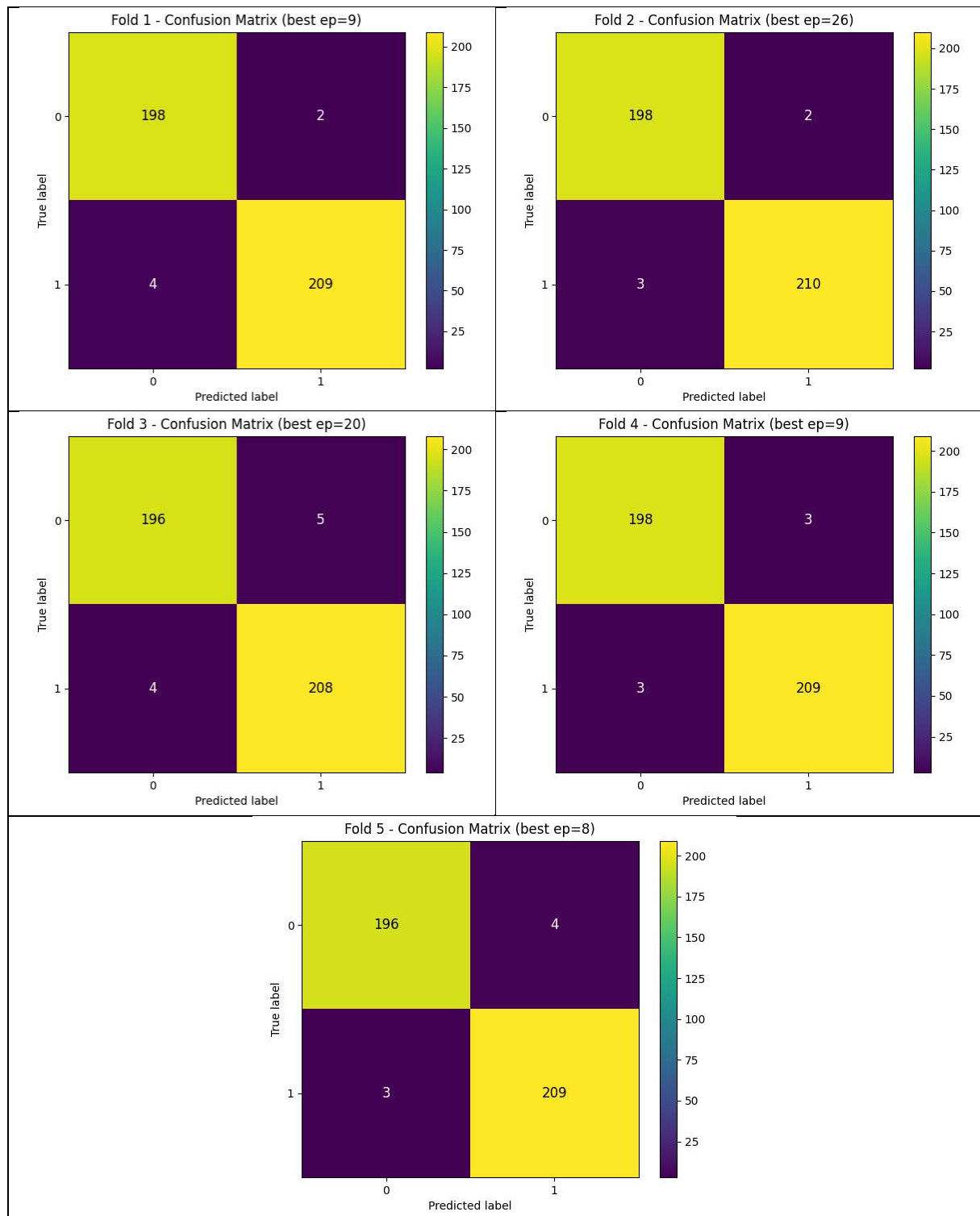


Fig. 3 Obtained confusion matrices for all folds

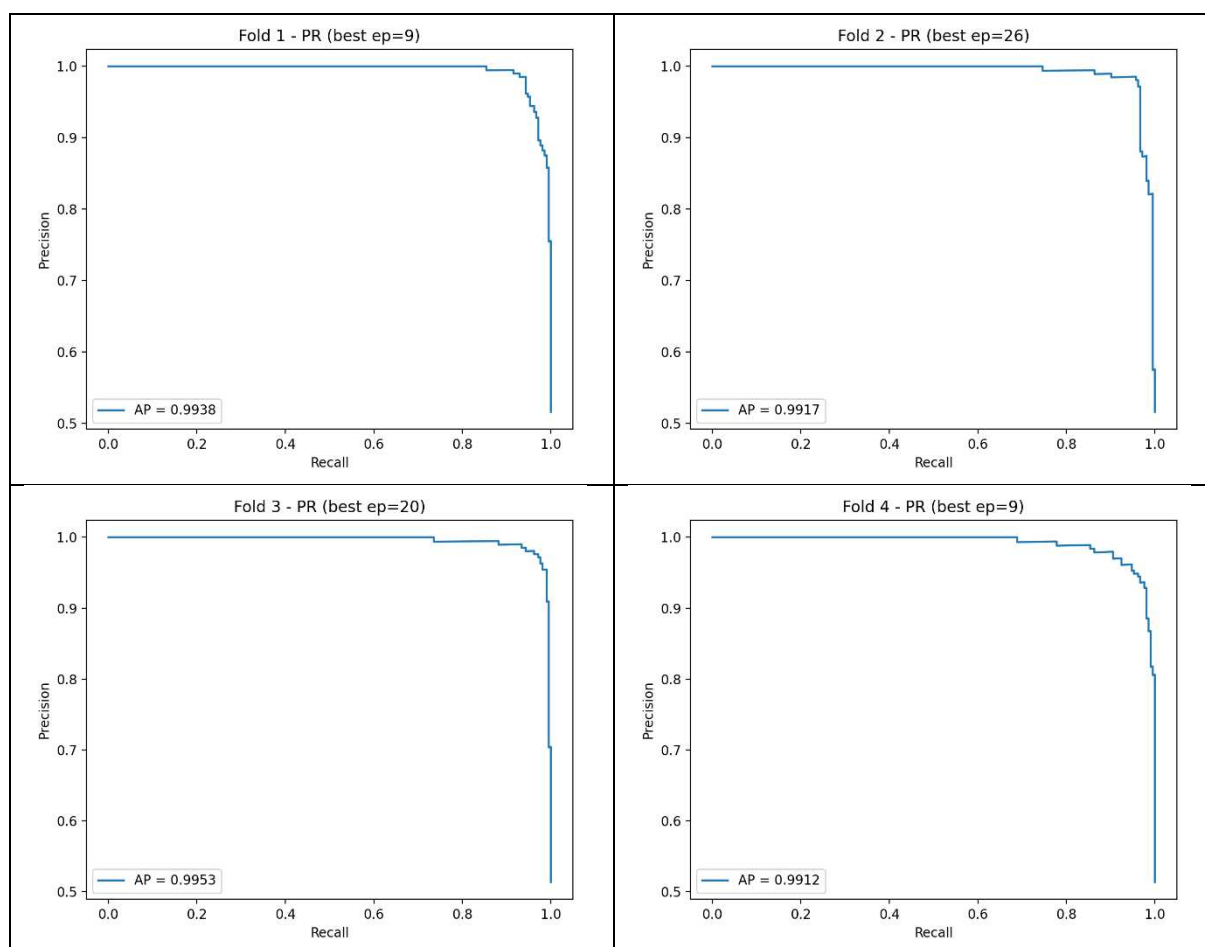
The performance evaluation results on each fold of the cross-validation process are provided in Table 1. In Fold 1, the model attained a very high accuracy of 96.61%, precision of 99.05%, and recall of 98.13%, while ensuring a specificity of 99.00%. The performance marginally increased in Fold 2, which attained an accuracy of 97.10% and an F1-score of 98.83%, along with a balance in precision and recall values. In the case of Fold 3, the model's accuracy continued to remain the same at 96.61%, although a slight decrease in precision and specificity was noticed compared to the first two folds. The results in Fold 4 continued to demonstrate equality in accuracy at 97.10%, along with a balance in the recall and

specificity values, which approached 98.6%. In Fold 5, the model's performance was the same, which recorded an accuracy of 97.10% in addition to the highest F1-score of 98.35%.

Table 1 Performance evaluation metrics for all folds

	Accuracy (%)	Precision(%)	Recall(%)	Specificity(%)	F1score(%)
Fold 1	96.61	99.05	98.13	99.00	98.58
Fold 2	97.10	99.06	98.60	99.00	98.83
Fold 3	96.61	97.66	98.11	97.51	97.88
Fold 4	97.10	98.59	98.59	98.51	98.59
Fold 5	97.10	98.12	98.59	98.00	98.35
Mean $\pm$ std	96.90 $\pm$ 0.23	98.50 $\pm$ 0.62	98.40 $\pm$ 0.22	98.40 $\pm$ 0.58	98.45 $\pm$ 0.35

The Precision-Recall curves plot precision against recall for a variety of threshold values. They are more informative and useful in assessing the performance of classifiers on medical datasets, where there is a possibility of class imbalance. A curve close to the top-right corner simultaneously represents a high value for precision and recall. In Fig. 4, it is observed that for each fold, the PR curves are consistently high, with precision close to 1.0 over a large number of recall values. This is an indication that high precision is obtained with effective discovery of most abnormal instances. Average precision scores are also high and with less variation among different folds. These indicate that the model is efficient in discriminating abnormal instances without a large number of instances being predicted as positive.



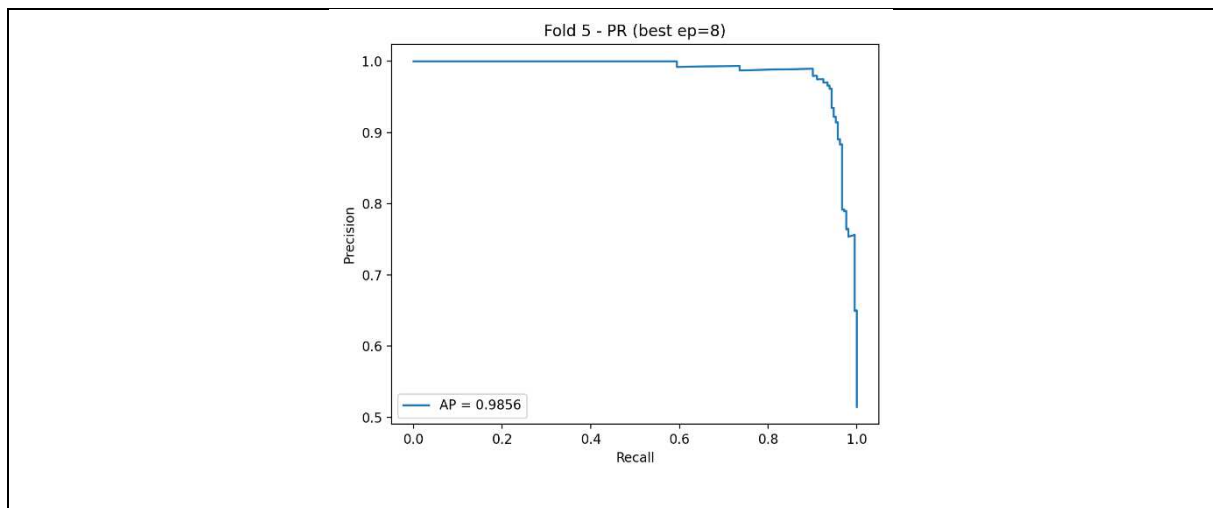


Fig. 4 The obtained PR curves for all folds

The trade-off between the true positive rates and false positive rates is shown through the receiver operating characteristic (ROC) curves, and they give a comparative measure of the model's discriminative ability. The curves that are closer to the top-left corner indicate better classification results, and the diagonal line represents random guessing. Figure 5 shows the results of the ROC curves for all five folds. These curves are well above the reference line and are close to the top-left corner of the graph, which shows that they have a good discriminative ability. Also, all curves are quite similar, which shows consistent results and a lack of sensitivity of the model to different splits of data. Additionally, a high true positive rate for lower false positive rates indicates that the model is very good at identifying normal and abnormal data.

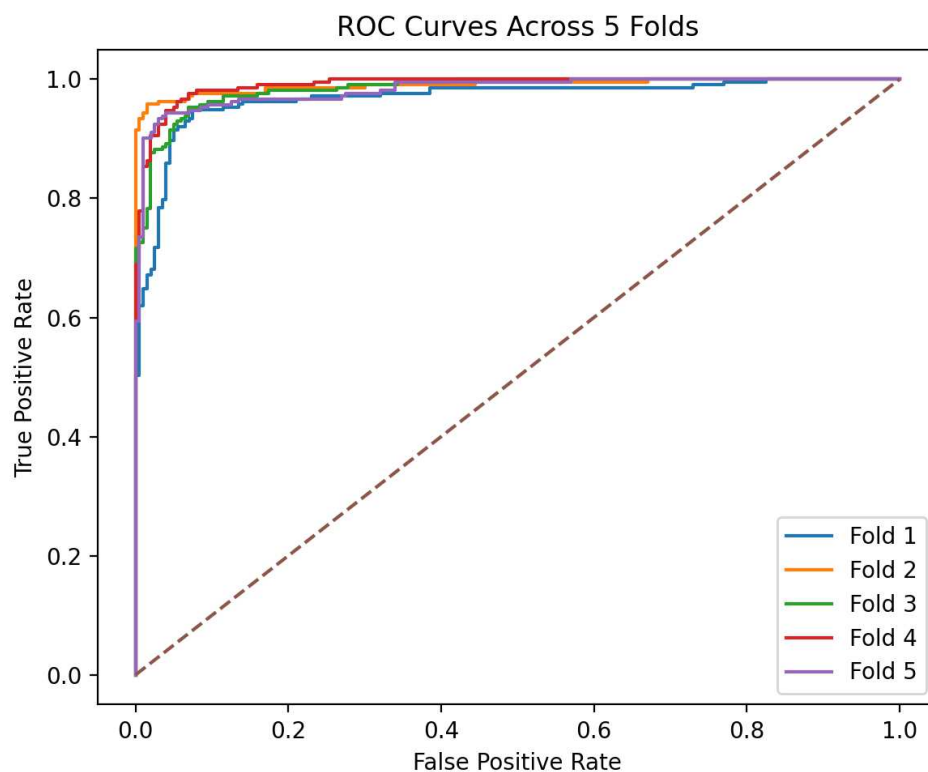


Fig. 5 The obtained ROC curves for all folds

3.1. Ablation study

We performed ablation tests to gauge the effectiveness of each component of the training approach [27]. The ResNet50-CervicalNet model was used each time to provide proper comparison, and all data was split into an equal stratified set for training and testing that was 80/20. The starting point of the comparison was a model that used cross-entropy loss function and SupCon learning with an alpha (α) weighting parameter. For investigating the role of the individual components surrounding SupCon, various ablations were proposed: SupCon for the classifier only (CE-only), SupCon removal via the projection head skip, contrastive learning directly in the features from the CervicalNet, as well as the classification to SupCon weight ratio adjustment through $\alpha = 0.25$, $\alpha = 0.75$. Various SupCon temperature factors were also considered, such that SupCon temperature = 0.07, SupCon temperature = 0.2. Various evaluation metrics such as accuracy, F1-score, precision, recall, area under the curve (AUC), and average precision (AP) were used, followed by the application of F1-score to identify the best epoch for the final evaluation.

Table 2 The obtained results for ablation study

Experiment	Best Ep	Acc	F1	Prec	Rec	AUC	AP
Baseline_ce+supcon	10	0.9709	0.9721	0.9631	0.9812	0.9975	0.9979
Ce_only	9	0.9685	0.9696	0.9673	0.9718	0.9942	0.9949
supcon_feat_no_proj	5	0.9709	0.9714	0.9855	0.9577	0.9953	0.9958
Alpha_0.25	10	0.9855	0.9858	0.9952	0.9765	0.9986	0.9988
Alpha_0.75	9	0.9685	0.9696	0.9673	0.9718	0.9949	0.9959
Temp_0.07	8	0.9758	0.9763	0.9856	0.9671	0.9949	0.9962
Temp_0.2	9	0.9782	0.9787	0.9857	0.9718	0.9962	0.9970

The result of the ablation study is presented above in Table 2. The base model performs well by integrating cross-entropy and supervised contrastive loss with 97.09% accuracy and 97.21% F1-score. By removing supervised contrastive loss and only relying on cross-entropy loss, the model performs marginally poorly to demonstrate the importance of supervised contrastive loss over cross-entropy loss. By integrating supervised contrastive loss with the backbone features but removing the projection head, accuracy is maintained but recall is marginally decreased. In weightage models, $\alpha=0.25$ is seen to be optimal as it performs well under all aspects: 98.55% accuracy, 98.58% F1-score, 99.86% AUC, and 99.88% AP. $\alpha=0.25$ retains a balance between classifier loss and contrastive loss. Increasing the weightage to $\alpha=0.75$ is suboptimal to specify the limit past which increased weightage to contrastive loss decreases the performance of the classifier. By trying different hyper-parameters to optimize temperature, 0.07 and 0.2 emerged as optimal choices; however, both were suboptimal to $\alpha=0.25$. By way of conclusion from this ablation study, supervised contrastive learning along with proper weightage to contrastive loss is essential to optimize performance. $\alpha=0.25$ is concluded to be the optimal weightage value.

3.2. Explainability analysis

This paper uses Grad-CAM++ to explain the explainability of our proposed model [28]. Grad-CAM++ is an explainability technique in machine learning, which uses a technique called heatmap to show which parts in an input image are most contributing towards a particular decision. It uses higher-order gradients and provides much finer results than Grad-CAM in circumstances when several

discriminative parts in an input image contribute towards a particular decision. The result uses a heatmap, where hotter parts contribute most to an outcome.

Figure 6 below highlights the explanation results for histopathological images of cervical conditions using Grad-CAM++. Note how the normal images in the top row have expanded regions of higher brightness in the activation map, rather than specific hotspots. The cervical cancer images in the row below demonstrate small, very bright regions of heavy activity pointed out by Grad-CAM++, consistent with the densely but randomly packaged tissue present in the regions indicative of cervical cancer. The easily recognizable differences in the activation maps of the normal versus abnormal images illustrate the presence of relevant features rather than background information being utilized by the computer program.

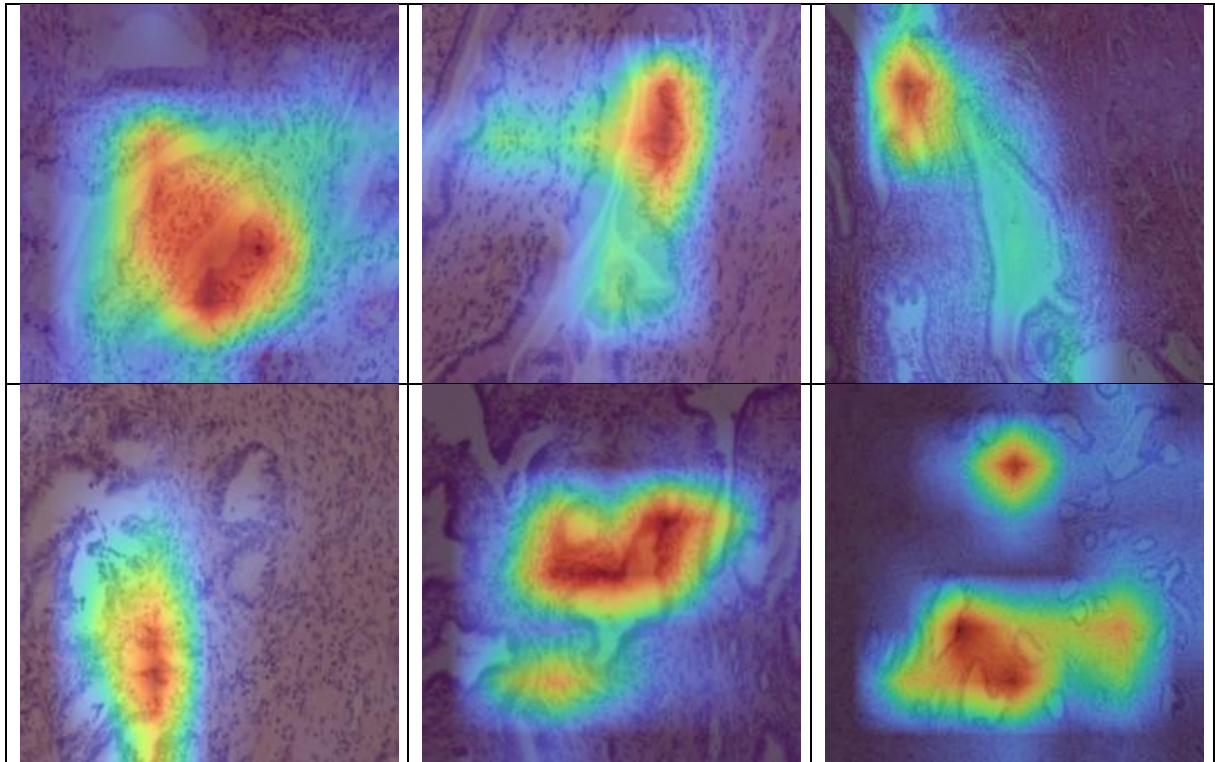


Fig. 6 Explanability analysis with Grad-CAM++

4. Discussions

In this paper, we propose a deep learning system capable of automatically classifying histopathological images of the cervix. The proposed system combines discriminative learning of feature representation with robust classification within an end-to-end training procedure. The system is built on top of ResNet-50 and leverages both contrastive learning and traditional cross-entropy loss functions with the goal of maximizing class separation and minimizing intra-class variations within the learned feature space. The design of this system is expected to encourage the learning of feature representation by focusing on clinically meaningful patterns within histopathological images, while distinguishing similar-looking structures in an image. Our framework has been validated through five-fold cross-validation to test its robustness and ability to generalize. On all five-folds, the framework performed well, and accuracy values are between 96.61% and 97.10%, and for F1-scores, they are between 98.13% and 98.83%. Precision values remain at higher levels on all five-folds, up to 99.05%, and recall values remain above 97.65%, showing that the detection of abnormal cervical tissue is accurate. The average accuracy on all fivefolds stands at around 97.00% with low values for standard deviation.

Apart from the above results based on the cross-validation approach, we further carried out an ablation study with an 80/20 stratified train and test split. For the ablation study, we observed that when we remove the supervised contrastive learning, there was a minor degradation in performance, and the CE-only model performed with 96.85% accuracy and 96.96% F1-score. To effectively classify and contrast the learning processes, we assigned the best weight value $\alpha = 0.25$ for contrastive learning, and with this, we obtained an accuracy of 98.55%, 98.58% F1-score, 99.86% AUC, and 99.88% average precision. However, by assigning an increased weight value $\alpha = 0.75$ to contrastive learning, there was a decrease in performance, indicating an overshoot in contrastive learning. From the above results, it can be concluded that supervised contrastive learning has made a prominent impact when properly balanced with classification.

Table 3. Performance comparison of the proposed method with the previous studies on histopathological datasets

Authors	Year	Method	Dataset	Validation	BCC (%)
Yang et al. [16]	2024	RF + Color Histogram features	CAISHI	Hold-out (classification)	90.18
Yang et al. [16]	2024	Linear SVM + Color Histogram features	CAISHI	Hold-out (classification)	77.14
Yang et al. [16]	2024	Linear SVM + LBP features	CAISHI	Hold-out (classification)	71.10
Yang et al. [16]	2024	Linear SVM + HOG features	CAISHI	Hold-out (classification)	64.37
Yang et al. [16]	2024	ResNet50	CAISHI	Hold-out (classification)	96.88
Yang et al. [16]	2024	Inception-V3	CAISHI	Hold-out (classification)	96.70
Yang et al. [16]	2024	VGG-16	CAISHI	Hold-out (classification)	92.59
Yang et al. [16]	2024	ViT	CAISHI	Hold-out (classification)	89.96
Yang et al. [16]	2024	Siamese Network (ResNet50 backbone), 1-shot	CAISHI	Few-shot (2-way 1-shot)	76.80
Proposed	2026	ResNet50 + SupCon ($\alpha = 0.25$)	CAISHI	5-fold cross-validation	96.90 $\pm$ 0.23

Published results in Table 3 summarize the comparison done for classification in histopathological images of the cervix using various approaches like existing classical machine learning algorithms, established deep learning models, a few shot learning techniques, and the proposed one. From existing classical models, for histopathological image classification in cervix images, the highest BCC value for Random Forest with color histogram is 90.18%, which is higher than the linear SVM approach results for color histogram, LBP, and HOG representations with BCC values 77.14%, 71.10%, and 64.37%, respectively, showing that simple color, texture, and gradient representations are not enough to efficiently model the morphological differences in histopathology cervix images. Deep learning models show a larger gap over existing simple handmade feature representations. Established CNN models like ResNet50 and Inception-V3 show BCC values of 96.88% and 96.70%, respectively, showing high representational power for histopathology image representations using CNN models. The BCC value for VGG-16 is also close to these with a value of 92.59%, while for Vision Transformers, the value slightly dropped to 89.96%, showing lower efficiency over CNN representations for histopathology image representations in similar conditions. The result obtained from a Siamese network with a 2-way

1-shot approach for histopathology image classification in cervix histopathology is much lower for BCC value with a measure of 76.80%, establishing low efficiency in learning robust class boundaries using a very small set of histopathology examples for cervix histopathology image representations. The proposed approach in histopathology image representations for cervix image classification is the most efficient one with a BCC value measure of 96.90 ± 0.23 using 5-fold cross-validation, establishing a high measure over existing models like CNN models despite similar measures from a few models with similar conditions for validation tests in cervix histopathology image representations.

The proposed approach has a number of distinct advantages over other approaches for image classification tasks. Firstly, incorporating the notion of supervised contrastive learning improves the way in which the features are discriminated across classes and in the same class, respectively. Secondly, a deep CNN is employed that captures complex morphological details in histopathological tissue images quite effectively.

On the negative side, the evaluation of the approach is based on one dataset alone, which could be the limiting factor for how well the approach may generalize to another environment.

5. Conclusions

To overcome the challenges involved in the process, a framework based on deep learning has been proposed in the classification task of cervical histopathological images. The framework employs joint optimization in classification performance and contrastive discriminative represents comprehension. This framework comprises a ResNet-50 convolutional feature extractor trained concurrently using contrastive supervised learning and cross-entropy loss. The joint optimization in the framework enables it to efficiently learn intra-class compact representations and improve inter-class discrimination, which has particular importance in the histopathological cervical images since they display subtle morphological variability and large intra-class variation. The framework has been tested thoroughly on the CAISHI histopathological image dataset using five-fold cross-validation. The framework showed a classification accuracy mean (BCC) of $96.90\% \pm 0.23$, and it depicted robust precision, recall, specificity, and F1-score on each fold. The standard deviation on each fold showed that the framework has a robust generalization capacity. The framework also depicted robust discriminative potential in the PR curve and ROC curve since it showed a higher true positive rate at a low false acceptance rate. A comparison between the framework and the S-CNN, TransformerNet, and few-shot learning framework indicated that the framework has the most robust and reliable performance.

Although the framework shows robust and strong performance, some areas are highlighted that could be targeted for future work. Firstly, testing the model on other histopathological samples obtained from different clinical settings would allow the estimation of the overall robustness and generality. Secondly, employing state-of-the-art data augmentation or stain normalization strategies could add to the overall stain and illumination variability robustness. A promising direction for future work could be exploring the use of explainability tools such as state-of-the-art explainable AI strategies that allow for an even deeper and more intuitive understanding of decision-making strategies developed by the models. Finally, the extension of the framework to two-class or whole slide level could allow for enhanced practical utility.

Acknowledgment

This Project was funded by the Deanship of Scientific Research (DSR) at King Abdulaziz University, Jeddah, Saudi Arabia under grant no. (IPP: 1594-135-2025). The authors, therefore, acknowledge with thanks DSR for technical and financial support.

Declarations

Funding This Project was funded by the Deanship of Scientific Research (DSR) at King Abdulaziz University, Jeddah, Saudi Arabia under grant no. (IPP: 1594-135-2025).

Competing interests The authors declare no competing interests.

References

- [1] World Health Organization. (2021). *WHO guideline for screening and treatment of cervical pre-cancer lesions for cervical cancer prevention*. World Health Organization.
- [2] Sopracordevole, F., Clemente, N., Alessandrini, L., Di Giuseppe, J., Cigolot, F., Buttignol, M., Ciavattini, A., & Canzonieri, V. (2017). Detection of occult endocervical glandular dysplasia in cervical conization specimens for squamous lesions. *Pathology – Research and Practice*, 213, 210–216. <https://doi.org/10.1016/j.prp.2016.12.015>
- [3] Li, C., Chen, H., Zhang, L., Xu, N., Xue, D., Hu, Z., Ma, H., & Sun, H. (2019). Cervical histopathology image classification using multilayer hidden conditional random fields and weakly supervised learning. *IEEE Access*, 7, 90378–90397. <https://doi.org/10.1109/ACCESS.2019.2924467>
- [4] Xue, D., Zhou, X., Li, C., Yao, Y., Rahaman, M. M., Zhang, J., Chen, H., Zhang, J., Qi, S., & Sun, H. (2020). An application of transfer learning and ensemble learning techniques for cervical histopathology image classification. *IEEE Access*, 8, 104603–104618. <https://doi.org/10.1109/ACCESS.2020.2999816>
- [5] Wang, C.-W., Chang, C.-C., Khalil, M. A., Lin, Y.-J., Liou, Y.-A., Hsu, P.-C., Lee, Y.-C., Wang, C.-H., & Chao, T.-K. (2022). Histopathological whole slide image dataset for classification of treatment effectiveness to ovarian cancer. *Scientific Data*, 9, Article 127. <https://doi.org/10.1038/s41597-022-01127-6>
- [6] Hemalatha, K., & Vetrivel, V. (2025). Self-supervised learning using diverse cell images for cervical cancer classification. *Measurement*, 243, 116413. <https://doi.org/10.1016/j.measurement.2024.116413>
- [7] Fei, M., Zhang, X., Chen, D., Song, Z., Wang, Q., & Zhang, L. (2025). Whole slide cervical cancer classification via graph attention networks and contrastive learning. *Neurocomputing*, 613, 128787. <https://doi.org/10.1016/j.neucom.2024.128787>
- [8] Pacal, I., & Attallah, O. (2025). Hybrid deep learning model for automated colorectal cancer detection using local and global feature extraction. *Knowledge-Based Systems*, 319, 113625. <https://doi.org/10.1016/j.knosys.2025.113625>
- [9] AbdulAzeem, Y., Balaha, H. M., ZainEldin, H., Abuain, W. A., Badawy, M., & Elhosseini, M. A. (2025). Challenging the status quo: Why artificial intelligence models must go beyond accuracy in cervical cancer diagnosis. *Biomedical Signal Processing and Control*, 105, 107620. <https://doi.org/10.1016/j.bspc.2025.107620>
- [10] AlZobi, F. I., Mansour, K., Nasayreh, A., Samara, G., Alsalman, N., Bashkami, A., Smerat, A., & Nahar, K. M. O. (2025). Optimized soft-voting CNN ensemble using particle swarm optimization for endometrial cancer histopathology classification. *Computer Methods and Programs in Biomedicine Update*, 8, 100217. <https://doi.org/10.1016/j.cmpbup.2025.100217>
- [11] Mohanta, A., Roy, S. D., Nath, N., Datta, A., & Bhowmik, M. K. (2025). A Comprehensive Survey on Diagnostic Microscopic Imaging Modalities, Challenges, Taxonomy, and Future Directions for

Cervical Abnormality Detection and Grading. *IEEE Transactions on Artificial Intelligence*. <https://doi.org/10.1109/TAI.2025.3551669>

[12] Sharma, A., & Parvathi, R. (2025). Enhancing Cervical Cancer Classification: Through a Hybrid Deep Learning Approach Integrating DenseNet201 and InceptionV3. *IEEE Access*, 13, 756–770. <https://doi.org/10.1109/ACCESS.2025.3527677>

[13] Ain, Q. U., Al-Sahaf, H., Xue, B., & Zhang, M. (2024). Genetic Programming for Malignancy Diagnosis From Breast Cancer Histopathological Images: A Feature Learning Approach. *IEEE Transactions on Emerging Topics in Computational Intelligence*. <https://doi.org/10.1109/TETCI.2024.3523769>

[14] Jeleń, Ł., Stankiewicz-Antosz, I., Chosia, M., & Jeleń, M. (2025). Optimizing Cervical Cancer Diagnosis with Feature Selection and Deep Learning. *Applied Sciences*, 15(3), 1458. <https://doi.org/10.3390/app15031458>

[15] Vazquez, B., Rojas-García, M., Rodríguez-Esquivel, J. I., Marquez-Acosta, J., Aranda-Flores, C. E., Cetina-Pérez, L. d. C., Soto-López, S., Estévez-García, J. A., Bahena-Román, M., Madrid-Marina, V., & Torres-Poveda, K. (2025). Machine and Deep Learning for the Diagnosis, Prognosis, and Treatment of Cervical Cancer: A Scoping Review. *Diagnostics*, 15(12), 1543. <https://doi.org/10.3390/diagnostics15121543>

[16] Yang, X., Li, C., He, R., Yang, J., Sun, H., Jiang, T., Grzegorzec, M., Li, X., & Liu, C. (2024). CAISHI: A benchmark histopathological H&E image dataset for cervical adenocarcinoma in situ identification, retrieval and few-shot learning evaluation. *Data in Brief*, 53, 110141. <https://doi.org/10.1016/j.dib.2024.110141>

[17] Rasheed, A., Shirazi, S. H., Khan, P., Aseere, A. M., & Shahzad, M. (2025). Techniques and challenges for nuclei segmentation in cervical smear images: a review. *Artificial Intelligence Review*, 58, 295. <https://doi.org/10.1007/s10462-025-11207-9>

[18] Gangrade, J., Kuthiala, R., Gangrade, S., Singh, Y. P., R, M., & Solanki, S. (2025). A deep ensemble learning approach for squamous cell classification in cervical cancer. *Scientific Reports*, 15, 91786. <https://doi.org/10.1038/s41598-025-91786-3>

[19] Chauhan, N. K., Singh, K., Kumar, A., Mishra, A., Gupta, S. K., Mahajan, S., Kadry, S., & Kim, J. (2025). A hybrid learning network with progressive resizing and PCA for diagnosis of cervical cancer on WSI slides. *Scientific Reports*, 15, 97719. <https://doi.org/10.1038/s41598-025-97719-4>

[20] Raza, M. A., Siddiqui, H. U. R., Saleem, A. A., Zafar, K., Zafar, A., Dudley, S., & Iqbal, M. (2025). Advanced Feature Extraction for Cervical Cancer Image Classification: Integrating Neural Feature Extraction and AutoInt Models. *Sensors*, 25(9), 2826. <https://doi.org/10.3390/s25092826>

[21] Deniz, E., Şengür, A., Kadiroğlu, Z., Guo, Y., Bajaj, V., & Budak, Ü. (2018). Transfer learning based histopathologic image classification for breast cancer detection. *Health information science and systems*, 6(1), 18.

[22] Koonce, B. (2021). ResNet 50. In *Convolutional neural networks with swift for tensorflow: image recognition and dataset categorization* (pp. 63-72). Berkeley, CA: Apress.

[23] Van Laarhoven, T. (2017). L2 regularization versus batch and weight normalization. *arXiv preprint arXiv:1706.05350*.

[24] Stein, R. A., Jaques, P. A., & Valiati, J. F. (2019). An analysis of hierarchical text classification using word embeddings. *Information Sciences*, 471, 216-232.

- [25] Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., ... & Krishnan, D. (2020). Supervised contrastive learning. *Advances in neural information processing systems*, 33, 18661-18673.
- [26] Le-Khac, P. H., Healy, G., & Smeaton, A. F. (2020). Contrastive representation learning: A framework and review. *Ieee Access*, 8, 193907-193934.
- [27] Meyes, R., Lu, M., De Puiseau, C. W., & Meisen, T. (2019). Ablation studies in artificial neural networks. *arXiv preprint arXiv:1901.08644*.
- [28] Selvaraju, R. R., Das, A., Vedantam, R., Cogswell, M., Parikh, D., & Batra, D. (2016). Grad-CAM: Why did you say that?. *arXiv preprint arXiv:1611.07450*.