Int. J. Appl. Math. Comput. Sci., 2025, Vol. 35, No. 4, 601–615
DOI: 10.61822/amcs-2025-0043

STOCHASTIC FEATURE SELECTION AND MACHINE LEARNING FOR OPTIMIZED CERVICAL CANCER CLASSIFICATION

ŁUKASZ JELEŃ [a,*], IZABELA STANKIEWICZ-ANTOSZ [b], MARIA CHOSIA [c], MICHAŁ JELEŃ [d]

[a]Department of Computer Engineering
Wroclaw University of Science and Technology
Wyb. Stanisława Wyspiańskiego 27, 50-370 Wrocław, Poland
e-mail: lukasz.jelen@pwr.edu.pl

[b]Faculty of Security Studies
General Tadeusz Kościuszko Military University of Land Forces
ul. Piotra Czajkowskiego 109, 51-147 Wrocław, Poland

[c]Department of Pathology
Pomeranian Medical University
ul. Unii Lubelskiej 1, 71-252 Szczecin, Poland

[d]Department of Immunopathology and Molecular Biology
Wroclaw Medical University
ul. Borowska 211, 50-556 Wrocław, Poland

Liquid-based cytology (LBC) is a widely used diagnostic tool for cervical cancer diagnosis. However, the accuracy and efficiency of LBC-based cervical cancer classification are still limited due to the lack of standardized, scalable, and objective cytological assessment protocols. To address these gaps, this study develops and evaluates a machine learning framework that integrates various feature extraction techniques, feature selection methods, and machine learning classifiers to improve cervical cancer detection. The results demonstrate that handcrafted and local binary pattern features achieve the best overall performance, with the SVM, gradient boosting and histogram-based gradient buffering reaching a 95.92% accuracy, highlighting the strength of combining morphological and texture descriptors to maximize their discriminative potential. Moreover, we provide a systematic comparison of different classification pipelines, offering insights into the feasibility of hybrid approaches, particularly in resource-constrained medical environments. The promising results obtained in this study highlight the potential impact of machine learning in modern medical diagnostics, providing a clinically relevant, highly accurate, and efficient classification method for LBC slides.

Keywords: cervical cancer classification, machine learning, feature selection, convolutional neural networks, stochastic models.

1. Introduction

Despite continuous advancements in diagnostic methodologies and the integration of artificial intelligence techniques, cervical cancer remains a significant challenge to global health, ranking as the fourth most common cancer in women with over 570,000 new cases and 311,000 deaths annually (Wild *et al.*, 2020). In Poland, the recorded five-year survival rate is 48.3%, compared to 62.1% across Europe (Polish National Cancer Registry, 2021). Despite HPV vaccines and enhanced screening, limitations in access to care, subjectivity in diagnosis, and human error persist, emphasizing the need for automated diagnostic solutions. Primary screening methods like Papanicolaou (Pap) smears and liquid-based cytology (LBC) have played a crucial role in detecting both precancerous and cancerous lesions (Stankiewicz, 2018). Despite their availability, these methods are still prone to subjectivity, human error,

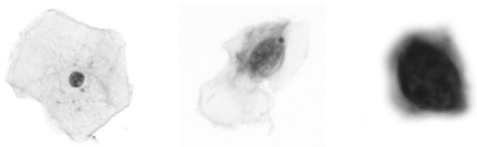
*Corresponding author

and variability in interpretation leading to variability in results and potential misdiagnoses. As a result, the development of automated and objective diagnostic approaches has become a priority in medical research to enhance accuracy and efficiency (Jeleń *et al*., 2025).

Today, it is a well known fact that machine learning has emerged as a powerful tool for automating the classification of cytological images, offering data-driven decision support in pathology (Kowal *et al*., 2013). However, the effectiveness of these models is highly dependent on the quality of extracted features, as well as the selection of relevant predictors for classification. In cervical cytology imaging, feature extraction typically follows two approaches:

- *handcrafted features*, where morphological and texture-based descriptors are designed based on domain knowledge;
- *deep learning-based features*, where CNNs automatically learn hierarchical patterns from raw images.

While CNNs have demonstrated state-of-the-art accuracy in medical image classification, they demand large datasets and high computational resources and often lack interpretability, which is crucial for clinical applications. Handcrafted features, on the other hand, offer greater transparency and interpretability but require rigorous selection methods to optimize classification performance. In this work, we develop and evaluate a novel cervical cancer classification framework that systematically integrates handcrafted morphological descriptors, LBP-based texture features, and CNN-derived deep features extracted from single-cell images prepared via liquid-based cytology. In contrast to recent studies that focus mainly on deep feature fusion or use conventional Pap smear datasets, our approach combines multiple levels of feature abstraction, ensuring both clinical interpretability and computational efficiency. For this purpose, we utilize images from liquid-based cytology slides as described in Section 2. Furthermore, we investigate the role of handcrafted features and deep learning-based feature representations in medical image classification, comparing their individual and combined impact on predictive performance. We also enhance model interpretability and computational efficiency by applying the Lasso, tree-based selection and SHAP methods to reduce feature dimensionality while maintaining high classification precision. This study also provides a systematic comparison of different classification pipelines, offering insights into the feasibility of hybrid approaches for cervical cancer diagnosis, particularly in resource-constrained medical environments—settings with limited access to computational resources. The findings from this work contribute to the ongoing advancements in artificial intelligence driven cytological analysis, supporting the development of scalable and reliable diagnostic tools for improved patient outcomes.



(a) (b) (c)

Fig. 1. Squamous intraepithelial lesions: healthy cell (NSIL) (a), low (LSIL) (b), high (HSIL) (c).

2. Cervical cancer diagnosis

Cervical cancer diagnosis relies on cytological screening, the primary method recommended by the World Health Organization. Traditionally, pathomorphologists examine cell samples, but subjectivity and variability in interpretation affect diagnostic accuracy. To enhance objectivity, HPV testing has been integrated into screening, improving sensitivity in detecting high-grade cervical lesions (Kitchener *et al*., 2009).

Liquid-based cytology (LBC) represents a major advancement over conventional smear-based methods, improving sample consistency and diagnostic reliability. It suspends collected samples in a preservation medium, isolates cellular material through automated filtration, and facilitates molecular testing, streamlining workflows and enhancing accuracy (Kitchener *et al*., 2009). To standardize classification, the Bethesda system (TBS) has been introduced, which provides diagnostic criteria for categorizing abnormalities such as low-grade squamous intraepithelial lesions (LSILs), which have a minimal malignancy risk, and high-grade squamous intraepithelial lesions (HSILs), indicating a higher likelihood of cancer progression, as shown in Fig. 1. TBS provides a structured approach to assessing nuclear enlargement, cytoplasmic ratio changes, and cell morphology, which are critical for accurate diagnosis (Stankiewicz, 2018). However, interobserver variability, subjectivity, and diagnostic inconsistencies remain challenges faced in cervical cancer screening.

Machine learning (ML) offers a promising solution by enabling automated cytological classification based on predefined diagnostic features. This study proposes an ML-driven framework that integrates clinical cytological criteria with computational feature extraction, improving classification accuracy and efficiency. To address these challenges, this work introduces a framework that integrates clinical cytological criteria with computational feature extraction, aiming to improve the accuracy and objectivity of cervical cancer diagnosis.

3. Cervical cancer classification

Given the increasing role of machine learning in medical image classification, numerous studies have explored various feature extraction techniques, classification algorithms, and hybrid approaches to optimize cervical cancer detection (Jeleń *et al*., 2025; Nithya and Ilango, 2019). The current section provides an overview of existing machine learning-based classification methods used in cervical cancer diagnosis, highlighting key advancements, challenges, and emerging trends in the field.

Over the years, numerous studies have explored computational approaches for cervical cancer classification. Rahmadwati *et al*. (2011) applied Gabor filter-based texture analysis to assess nuclear structures, achieving a 89% accuracy. Similarly, Mariarputham and Stephen (2015) used GLCM and LBP features with the SVM and neural networks, reporting 90% of accuracy. In 2014, Mazurek and Oszutowska-Mazurek introduced a novel, fractal-based area–perimeter method for characterizing cervical cell nuclei in Pap smears. More recently, deep learning models have been widely adopted, with fuzzy clustering methods achieving over a 96% accuracy (Qu *et al*., 2021; Rajeev, 2021).

Deep learning-based techniques have shown high classification performance, as demonstrated by Khoulqi and Idrissi (2022). Hybrid approaches, incorporating edge detection, wavelet transformations, and CNNs, have further enhanced classification accuracy (Subarna and Sukumar, 2022). However, most studies focus on Pap smear-based classification, whereas liquid-based cytology remains underexplored. Research by Klug *et al*. (2013) confirmed LBC's superior diagnostic accuracy, motivating the development of LBC-specific machine learning frameworks.

Recent work by Sornapudi *et al.* (2020) (95% accuracy), Hut *et al.* (2022) (70.8% accuracy with VGG16), and Kanavati *et al.* (90.7% accuracy with deep learning) demonstrates the potential of machine learning for LBC classification. Wong *et al*. (2023) applied ResNet architectures, emphasizing the importance of balanced datasets. Current research (e.g., Mathivanan *et al*., 2024) focuses on fusing deep learning models for the classification of cervical cancer. In contrast, our approach integrates deep features with handcrafted and LBP-based descriptors to improve both accuracy and clinical interpretability. Other recent studies (Sabeena *et al*., 2022; Alquran *et al.*, 2022; Sharma *et al*., 2024; Uthayakumar *et al*., 2018) further explore Pap smears or cervical cancer classification using CNN-based and ensemble architectures, attention mechanisms, or handcrafted descriptors. Compared to these works, our framework addresses LBC-based single-cell grading and emphasizes feature interpretability through hybrid fusion.

Beyond deep learning, feature-based methods remain an active area of research, with LBP and advanced feature extraction techniques improving classification efficiency (Su *et al*., 2023; Tarawneh *et al*., 2020). Awais *et al*. (2024) recently proposed a probabilistic feature selection framework for leukocyte classification, demonstrating that optimization of deep features can significantly reduce dimensionality while maintaining high accuracy. Feature selection has also been explored, with Nithya and Ilango (2019) optimizing feature selection for cervical cancer prediction, and Jeleń *et al*. (2025) showing that optimized feature vectors can achieve CNN-like classification accuracy. These findings reinforce the importance of feature optimization in enhancing cervical cancer classification performance. Although previous studies have explored various feature extraction methods and classification approaches for cervical cancer detection, much of the existing research has focused on Pap smear-based cytology. Furthermore, while deep learning models have demonstrated high accuracy, their computational demands and lack of interpretability pose challenges for practical clinical implementation, particularly in resource-constrained environments. To address these gaps, this study develops and evaluates an optimized cervical cancer classification framework that integrates multiple feature extraction techniques, feature selection methods, and machine learning classifiers to enhance diagnostic precision.

By bridging the gap between deep learning-based feature representations and traditional machine learning classifiers, this study highlights how modern feature selection and interpretability techniques can enhance the performance of well-established models. The integration of explainable AI (XAI) techniques, such as SHAP values, ensures that the classification process remains interpretable and clinically relevant. By evaluating hybrid feature extraction techniques, feature selection strategies, and a diverse set of classifiers, we provide insights into model generalization, inference efficiency, and interpretability, supporting the adoption of scalable, clinically viable AI-driven diagnostic tools.

4. Materials and methods

This section describes the methodology for developing and evaluating a machine learning framework for the classification of cervical cancer from liquid-based cytology slides. We explore five different feature extraction approaches: manual features, hand-made features, local binary patterns (LBPs), deep learning-based features (CNN) and hybrid combinations to assess their impact on classification performance. The study applies feature selection methods like Lasso, tree-based, RFE, or SHAP to improve interpretability and reduce feature dimensionality. Seven classifiers are

evaluated, including traditional models (SVM, LDA, k-NN), ensemble methods (gradient boosting, extra trees), and an MLP neural network.

To ensure robust and generalizable results, all models are trained and tested using 5-fold cross-validation. In each fold, grid search hyperparameter tuning is performed on the training set with internal cross-validation, ensuring that the test data remain unseen during model optimization. Performance is evaluated using accuracy, precision, and recall, which provide a comprehensive measure of the classifiers' effectiveness. Additionally, explainable AI techniques, such as SHAP values, are used to interpret model decisions, ensuring alignment with clinically relevant cytological markers. The following subsections describe the dataset, feature extraction and selection techniques, classification models, and evaluation metrics used in this study.

4.1. Cervical cancer database. To assess a proposed framework, a database of cytological images derived from cervical smears, prepared via the LBC technique, was assembled. Before digitization, smears were stained following the Papanicolaou method, which enhances the contrast of cellular structures and nuclei through specialized staining agents. The stained slides were obtained from the Department of Pathomorphology at the Pomeranian Medical University in Szczecin, Poland, and subsequently digitized at the Department of Pathology and Oncological Cytology of the Medical University of Wroclaw, Poland. Image acquisition was conducted using an Olympus BX 51 microscope equipped with an Olympus DP72 camera, with the images captured at a 300 dpi resolution and the dimensions of 2510×1540 pixels using the Olympus CellD software.

The dataset consists of 263 digitized LBC microscopic images, categorized as follows: 158 normal squamous intraepithelial lesions (NSILs), 70 high-grade squamous intraepithelial lesions (HSILs), and 35 low-grade squamous intraepithelial lesions (LSILs). For further cell shape analysis, the ImageJ software was used to extract 428 individual cell images, including 233 NSILs, 110 HSILs, and 85 LSILs. To our knowledge, no public LBC dataset provides expert-labeled single-cell images with the NSIL/LSIL/HSIL grading, limiting direct external validation. All cytological cases included in our dataset were validated by an expert pathomorphologist. The dataset is based on a retrospective study, and all images were irreversibly anonymized to ensure compliance with ethical guidelines.

4.2. Feature extraction. In this paper, we discuss four feature extraction approaches to evaluate their individual and combined effectiveness in cervical cancer classification. Their classification is then compared with features derived manually by an expert diagnostician. These features refer to cytomorphological descriptors defined and selected according to established cytology criteria (see Stankiewicz, 2018). Furthermore, all feature extraction methods are applied to single-cell images, ensuring that each extracted feature represents individual cellular characteristics rather than aggregated slide-level data.

4.2.1. Handcrafted features. The concept of handcrafted features was developed to resemble the traditional manual assessment of cytological samples, where pathologists evaluate cell morphology, nucleus size, shape irregularities, and texture variations to distinguish between normal and abnormal cells. In manual diagnosis, characteristics such as nuclear enlargement, an increased nucleus-to-cytoplasm ratio, irregular nuclear contours, and chromatin texture alterations serve as key indicators of malignancy. Inspired by this visual and morphological analysis, handcrafted features were designed to quantify these characteristics in an objective and reproducible manner.

The extraction of handcrafted features requires a binary image representation, where the cell and nucleus are clearly segmented to ensure accurate computation of morphological descriptors. To achieve this, K-means clustering was employed to segment the images into distinct regions. This method effectively separates the nucleus, cytoplasm, and background, providing a clean binary mask for feature computation. K-means clustering was applied in the RGB color space, segmenting each image into three clusters ($k = 3$), corresponding to the nucleus, cytoplasm, and background. The segmented image was then thresholded into binary masks, ensuring that handcrafted features were extracted exclusively from the relevant cellular structures.

Based on the binary nucleus and cell images, 27 handcrafted features were extracted to capture morphological properties in single-cell images. The following features were computed:

- *morphological features*: nucleus size (N), cell size (C), nucleus-to-cell ratio (NCr), nucleus perimeter (Np), cell perimeter (Cp), nucleus-to-cell perimeter ratio (NCp), minimum axis length (MinA), maximum axis length (MaxA), and their respective ratios (MinAr, MaxAr);
- *shape and compactness features*: nucleus aspect ratio (Nar), cell aspect ratio (Car), nucleus-to-cell aspect ratio (NCar), extent (Ext), nucleus-to-cell extent ratio (NCExt), solidity (Sol), nucleus-to-cell solidity ratio (NCs), equivalence diameter (Eq), and nucleus-to-cell equivalence diameter ratio (NCEq);
- *orientation and spatial distribution features*: orientation (Or) and nucleus-to-cell orientation ratio (NCOr).

A detailed description of these features, including their mathematical formulations and clinical significance, can be found in the work of Jeleń *et al*. (2025).

4.2.2. Local binary pattern features. The local binary pattern (LBP) feature set was developed to quantify textural variations in cytological images, particularly those related to chromatin distribution and nuclear irregularities, which are key indicators in manual cytological assessments. In traditional pathology, cytologists assess nuclear texture, looking for coarse chromatin, granularity, and subtle intensity variations that may signal cellular abnormalities. However, manual assessment of texture is inherently subjective and difficult to standardize, making automated texture analysis a crucial addition to machine learning-based classification.

The LBP method was introduced as a robust descriptor for capturing micro-patterns in grayscale images, translating local intensity differences into a compact numerical representation (Ojala *et al*., 1996). The method works by comparing each pixel to its neighboring pixels within a predefined radius and encoding the result into a binary pattern, which is then converted into a decimal value. This transformation effectively captures fine textural details, allowing the model to differentiate between homogeneous and heterogeneous nuclear structures, which is an important aspect of cervical cytology grading.

By applying local binary patterns to single-cell images, this study ensures that nuclear texture features are captured in an interpretable and reproducible manner, providing valuable information for distinguishing between NSIL, LSIL, and HSIL cell samples. The extracted LBP descriptors complement morphological features, offering a more comprehensive representation of cellular abnormalities.

4.2.3. Deep learning-based CNN features. Unlike handcrafted and texture-based feature extraction methods, convolutional neural networks (CNNs) provide an automated, data-driven approach to extracting hierarchical representations from cytological images. In traditional cytological assessment, pathologists visually identify subtle differences in nuclear shape, chromatin distribution, and cytoplasmic texture, making deep learning-based features particularly valuable in capturing complex spatial patterns that may be difficult to quantify manually. In this study, high-dimensional feature representations were extracted from single-cell images using a pre-trained VGG16 model, a widely used deep CNN architecture known for its strong feature extraction capabilities in medical imaging tasks (Devi and Neelambary, 2022). Due to the relatively small dataset, we used a pre-trained VGG16 model as a fixed feature extractor without fine-tuning. For comparison purposes, we also performed limited fine-tuning using 10% of the dataset. This proportion was selected because of the limited size of the database. The goal was to evaluate whether even slight adjustments to CNN weights enhance classification performance while maintaining a fair comparison with traditional and hybrid feature sets. Specifically, we unfroze convolutional blocks 4 and 5 of the VGG16 model and trained them for 15 epochs using the Adam optimizer, with early stopping based on validation loss. Earlier layers remained frozen to preserve general visual features. The features were extracted from the last convolutional block, producing activation maps of size $7\times7\times512$, which were flattened into 25,088 dimensional feature vectors. These were then used as input to traditional machine learning classifiers. This approach enables the model to capture hierarchical spatial features while minimizing data annotation requirements. To perform the feature extraction, we followed a three-step procedure described below:

1. *Image normalization and resizing*: in this step, single-cell images were resized to 224×224 pixels to match the input dimensions of the VGG16 model, ensuring consistency across samples.

2. *Convolutional feature extraction*: to create a feature vector, in this step, features were extracted from the last convolutional layer of the network before the fully connected layers. This ensures that spatial and structural patterns are captured while discarding classification-specific information.

3. *Flattening and dimensionality reduction*: for further classification, the extracted feature maps were flattened into 1D vectors. This step assures that the feature vector can be later used as input for the classification models.

VGG16 was chosen due to its proven effectiveness in cytological image analysis, its computational efficiency, and its capability to extract features without requiring large datasets or high computational resources. This choice ensures a fair comparison with handcrafted features and maintains clinical feasibility in resource-constrained environments.

Compared with handcrafted and LBP-based features, the deep learning-based features have an advantage in the sense that they automatically learn relevant features from the data, rather than relying on predefined descriptors. Furthermore, the hierarchical nature of CNNs allows them to detect low-level textures like edges or gradients, mid-level structures like shapes and nuclei, and high-level patterns like spatial arrangements and relationships between nuclei and cytoplasm.

However, while CNN-based features offer strong discriminatory power, they also come with increased

computational cost and reduced interpretability compared to handcrafted and LBP-based features. To mitigate this, we applied an explainable AI technique, i.e., SHAP (Shapley additive explanations), to analyze the most valuable features and ensure alignment with clinically relevant cytological markers.

Incorporating the deep learning-based features, this study enhances LBC-based cervical cancer classification by allowing machine learning models to use automatically learned representations, complementing the handcrafted and texture-based features.

4.2.4. Hybrid feature evaluation. While individual feature sets provide valuable information for cervical cancer classification, each captures distinct aspects of cellular morphology and structure. To assess the complementary nature of these feature sets, we explored the following four hybrid feature combinations derived from single-cell images, evaluating their impact on classification performance:

1. *Handcrafted and LBP features.* Handcrafted morphological features provide quantitative measurements of nuclear size, shape, and structural irregularities, while local binary patterns capture local textural variations that may be indicative of chromatin abnormalities. Combining these two feature sets enables the classifier to leverage both morphological changes and texture differences, which are crucial for distinguishing between low- and high-grade squamous intraepithelial lesions.

2. *Handcrafted and CNN features*. While handcrafted features focus on explicitly defined morphometric properties, CNN-derived ones capture higher-order spatial relationships that are difficult to encode manually. This combination balances interpretable, clinically relevant morphological metrics with deep features allowing for a hybrid representation that bridges traditional cytology with the AI-driven approach.

3. *LBP and CNN features*. These feature types both focus on texture and spatial information, but at different levels of abstraction. LBPs emphasize localized micro-patterns within the nucleus and cytoplasm, whereas CNN features capture multi-scale hierarchical patterns across the entire cell. Combining these methods enhances the model's ability to detect subtle chromatin variations and complex spatial structures, potentially improving the classification of borderline cases.

4. *Handcrafted, LBP and CNN features*. The most comprehensive feature representation integrates all three feature sets using explicit morphological measurements, localized texture information, and deep hierarchical patterns. This approach ensures that the classifier has access to the widest range of descriptive attributes, which potentially maximizes classification performance. However, this fusion also introduces higher dimensionality, requiring effective feature selection methods like Lasso, tree-based selection or SHAP to reduce redundancy and enhance model efficiency.

Systematic evaluation of these hybrid feature combinations aims to determine which fusion strategy provides the most discriminative power for cervical cancer classification. The results of these experiments will offer insights into the synergy between handcrafted, texture-based, and deep learning based features, guiding the development of optimal feature selection and classification strategies.

4.3. Feature selection. As mentioned by Guyon and Elisseeff (2003) excessive number of features does not necessarily improve model performance. According to their study, when classification accuracy remains largely unaffected and the error rate does not increase, reducing the feature vector size is a more efficient approach. In this work, we embrace this idea by applying feature selection methods that allow elimination of redundant, non-informative, or correlated features while preserving those that contribute most to cancer classification. To achieve this, we employ four feature selection techniques, each with unique advantages in dimensionality reduction, classifier performance enhancement, and interpretability.

1. *Lasso.* Lasso regression is a standard feature selection method in high-dimensional data analysis, frequently used in biomedical applications. It is a linear feature selection technique that enforces sparsity by penalizing the absolute magnitude of feature coefficients. It solves the optimization problem described with Eqn. (1),

$$\hat{\beta} = \arg\min_{\beta} \left(||y - X\beta||^2 + \lambda ||\beta||_1 \right) \qquad (1)$$

where X represents the feature matrix, y is the target variable, and λ is the regularization parameter controlling sparsity. Features with coefficients minimized to zero are considered irrelevant and removed, ensuring that only the most predictive features are retained.

2. *Tree-based feature selection.* Tree-based ensemble methods naturally calculate feature importance scores by evaluating how splitting on each feature improves classification performance. The feature importance score for a given feature f_i is determined as:

$$Importance(f_i) = \sum_{t=1}^{T} \frac{\Delta Gini(f_i)}{T}, \qquad (2)$$

where $\Delta Gini(f_i)$ represents the decrease in Gini impurity resulting from splits on feature f_i aggregated over T trees in the ensemble. Features with the highest importance scores are retained, while those with minimal contribution are discarded.

3. *Recursive feature elimination with cross-validation.* Recursive feature elimination is an iterative backward feature selection technique that systematically removes the least significant features based on a trained estimator (Freytes *et al*., 2023). In this study, we applied RFECV using a random forest classifier, where feature selection is guided by the classifier's internal importance metrics. Cross-validation ensures that the selected feature set generalizes well, reducing the risk of overfitting to specific cytology datasets.

4. *Shapley additive explanations.* To enhance explainability and interpretability, we employed SHAP-based feature selection, a model-agnostic approach that assigns a Shapley value to each feature, representing its contribution to the classification outcome (Molnar *et al*., 2020). The SHAP value for a given feature f_i is calculated as

$$SHAP(f_i) = \sum_{S \subseteq F \setminus \{f_i\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} \times (score(S \cup \{f_i\}) - score(S)), \tag{3}$$

where F is the set of all features, S is a subset of features excluding f_i, and $score(S)$ represents the model's output accuracy using subset S. This method evaluates how adding or removing a specific feature impacts classification performance and ensures that only the most influential features are retained while maintaining interpretability. Here, SHAP was applied to all feature types, including CNN vectors, to identify the most impactful features for further feature vector reduction.

4.4. Feature classification. To evaluate the effectiveness of different machine learning approaches for LBC cervical cancer classification, we trained and tested seven classifiers using extracted single-cell feature vectors. These classifiers were selected to represent a diverse range of learning paradigms, including linear models, instance-based methods, ensemble techniques, and neural networks. Each classifier was assessed across various feature extraction and selection strategies to determine the most effective approach for single-cell classification.

1. *Support vector machine (SVM).* This is a widely used supervised learning algorithm that is particularly effective in high-dimensional spaces. It works by identifying an optimal hyperplane that maximally separates different classes while minimizing misclassification (Cristianini and Shawe-Taylor, 2000). In this study, we employed a radial basis function kernel, which enables the model to capture non-linear relationships between features. Given the complexity of cytological image data, the SVM provides robust decision boundaries for distinguishing between NSIL, LSIL, and HSIL cells.

2. *K-nearest neighbors (KNNs).* This is a popular and simple non-parametric, instance-based learning algorithm that classifies a sample based on the majority class among its k-nearest neighbors in feature space (Duda *et al*., 2000). In this study, we used five nearest neighbors, chosen empirically for optimal performance. The KNN is particularly effective when dealing with localized decision boundaries, making it useful for identifying subtle morphological variations in single-cell images. However, its sensitivity to high-dimensional data and potential for computational inefficiency requires a careful feature selection.

3. *Linear discriminant analysis (LDA).* This is a probabilistic model that seeks to find a linear combination of features that best separates different classes (Duda *et al*., 2000). It assumes that the data follows a Gaussian distribution and maximizes the between-class variance while minimizing within-class variance. Given its strong interpretability and effectiveness in low-dimensional spaces, LDA serves as a useful baseline for morphological and handcrafted feature sets, where class separability is well-defined.

4. *Gradient boosting (GB).* This is an ensemble classifier that builds multiple weak learners (decision trees) in a sequential manner, optimizing classification performance by minimizing residual errors. It is highly effective for handling complex feature interactions, making it well-suited for hybrid feature sets. While computationally expensive, GB is known for its high accuracy and robustness in medical imaging tasks (Ramkumar and Sajiv, 2023).

5. *Histogram-based gradient boosting (HGB).* This is an optimized version of the gradient boosting classifier that discretizes continuous features into histograms. This significantly improves computational efficiency for large datasets. This model is useful for deep learning-derived feature sets, where traditional gradient boosting might be computationally excessive. By reducing memory usage and training time, this method provides a scalable solution for a real-world clinical scenarios (Liew *et al*., 2021).

6. *Extra trees classifier (ETC).* This is a variant of the random forest classifier that introduces additional

Table 1. Classification results for different feature extraction methods: best results in bold, second-best underlined (per group).

	Manual			Handcrafted			LBP		
Classifier	*Accuracy*	*Precision*	*Recall*	*Accuracy*	*Precision*	*Recall*	*Accuracy*	*Precision*	*Recall*
Traditional feature sets									
SVM	**0.8958**	**0.8995**	**0.8958**	**0.9592**	**0.9592**	**0.9592**	0.9592	0.9592	0.9592
KNN	**0.8958**	**0.8995**	**0.8958**	0.8980	0.9042	0.8980	**0.9592**	**0.9640**	**0.9592**
LDA	**0.8958**	**0.8995**	**0.8958**	0.8980	0.9102	0.8980	0.7347	0.7678	0.7347
GB	0.8750	0.8778	0.8750	0.9184	0.9242	0.9184	0.8776	0.8844	0.8776
HGB	0.8958	0.8942	0.8958	0.9388	0.9413	0.9388	0.8776	0.9125	0.8776
ET	0.8750	0.8741	0.8750	**0.9592**	**0.9592**	**0.9592**	0.8776	0.9125	0.8776
MLP	0.8750	0.8732	0.8750	0.8163	0.8200	0.8163	0.8776	0.8872	0.8776

Deep learning-based features						
	CNN			CNN (fine-tuned)		
Classifier	*Accuracy*	*Precision*	*Recall*	*Accuracy*	*Precision*	*Recall*
SVM	0.3061	0.7876	0.3061	0.5918	0.8123	0.5918
KNN	0.4694	0.6162	0.4694	0.6735	0.6941	0.6735
LDA	0.4286	0.4157	0.4286	0.6327	0.6214	0.6327
GB	0.7959	0.8197	0.7959	0.8571	0.8722	0.8571
HGB	**0.8571**	**0.8629**	**0.8571**	**0.8980**	**0.8932**	**0.8980**
ET	0.7959	0.8085	0.7959	0.8673	0.8694	0.8673
MLP	0.5510	0.5732	0.5510	0.7347	0.7419	0.7347

randomness in the construction of decision trees. Unlike traditional decision trees, where optimal split points are chosen based on information gain or Gini impurity, the selects split points randomly, reducing variance and improving generalization. This classifier is particularly effective for heterogeneous feature sets, where handcrafted and CNN-derived features need to be combined without overfitting (Sajiv and Ramkumar, 2023).

7. *Multilayer perceptron (MLP).* This is a feedforward artificial neural network that learns feature representations through multiple hidden layers and non-linear activation functions. In this study, we used an MLP architecture with two hidden layers (100,100 neurons each) and a ReLU activation function. The MLP is particularly beneficial when working with CNN-extracted features, as it can learn complex non-linear relationships between high-dimensional representations. However, it requires careful hyperparameter tuning to prevent overfitting, especially when dealing with limited training data (Jeleń *et al.*, 2025).

The above classifiers were tested using the described feature extraction and selection strategies, ensuring a comprehensive comparison of classification performance. Our research was aimed at evaluating the accuracy and effectiveness of individual feature combinations in enhancing classifier performance. We also analyzed computational efficiency, training and inference as well as model interpretability using SHAP-based feature selection for understanding classifier decisions.

To assess classifier performance, we used a grid search optimization for hyperparameter tuning and a five-fold cross-validation strategy to enhance model generalizability. For comprehensive evaluation of the results, the multiclass variant of the confusion matrix was used. Based on the matrix values, the accuracy, precision and recall evaluation metrics were used. Accuracy provides the proportion of correctly classified cells, while precision is a proportion of true positive predictions among predicted positives, and recall being a proportion of correctly identified abnormal cases.

5. Results

In this section, we present the results of the liquid-based cytology single-cell classification experiments, evaluating the impact of different feature selection strategies and classification models. The results provide insights into the effectiveness of handcrafted, local binary patterns and CNN-based features, as well as hybrid feature combinations in enhancing classification performance. We analyze the impact of feature selection techniques and evaluate the performance of the seven classifiers described in Section 4.4.

To assess the contribution of each feature extraction method, we first evaluated the classification accuracy of models trained separately on manual features, handcrafted features, LBP-based features, and CNN-extracted features. The results, summarized in Table 1, highlight the relative performance of each feature type across different classifiers.

The results show that manual features provided

Table 2. Classification results for hybrid features: best results in bold, second-best underlined. Top: hybrids using CNN features, bottom: hybrids using fine-tuned CNN features.

	Handcrafted + LBP			Handcrafted + CNN			LBP + CNN			Handcrafted + LBP + CNN		
	Accuracy	*Precision*	*Recall*	*Accuracy*	*Precision*	*Recall*	*Accuracy*	*Precision*	*Recall*	*Accuracy*	*Precision*	*Recall*
Hybrid features using CNNs (No fine-tuning)												
SVM	0.9592	0.9592	0.9592	0.3061	0.7876	0.3061	0.3061	0.7876	0.3061	0.3061	0.7876	0.3061
KNN	0.9388	0.9397	0.9388	0.8571	0.8741	0.8571	0.8571	0.8864	0.8571	0.8980	0.9042	0.8980
LDA	0.8367	0.8762	0.8367	0.6939	0.7324	0.6939	0.5714	0.6502	0.5714	0.8571	0.8703	0.8571
GB	0.9388	0.9413	0.9388	0.8980	0.9004	0.8980	0.9184	0.9208	0.9184	0.8980	0.9086	0.8980
HGB	**0.9592**	**0.9640**	**0.9592**	**0.9388**	**0.9384**	**0.9388**	**0.9388**	**0.9384**	**0.9388**	0.9388	0.9384	0.9388
ET	0.9388	0.9413	0.9388	0.8776	0.8945	0.8776	0.8776	0.8781	0.8776	**0.9592**	**0.9615**	**0.9592**
MLP	0.8980	0.9086	0.8980	0.8980	0.9116	0.8980	0.8980	0.9116	0.8980	0.9184	0.9256	0.9184
Hybrid features using fine-tuned CNNs												
SVM	—	—	—	0.3267	0.8169	0.3321	0.3267	0.8169	0.3321	0.3267	0.8169	0.3321
KNN	—	—	—	0.8811	0.8914	0.8744	0.8811	0.9037	0.8744	0.9220	0.9215	0.9153
LDA	—	—	—	0.7098	0.7604	0.7179	0.5873	0.6782	0.5954	0.8730	0.8983	0.8811
GB	—	—	—	0.9236	0.9157	0.9275	0.9440	0.9361	0.9479	0.9236	0.9239	0.9275
HGB	—	—	—	**0.9663**	**0.9566**	**0.9565**	**0.9663**	**0.9566**	**0.9565**	**0.9663**	**0.9566**	**0.9565**
ET	—	—	—	0.9388	0.9473	0.9388	0.9388	0.9473	0.9388	**0.9663**	**0.9566**	**0.9565**
MLP	—	—	—	0.9388	0.9463	0.9388	0.9388	0.9463	0.9388	**0.9663**	**0.9566**	**0.9565**

moderate accuracy of 89.58% with the SVM, KNN, and LDA, but lacked strong discriminatory power, particularly in ensemble models. Handcrafted features performed best, reaching a 95.92% accuracy with the SVM and extra trees, though the MLP achieving a 81.63% accuracy suggests that deep learning models may not fully exploit these descriptors. LBP features were highly effective in the SVM and KNN (95.92% accuracy) but performed poorly with the LDA (73.47%), indicating that they are not well-suited for linear separation. CNN-derived features showed mixed results, with the SVM and LDA performing poorly with a 30.61% and a 42.86% accuracy, while histogram-based gradient boosting (85.71%) and gradient boosting (79.59%) handled them more effectively. These findings highlight the strengths of handcrafted and LBP features in traditional classifiers and the need for hybrid approaches to fully leverage CNN-based features.

In Table 2 the classification results for hybrid feature sets are presented. They reveal that handcrafted and local binary pattern features achieved the best overall performance, with the SVM, gradient boosting and histogram-based gradient boosting reaching a 95.92% accuracy, highlighting the strength of combining morphological and texture descriptors. In contrast, handcrafted and CNN features showed inconsistent results, with the SVM performing poorly while gradient boosting and the MLP performed well (89.80% accuracy), suggesting that CNN-derived features may introduce noise when combined with structured handcrafted features. A similar pattern was observed for the LBP and CNN, where the SVM again struggled but gradient boosting and the KNN showed improvements, achieving 91.84% and 85.71% of accuracy, respectively. The full feature combination performed best in extra trees, with a 95.92% accuracy, and the MLP with a 91.84% accuracy, while the SVM remained ineffective. These results suggest that handcrafted and LBP features are most effective for traditional classifiers, while CNN features require more flexible models like ensembles or deep learning-based approaches for optimal classification.

The next part of the research demonstrates that combining handcrafted, LBP, and CNN-derived features significantly improves classification performance for all classifiers. The results are shown in Tables 3 and 6, where it can be noted that this approach consistently achieved the highest accuracy, particularly when the SHAP method was applied. From the results we can see that the best classifiers were histogram-based gradient boosting and extra trees, reaching a 99.10% accuracy, 99.75% precision, and 98.17% recall. This confirms that integrating morphological, texture-based, and deep learning features enhances model robustness and classification power. Among individual feature sets, the LBP and CNN outperformed standalone CNN or LBP features, demonstrating the complementary nature of local texture descriptors and deep feature representations. While CNN features alone struggled in linear classifiers such as the SVM and LDA, their combination with handcrafted and LBP features improved performance in ensemble classifiers and the MLP. These results suggest that CNN-based deep learning features provide valuable abstract representations, but their effectiveness improves significantly when fused with handcrafted and texture-based descriptors.

In conclusion, the presented results confirm that deep learning-based features are highly informative but require additional morphological or texture-based descriptors to

Table 3. Classification results after feature selection across all classifiers and feature sets: best results in bold, second-best underlined.

	Lasso			Tree-based			RFECV			SHAP		
	Accuracy	*Precision*	*Recall*	*Accuracy*	*Precision*	*Recall*	*Accuracy*	*Precision*	*Recall*	*Accuracy*	*Precision*	*Recall*
Handcrafted features												
SVM	0.8891	0.9023	0.8827	0.9215	0.9312	0.9150	0.8702	0.8849	0.8613	0.9423	0.9516	0.9321
KNN	0.8725	0.8874	0.8638	0.9098	0.9205	0.9032	0.8589	0.8713	0.8405	0.9281	0.9397	0.9214
LDA	0.8294	0.8452	0.8136	0.8541	0.8720	0.8392	0.7982	0.8156	0.7805	0.8697	0.8821	0.8602
GB	0.8957	0.9104	0.8798	0.9265	0.9382	0.9171	0.8749	0.8897	0.8543	0.9482	0.9579	0.9354
HGB	0.9108	0.9243	0.8954	0.9416	0.9517	0.9278	0.8893	0.9035	0.8702	**0.9635**	**0.9714**	**0.9493**
ET	0.9023	0.9152	0.8857	0.9367	0.9469	0.9223	0.8816	0.8958	0.8627	0.9571	0.9653	0.9437
MLP	0.8852	0.8994	0.8679	0.9189	0.9307	0.9065	0.8647	0.8795	0.8413	0.9358	0.9451	0.9235
LBP features												
SVM	0.8794	0.8952	0.8698	0.9087	0.9204	0.8985	0.8542	0.8671	0.8408	0.9245	0.9367	0.9103
KNN	0.8601	0.8765	0.8473	0.8973	0.9091	0.8829	0.8324	0.8467	0.8123	0.9118	0.9234	0.8967
LDA	0.8192	0.8347	0.8004	0.8452	0.8609	0.8293	0.7793	0.7921	0.7523	0.8642	0.8776	0.8541
GB	0.8903	0.9051	0.8705	0.9223	0.9345	0.9120	0.8689	0.8835	0.8481	0.9431	0.9527	0.9309
HGB	0.9056	0.9192	0.8894	0.9381	0.9489	0.9235	0.8827	0.8974	0.8652	**0.9593**	**0.9674**	**0.9458**
ET	0.8962	0.9103	0.8768	0.9327	0.9432	0.9189	0.8741	0.8886	0.8549	0.9546	0.9625	0.9404
MLP	0.8805	0.8947	0.8609	0.9145	0.9261	0.9018	0.8573	0.8729	0.8357	0.9324	0.9417	0.9203
CNN features												
SVM	0.7398	0.7552	0.7201	0.7849	0.8017	0.7683	0.7023	0.7210	0.6894	0.8056	0.8187	0.7923
KNN	0.7564	0.7721	0.7386	0.8127	0.8274	0.8032	0.7310	0.7465	0.7103	0.8294	0.8412	0.8183
LDA	0.7204	0.7385	0.6998	0.7608	0.7779	0.7402	0.6931	0.7084	0.6732	0.7823	0.7967	0.7681
GB	0.8052	0.8198	0.7837	0.8503	0.8642	0.8315	0.7805	0.7963	0.7509	0.8727	0.8853	0.8604
HGB	0.8237	0.8385	0.8012	0.8701	0.8847	0.8523	0.7985	0.8142	0.7691	0.8956	0.9071	0.8825
ET	0.8105	0.8251	0.7894	0.8604	0.8743	0.8401	0.7852	0.8008	0.7557	0.8829	0.8954	0.8708
MLP	0.8456	0.8597	0.8249	0.8889	0.9025	0.8694	0.8251	0.8398	0.7984	**0.9112**	**0.9238**	**0.8997**
CNN features (fine-tuned)												
SVM	0.7555	0.7784	0.7428	0.8028	0.8186	0.7931	0.7214	0.7436	0.7053	0.8237	0.8399	0.8119
KNN	0.7736	0.7937	0.7604	0.8372	0.8452	0.8244	0.7498	0.7655	0.7347	0.8537	0.8657	0.8371
LDA	0.7388	0.7601	0.7152	0.7781	0.7972	0.7560	0.7158	0.7319	0.6897	0.7983	0.8144	0.7834
GB	0.8286	0.8408	0.8080	0.8703	0.8830	0.8489	0.8015	0.8146	0.7705	0.8949	0.9084	0.8853
HGB	0.8410	0.8592	0.8166	0.8945	0.9024	0.8731	0.8172	0.8318	0.7860	0.9162	0.9310	0.8980
ET	0.8264	0.8409	0.8072	0.8784	0.8912	0.8592	0.8054	0.8227	0.7709	0.9076	0.9172	0.8952
MLP	0.8675	0.8764	0.8422	0.9135	0.9228	0.8900	0.8494	0.8572	0.8163	**0.9339**	**0.9398**	**0.9164**

Note: Standard deviations (σ) for accuracy, precision, and recall across five-fold cross-validation are low for all classifiers. For example, for the best-performing configuration, accuracy is 99.01% ± 0.56%, precision is 99.54% ± 0.43%, and recall is 97.86% ± 0.61%. For other classifiers, σ values are ranging between 0.4% and 1.2%.

maximize their discriminative potential.

The framework was tested on an Intel Core i5-8500 CPU computer with 16 GB RAM and an NVIDIA GeForce RTX 1080 GPU running Ubuntu. Single-cell pre-processing and classification were completed in less than 1 second per image, with CNN feature extraction taking approximately 3 seconds on CPU and 0.2 seconds using GPU acceleration. These results confirm the suitability of the framework for near-real-time clinical applications using standard hardware.

6. Discussion and conclusions

The results presented in this study demonstrate that integrating handcrafted, textural and CNN-derived features significantly enhances the classification of liquid-based cytology images for cervical cancer diagnosis. Our framework effectively combines traditional morphological descriptors, texture-based LBP features, and deep learning-based representations, leading to improved classification performance with multiple machine learning models. Among the tested classifiers, ensemble-based models consistently achieved the highest accuracy, reaching up to 99.10% when combined with SHAP feature selection. This confirms that feature fusion, combined with an optimal feature selection strategy, is important in achieving robust and clinically relevant predictions.

Furthermore, the study highlights that CNN-derived features alone struggle in traditional classifiers achieving lower accuracies when not combined with other feature sets. However, CNN features significantly improved classification performance when fused with handcrafted or LBP descriptors, particularly in ensemble-based models and the MLP. These findings suggest that deep learning features, while highly informative, require complementary morphological or texture-based descriptors to maximize their diagnostic potential. As for the impact of feature selection methods on classification performance, the SHAP method consistently preserved the highest accuracy across all classifiers. Fine-tuning improved CNN classification, but hybrid sets still outperform the individual datasets. The results demonstrate that tree-based selection is highly effective for ensemble models, whereas Lasso and RFECV can sometimes remove relevant descriptors, leading to minor performance drops.

In this paper, we depicted a robust and clinically relevant framework for automated cervical cancer classification using liquid-based cytology slides. By integrating various types of features, the proposed approach enhances classification accuracy and reliability, surpassing traditional feature extraction methods. Recent studies, such as that by Fang *et al*. (2024), focus on fusing CNN-based features for cervical cell classification. In contrast, our approach integrates handcrafted morphological, LBP texture, and CNN-derived features for LBC single-cell analysis, using data from true liquid-based cytology preparations aligned with current clinical practice. Moreover, this study confirms that feature selection plays a crucial role in optimizing classification accuracy while reducing feature vector size, ensuring a balance between interpretability and computational efficiency. The proposed SHAP selection strategy provides an optimal trade-off, retaining high classification performance while significantly reducing dimensionality.

In conclusion, the proposed framework represents a significant step toward automated cytological diagnosis, offering a clinically relevant, highly accurate, and efficient classification method for LBC slides. The ability to accurately predict cervical cancer cases while maintaining interpretability makes this approach suitable for real-world clinical applications. Future work will focus on expanding the dataset, incorporating additional deep learning architectures, and evaluating the generalizability of the model to other cytological screening tasks. As shown in Section 5, fine-tuning led to modest performance improvements, but the small dataset remained a limiting factor. For that reason, extending the dataset and including external data sources will be essential to fully exploit these models. The promising results achieved in this study highlight the potential impact of machine learning on modern medical diagnostics, paving the way for more standardized, scalable, and objective cytological assessment protocols.

References

Alquran, H., Alsalatie, M., Mustafa, W. A., Abdi, R. A., and Ismail, A. (2022). Cervical Net: A novel cervical cancer classification using feature fusion, *Bioengineering* **9**(10): 578, DOI: 10.3390/bioengineering9100578.

Awais, M., Akram, T., Alasiry, A., Marzougui, M., Park, J. and Chang, B. (2024). An expert system for leukocyte classification using probabilistic deep feature optimization via distribution estimation, *International Journal of Applied Mathematics and Computer Science* **34**(4): 579–595, DOI: 10.61822/amcs-2024-0039.

Cristianini, N. and Shawe-Taylor, J. (2000). *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge University Press, Cambridge.

Devi, G.M. and Neelambary, V. (2022). Computer-aided diagnosis of white blood cell leukemia using VGG16 convolution neural network, *2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India*, pp. 1064–1068.

Duda, R., Hart, P. and Stork, D. (2000). *Pattern Classification*, 2nd Edn, Wiley Interscience Publishers, New York.

Fang, M., Fu, M., Liao, B., Lei, X. and Wu, F.-X. (2024). Deep integrated fusion of local and global features for cervical

Table 4. Classification results for hybrid feature sets after feature selection: best in bold, second-best underlined (continued on the next page).

	Lasso			Tree-based			RFECV			SHAP		
	Accuracy	*Precision*	*Recall*	*Accuracy*	*Precision*	*Recall*	*Accuracy*	*Precision*	*Recall*	*Accuracy*	*Precision*	*Recall*
Handcrafted and LBP features												
SVM	0.9243	0.9362	0.9178	0.9517	0.9604	0.9429	0.9031	0.9183	0.8897	0.9678	0.9741	0.9587
KNN	0.9105	0.9236	0.9001	0.9442	0.9549	0.9328	0.8924	0.9067	0.8743	0.9601	0.9673	0.9492
LDA	0.8829	0.8992	0.8645	0.9237	0.9356	0.9102	0.8643	0.8798	0.8456	0.9445	0.9531	0.9308
GB	0.9264	0.9392	0.9125	0.9603	0.9681	0.9504	0.9098	0.9235	0.8921	0.9745	0.9803	0.9657
HGB	0.9387	0.9521	0.9209	0.9704	0.9776	0.9613	0.9193	0.9334	0.8995	**0.9812**	**0.9874**	**0.9721**
ET	0.9312	0.9453	0.9128	0.9657	0.9725	0.9551	0.9145	0.9289	0.8942	0.9783	0.9842	0.9694
MLP	0.9208	0.9341	0.9063	0.9556	0.9638	0.9467	0.9042	0.9187	0.8815	0.9701	0.9765	0.9604
Handcrafted and CNN features												
SVM	0.9185	0.9311	0.9102	0.9487	0.9578	0.9393	0.8943	0.9116	0.8784	0.9642	0.9703	0.9531
KNN	0.9042	0.9180	0.8904	0.9395	0.9501	0.9278	0.8807	0.8952	0.8604	0.9573	0.9645	0.9456
LDA	0.8721	0.8897	0.8552	0.9206	0.9325	0.9068	0.8584	0.8736	0.8413	0.9437	0.9519	0.9298
GB	0.9284	0.9423	0.9139	0.9621	0.9698	0.9516	0.9127	0.9268	0.8945	0.9758	0.9814	0.9663
HGB	0.9398	0.9536	0.9218	0.9715	0.9789	0.9632	0.9223	0.9364	0.9012	**0.9826**	**0.9881**	**0.9735**
ET	0.9327	0.9468	0.9143	0.9664	0.9737	0.9563	0.9181	0.9325	0.8974	0.9798	0.9856	0.9707
MLP	0.9235	0.9372	0.9095	0.9583	0.9659	0.9492	0.9089	0.9234	0.8861	0.9724	0.9787	0.9621
Handcrafted and CNN features (fine-tuned)												
SVM	0.9403	0.9510	0.9306	0.9359	0.9538	0.9337	0.9410	0.9553	0.9345	0.9672	0.9680	0.9563
KNN	0.9208	0.9406	0.9075	0.9229	0.9369	0.9092	0.9257	0.9407	0.9065	0.9699	0.9607	0.9610
LDA	0.8931	0.9117	0.8721	0.8929	0.9116	0.8758	0.8931	0.9115	0.8738	0.9698	0.9612	0.9569
GB	0.9522	0.9623	0.9337	0.9480	0.9611	0.9343	0.9523	0.9644	0.9340	0.9809	0.9705	0.9695
HGB	0.9565	0.9761	0.9427	0.9596	0.9752	0.9424	0.9618	0.9741	0.9447	**0.9856**	**0.9765**	**0.9732**
ET	0.9524	0.9655	0.9326	0.9519	0.9661	0.9333	0.9547	0.9657	0.9334	0.9793	0.9700	0.9682
MLP	0.9464	0.9560	0.9299	0.9426	0.9607	0.9316	0.9441	0.9580	0.9325	0.9679	0.9602	0.9567
LBP and CNN features												
SVM	0.9104	0.9235	0.9028	0.9423	0.9508	0.9335	0.8821	0.8984	0.8605	0.9587	0.9652	0.9480
KNN	0.8972	0.9106	0.8864	0.9358	0.9464	0.9232	0.8724	0.8879	0.8503	0.9542	0.9614	0.9426
LDA	0.8681	0.8852	0.8495	0.9173	0.9292	0.9037	0.8517	0.8671	0.8312	0.9405	0.9487	0.9274
GB	0.9237	0.9376	0.9098	0.9584	0.9665	0.9481	0.9072	0.9218	0.8892	0.9726	0.9784	0.9635
HGB	0.9345	0.9486	0.9162	0.9681	0.9756	0.9594	0.9165	0.9309	0.8967	**0.9813**	**0.9875**	**0.9722**
ET	0.9283	0.9427	0.9118	0.9635	0.9708	0.9535	0.9121	0.9268	0.8923	0.9781	0.9843	0.9690
MLP	0.9198	0.9336	0.9054	0.9561	0.9634	0.9460	0.9038	0.9182	0.8797	0.9709	0.9772	0.9610

Continuation of Table 4.

	Lasso			Tree-based			RFECV			SHAP		
	Accuracy	*Precision*	*Recall*	*Accuracy*	*Precision*	*Recall*	*Accuracy*	*Precision*	*Recall*	*Accuracy*	*Precision*	*Recall*
LBP and CNN features (fine-tuned)												
SVM	0.9223	0.9336	0.9177	0.9572	0.9626	0.9483	0.8965	0.9092	0.8725	0.9750	0.9700	0.9600
KNN	0.9090	0.9209	0.8982	0.9492	0.9597	0.9354	0.8846	0.8993	0.8628	0.9748	0.9700	0.9600
LDA	0.8783	0.8982	0.8630	0.9308	0.9398	0.9154	0.8651	0.8775	0.8444	0.9598	0.9694	0.9474
GB	0.9363	0.9480	0.9231	0.9697	0.9805	0.9587	0.9192	0.9353	0.9019	0.9781	0.9699	0.9654
HGB	0.9461	0.9631	0.9289	0.9796	0.9862	0.9730	0.9286	0.9448	0.9086	**0.9835**	**0.9772**	**0.9738**
ET	0.9432	0.9565	0.9259	0.9773	0.9858	0.9681	0.9264	0.9370	0.9063	0.9781	0.9699	0.9654
MLP	0.9325	0.9485	0.9164	0.9679	0.9749	0.9594	0.9185	0.9290	0.8933	0.9750	0.9700	0.9600
Handcrafted, LBP and CNN												
SVM	0.9423	0.9554	0.9331	0.9678	0.9751	0.9568	0.9312	0.9458	0.9126	0.9792	0.9850	0.9683
KNN	0.9285	0.9417	0.9173	0.9604	0.9692	0.9497	0.9183	0.9321	0.8987	0.9765	0.9827	0.9648
LDA	0.8997	0.9158	0.8732	0.9452	0.9561	0.9294	0.8851	0.9014	0.8603	0.9682	0.9756	0.9553
GB	0.9483	0.9614	0.9298	0.9726	0.9794	0.9617	0.9367	0.9505	0.9182	0.9847	0.9903	0.9725
HGB	0.9591	0.9728	0.9365	0.9793	0.9859	0.9704	0.9456	0.9592	0.9268	**0.9901**	**0.9954**	**0.9786**
ET	0.9527	0.9653	0.9317	0.9754	0.9828	0.9662	0.9409	0.9547	0.9225	0.9876	0.9932	0.9758
MLP	0.9442	0.9582	0.9243	0.9698	0.9773	0.9585	0.9284	0.9432	0.9105	0.9819	0.9878	0.9694
Handcrafted, LBP and CNN (fine-tuned)												
SVM	0.9812	0.9863	0.9712	0.9822	0.9873	0.9722	0.9817	0.9868	0.9717	0.9828	0.9883	0.9719
KNN	0.9805	0.9846	0.9664	0.9815	0.9856	0.9674	0.9810	0.9851	0.9669	0.9814	0.9865	0.9690
LDA	0.9696	0.9781	0.9584	0.9706	0.9791	0.9594	0.9701	0.9786	0.9589	0.9705	0.9795	0.9594
GB	0.9872	0.9929	0.9745	0.9882	0.9939	0.9755	0.9877	0.9934	0.9750	0.9901	0.9938	0.9769
HGB	0.9899	0.9989	0.9824	0.9909	0.9999	0.9834	0.9904	0.9994	0.9829	**0.9910**	**0.9975**	**0.9817**
ET	0.9887	0.9944	0.9782	0.9897	0.9954	0.9792	0.9892	0.9949	0.9787	0.9903	0.9964	0.9799
MLP	0.9835	0.9899	0.9706	0.9845	0.9909	0.9716	0.9840	0.9904	0.9711	0.9845	0.9908	0.9729

Note: Standard deviations (σ) for accuracy, precision, and recall across five-fold cross-validation are low for all classifiers. For example, for the best-performing configuration, accuracy is 99.01% ± 0.56%, precision is 99.54% ± 0.43%, and recall is 97.86% ± 0.61%. For other classifiers, σ values are ranging between 0.4% and 1.2%.

cell classification, *Computers in Biology and Medicine* **171**(C): 108153.

Freytes, C.Y., Perry Mayrand, R., Sawada, L.O., Yan Liang, T., Curiel Cid, R.E., Burke, S., Loewenstein, D., Duara, R. and Adjouadi, M. (2023). Recursive feature elimination with cross validation for Alzheimer's disease classification using cognitive exam scores, *2023 Intelligent Methods, Systems, and Applications (IMSA) Conference, Giza, Egypt*, pp. 327–332.

Guyon, I. and Elisseeff, A. (2003). An introduction to variable and feature selection, *Journal of Machine Learning Research* **3**(3): 1157–1182.

Hut, I., Jeftic, B., Dragicevic, A., Matija, L. and Koruga, D. (2022). Computer-aided diagnostic system for whole slide imaging of liquid-based cervical cytology sample classification using convolutional neural networks, *Contemporary Materials* **13**(2): 169–177.

Jeleń, Ł., Stankiewicz-Antosz, I., Chosia, M. and Jeleń, M. (2025). Optimizing cervical cancer diagnosis with feature selection and deep learning, *Applied Sciences* **15**(3): 1458.

Kanavati, F., Hirose, N., Ishii, T., Fukuda, A., Ichihara, S. and Tsuneki, M. (2022). A deep learning model for cervical cancer screening on liquid-based cytology specimens in whole slide images, *Cancers* **14**(5): 1159.

Khoulqi, I. and Idrissi, N. (2022). Deep learning-based cervical cancer classification, *2022 International Conference on Technology Innovations for Healthcare, Magdeburg, Germany*, pp. 30–33.

Kitchener, H., Almonte, M., Thomson, C., Wheeler, P., Sargent, A., Stoykova, B., Gilham, B., Baysson, H., Roberts, C., Dowie, R., Desai, M., Mather, J., Bailey, A., Turner, A., Moss, S. and Peto, J. (2009). HPV testing in combination with liquid-based cytology in primary cervical screening (artistic): A randomised controlled trial, *The Lancet Oncology* **10**(7): 672–682.

Klug, S., Neis, K., Harlfinger, W., Malter, A., König, J., Spieth, S., Brinkmann-Smetanay, F., Kommoss, F., Weyer, V. and Ikenberg, H. (2013). A randomized trial comparing conventional cytology to liquid-based cytology and computer assistance, *International Journal of Cancer* **132**(12): 2849–2857.

Kowal, M., Filipczuk, P., Obuchowicz, A., Korbicz, J. and Monczak, R. (2013). Computer-aided diagnosis of breast cancer based on fine needle biopsy microscopic images, *Computers in Biology and Medicine* **43**(10): 1563–1572.

Liew, X.Y., Hameed, N. and Clos, J. (2021). An investigation of XGBOOST-based algorithm for breast cancer classification, *Machine Learning with Applications* **6**: 100154.

Mariarputham, E.J. and Stephen, A. (2015). Nominated texture based cervical cancer classification, *Computational and Mathematical Methods in Medicine* **2015**(1): 586928, DOI: 10.1155/2015/586928.

Mathivanan, S.K., Francis, D., Srinivasan, S., Khatavkar, V.P.K. and Shah, M.A. (2024). Enhancing cervical cancer detection and robust classification through a fusion of deep learning models, *Scientific Reports* **14**(1): 10812.

Mazurek, P. and Oszutowska-Mazurek, D. (2014). From the slit-island method to the Ising model: Analysis of irregular grayscale objects, *International Journal of Applied Mathematics and Computer Science* **24**(1): 49–63, DOI: 10.2478/amcs-2014-0004.

Molnar, C., Casalicchio, G. and Bischl, B. (2020). Interpretable machine learning—A brief history, state-of-the-art and challenges, *in* I. Koprinska *et al.* (Eds) *ECML PKDD 2020 Workshops, Communications in Computer and Information Science*, Vol. 1323, Springer, Cham, pp. 417–431.

Polish National Cancer Registry (2021). Incidence statistics for 2021, `http://onkologia.org.pl/nowotwory-szyjki-macicy/`.

Nithya, B. and Ilango, V. (2019). Evaluation of machine learning based optimized feature selection approaches and classification methods for cervical cancer prediction, *SN Applied Sciences* **1**(6): 641.

Ojala, T., Pietikäinen, M. and Harwood, D. (1996). A comparative study of texture measures with classification based on featured distributions, *Pattern Recognition* **29**(1): 51–59.

Qu, H., Yan, Z., Wu, W., Chen, F., Ma, C., Chen, Y., Wang, J. and Lu, X. (2021). Rapid diagnosis and classification of cervical lesions by serum infrared spectroscopy combined with machine learning, *in* X. Wei and L. Liu (Eds), *AOPC 2021: Biomedical Optics*, International Society for Optics and Photonics, Bellingham, p. 120670A.

Rahmadwati, Naghdy, G., Ros, M., Todd, C. and Norahmawati, E. (2011). Cervical cancer classification using Gabor filters, *2011 1st IEEE International Conference on Healthcare Informatics, Imaging and Systems Biology (HISB 2011), San Jose, USA*, pp. 48–52.

Rajeev, M.A. (2021). A framework for detecting cervical cancer based on UD-MHDC segmentation and MBD-RCNN classification techniques, *2021 2nd Global Conference for Advancement in Technology, Bangalore, India*, pp. 1–9.

Ramkumar, G. and Sajiv, G. (2023). Histopathological image analysis for breast cancer identification using gradient boosting classifier, *2023 International Conference on Self Sustainable Artificial Intelligence Systems, Erode, India*, pp. 789–796.

Sabeena, K. and Gopakumar, C. (2022). A hybrid model for efficient cervical cell classification, *Biomedical Signal Processing and Control* **72**(Part A): 103288, DOI: 10.1016/j.bspc.2021.103288.

Sajiv, G. and Ramkumar, G. (2023). A robust breast cancer classification model using extra-trees classifier for histopathological image, *2023 International Conference on Advances in Computing, Communication and Applied Informatics, Chennai, India*, pp. 1–7.

Sharma, P., Verma, G. and Sinha, R. (2024). Deep ensemble attention framework for cervical cytology image classification, *SN Computer Science* **5**(3): 185.

Sornapudi, S., Brown, G., Xue, Z., Long, R., Allen, L. and Antani, S. (2020). Comparing deep learning models for multi-cell classification in liquid-based cervical

cytology images, *AMIA Annual Symposium Proceedings* **2019**: 820–827.

Stankiewicz, I. (2018). *Using a Computer Program to Evaluate Cytological Smears Received from the Vaginal Part of the Cervix Using the LBC Method*, Master's thesis, Wroclaw Medical University, Wroclaw.

Su, Z., Pietikäinen, M. and Liu, L. (2023). From local binary patterns to pixel difference networks for efficient visual representation learning, *in* R. Gade and M. Felsberg (Eds), *Image Analysis*, Springer Nature Switzerland, Cham, pp. 138–155.

Subarna, T. and Sukumar, P. (2022). Detection and classification of cervical cancer images using CEENET deep learning approach, *Journal of Intelligent & Fuzzy Systems* **43**(3): 3695–3707.

Tarawneh, A.S., Celik, C., Hassanat, A.B. and Chetverikov, D. (2020). Detailed investigation of deep features with sparse representation and dimensionality reduction in CBIR: A comparative study, *Intelligent Data Analysis: An International Journal* **24**(1): 47–68, DOI: 10.3233/IDA-184411.

Uthayakumar, J., Easwaramoorthy, V. and Kannan, R. (2018). Pap smear image classification using local binary patterns and support vector machine, *Journal of Medical Systems* **42**(9): 163.

Wild, C.P., Weiderpass, E. and Stewart, B.W. (Eds) (2020). *World Cancer Report: Cancer Research for Cancer Prevention*, IARC, Lyon.

Wong, L., Ccopa, A., Diaz, E., Valcarcel, S., Mauricio, D. and Villoslada, V. (2023). Deep learning and transfer learning methods to effectively diagnose cervical cancer from liquid-based cytology Pap smear images, *International Journal of Online and Biomedical Engineering* **19**(04): 77–93.

Łukasz Jeleń received his MS in biomedical engineering from the Wroclaw University of Science and Technology in 2001 and his PhD in computer science from Concordia University, Montreal, in 2009. His research focuses on AI applications in cancer classification and medical image analysis. He has authored numerous papers and contributed to commercial projects involving artificial intelligence in healthcare.

Izabela Stankiewicz-Antosz received her MS in laboratory diagnostics from the Wroclaw Medical University in 2018. She has worked as a laboratory diagnostician and currently serves as an assistant at the Military University of Land Forces in Wrocław. Her research interests include CBRN threats, emerging technologies, as well as their medical and defense applications.

Maria Chosia received her PhD in medical sciences in 1981 and her DSc (habilitation) in 1991 from Pomeranian Medical University in Szczecin. In 2003, she was awarded the title of a professor of medical sciences. Her research is focused on oncology and immunocytochemical diagnostics. Professor Chosia has authored over 300 scientific publications, including books and abstracts.

Michał Jeleń received his PhD in medical sciences in 1981 and his DSc (habilitation) in 1999 from the Medical University of Wroclaw. In 2004, he was awarded the title of a professor of medical sciences. He had founded and then led the Department of Pathomorphology and Oncological Cytology at the same university. His research focuses on breast, gastrointestinal, and hematopoietic cancers. He has pioneered the use of neural networks in breast cancer assessment. Professor Jeleń is a member of several scientific societies, including the European and Polish Societies of Pathology.

Received: 1 March 2025
Revised: 6 June 2025
Re-revised: 29 & 30 July 2025
Accepted: 1 August 2025