Pertanika J. Sci. & Technol. 34 (S1): 133 - 159 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Position-Aware Normalized Discounted Cumulative Gain (NDCG) and Multi-Task TabNet for Cervical Precancerous Lesion Classification

P. Leela[1]*, K. Hari Krishna[1], and K. K. Baseer[2]

[1]Department of Computer Science and Engineering, School of Computing, Mohan Babu University, Sree Sainath Nagar, 517102 Tirupati, Andhra Pradesh, India
[2]Department of Computer Science and Systems Engineering, GITAM School of CSE, GITAM (Deemed to be University), Nagadenahalli, 562163 Bengaluru, Karnataka, India

ABSTRACT

Cervical cancer is still one of the major causes of cancer death in women around the world. Accurate and early classification of precancerous lesions from images of cervical cells is crucial for the enhancement of patient outcomes. Based on MMT, this paper introduces a new hybrid model named TabNet-PANDCG for improving the classification performance, severity-aware ranking, and interpretability of models. This paper presents TabNet-PANDCG, a novel hybrid model built on top of MMT to boost classification performance, severity-aware ranking, and model interpretability. The proposed method extracts the morphological features, intensity features and texture features from cervical cell images to generate tabular structured images, in contrast with the conventional method, which processes raw images directly. The image-derived features are then passed into Multi-Task TabNet with a sparse attention mechanism that performs the task of dynamically selecting the relevant features and uses it to simultaneously classify the lesion and predict the lesion's severity. The proposed PANDCG metric is a ranking evaluation that assigns greater weight to the clinically relevant, high-severity lesions that rank highly, giving a more meaningful evaluation that fits more with the clinical focus. The framework was tested using a publicly available dataset containing 917 Pap images from single-cell images, across seven diagnostic categories, namely Herlev. Experimental results show that TabNet-PANDCG has the greatest performance with 97.39% accuracy, 93.66% macro precision, 92.17% macro recall and 92.14% macro F1 score, outperforming some of the state-of-the-art techniques. Additionally, the incorporation of PANDCG facilitates clinically relevant ranking assessments, and TabNet's attention-based

ARTICLE INFO

Article history:
Received: 07 March 2026
Accepted: 14 May 2026
Published: 24 July 2026

DOI: https://doi.org/10.47836/pjst.34.S1.07

E-mail addresses:
leelapachappa@gmail.com (P. Leela)
khk396@gmail.com (K. Hari Krishna)
drkkbaseer@gmail.com (K. K. Baseer)
* Corresponding author

e-ISSN: 2231-8526

feature selection further promotes transparency and trust in the model. The framework exhibits great promise for computer-aided diagnosis systems in cervical cancer screening.

Keywords: Cervical cancer, clinical diagnostics, deep learning, interpretability, medical data analysis, multi-task learning, multi-task TabNet, position-aware NDCG, sparse attention

INTRODUCTION

Cervical cancer often develops in the cells of the cervix, the lower section of the uterus. It is widely known that a sexually transmitted infection, namely human papillomavirus (HPV), is the leading cause of cervical cancer. Research studies have shown that cervical cancer is one of the leading causes of death among women with cancer (Allogmani et al., 2024). A significant number of cases of cervical cancer continue to be reported in both low- and high-income countries. Cancer of the cervix is one of the top cancers affecting women, accounting for about 6.5 per cent of all cancer cases in women. It is estimated that the number of deaths caused by cervical cancer and the new incidences in 2020 were 342,000 and 604,127 globally, respectively (Attallah, 2023). Cervical cancer is a serious problem in the world, especially in areas lacking screening, vaccination, diagnostic facilities and early detection services. Approximately 54,517 new cases of invasive cervical cancer have been detected in Europe every year, and 24,874 women die of the disease (Bentéjac et al., 2021). The anatomical site and pathological trajectory of cervical cancer are shown in Figure 1.

Vaccinations are common in ordinary immunisation programs. World Health Organisation (2022) have set a target to vaccinate 90% of girls to be fully vaccinated against HPV by the age of 15 years by 2030 (Chandran et al., 2021). It is also emphasised that the European authorities must employ the existing preventive strategies to intensify the initiative to eliminate cervical cancer, as highlighted by the World Health Organisation (2023). Like other health issues, mortality rates can be diminished to a considerable extent by conducting regular check-ups and diagnosing the disease at its early stages (Darwish et al., 2023). Due to the absence of symptoms during the early development of cervical cancer, the diagnosis may become challenging. So, an automated and accurate diagnostic system will help clinicians in the screening and classification of cervical lesions.

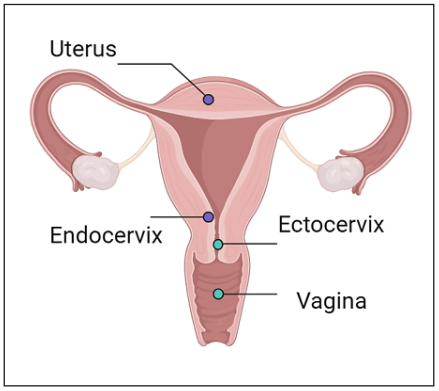

Figure 1. Cervix cancer image

Abnormal bleeding is the most common symptom of cervical cancer, and it is likely to become more serious with the development

of the disease. In later stages, almost 90% of the patients have distinct symptoms. The typical signs of cervical cancer are reddish discharge, postmenopausal bleeding, contact bleeding, and spotting. Moreover, foul-smelling bloody discharge is another important sign of the disease.

The General Article of this Study

The proposed study introduces a new method that integrates Position-aware NDCG, a metric that aims to focus on the quality of diagnostics according to the severity of the lesions, with Multi-Task TabNet, a deep learning model with the ability to process tabular data and perform multiple related prediction tasks. With Position-aware NDCG, the model prioritises clinically important cases, and in the case of precancer lesions, the sensitivity is high. At the same time, the Multi-Task TabNet is a platform that incorporates a wide range of diagnostic features, which makes the model both more robust and interpretable. In this study, the proposed multi-task TabNet is applied to image-derived feature vectors representing cervical cell images, due to TabNet's demand for structured tabular input. The given work has great opportunities to contribute to the development of early cervical lesion detection and classification and introduce more specific measures and better patient outcomes.

Motivations of this Research

The motivation behind this study stems from several critical challenges in current cervical precancerous lesion classification systems:

- Existing evaluation metrics treat all misclassifications equally, which is clinically inadequate. In cervical cancer screening, misclassifying a high-severity lesion (such as severe dysplasia or carcinoma in situ) as a low-risk class has far more serious consequences than the reverse. There is a strong need for severity-aware ranking metrics that prioritise clinically critical cases.
- Most deep learning models for cervical cell classification operate as black boxes, offering limited interpretability. Clinicians require transparent models that can reveal which diagnostic features influence predictions, thereby increasing trust and facilitating adoption in clinical workflows.
- Current approaches often fail to leverage the complementary strengths of advanced tabular learning architectures and clinically meaningful evaluation strategies. There is a clear opportunity to develop hybrid frameworks that combine powerful feature selection mechanisms with ranking metrics aligned with medical priorities.
- Early and accurate detection of precancerous lesions remains essential for reducing cervical cancer mortality, particularly in resource-limited settings where automated,

reliable computer-aided diagnosis tools can significantly improve screening outcomes.

Contributions of this Study

- We propose PANDCG (Position-aware Normalised Discounted Cumulative Gain), a novel severity-aware ranking metric specifically designed for cervical precancerous lesion classification. Unlike standard metrics that assign equal weight to all errors, PANDCG incorporates position-aware weighting to prioritise high-severity lesions at top-ranked positions, thereby providing a more clinically meaningful evaluation of model predictions.
- We develop a hybrid TabNet-PANDCG framework that transforms cervical cell images into structured tabular feature representations through comprehensive feature engineering. These image-derived features are processed by Multi-Task TabNet, which employs sparse attention mechanisms for interpretable feature selection and performs simultaneous lesion classification and severity prediction.
- We demonstrate through extensive experiments on the Herlev dataset (917 single-cell cervical images across seven diagnostic classes) that the proposed framework achieves state-of-the-art performance (97.39% accuracy) while offering improved interpretability and clinically aligned ranking evaluation. The integration of PANDCG and TabNet's attention mechanism addresses key limitations of existing black-box models and conventional evaluation approaches.
- We provide a transparent and reproducible pipeline that bridges image-based cervical cell analysis with tabular deep learning, offering a practical pathway toward trustworthy computer-aided diagnosis systems for cervical cancer screening.

Related Work

Chandran et al. (2021) introduced a deep learning method used to identify cervical cancer by classifying colposcopy images into three classes, i.e. type 1, type 2 and type 3. The researchers used a popular cervical screening dataset to train their algorithm and test it. They introduced a powerful network known as the Colposcopy Ensemble Network (CYENET) that was more accurate than the preceding models such as VGG16 and VGG19. CYENET posted an accuracy of 92.3% on the test (Lebanova et al., 2023). However, this high accuracy could have been affected by the fact that the quantity of testing images used in this study (1884) was rather small.

According to Tomko et al. (2023), there were considerable effects of cervical cancer on the physical, psychological, and social well-being of women throughout their lives.

This research concentrates on the data employed throughout the Kaggle competition in the Intel and MobileODT Cervical Cancer Screening, in which the challenge is to deal with multi-class and multi-label organisation. They also included optimisation of image size in their strategy.

Attallah (2023) introduced a new CAD model known as CerCan-Net to be used in the automation of detecting cervical cancer. The CerCan-Net is built on three CNN architectures with lower depth and fewer parameters (ResNet-18, MobileNet, and DarkNet-19) in comparison to the traditional models. This is a calculated decision that will ease and hasten the classification process. An essential element in the effectiveness of CerCan-Net is transferring learning, and it relies on the deep features in the final three layers of every CNN, as opposed to relying on features in just one layer. The approach allows describing the complexity of the information in more detail.

To automatically classify cervical cytology images, Fang et al. (2022) suggested a deep convolutional neural network called DeepCELL. This network learns feature representations through multiple kernels of varying sizes. Primarily, three distinct core modules of DeepCELL were designed to extract feature information utilising these kernels. Subsequently, the cervical cell classification model was structured by stacking several of these core modules in a top-to-bottom manner.

Xu et al. (2022) presented a new risk assessment network for cervical lesions in colposcopic images, which was called the RACNet. The RACNet has two major components. The first is a multi-scale detection network that is used to detect the location and severity of the lesions at different scales. Second, the development of a multi-branch convolutional neural network (CNN) that can classify these lesions is done to enhance the performance of risk assessment (Okunade, 2020). RACNet was linked with the most suggested methods and colposcopists under the same circumstances. The results of the experiment conducted in this paper show that RACNet has the best performance, as it has an accuracy ratio of 84.5, which is 15 percent higher than the average accuracy of colposcopists.

Shakil et al. (2024) came up with a precise prediction algorithm for cervical cancer by employing six ML models. The approach was used following evaluation of 36 risk factor variables among 858 participants. The Adaptive Synthetic Sampling and the Synthetic Minority Oversampling Technique were used to balance the data. Also, feature selection approaches such as chi-square and Least Absolute Shrinkage and Selection Operator (LASSO) were used to assess the ranking of features, especially in the detection of illness. The results of the model were explained using Shapley Additive Explanations, an explainable AI model.

Limitations of Existing Systems

- The current systems are not resistant to the nature of being used on varied data. The difference in the image acquisition settings, patient demographics and lesion characteristics may result in differing performance between clinical settings.
- Most algorithms are trained in small and homogeneous datasets, which do not cover a broad range of cervical pathologies. This weakness does not allow them to generalise to unseen data, especially in the real world.
- The systems usually use the pathologist annotation, which may be subjective and prone to inter- and intra-observer variability. This brings noise to the training data, which eventually decreases the accuracy of the model.

Research Gap

The lack of research in the development of cervical cancer screening tools is the inability to find cervical precancerous lesions correctly and efficiently, which is a significant challenge. Existing methods, including cytology and histology, are manual-based and therefore prone to error and subjectivity. Moreover, the fact that the adoption of new high-level technologies, including AI, has not become an essential part of daily medical activity has prevented the possibility of automated and high-quality diagnosis. The most important research gap is the formulation of strong, explainable, and scalable AI-based models that can address these shortcomings by offering stable, trustworthy, and early detection of precancer lesions, primarily in resource-strained cases in which cervical cancer is overrepresented.

Proposed System

In this paper, we suggest a hybrid interpretable model, TabNet-PANDCG, which combines Position-aware Normalised Discounted Cumulative Gain (PANDCG) with Multi-Task TabNet for the classification of cervical precancerous lesions. The framework introduces a "severity-aware" ranking evaluation and "attention" tabular learning on features derived from images, to enhance classification accuracy and clinical usability. The proposed approach differs from traditional image classification pipelines, starting with segmenting cervical cell images for morphological, intensity and texture features to create structured tabular representations of these images. These feature vectors are provided as input to Multi-Task TabNet, which uses sequential sparse attention mechanisms to decide which features are the most discriminative without discarding any information, while also performing lesion classification and severity prediction. By enriching standard ranking measures with the importance of having a high severity lesion (severe dysplasia, carcinoma in situ) at top-ranked positions, PANDCG complements this and can bring these features into the ranking process. The proposed TabNet-PANDCG framework overall architecture is

shown in Figure 2. The following subsections describe the dataset, pre-processing pipeline, PANDCG formulation and Multi-Task TabNet architecture.

Dataset

The Herlev dataset contains 917 isolated single-cell cervical cell images (Peng et al., 2021), with each image featuring a single cervical cell. Figure 3 shows sample images from the Herlev dataset. The seven classes in the dataset are columnar epithelia, superficial squamous epithelia, mild squamous non-keratinising dysplasia, moderate squamous non-keratinising dysplasia, severe squamous non-keratinising dysplasia, intermediate squamous cells, and squamous cell carcinoma in situ. Table 1 shows the summary of class-wise distribution of the Herlev dataset. The raw images were not used as input to TabNet in this study; first, structured feature vectors were extracted from the images and then used as inputs into TabNet with the proposed Multi-Task TabNet classifier.

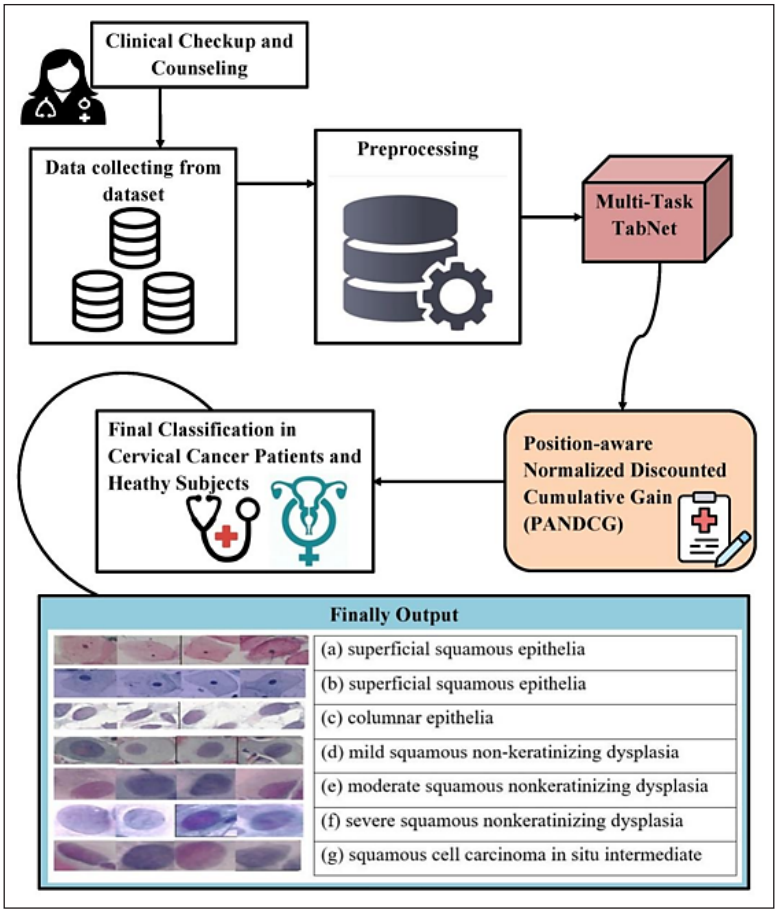

Figure 2. Architecture of TabNet-PANDCG

There are 917 single-cell cervical pictures in the Herlev dataset, which are divided into groups for normal and pathological conditions. 98 columnar epithelial cells, 70 intermediate squamous epithelial cells, and 74 superficial squamous epithelial cells make up the typical category. The pathological category includes 146 photographs of moderate squamous non-keratinising dysplasia, 197 images of severe squamous non-keratinising dysplasia, 182 images of mild squamous non-keratinising dysplasia, and 150 images of intermediate squamous cell carcinoma. Models for classifying cervical cells and identifying precancerous and malignant states require this varied dataset for training and evaluation.

Data Preprocessing

To learn with Multi-Task TabNet, the algorithm is designed for learning with tabular data; the raw images of cervical cells in the Herlev dataset had to be systematically transformed into quantitative representations of the data (Shakil et al 2024). The process is fundamental, as TabNet can't work with pixel values as an input but instead requires structured input.

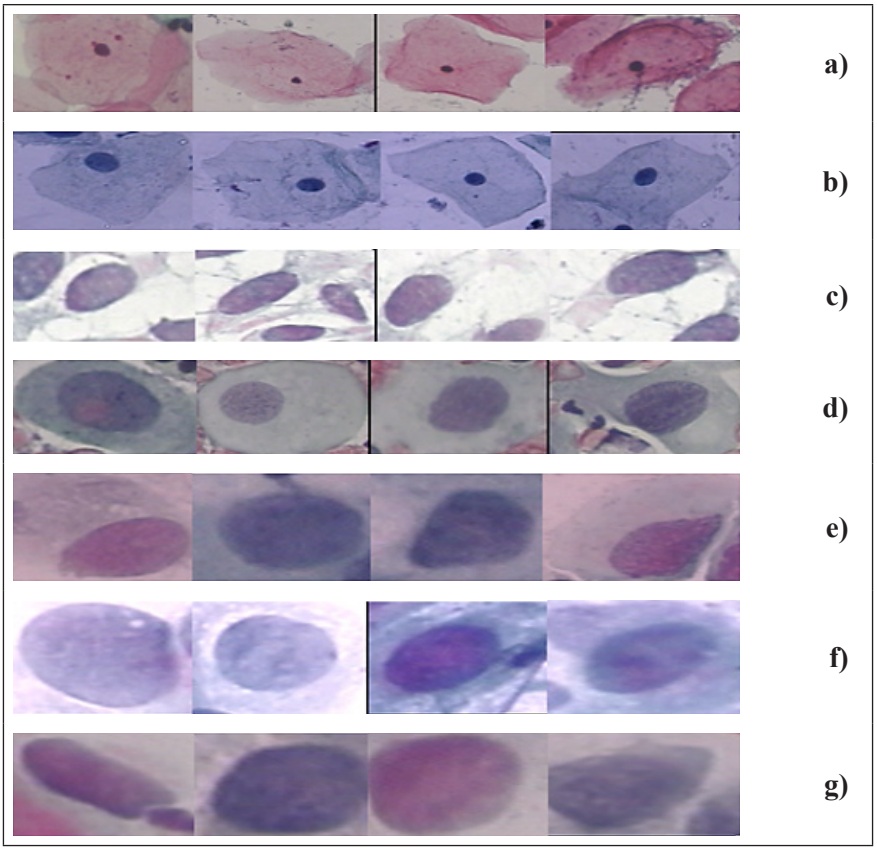

Figure 3. Sample images for the Herlev dataset

Table 1
Provides the statistical data from the Herlev dataset

Dataset type	Category	Quantity
Normal	Superficial squamous epithelial	74
Normal	Intermediate squamous epithelial	70
Normal	Columnar epithelial	98
Abnormal	Squamous cell carcinoma in situ intermediate	150
Abnormal	Mild squamous non-keratinising dysplasia	182
Abnormal	Moderate squamous non-keratinising dysplasia	146
Abnormal	Severe squamous non-keratinising dysplasia	197
Total image		917

We derived an extensive feature suite for each of the 917 single-cell images, covering three broad classes of hand-crafted features:

- Morphological features: nucleus area, cytoplasm area, nucleus-to-cytoplasm ratio, eccentricity, solidity, perimeter and equivalent diameter.
- Intensity-based features: mean, standard deviation, skewness and kurtosis of the pixel intensity distribution in the nucleus region and cytoplasm region of the cell.
- Texture features: Gray-Level Co-occurrence Matrix (GLCM) statistics (at 8 different orientations and 4 distances).

These features were concatenated into a high-dimensional tabular feature matrix with one row being an image of a cervical cell, and each column a quantitative feature which describes the image. Data were checked for obvious outliers and missing values and eliminated during quality control. To reduce the number of input variables that could lead to overfitting and to remove potential overfitting, a Principal Component Analysis (PCA) was performed and the principal components that explained the largest percentage of cumulative variance (95%) were used. To have similar scales among the different types of features before training the models, the features were standardised in advance by z-score normalisation.

This approach to feature engineering uses the image classification task as a structured tabular learning task, allowing access to Multi-Task TabNet, which is highly effective at solving tabular learning problems, while revealing inherent interpretability thanks to sparse attention mechanisms.

Position-Aware Normalised Discounted Cumulative Gain (PANDCG)

Standard ranking metrics such as Normalised Discounted Cumulative Gain (NDCG) treat all misrankings equally, which is suboptimal in medical diagnosis where the clinical cost of errors varies significantly with lesion severity (Sompawong et al., 2019). To address

this limitation, we propose Position-aware Normalised Discounted Cumulative Gain (PANDCG), a severity-sensitive ranking metric designed to prioritise high-severity cervical lesions at top-ranked positions during evaluation.

PANDCG modifies the standard NDCG formulation by introducing position-aware weighting and clinically defined relevance scores. Let denote the ranked list of predicted lesion classes for a given sample, where higher ranks correspond to more severe lesions according to the model output. We first assign a relevance score $rel(i)$ to each class based on its clinical severity in Equation 1:

$$rel(i) = \begin{cases} 0 & \text{if class } i \text{ is Normal} \\ 1 & \text{if class } i \text{ is Mild dysplasia} \\ 2 & \text{if class } i \text{ is Moderate dysplasia} \\ 3 & \text{if class } i \text{ is Severe dysplasia} \\ 4 & \text{if class } i \text{ is Carcinoma in situ} \end{cases} \quad [1]$$

Discounted Cumulative Gain (DCG)

The DCG score reduces the influence of lower-ranked items while calculating the relevance of elements in a ranked list. For the top-K ranked predictions, DCG is defined as in Equation 2:

$$DCG = \sum_{i=1}^{k} \frac{rel_i^*}{\log_2(i+1)} \quad [2]$$

where rel_i represents the relevance score of the prediction at rank position i, and K denotes the number of ranked predictions considered. The logarithmic discount term $\log_2(i+1)$ reduces the contribution of lower-ranked predictions.

Ideal DCG (IDCG)

To normalise DCG, the ideal DCG score is calculated in Equation 3, representing the DCG of a perfect ranking.

$$IDCG = \sum_{i=1}^{K} \frac{rel_i^*}{\log_2(i+1)} \quad [3]$$

where rel_i^* represents the ideal relevance score at position i after sorting the lesion classes according to the clinically preferred ranking. IDCG represents the maximum possible DCG value for a given sample.

Position-Aware Weights

In PANDCG, position weights $w(i)$ are introduced to amplify the importance of earlier positions. The weights are often assigned based on clinical priorities, such as imposing higher penalties for misclassifying severe lesions. The position-aware weight is defined as in Equation 4:

$$\mathrm{w_i} = \frac{1}{\sqrt{\mathrm{i}}} \quad [4]$$

where w_i denotes the position-aware weight for rank position i. This weighting strategy assigns a larger weight to earlier-ranked predictions and gradually reduces the weight for lower-ranked predictions.

The DCG is modified using position-aware weights in Equation 5:

$$PDCG = \sum_{i=1}^{k} w_i \times \frac{rel_i}{\log_2(i+1)} \quad [5]$$

where $PDCG$ represents the position-aware discounted cumulative gain. By multiplying the relevance score with w_i, the metric gives additional emphasis to clinically relevant predictions appearing near the top of the ranked output.

PANDCG

Finally, PANDCG is normalised using the ideal position-aware DCG. The final $PANDCG$ score is calculated as in Equation 6:

$$PANDCG = \frac{PDCG}{IPDCG} \quad [6]$$

where $IPDCG@K$ represents the ideal position-aware discounted cumulative gain. It is calculated as in Equation 7:

$$IPDCG = \sum_{i=1}^{k} w_i \times \frac{rel_i^*}{\log_2(i+1)} \quad [7]$$

PANDCG are between 0 and 1, with higher scores placing the clinically important classes higher on the prediction list. For this reason, PANDCG offers a severity-dependent effectiveness assessment approach to the classification of cervical precancerous lesions.

Multi-Task TabNet for Cervical Precancerous Lesion Classification

Following the feature engineering process described in Section 3.2, each cervical cell image is represented as a structured tabular feature vector. Multi-Task TabNet to learn from this tabular representation because it is specifically designed for structured data and offers built-in interpretability through sparse attention. The model simultaneously performs two related tasks: (1) multi-class lesion classification and (2) severity regression, allowing shared representations to improve generalisation across both tasks.

TabNet processes the input through a sequence of decision steps, each consisting of a Feature Transformer and an Attentive Transformer. This sequential attention mechanism enables the model to focus on the most relevant features at each step while producing interpretable feature masks. The overall architecture comprises an input layer, multiple decision steps with shared and step-dependent feature transformers, attentive transformers with sparsemax normalisation, and task-specific output heads for classification and regression.

TabNet Feature Transformer

The output F_t of a feature transformer layer is calculated as in Equation 8:

$$F_t = f_t(M_t \odot X) \quad [8]$$

where F_t represents the transformed features at decision step t, X denotes the input feature vector, M_t is the sparse attention mask generated at step t, $\odot$ represents element-wise multiplication, and $f_t(\cdot)$ denotes the feature transformation function.

Attentive Transformer

The sparse attention mask M_t is derived as in Equation 9:

$$M_t = Sparsemax\ (A_t) \quad [9]$$

The attention coefficient of a block A_t at the decision step t and a sparse feature-selection mask (Sparsemax) are used to generate. This sparse mask enables TabNet to only focus on features that are the most informative and minimise the effects of irrelevant and/or redundant features.

Multi-Task Loss Function

For cervical precancerous lesion classification, the total loss L_{total} is a combination of losses for all tasks as represented in Equation 10.

$$L_{total} = \sum_{j=1}^{T} \lambda_j L_j \quad [10]$$

where T represents the number of learning tasks, L_j denotes the loss for the j^{th} task, and λ_j represents the weight assigned to that task. In this study, the task losses include lesion classification loss and severity-aware prediction loss.

For classification, the prediction is represented by Equation 11:

$$L_{cls} = -\sum_{c=1}^{C} y_c \log(\hat{y}_c) \quad [11]$$

where C is the number of lesion classes, y_c is the true class label, and $\hat{y}_c$ is the predicted probability for class c.

For severity prediction (regression), the prediction function is defined as follows in Equation 12:

$$L_{sev} = \frac{1}{N}\sum_{n=1}^{N} (s_n - \hat{s}_n)^2 \quad [12]$$

where N represents the number of samples, s_n is the true severity score of the n^{th} sample, and $\hat{s}_n$ is the predicted severity score.

Final Prediction

The final prediction $\hat{y}$ for each task combines all features after the decision step, as shown in Equation 13:

$$\hat{y} = \sigma(WF + b) \quad [13]$$

The last layer of the network, where W is the weights of the last layer, F is the last feature representation from TabNet decision steps, bis the bias term, and σ denotes the activation function. In binary prediction, the sigmoid activation function can be applied, while for multi-class lesion classification the softmax activation function can be applied. The proposed Multi-Task TabNet model learns discriminative cervical cell features, selects crucial features for cancer diagnosis from many features in a sparse fashion and can classify cervical cell lesions. The PANDCG module then checks if clinically relevant categories

of lesions are ranked in the predictions correctly. This architecture enables TabNet to dynamically select important diagnostic features while learning shared representations beneficial for both classification and severity estimation, thereby improving both accuracy and interpretability.

Application of Proposed Method

The TabNet-PANDCG framework is specifically designed to address the dual challenges of accurate lesion classification and clinically meaningful evaluation in cervical cancer screening. By transforming cervical cell images into structured tabular feature representations and processing them through Multi-Task TabNet, the framework enables simultaneous lesion classification and severity estimation while maintaining interpretability through sparse attention mechanisms.

The integration of PANDCG further enhances the framework's clinical utility by ensuring that high-severity lesions are prioritised in the ranked output. This severity-aware evaluation strategy aligns model assessment with real-world diagnostic priorities, where misranking critical cases carries significant clinical consequences. The proposed approach is therefore particularly suitable for computer-aided diagnosis systems that require both high accuracy and transparent, clinically aligned decision support.

Performance Evaluation Metrics

Macro Precision: The average of the precision scores calculated separately for each class in a multi-class classification task is known as macro precision, as shown in Equation 14. All classes, regardless of size, are treated equally.

$$Macro\,\mathrm{Pr}ecision = \frac{1}{N}\sum_{i=1}^{N}\mathrm{Pr}ecision_i \qquad [14]$$

Macro Recall: Macro Recall is a metric used in classification problems to calculate the average recall across all classes as shown in Equation 15. It treats all classes equally by computing the recall for each class independently and then averaging the results.

$$Macro\,\mathrm{Re}call = \frac{1}{N}\sum_{i=1}^{N}\frac{True\ Positives_i}{True\ Positives_i + False\ Negatives_i} \qquad [15]$$

Macro F1-Score: A statistic used for evaluation in multi-class classification tasks is the Macro F1-Score. Each class is treated equally, regardless of its frequency, and the F1-Score is calculated individually for each class. Finally, the unweighted mean of these scores is computed as shown in Equation 16.

$$Macro\ F1-Score=\frac{1}{N}\sum_{i=1}^{N}F1_i \quad [16]$$

Accuracy: The proportion of properly recognised occurrences (including true positives and true negatives) relative to total examples in a dataset is known as accuracy. It is a widely utilised metric to estimate the performance of classification models as shown in Equation 17.

$$Accuracy=\frac{TP+TN}{Total\ Instances} \quad [17]$$

Root Mean Square Error (RMSE): The average magnitude of the discrepancies between anticipated and observed values in a dataset is typically measured using the Root Mean Square Error (RMSE) metric as shown in Equation 18. Like the observed data, RMSE provides a clear measurement of these differences in the same units.

$$RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2} \quad [18]$$

Processing Time: Processing time refers to the duration required to complete a specific task or process within a system. It encompasses the time taken to execute all necessary steps in the workflow as shown in Equation 19.

$$T_p=T_e+T_w \quad [19]$$

Advantages of Proposed Method

The proposed TabNet-PANDCG framework offers several key advantages over existing approaches for cervical precancerous lesion classification:

- **Severity-aware ranking evaluation**: PANDCG incorporates clinical severity into the ranking metric, ensuring that high-grade lesions (severe dysplasia and carcinoma in situ) are prioritised at top-ranked positions, thereby providing a more meaningful assessment than standard accuracy-based or ranking metrics.
- **Interpretable feature selection**: Multi-Task TabNet's sparse attention mechanism dynamically identifies the most discriminative diagnostic features at each decision step, offering inherent interpretability without requiring post-hoc explanation techniques.
- **Multi-task learning capability**: By jointly optimising lesion classification and severity regression, the model learns shared representations that improve generalisation and provide complementary diagnostic information in a single framework.

- **Suitability for image-derived tabular data**: The framework bridges image-based analysis and tabular deep learning by converting cervical cell images into structured feature vectors, enabling the use of TabNet's efficient and interpretable architecture.
- **Clinical alignment and practicality**: The combination of high classification performance, severity-aware ranking, and built-in interpretability makes the framework well-suited for supporting reliable and trustworthy computer-aided diagnosis in clinical cervical cancer screening workflows.

RESULTS AND DISCUSSIONS

The proposed TabNet-PANDCG framework was evaluated on the publicly available Herlev dataset, consisting of 917 single-cell cervical Pap smear images across seven diagnostic classes. To comprehensively assess model performance, we conducted both seven-class multi-class classification and binary classification (normal vs. abnormal) experiments. This dual evaluation strategy allows us to examine the model's ability to distinguish fine-grained lesion categories as well as its effectiveness in clinically relevant binary screening scenarios.

The TabNet-PANDCG model was implemented using Keras with the TensorFlow 2.8.2 backend and trained on Google Colab equipped with an NVIDIA Tesla T4 GPU (CUDA 11.2, driver version 460.32.03). All experiments were conducted using consistent data splits and hyperparameter settings to ensure fair and reproducible comparisons. Model performance was assessed using accuracy, macro precision, macro recall, macro F1-score, Root Mean Square Error (RMSE), processing time, confusion matrix analysis, ablation studies, and k-fold cross-validation.

Comparative Methods

To evaluate the effectiveness of the proposed TabNet-PANDCG framework, we compared it against several recent state-of-the-art methods for cervical cancer classification. The baseline methods are briefly described below.

- **CPLDC-AOATL** (Allogmani et al., 2024): This method employs bilateral filtering as a preprocessing step to reduce noise in cervical images, followed by an Archimedes Optimisation Algorithm combined with transfer learning for classification.
- **CACCD-GOADL** (Nour et al., 2024): This approach utilises an enhanced MobileNetV3 architecture for feature extraction from cervical images, with the Gazelle Optimisation Algorithm (GOA) employed for hyperparameter tuning of the model.
- **CYENET**: This method introduces a deep ensemble network (CYENET) that integrates multiple convolutional neural network backbones for the classification of colposcopy images into different cervical cancer types.

- **Mask R-CNN** (Xu, T et al., 2022): This method applies Mask R-CNN for instance segmentation of cervical cell nuclei in liquid-based cytology slides, enabling the detection and analysis of nuclear morphological features for distinguishing normal and abnormal cells.

Confusion Matrix

The effectiveness of the classification for the dataset's seven categories is illustrated in the confusion matrix in Figure 4.

The confusion matrix for the seven-class classification task is presented in Figure 4. In this matrix, the rows represent the ground-truth classes, while the columns correspond to the classes predicted by the model. The diagonal elements indicate correct classifications, whereas the off-diagonal elements represent misclassifications between different lesion categories.

The matrix reveals strong overall classification performance, with most samples correctly assigned to their respective classes, as evidenced by the prominent diagonal

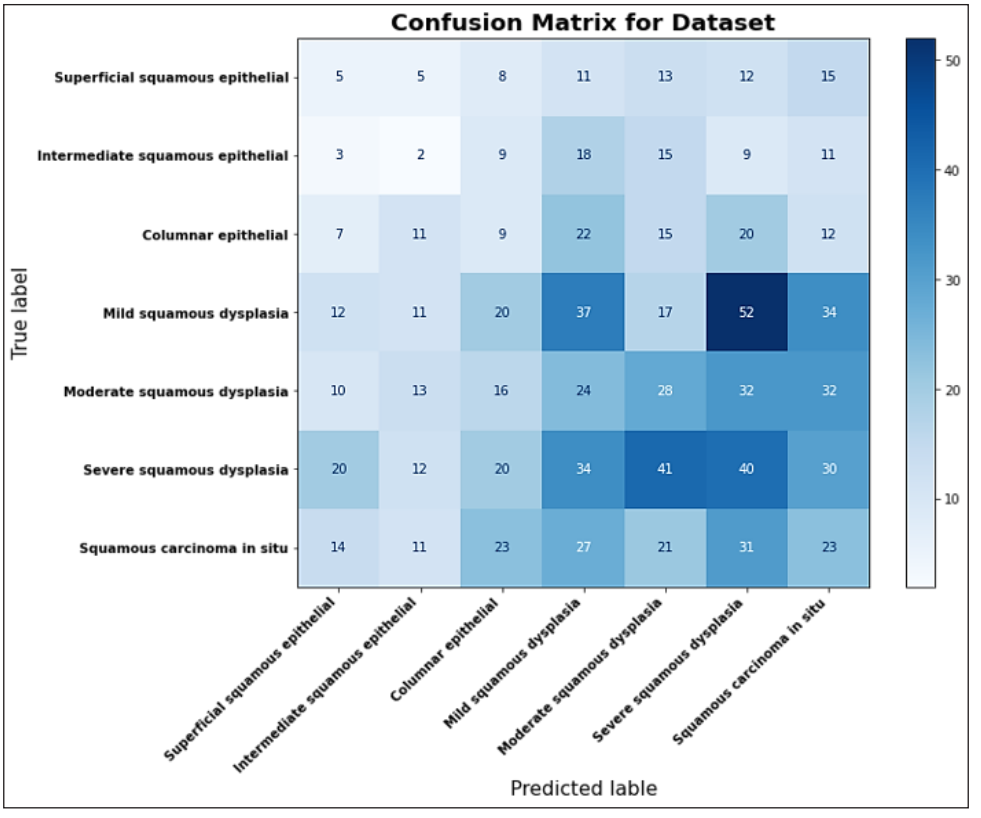

Figure 4. Confusion matrix of the proposed TabNet-PANDCG model on the Herlev dataset

entries. However, notable confusion persists among the mild, moderate, and severe squamous non-keratinising dysplasia classes. This pattern is expected, given the substantial overlap in morphological and textural characteristics among these progressive stages of dysplasia. Such inter-class similarity poses a significant challenge for fine-grained classification and highlights the need for more discriminative feature representations or advanced class-separation techniques in future work.

Binary Class-Wise Comparison

Table 2 shows the comparison of various methods for the binary evaluation setting of normal and abnormal classes. Case "0" = "Normal" class and case "1" = "Abnormal" class. MP, MR and MF1 represent macro precision, macro recall and macro F1-score, respectively.

The proposed TabNet-PANDCG framework consistently outperforms all baseline methods in both normal and abnormal classes across all evaluation metrics. Notably, it achieves the highest accuracy of 97.14% for normal cases and 97.65% for abnormal cases, along with superior macro-level performance. These results demonstrate the effectiveness of combining Multi-Task TabNet with the PANDCG evaluation strategy for robust binary cervical cell classification.

The results obtained from the proposed model are compared with the existing methods in terms of MP, MR, MF1 and accuracy in the proposed model for both normal and abnormal cases as shown in Figure 5 and Table 2, respectively. For case "0", the proposed TabNet-PANDCG achieved 94.15 MP, 92.16 MR, 91.11 MF1, and 97.14% accuracy. For case "1", it achieved 93.17 MP, 92.18 MR, 93.17 MF1, and 97.65% accuracy. The obtained results show a good performance of the proposed model for normal and abnormal cervical cell classification. The proposed approach outperforms existing methods such as CPLDC-AOATL, CACCD-GOADL, and Mask R-CNN in terms of overall performance, including accuracy and weighted performance at the macro level**.**

Table 2
Comparison of methods for binary normal and abnormal cases

Methods	Case 0 MP	Case 0 MR	Case 0 MF1	Case 0 Accuracy	Case 1 MP	Case 1 MR	Case 1 MF1	Case 1 Accuracy
CPLDC-AOATL	87.34	62.25	77.45	88.28	64.81	75.56	88.23	80.34
CACCD-GOADL	76.12	81.87	71.22	81.23	79.45	85.23	76.14	87.34
CYENET	66.23	88.13	89.55	87.34	81.28	87.23	87.98	89.22
Mask R-CNN	89.14	77.34	88.16	89.91	88.76	79.13	89.32	87.11
Proposed TabNet-PANDCG	94.15	92.16	91.11	97.14	93.17	92.18	93.17	97.65

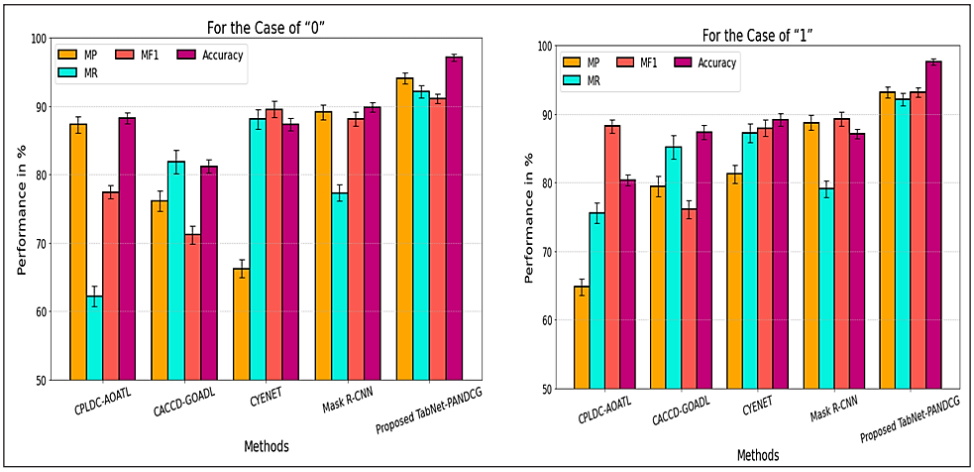

Figure 5. Comparative performance of different methods for normal and abnormal cases

RMSE and Processing Time Analysis

Table 3 and Figure 6 show the comparative results between the proposed model and the comparative models. Repair error was evaluated using the root mean square error (RMSE), and the closer the RMSE was to a minimum, the more consistent the prediction.

Table 3
RMSE Analysis of proposed method

Methods	RMSE
CPLDC-AOATL	39.91
CACCD-GOADL	36.16
CYENET	31.55
Mask R-CNN	23.87
Proposed TabNet-PANDCG	18.24

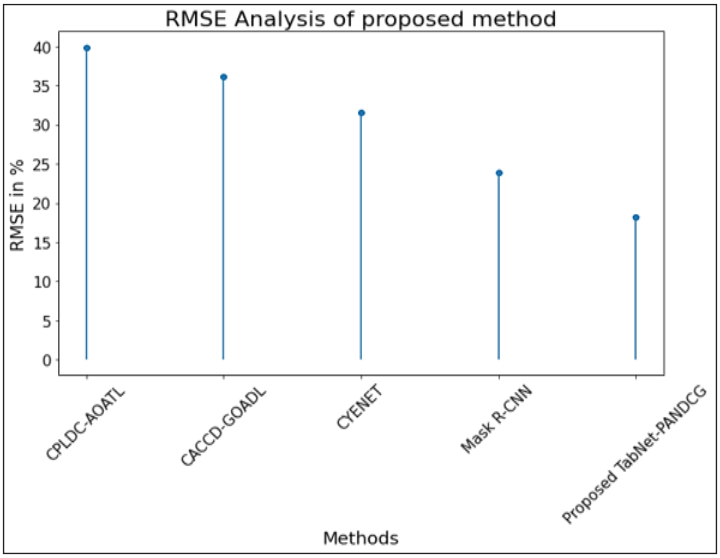

Figure 6. RMSE analysis of proposed method

TabNet-PANDCG could obtain the lowest RMSE of 18.24 in predicting, which is better than other comparative models. Mask R-CNN had the second smallest RMSE value of 23.87, and CYENET and CACCD-GOADL had relatively larger error values. CPLDC-AOATL achieved the largest RMSEs, indicating less consistency in prediction. The overall RMSE comparison indicates the effectiveness of the proposed model, TabNet-PANDCG.

TabNet-PANDCG method had the smallest RMSE value of 18.24, which is better than the other comparative methods in terms of consistency of the predictions. Among the three methods, Mask R-CNN achieved the second-best RMSE of 23.87, and CYENET and CACCD-GOADL had relatively higher errors. CPLDC-AOATL produced the highest RMSE value, indicating poor consistency of predictions. For overall analysis, RMSE analysis shows the effectiveness of the proposed TabNet-PANDCG model.

The comparison of processing time of different methods is presented in Table 4 and Figure 7. The processing time is used as a criterion for the computational efficiency of each model.

The proposed TabNet-PANDCG achieved the lowest processing time of 3.765, indicating that it is computationally more efficient than the comparative approaches.

Table 4
Processing time analysis of the proposed method

Methods	Processing Time
CPLDC-AOATL	9.234
CACCD-GOADL	4.116
CYENET	6.154
Mask R-CNN	8.113
Proposed TabNet-PANDCG	**3.765**

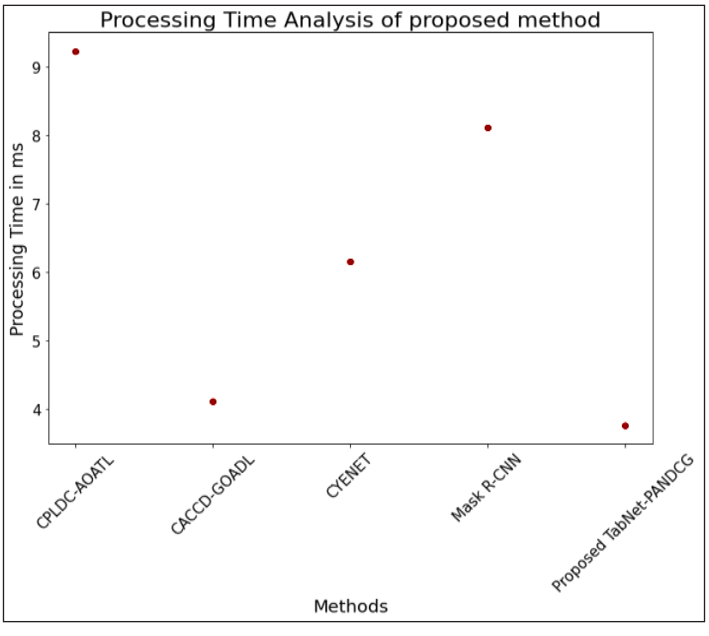

Figure 7. Processing time analysis of the proposed method

CACCD-GOADL achieved the second-lowest processing time, while CYENET, Mask R-CNN, and CPLDC-AOATL required comparatively longer processing times. This result suggests that the proposed framework can provide improved classification performance while maintaining efficient computation.

Training and Validation Analysis

The performance of the proposed TabNet-PANDCG is shown in the training and validation graph shown in Figure 8. The model's learning behaviour was checked while it was training by using embedded training and validation curves.

The training and validation loss seem to be decreasing over time, suggesting that the model is learning discriminative information about the cervical cells. Concurrently, the accuracy of validation demonstrates the improvement of the model's classification capability for cervical pre-cancerous lesions. A close correlation between the trends of training and validation showed that there was no severe overfitting in the training step. Therefore, the reliability of the proposed TabNet-PANDCG model is confirmed by the training and validation analysis.

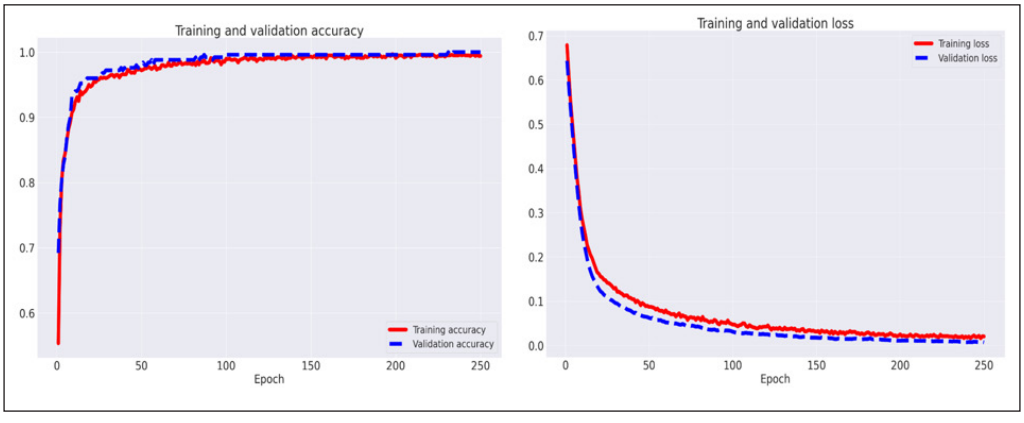

Figure 8. Training and validation loss and accuracy analysis

Ablation Study

Table 5 and Figure 9 present the ablation analysis of the proposed framework. The purpose of the ablation study was to evaluate the individual and combined contribution of PANDCG and Multi-Task TabNet**.**

The ablation results show that PANDCG alone achieved 89.34% accuracy, while Multi-Task TabNet achieved 93.65% accuracy. The combined TabNet-PANDCG model achieved the highest performance with 93.66 MP, 92.17 MR, 92.14 MF1, and 97.39% accuracy. This indicates that the integration of Multi-Task TabNet with PANDCG improves both

Table 5
Ablation experiment of PANDCG and multi-task TabNet on cervical cancer

Methods	MP	MR	MF1	Accuracy
PANDCG	87.87	86.17	77.34	89.34
Multi-Task TabNet	91.17	89.81	88.17	93.65
TabNet-PANDCG	**93.66**	**92.17**	**92.14**	**97.39**

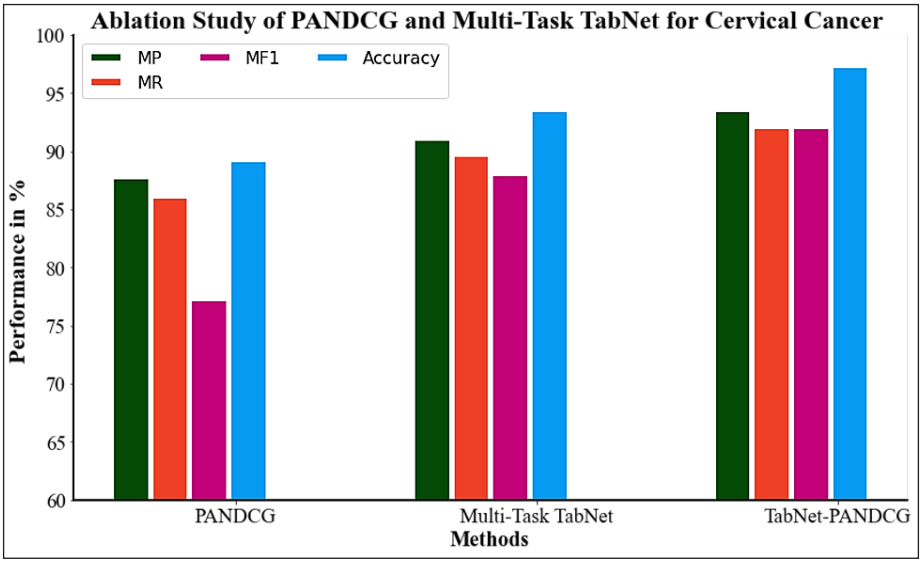

Figure 9. Ablation study of PANDCG and multi-task TabNet for cervical cancer

classification performance and severity-aware ranking quality. Therefore, both components contribute meaningfully to the final model performance.

K-Fold Matrix

The Herlev dataset was used for 5-fold cross-validation to evaluate the stability and generalisation ability of the proposed TabNet-PANDCG. This is specifically good for smaller-sized data sets as it takes the whole data set for testing and training by splitting the data set into 5 parts; each part is tested once, training with the remaining parts. The values of each of the performance metrics for each fold are summarised in Table 6, with their respective mean and standard deviation computed. The model showed consistent results with a low variance between the folds, with an average accuracy of 97.38 ± 0.68%, macro precision of 93.66 ± 1.07%, macro recall of 92.37 ± 1.29%, and macro F1-score of 92.74 ± 1.32%. The validation results validate the robustness and reliability of the proposed TabNet-PANDCG framework. The cross-validation process is shown in Figure 10.

Table 6
5-fold cross-validation results of TabNet-PANDCG

Fold	Accuracy (%)	Macro Precision (%)	Macro Recall (%)	Macro F1-score (%)	RMSE	Processing Time (s)
1	97.82	94.15	93.08	93.45	17.85	3.71
2	96.72	92.87	91.45	91.98	18.92	3.82
3	97.27	93.42	92.31	92.67	18.41	3.68
4	98.36	95.21	94.12	94.56	17.12	3.79
5	96.72	92.65	90.89	91.04	19.05	3.85
Mean ± Std	**97.38 ± 0.68**	**93.66 ± 1.07**	**92.37 ± 1.29**	**92.74 ± 1.32**	**18.27 ± 0.78**	**3.77 ± 0.07**

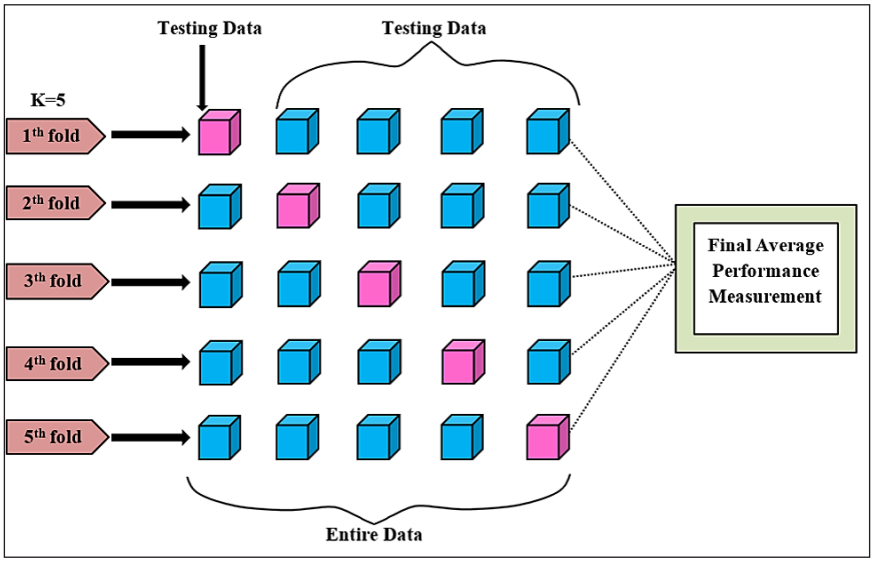

Figure 10. K-fold cross-validation matrix of TabNet-PANDCG

Comparison with Proposed Method

Table 6 and Figure 11 compare the proposed TabNet-PANDCG model with existing cervical cancer classification methods reported in the literature.

The effectiveness of different methods in comparison with each other is shown in Table 7 and Figure 11 using several datasets. Yuan et al. (2020) used ResNet to process histopathology images and yielded an accuracy of 84.1%. ResNet led to slightly less accuracy of 80.7% when applied to raw colposcopy pictures. Peng et al. (2021) used DenseNet on raw colposcopy images and achieved an accuracy rate of 76.4%. Darwish et al. (2023) applied ViT with SPT to raw colposcopy images and obtained an accuracy of 91. TabNet-PANDCG in the present study was run on the Herlev dataset, which gave the largest accuracy of 97.39, exceeding all other models in the analysis. This shows that TabNet-PANDCG is more effective on a range of datasets and approaches.

Table 7
Comparative with proposed method

Authors	Dataset	Year	Method	Accuracy
Yuan et al., 2020	Histopathological	(2020)	ResNet	84.1%
Liu et al., 2021	Raw colposcopy images	(2021)	ResNet	80.71%
Peng et al., 2021	Raw colposcopy images	(2020)	DenseNet	76.4%
Darwish et al., 2023	Raw colposcopy images	(2023)	ViT with SPT	91%
Our Research	**Herlev dataset**	**-**	**TabNet-PANDCG**	**97.39%**

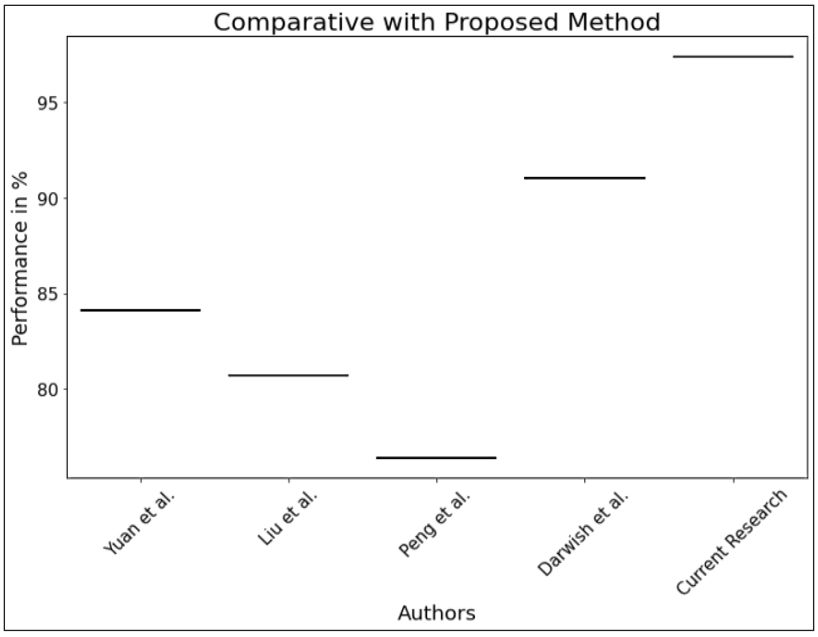

Figure 11. Comparative analysis of the proposed method with existing methods

It should be noted that differences in datasets, imaging modalities, preprocessing steps, and class distributions may influence direct comparison. Therefore, the comparative analysis is presented as an indicative performance comparison rather than a strictly controlled benchmark.

Limitations

Although some of the outcomes were highly favourable, the study did have a few limitations:

- The proposed framework has been tested only on the Herlev dataset with a comparatively low number of images (917 single-cell samples). The approach needs further validation with larger and more diverse datasets to establish it as generalisable.

- The quality and relevance of the handcrafted image features extracted during the preprocessing stage significantly affect the performance of the model, as TabNet has been specifically tailored for structured tabular data.
- While the usability of TabNet Zhang et al. (2024) is built on the principle of interpretability through sparse attention, the current study does not include a detailed study of explainability measures, including attention mask visualisation, ranking of feature importance or post-hoc interpretability methods (e.g., SHAP or LIME). The use of such techniques would indefinitely boost the clinical relevance of the decisions made by the model.
- The framework performance might be sensitive to the hyperparameters and may perform differently in different datasets, possibly because of variations in image conditions, staining techniques, class distribution or the quality of annotation provided.
- Therefore, external validation to ensure the reliability of the proposed system in computer-aided cervical cancer screening workflows is important for independent, real-world clinical datasets.

CONCLUSION

This research introduced TabNet-PANDCG, a hybrid model that combines Position-aware Normalised Discounted Cumulative Gain (PANDCG) with Multi-Task TabNet for the classification of cervical precancerous lesions. The framework aims to transform raw cervical cell images into structured tabular representations of the features, allowing Multi-Task TabNet to simultaneously classify the cell as a lesion and predict the lesion severity, with existing cell-level attention filtered to extract interpretable features. New to PANDCG is also the possibility to provide severity-aware ranking evaluation, to attach more weight to clinically relevant high-grade lesions in the model evaluation process. Experimental results over the Herlev dataset showed that the proposed framework can attain the strong performance of 97.39% accuracy, with competitive macro-level descriptors, and better than several of the other existing methods. Considering the attention-based interpretability of TabNet and the clinically aligned ranking of PANDCG, these two approaches could prove to be a viable direction to build trustworthy computer-aided diagnosis systems for screening cervical cancer. Moving forward, the framework will be validated on wider datasets from multiple centres to improve generalizability. In addition, we will attempt to use advanced explainability techniques, such as visualisation of attention and interpretation through the SHAP scores, to enhance model transparency. Based on these findings, the use of a multimodal framework including cytology, colposcopy, histopathology, and data on clinical risk factors is an important additional area of research for multimodal settings.

ACKNOWLEDGEMENT

Author Contributions: All authors contributed equally, and all authors read and approved the final version of the paper.

LIST OF ABBREVIATIONS

PANDCG	:	Position-aware Normalised Discounted Cumulative Gain
RMSE	:	Root Mean Square Error
CACCD-GOADL	:	cervical cancer diagnosis utilising the gazelle optimiser algorithm with deep learning
Mask R-CNN	:	Mask Regional Convolutional Neural Network
CPLDC-AOATL	:	Cervical precancerous lesions detection and classification using the Archimedes optimisation algorithm with transfer learning

REFERENCES

Allogmani, A. S., Mohamed, R. M., Al-Shibly, N. M., & Ragab, M. (2024). Enhanced cervical precancerous lesions detection and classification using Archimedes optimisation algorithm with transfer learning. *Scientific Reports, 14*, Article 12076. https://doi.org/10.1038/s41598-024-62773-x

Attallah, O. (2023). CerCan-Net: Cervical cancer classification model via multi-layer feature ensembles of lightweight CNNs and transfer learning. *Expert Systems with Applications, 229*, Article 120624. https://doi.org/10.1016/j.eswa.2023.120624

Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review, 54*(3), 1937-1967. https://doi.org/10.1007/s10462-020-09896-5

Chandran, V., Sumithra, M. G., Karthick, A., George, T., Deivakani, M., Elakkiya, B., & Manoharan, S. (2021). Diagnosis of cervical cancer based on ensemble deep learning network using colposcopy images. *BioMed Research International, 2021*, Article 5584004. https://doi.org/10.1155/2021/5584004

Darwish, M., Altabel, M. Z., & Abiyev, R. H. (2023). Enhancing cervical pre-cancerous classification using advanced vision transformers. *Diagnostics, 13*(18), Article 2884. https://doi.org/10.3390/diagnostics13182884

Fang, M., Lei, X., Liao, B., & Wu, F. X. (2022). A deep neural network for cervical cell classification based on cytology images. *IEEE Access, 10*, 130968-130980. https://doi.org/10.1109/ACCESS.2022.3230280

Herlev Dataset. (n.d.). *Herlev dataset* [Data set]. Kaggle. https://www.kaggle.com/datasets/yuvrajsinhachowdhury/herlev-dataset

Lebanova, H., Stoev, S., Naseva, E., Getova, V., Wang, W., Sabale, U., & Petrova, E. (2023). Economic burden of cervical cancer in Bulgaria. *International Journal of Environmental Research and Public Health, 20*(3), Article 2746. https://doi.org/10.3390/ijerph20032746

Liu, L., Wang, Y., Liu, X., Han, S., Jia, L., Meng, L., & Qiao, X. (2021). Computer-aided diagnostic system based on deep learning for classifying colposcopy images. *Annals of Translational Medicine, 9*(13), Article 1045. https://doi.org/10.21037/atm-21-885

Nour, M. K., Issaoui, I., Edris, A., Mahmud, A., Assiri, M., & Ibrahim, S. S. (2024). Computer-aided cervical cancer diagnosis using gazelle optimisation algorithm with a deep learning model. *IEEE Access, 12*, 13046-13054. https://doi.org/10.1109/ACCESS.2024.3351883

Okunade, K. S. (2020). Human papillomavirus and cervical cancer. *Journal of Obstetrics and Gynaecology, 40*(5), 602-608. https://doi.org/10.1080/01443615.2019.1634030

Peng, G., Dong, H., Liang, T., Li, L., & Liu, J. (2021). Diagnosis of cervical precancerous lesions based on multimodal feature changes. *Computers in Biology and Medicine, 130*, Article 104209. https://doi.org/10.1016/j.compbiomed.2021.104209

Shakil, R., Islam, S., & Akter, B. (2024). A precise machine learning model: Detecting cervical cancer using feature selection and explainable AI. *Journal of Pathology Informatics, 15*, Article 100398. https://doi.org/10.1016/j.jpi.2024.100398

Sompawong, N., Mopan, J., Pooprasert, P., Himakhun, W., Suwannarurk, K., Ngamvirojcharoen, J., & Tantibundhit, C. (2019). Automated Pap smear cervical cancer screening using deep learning. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 7044-7048). IEEE. https://doi.org/10.1109/EMBC.2019.8856369

Tomko, M., Pavliuchenko, M., Pavliuchenko, I., Gordienko, Y., & Stirenko, S. (2023). Multi-label classification of cervix types with image size optimisation for cervical cancer prescreening by deep learning. In S. Smys, V. E. Balas, & R. Palanisamy (Eds.), *Inventive computation and information technologies: Proceedings of ICICIT 2022* (pp. 885-902). Springer Nature Singapore. https://doi.org/10.1007/978-981-19-7402-1_63

World Health Organisation. (2022). *Vaccine offers solid protection against cervical cancer*. https://www.who.int/news/item/2022-vaccine-offers-solid-protection-against-cervical-cancer

World Health Organisation. (2023). *Behavioural and cultural insights at the WHO Regional Office for Europe: Annual progress report 2022* (WHO/EURO: 2023-6987-46753-68110). World Health Organisation Regional Office for Europe.

World Health Organisation. (2023) (n.d.). *Cervical cancer elimination initiative*. https://www.who.int/initiatives/cervical-cancer-elimination-initiative

Xu, T., Liu, P., Li, P., Wang, X., Xue, H., Guo, J., & Sun, P. (2022). RACNet: Risk assessment net of cervical lesions in colposcopic images. *Connection Science, 34*(1), 2139-2157. https://doi.org/10.1080/09540091.2022.2085665

Yuan, C., Yao, Y., Cheng, B., Cheng, Y., Li, Y., Li, Y., & Lu, W. (2020). The application of deep learning-based diagnostic system to cervical squamous intraepithelial lesions recognition in colposcopy images. *Scientific Reports, 10*, Article 11639. https://doi.org/10.1038/s41598-020-68252-3

Zhang, B., Jin, X., Liang, W., Chen, X., Li, Z., Panoutsos, G., & Tang, Z. (2024). TabNet: Locally interpretable estimation and prediction for advanced proton exchange membrane fuel cell health management. *Electronics, 13*(7), Article 1358. https://doi.org/10.3390/electronics13071358