

Research Article

CerviDBF-Net: Leveraging Spatial and Temporal Features for Accurate Cervical Cancer Diagnosis

Sathupati Lydia* , Nagadasari Merrin Prasanna 

Department of Electronics and Communication Engineering, Annamacharya University, Rajampet, Andhra Pradesh, India

*lydia.ece@gmail.com

Abstract

Cervical cancer screening increasingly relies on automated image analysis to support colposcopic assessment and improve the consistency of clinical decision-making. This study proposes CerviDBF-Net, a novel deep learning framework for automated cervix-type classification from colposcopic images. The proposed architecture combines DenseNet121 for spatial feature extraction, a brain-inspired Liquid Neural Network (LNN) for contextual and temporal feature modelling, and a fully connected Softmax layer for final classification. To improve image quality and feature representation, the preprocessing pipeline incorporates bicubic interpolation, Bi-Histogram Equalization with Plateau Limit (BHEPL) for contrast enhancement, Kapur entropy-based thresholding, and the Roberts Cross operator for edge detection. The framework is evaluated on the Intel & MobileODT Cervical Cancer Screening dataset, which contains three cervix types (Type 1, Type 2, and Type 3) defined by the location and visibility of the transformation zone. CerviDBF-Net achieves an accuracy of 98.92%, sensitivity of 99.13%, specificity of 98.38%, precision of 98.69%, and an F1-score of 98.27%, demonstrating strong classification performance compared with the evaluated existing models. The proposed approach provides an automated baseline tool for cervix-type classification that may support triage in colposcopy-assisted screening by helping clinicians assess whether adequate visualization of the squamocolumnar junction has been achieved.

Keywords: Cervical Cancer Detection; Colposcopic Images; Deep Learning; DenseNet121; Liquid Neural Network

INTRODUCTION

Cervical cancer remains a global public health problem, particularly in low- and middle-income countries (LMICs), where routine screening and early intervention are often limited [1, 2]. Early detection is associated with a more favourable prognosis and better survival, and depends on accurate and timely diagnosis [3, 4]. Colposcopy is a routine clinical procedure that allows visual detection of potentially abnormal changes in

cervical tissue [5–7]. However, the procedure is highly dependent on clinician expertise [8], and its visual interpretation is subject to inter-observer variability [9].

Convolutional models have been applied to cervical-image analysis in several settings. Zhang et al. used a pretrained densely connected network for colposcopy classification [10], Sato et al. evaluated deep classification of colposcopy images [11], and Lee and Choi compared artificial-intelligence methods for digital colposcopy [12]. Recent work has also examined transformer-based cervix-type classification [13, 14] and explainable attention for colposcopic risk assessment [15]. These studies establish the value of learned spatial representations, but they do not evaluate the DenseNet121-LNN feature-evolution mechanism examined in this study.

DenseNet121 was selected because its dense connections reuse fine-scale features and support gradient flow in deep image models [10]. The LNN component follows the continuous-time recurrent formulation introduced by Hasani et al. [16], in which hidden states evolve through differential-equation updates. CerviDBF-Net applies this mechanism to a static 1024-dimensional DenseNet feature vector by iterating an internal state over 15 pseudo-time steps. The term pseudo-time describes computation within the network and does not represent longitudinal lesion development. Based on these observations, the major contributions of this research are summarized as follows.

This work proposes a novel hybrid deep learning model, namely CerviDBF-Net, which combines the DenseNet121 network for spatial feature representation and a brain-inspired Liquid Neural Network (LNN) for contextual and temporal modelling.

A detailed preprocessing pipeline comprising bicubic interpolation, BHEPL contrast enhancement, Kapur entropy thresholding, and Roberts edge detection is developed to enhance the visibility and quality of lesion features. A feature extraction approach that integrates handcrafted LBP texture features with a deep spatial–temporal representation is proposed to improve discriminative power for cervix-type classification, motivated by prior evidence that combining handcrafted and deep features benefits cervical image analysis [17]. Extensive experiments on the Intel & MobileODT Cervical Cancer Screening dataset show that CerviDBF-Net outperforms the internally evaluated baseline variants in terms of accuracy, sensitivity, specificity, precision, and F1-score.

The rationale for applying Liquid Neural Network dynamics to static colposcopic features is stated explicitly, and ablation experiments show that processing the DenseNet121 feature vector through an ODE-driven reservoir over $T = 15$ pseudo-time steps yields a 4.71 percentage-point accuracy gain over the DenseNet121-only variant, consistent with the modelling of non-linear feature interactions that a feed-forward classification head does not capture. The remainder of this paper is organized as follows. The Related Works section reviews recent deep learning studies on cervical cancer diagnosis and summarizes their advantages and limitations. The Proposed System section describes the CerviDBF-Net architecture, including the preprocessing strategy, feature extraction pipeline, and hybrid spatial–temporal modelling mechanism. The Results and Analysis section presents the dataset, experimental setup, evaluation criteria, and results,

which are then examined in the Discussion. The final section summarizes the main findings and outlines directions for future research.

RELATED WORKS

In recent years, several studies have sought to improve the diagnosis of cervical cancer with deep learning, but important limitations remain. The study in [18] explored multimodal baseline deep learning models for cervical cancer detection using colposcopic images. The authors of [19] used a VGG19 model and achieved 92% accuracy for the classification of cervical conditions. The authors of [20] used ResNet50 for cervical cancer detection and achieved an accuracy of 91%. The study in [21] introduced a Dense-U-Net hybrid that merges DenseNet and U-Net. The authors of [22] studied cervical image preprocessing and segmentation with a U-Net-based model. The encoder-decoder structure of U-Net enables efficient lesion localization, but its lower discriminative power limits performance on complex classification tasks, particularly when healthy and diseased regions share similar visual features. The study in [23] used ResNet152 to classify cervical cancer with a high accuracy of 98.08%. The deeper ResNet152 architecture learned complex spatial patterns successfully, but at a high computational cost and with possible inference delay.

The authors of [24] presented the ResNet101-based DepDec model; however, it relies on static feed-forward feature mining and does not integrate contextual feature modelling. The study in [25] reported a sensitivity of 63% and a specificity of 74%. Authors in [13] combined a Vision Transformer with shifted patch tokenization and locality self-attention for three-class cervix-type classification on Intel & MobileODT colposcopy images.

The authors of [26] proposed a model that combines deep features with a Random Forest classifier (DeepColpo + RF). Although the model showed good results, it relied on manual feature fusion and conventional classification. Hong et al. [14] applied low-rank adaptation to a Vision Transformer for the same three-class Intel & MobileODT task, reducing the number of trainable parameters while preserving the pretrained model.

Table 1 distinguishes studies with comparable colposcopy inputs from studies that use different labels, cohorts, and evaluation designs. The table therefore compares methodological scope rather than treating values obtained under different protocols as a controlled ranking.

The comparison identifies three questions that remain relevant to the present experimental design. Table 2 links each gap to the studies that define its scope and to the corresponding evidence examined in CerviDBF-Net.

- RQ1: Does incorporating an LNN with DenseNet121 significantly improve the sensitivity of cervix-type classification on the Intel & MobileODT dataset compared with spatial-only CNN baselines?

- RQ2: Does the complete preprocessing pipeline (BHEPL, Kapur thresholding, and Roberts edge detection) provide a measurable accuracy improvement over the classification of unprocessed images?
- RQ3: Does CerviDBF-Net perform stably across the cross-validation folds and across all three cervix types?

Table 1. Structured Comparison with Closely Related Colposcopy Studies

Study	Data and task	Core method	Evidence relevant to this study	Difference from CerviDBF-Net
[21]	Institutional colposcopy cohort; cervical-lesion classification and localization	Dense-U-Net (DenseNet-U-Net hybrid)	Joint spatial classification and lesion-region analysis	Different clinical endpoint and cohort; no LNN feature evolution
[13]	Intel & MobileODT; three cervix types	Vision Transformer with shifted patch tokenization and locality self-attention	Same public benchmark and cervix-type task	Transformer-based spatial modelling; no continuous-time reservoir
[14]	Intel & MobileODT; cervix-type classification	Low-rank adaptation of a Vision Transformer	Same benchmark with parameter-efficient adaptation	Adapter-based transformer tuning; no DenseNet121-LNN coupling
[15]	DYSIS colposcopy images; high- and low-risk classification	Attention-based CNN with Grad-CAM and complementary explanation methods	Explicit model-attribution analysis for colposcopy	Different cohort and labels; no pseudo-time feature transformation
CerviDBF-Net	Intel & MobileODT; three cervix types	DenseNet121 with a trainable LNN reservoir	Preprocessing ablation, five-fold evaluation, calibration, and controlled paired comparison	Present study

G1 concerns contextual feature modelling. The closest Dense-U-Net, deep semantic, ViT, LoRA-ViT, and attention-CNN studies [13–15, 21, 24] learn spatial or attention-based representations, whereas the present study examines whether repeated continuous-time state updates refine a static DenseNet121 embedding.

G2 concerns attribution of preprocessing gains. Studies using segmentation or image enhancement [21, 22] and the closely related cervix-type classifiers [13, 14] do not provide

the same component-controlled comparison of the complete preprocessing sequence used here.

G3 concerns same-sample statistical comparison. Results from [13, 14, 21, 24] arise from different cohorts, partitions, or model protocols and therefore cannot provide a paired test for CerviDBF-Net. The paired analysis in Table 16 is restricted to internally evaluated variants that share the same test samples.

Table 2. Literature Gaps and Their Experimental Linkage

Gap	Studies defining the gap	Evidence boundary	Evidence examined in this study
G1 Contextual feature modelling	[13–15, 21, 24]	Spatial, transformer, or attention representations without a continuous-time reservoir applied to DenseNet features	DenseNet121-only, DenseNet121-LSTM, and DenseNet121-LNN variants in Table 13
G2 Preprocessing attribution	[13, 14, 21, 22]	No directly comparable component-controlled test of the complete BHEPL-Kapur-Roberts sequence	No-preprocessing and full-pipeline variants in Table 13
G3 Paired statistical evidence	[13, 14, 21, 24]	Different cohorts or partitions prevent a same-sample significance test against CerviDBF-Net	McNemar comparisons for internally evaluated variants on the shared test set in Table 16

To address these gaps, the following hypotheses are tested. H1: An LNN trained on top of DenseNet121 will improve classification sensitivity over the DenseNet121 baseline in a statistically significant manner by capturing internal pseudo-temporal feature dynamics. H2: CerviDBF-Net will achieve similarly high accuracy in all five cross-validation folds (standard deviation of accuracy below 1 percentage point), providing evidence that the hybrid design is stable and reproducible. H3: The preprocessing pipeline of BHEPL contrast enhancement, Kapur entropy thresholding, and Roberts edge detection will measurably improve accuracy compared with unprocessed input, as assessed by the component-wise ablation study.

The validation results support all three hypotheses. H1 is supported: relative to the DenseNet121-only model, the LNN improves sensitivity by 4.55 percentage points (94.58% to 99.13%) and accuracy by 4.71 percentage points, and the paired accuracy difference is statistically significant (McNemar’s test, $p < 0.001$; Table 16). H2 is supported, as the standard deviation of accuracy across the five cross-validation folds is 0.23 percentage points (Table 11). H3 is supported, as the preprocessing ablation shows a 5.45 percentage-point accuracy gain over unprocessed input (Table 13).

PROPOSED SYSTEM

The proposed system, depicted in Figure 1, classifies colposcopy images through a systematic pipeline comprising dataset acquisition, image preprocessing, segmentation,

feature extraction, and classification with the proposed hybrid deep learning architecture, which combines DenseNet121, a brain-inspired Liquid Neural Network (LNN), and a fully connected classification layer.

The colposcopic images are taken from the publicly available Intel & MobileODT Cervical Cancer Screening dataset hosted on Kaggle [27], which contains three labelled classes (Type 1, Type 2, and Type 3); these labels enable supervised learning and performance evaluation.

Images are initially resized to a fixed dimension using bicubic interpolation, which utilizes the 16 nearest neighbouring pixels to estimate the new pixel intensity:

$$g(x', y') = \sum_{i=-1}^2 \sum_{j=-1}^2 a_{ij} \cdot f(x + i, y + j) \quad (1)$$

where $g(x', y')$ represents the interpolated pixel intensity at the resized location (x', y') , $f(x + i, y + j)$ denotes the intensity of the neighboring pixel at offset (i, j) from the original location (x, y) , and a_{ij} are the bicubic weighting coefficients determined by the distance between original and target pixel positions, where i and j range from -1 to 2 , covering the 16 nearest neighbouring pixels.

Colour images are transformed into grayscale using a weighted sum of the RGB components:

$$f_{gray}(x, y) = 0.2989f_R(x, y) + 0.5870f_G(x, y) + 0.1140f_B(x, y) \quad (2)$$

where $f_{gray}(x, y)$ denotes the grayscale intensity at pixel location (x, y) , $f_R(x, y)$, $f_G(x, y)$, and $f_B(x, y)$ represent the red, green, and blue channel intensities, respectively, and the coefficients 0.2989, 0.5870, and 0.1140 correspond to standard luminance weights based on human visual sensitivity to each colour channel.

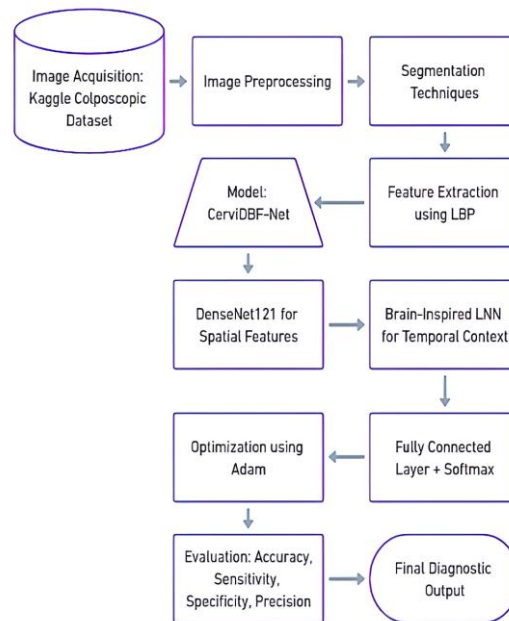


Figure 1. Proposed System Block Diagram

Contrast enhancement and segmentation are then performed to improve visibility and localize candidate lesion regions in the grayscale images. In BHEPL, the histogram of the grayscale image is divided into two sub-histograms based on the mean intensity m :

$$H_L = \{f_{gray}(x, y) \leq m\} \quad (3)$$

$$H_U = \{f_{gray}(x, y) > m\} \quad (4)$$

where H_L represents the lower sub-histogram containing pixels with intensity values less than or equal to the mean value m , H_U denotes the upper sub-histogram containing pixels with intensity values greater than m , $f_{gray}(x, y)$ represents grayscale intensity at pixel location (x, y) , and m is the mean intensity of the entire grayscale image used as the splitting criterion.

Each sub-histogram is independently equalized to preserve critical visual information while improving lesion visibility. Kapur's entropy-based thresholding determines the optimal threshold for lesion segmentation by maximizing foreground and background entropy:

$$T^* = \arg \max [H_1(T) + H_2(T)] \quad (5)$$

where T^* is the optimal global threshold that separates lesion regions from the background, $H_1(T)$ and $H_2(T)$ denote the Shannon entropy values of foreground and background regions at threshold T , respectively, and $\arg \max$ selects the threshold value that maximizes the combined entropy.

Edge detection is further performed using the Roberts cross operator to identify lesion boundaries:

$$G_x = f(x, y) - f(x + 1, y + 1) \quad (6)$$

$$G_y = f(x + 1, y) - f(x, y + 1) \quad (7)$$

$$E(x, y) = \sqrt{G_x^2 + G_y^2} \quad (8)$$

where G_x and G_y represent gradient magnitudes in diagonal and anti-diagonal directions, respectively, $f(x, y)$ denotes pixel intensity at location (x, y) , and $E(x, y)$ represents the final edge magnitude used to highlight lesion boundaries.

Texture information is extracted from segmented images using the Local Binary Pattern (LBP) method, which captures local spatial texture structures:

$$LBP(x, y) = \sum_{n=0}^{P-1} s(i_n - i_c) \cdot 2^n \quad (9)$$

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (10)$$

where $LBP(x, y)$ denotes the Local Binary Pattern code at pixel (x, y) , P represents the total number of neighboring pixels sampled on radius R , i_n denotes the intensity of the n^{th} neighboring pixel, i_c represents the center pixel intensity, $s(\cdot)$ is the thresholding function, and 2^n represents the binary weighting factor used to encode local texture

patterns. These binary patterns are converted into histograms to represent texture descriptors of cervical images.

The extracted LBP feature vectors are then passed to the proposed CerviDBF-Net model, which consists of DenseNet121, Liquid Neural Network, and Softmax classification layers. DenseNet121 extracts deep spatial features through dense connectivity using convolution operations defined as:

$$(f * g)(i, j) = \sum_m \sum_n f(m, n) \cdot g(i - m, j - n) \quad (11)$$

where $(f * g)(i, j)$ represents the output feature map at location (i, j) , $f(m, n)$ denotes the input feature value, and $g(i - m, j - n)$ represents convolution kernel weights.

The activation function is expressed as:

$$f(x) = \max(0, x) \quad (12)$$

where $f(x)$ denotes the activated output and x represents the input value. ReLU introduces non-linearity by preserving positive values while suppressing negative values.

Spatial dimensionality reduction is performed through max pooling:

$$y(i, j) = \max_{0 \leq m < k, 0 \leq n < k} x(i + m, j + n) \quad (13)$$

where $y(i, j)$ denotes the pooled output at location (i, j) , $x(i + m, j + n)$ represents values within the pooling window, and k denotes the pooling kernel size.

The extracted DenseNet121 feature vector is forwarded to the Liquid Neural Network module, which captures dynamic contextual dependencies through adaptive neuron states:

$$\tau \frac{du(t)}{dt} = -u(t) + W_{in}x(t) + W_{res}r(t) \quad (14)$$

$$r(t) = \sigma(u(t)) \quad (15)$$

where τ represents the time constant governing neuron state decay, $u(t)$ denotes the internal neuron state at pseudo-time step t , W_{in} is the input weight matrix, $x(t)$ represents the DenseNet121 feature vector ($f_{dense} \in \mathbb{R}^{1024}$), W_{res} denotes the recurrent reservoir weight matrix with spectral radius $\rho = 0.9$ and 10% sparsity, $r(t)$ represents neuron response output, and σ denotes the leaky ReLU activation function with slope $\alpha = 0.01$.

The geometric interpretation of the LNN is treated as a working hypothesis rather than a convergence theorem. The DenseNet121 vector lies in a high-dimensional feature space, and the recurrent ODE update applies a sequence of non-linear transformations over 15 pseudo-time steps. A spectral radius of 0.90 constrains the recurrent state changes within the selected operating regime [16], but it does not by itself prove contraction, improved separability, or generalization for this dataset. The gain over the DenseNet121-only variant

in Table 13 is consistent with useful feature refinement; it is empirical evidence for the selected configuration rather than a general theoretical result.

The LNN adds $O(T \cdot N^2)$ operations per inference pass for the reservoir update, where $T=15$ and $N=128$, contributing 245,760 multiply-accumulate operations, which is less than 0.008% of the 3.12G FLOPs of the full CerviDBF-Net pipeline (Table 15). End-to-end convergence via Adam follows the standard $O(P/B)$ per-epoch complexity where P is the parameter count (9.14M) and B is the batch size (32); empirically, CerviDBF-Net converges within 30–50 epochs under early stopping with patience=10, consistent with the batch size and learning rate schedule in Table 14.

The Liquid Neural Network treats the 1024-dimensional DenseNet feature vector as the initial state of a dynamic reservoir and iterates it over $T = 15$ pseudo-time steps using Euler integration with $\Delta t = 0.1$. The reservoir consists of 128 neurons, enabling efficient modeling of higher-order contextual dependencies in static medical images without requiring true sequential inputs.

The 1024-dimensional DenseNet121 representation is projected to the 128-neuron reservoir and updated for 15 internal steps. At each step, the input projection preserves information from the spatial descriptor, the recurrent matrix mixes information across reservoir dimensions, and the leaky activation controls how rapidly the hidden state changes. The spectral radius of 0.90 and the integration step of 0.1 limit abrupt state growth. The final reservoir state is therefore a context-refined representation of one static image, not a temporal record of lesion progression. The ablation gain in Table 13 supports the value of this internal feature transformation for the present dataset, although it does not identify a biological time process.

The final fully connected layer maps the learned features into three cervical classes:

$$y = Wx + b \quad (16)$$

$$\text{Softmax}(z_i) = \frac{e^{z_i}}{\sum_j e^{z_j}} \quad (17)$$

where y denotes the output vector of the fully connected layer, W represents the learnable weight matrix of size 3×128 , x denotes the final LNN state vector, b represents the bias vector, and z_i denotes the class logit for class i . The Softmax function converts logits into probability scores corresponding to Type 1, Type 2, and Type 3 cervix classes.

The model is trained using the Adam Optimizer, which adaptively updates network parameters according to:

$$\theta_{t+1} = \theta_t - \frac{\alpha m_t}{\sqrt{v_t} + \epsilon} \quad (18)$$

where θ_t and θ_{t+1} represent model parameters at iterations t and $t + 1$, respectively, α denotes the learning rate (1×10^{-4}), m_t represents the first moment estimate ($\beta_1 = 0.9$), v_t denotes the second moment estimate ($\beta_2 = 0.999$), and $\epsilon = 10^{-8}$ ensures numerical stability during optimization.

Evaluation Metrics

The main evaluation metrics used in our research work are as follows:

- Accuracy

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (19)$$

- Sensitivity (Recall)

$$Sensitivity (Recall) = \frac{TP}{TP + FN} \quad (20)$$

- Specificity

$$Specificity = \frac{TN}{TN + FP} \quad (21)$$

- Precision

$$Precision = \frac{TP}{TP + FP} \quad (22)$$

- F1-score

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (23)$$

where TP, TN, FP, and FN denote the numbers of true positives, true negatives, false positives, and false negatives, respectively.

Proposed Model

The proposed CerviDBF-Net architecture (Figure 2) is designed as a unified framework that integrates spatial, temporal, and classification capabilities for automated cervix type classification in colposcopy-based screening. It uses DenseNet121 as the backbone for spatial feature extraction, which has 121 densely connected layers with four dense blocks consisting of 1×1 and 3×3 convolutions, bottleneck layers, transition layers, batch normalization, ReLU activation and average pooling functions to learn rich and high-resolution features from colposcopic images.

DenseNet121 extracts spatial descriptors from the preprocessed image and compresses them into a 1024-dimensional feature vector. The LNN then performs an internal sequence of state updates that mixes relations among these feature dimensions before classification. This operation can change the relative emphasis of texture, boundary, and contrast evidence when the spatial embedding alone is ambiguous. The final state is passed to the fully connected Softmax layer for Type 1, Type 2, or Type 3 prediction. The pseudo-time trajectory is a computational representation and is not interpreted as lesion evolution in the patient.

CerviDBF-Net differs from the closest colposcopy studies in Table 1 through the trainable coupling of a DenseNet121 spatial encoder and a continuous-time LNN reservoir for three-class cervix-type classification. The contribution lies in this task-specific integration and its controlled evaluation. Table 13 isolates the effect of the LNN from the DenseNet121-only and DenseNet121-LSTM variants, while Table 16 compares internally evaluated variants on the same test samples. The study extends neither DenseNet theory nor liquid time-constant network theory; it evaluates their combined use for contextual refinement of static colposcopy features.

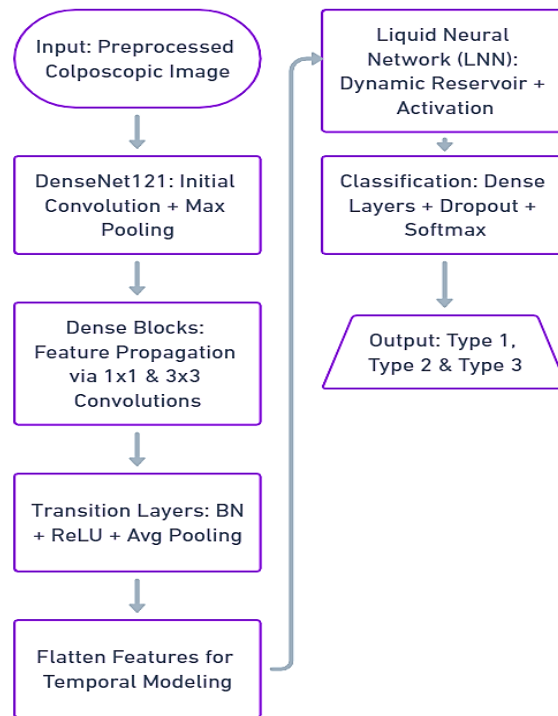


Figure 2. CerviDBF-Net Model Architecture

Algorithm of the Proposed Model

The CerviDBF-Net algorithm takes input features extracted from cervical images. It uses DenseNet121 to learn deep spatial features, the Liquid Neural Network (LNN) to capture pseudo-temporal patterns, and a Softmax layer to classify the image as Type 1, Type 2, or Type 3.

Algorithm: CerviDBF-Net Model

Input: Colposcopic image $I \in \mathbb{R}^{(224 \times 224 \times 3)}$

Output: Class probability vector $p \in \mathbb{R}^3$ (Type 1, Type 2, Type 3)

Step 1 [Preprocessing]: Apply bicubic resizing $\rightarrow$ BHEPL enhancement $\rightarrow$ Kapur thresholding $\rightarrow$ Roberts edge detection to produce enhanced image I_{proc} .

Step 2 [LBP Extraction]: Compute uniform LBP histogram ($P=8, R=1$) from $I_{\text{proc}} \rightarrow$ feature vector $x_{\text{lbp}} \in \mathbb{R}^{256}$.

Step 3 [DenseNet121 Spatial Features]: Pass I_{proc} through DenseNet121 (pretrained, fine-tuned) $\rightarrow$ global average pooling $\rightarrow$ $f_{\text{dense}} \in \mathbb{R}^{1024}$.

Step 4 [LNN Initialization]: Set initial neuron state $u(0) = W_{in} \cdot f_{dense}$, where $W_{in} \in \mathbb{R}^{(128 \times 1024)}$.

Step 5 [LNN Dynamics]: For $t = 0, 1, \dots, 14$ ($T=15$ steps): $u(t+1) = u(t) + \Delta t \times (-u(t)/\tau + W_{in} \cdot f_{dense} + W_{res} \cdot \sigma(u(t)))$, where $\Delta t=0.1$, $\tau=1.0$, $\sigma = \text{leaky ReLU}$ ($\alpha=0.01$).

Step 6 [LNN Output]: Extract terminal state $r = \sigma(u(T)) \in \mathbb{R}^{128}$.

Step 7 [Classification]: $p = \text{Softmax}(W_{fc} \cdot r + b)$, $W_{fc} \in \mathbb{R}^{(3 \times 128)}$.

Step 8 [Prediction]: $\text{class} = \text{argmax}(p)$.

Training Protocol

CerviDBF-Net is implemented in Python 3.9 using the PyTorch 2.0 deep learning framework and trained on an NVIDIA GPU (16 GB VRAM). The DenseNet121 backbone is initialized with ImageNet pretrained weights and fine-tuned end-to-end. The LNN reservoir weights (W_{in} and W_{res}) are initialized using He initialization and trained from scratch via automatic differentiation through the Euler ODE solver.

The Adam optimizer is employed with learning rate $lr = 1 \times 10^{-4}$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 1 \times 10^{-8}$, and weight decay $= 1 \times 10^{-5}$. Training is conducted for a maximum of 100 epochs with early stopping (patience = 10 epochs monitored on validation cross-entropy loss). Batch size is set to 32. Cross-entropy loss is used as the training objective. A ReduceLROnPlateau scheduler reduces the learning rate by a factor of 0.5 when validation loss does not improve for 5 consecutive epochs (minimum $lr = 1 \times 10^{-6}$).

Data augmentation applied during training includes: random horizontal flipping ($p = 0.5$), random rotation ($\pm 15^\circ$), random brightness and contrast jitter (factor range: 0.8–1.2), and random Gaussian noise injection ($\sigma = 0.01$) to simulate illumination variability and imaging artefacts characteristic of real clinical colposcopy. All random operations are seeded with seed = 42. Input images are resized to 224×224 pixels prior to feeding into DenseNet121.

RESULTS AND ANALYSIS

The Intel & MobileODT Cervical Cancer Screening dataset [27], summarized in Table 3, is a valuable benchmark for AI-based cervical cancer screening technologies. It consists of high-resolution images acquired with typical colposcopy equipment under real clinical conditions. The images vary considerably in illumination, zoom, specular reflection, and field of view, which reflects practical screening situations. Each image is assigned to one of three cervix types according to the location and visibility of the transformation zone, which adds clinically relevant complexity. The variety of texture, colour, and boundary characteristics across cervix types makes the dataset a challenging but reliable benchmark for evaluating the robustness and generalization of automated cervical screening models.

Cervix Type 1 is the simplest, with a transformation zone which is completely ectocervical and easily seen. Type 2 involves a transformation zone extending into both the ectocervical and endocervical regions, but it is still completely visible and therefore can be fully evaluated. Type 3 is more problematic as the transformation zone is primarily endocervical and not entirely visible and may require additional diagnostic tests.

In total, there are 7,771 labeled images, taken from the main training set as well as an extra supplementary set. The images are divided into the 3 cervix types as follows: 2,379 are labelled as Type 1, 2,907 are labelled as Type 2, and 2,485 are labelled as Type 3. This diverse and clinically relevant dataset shown in Table 3 provides a solid foundation for training and testing computer vision models.

Table 3. Cervix-Type Distribution in the Intel & MobileODT Dataset [27]

Cervix Type	Training Images	Additional Images	Total Images
Type 1	866	1,513	2,379
Type 2	1,437	1,470	2,907
Type 3	1,129	1,356	2,485
Total	3,432	4,339	7,771

The data were split using an 80–10–10 strategy (Table 4) to support balanced learning and unbiased evaluation: 80% of the images were assigned to training, 10% to validation for hyperparameter tuning and overfitting monitoring, and 10% to final testing for an objective measure of model performance. The split was stratified so that the class distribution was preserved across all three cervix types.

Table 4. Dataset Split

Cervix Type	Train (80%)	Validation (10%)	Test (10%)
Type 1	1,903	238	238
Type 2	2,325	291	291
Type 3	1,988	249	248
Total	6,216	778	777

To avoid data leakage, the 10% test set is withheld from training and used only for the final evaluation, while the 80% training set is used for fold selection, hyperparameter tuning, and model selection. The validation set (10%) is used for early stopping and learning-rate scheduling. Stratification maintains the same class proportions across all partitions. To further validate the results, future work will involve external validation on an independent colposcopy dataset, such as an institutional colposcopy archive collected with IRB approval or another publicly available colposcopy image collection.

All experiments use a fixed random seed of 42 for PyTorch random number generation, NumPy operations, and dataset splitting to ensure reproducibility. The 5-fold stratified cross-validation reported later in this section shows that the reported performance is stable,

with five model evaluations on non-overlapping test partitions yielding a mean accuracy of $98.84\% \pm 0.23\%$.

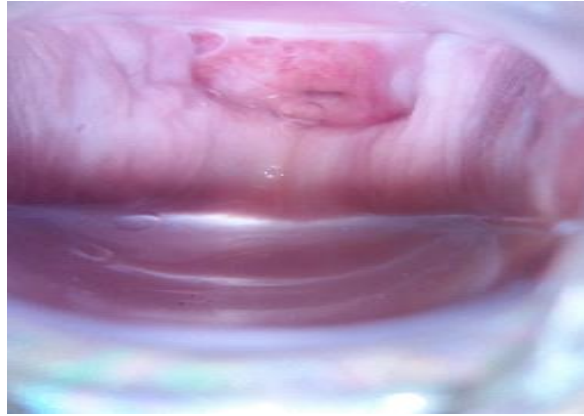


Figure 3. Sample Colposcopy Image from the Intel & MobileODT Dataset [27]

Figure 3 shows an example colposcopy image from the Intel & MobileODT Cervical Cancer Screening dataset [27]. The image is resized to a standard size by bicubic interpolation (Figure 4) and then converted to grayscale (Figure 5).

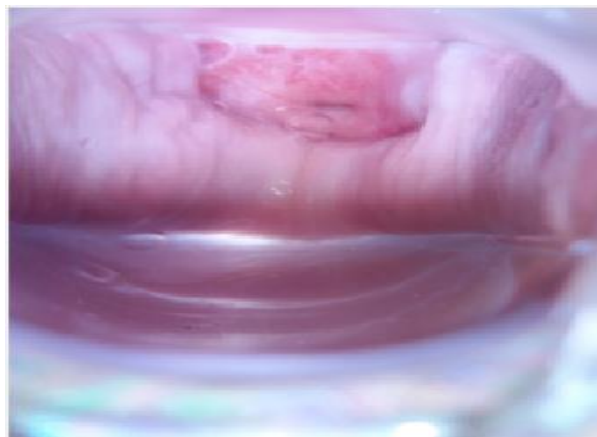


Figure 4. Resized Sample Image



Figure 5. Grayscale Version of Resized Sample

The contrast enhanced version of the grayscale image by applying BHEPL is illustrated in Figure 6. The result of this step is to split the histogram into two sub-ranges, according to the mean intensity, and to use equalization with clipping to avoid over-enhancement, thus enhancing structures and lesions in cervical tissue. The improved image is then thresholded by Kapur's entropy method shown in Figure 7. This is a global thresholding method that maximizes the sum of the entropies of the foreground and background areas of the image, which helps to isolate the lesion areas from the remaining areas of the cervix.



Figure 6. Contrast Enhanced Image

After segmentation, the Roberts cross operator, a gradient-based edge detection method that emphasizes abrupt intensity changes, is applied; its output is shown in Figure 8. This step delineates candidate lesion boundaries and highlights the structures of interest in the segmented region. The edge-detected image is then used as an input to the subsequent feature extraction stage.



Figure 7. Thresholded Image

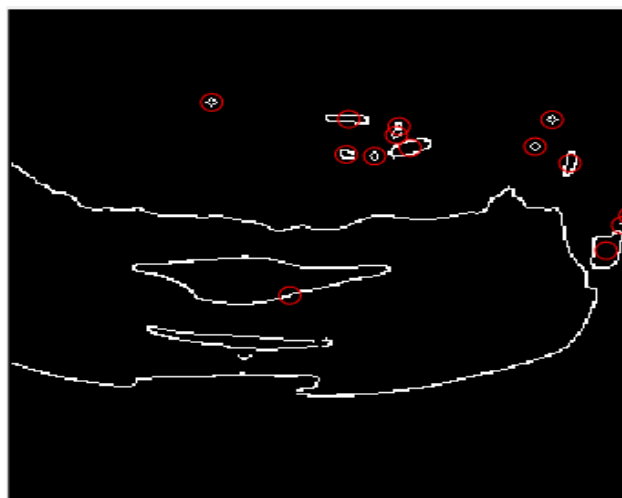


Figure 8. Edge-Detected Image

Figures 7 and 8 delineate candidate regions from image intensity and edge structure. Kapur thresholding separates foreground and background intensity distributions, while the Roberts operator emphasizes abrupt local transitions that may coincide with visible tissue boundaries. These outputs are preprocessing maps rather than clinically annotated lesion masks. The Intel & MobileODT cervix-type labels do not provide pixel-level lesion contours, so Dice, intersection over union, or boundary-distance measures cannot be calculated from this dataset. The maps therefore support feature preparation but do not establish lesion-localization accuracy.

The morphological region properties and colour statistics obtained from the segmented cervical lesion regions detected by the Roberts edge detection stage are presented in Table 5. The primary features used for characterization of lesion boundaries are: centroid coordinates (lesion spatial location), shape descriptors (aspect ratio, solidity, eccentricity), and colour statistics (mean intensity, colour variance per channel).

Simultaneously, a uniform Local Binary Pattern (LBP) Histogram (with 8 sampling points and 1 pixel radius, resulting in a 256-dimensional histogram per image) is independently calculated from the grayscale images obtained from the BHEPL contrast-enhanced images, with the aim of encoding fine-grained texture information of the cervical mucosa. The LBP histogram is utilized as the main deep learning input, and morphological features are used as complementary structural context in both pipelines used in CerviDBF-Net.

Figure 9 shows the output interface of CerviDBF-Net, which displays the predicted cervix type after classification.

Table 6 and Figure 10 compare CerviDBF-Net with two transformer-based studies that also addressed the three-class Intel & MobileODT task. The image counts and data partitions are included so that each accuracy can be read in relation to its own experimental setting.

Table 7 and Figure 11 present the sensitivity comparison. CerviDBF-Net achieves the highest sensitivity (99.13%), compared with 84% for Dense-U-Net [21], 96.38% for DepDec [24], and 98.68% for DeepColpo + RF [26]. Because these values were obtained with different datasets and protocols, they serve as cross-study reference points rather than a controlled ranking. High sensitivity is important in medical diagnostics to ensure that abnormal cases are not missed.

Table 5. Morphological and Colour Features from Segmented Cervical Regions

Feature Type	Feature Index	Value 1	Value 2
Centroid (X, Y)	1	116.1480	149.5334
	2	100.0740	99.2667
	3	113.5096	85.4238
	4	119.9656	83.1012
	5	126.1724	81.8810
	...	...	...
Shape Features	1	0.1675	
	2	0.1609	
	3	0.6513	
	...	...	
Color Features	1	18566	
	2	1	
	3	42	

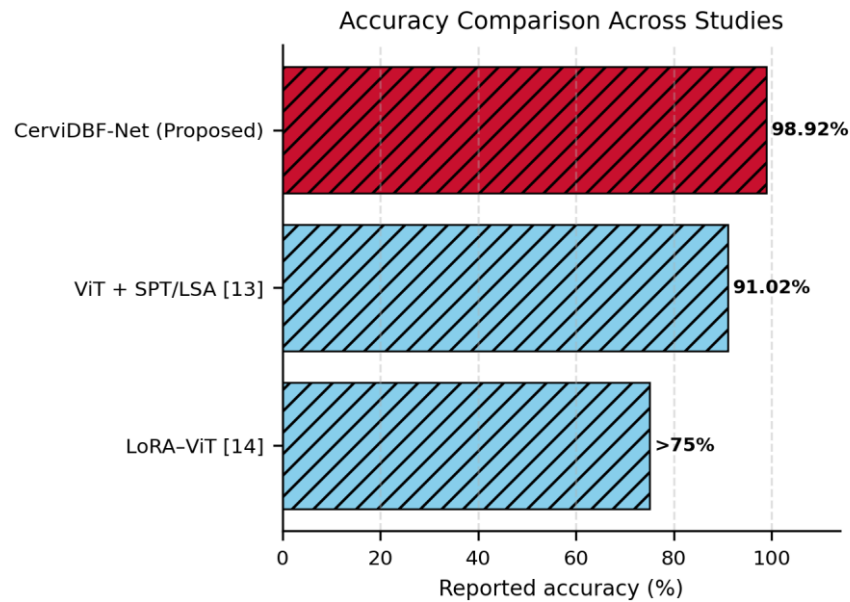
Classification Result

Type 1 Cervix

Figure 9. Output Dialogue Box Displaying the Predicted Cervix Type

Table 6. Accuracy Comparison under the Reported Experimental Settings

Method and experimental setting	Accuracy (%)
ViT with SPT and LSA [13] — Intel & MobileODT, three cervix types; 70:30 split; 5,750 train and 2,464 test images	91.02
LoRA-ViT [14] — Intel & MobileODT, three cervix types; 1,481 train and 512 held-out test images	>75
CerviDBF-Net (Proposed) — Intel & MobileODT, three cervix types; 80:10:10 split; 6,216 train, 778 validation and 777 test images	98.92

**Figure 10.** Accuracy Comparison under the Reported Experimental Settings**Table 7.** Sensitivity Value Comparison

Method/Model Used	Sensitivity (%)
Dense-U-Net [21]	84.0
DepDec (ResNet101) [24]	96.38
Deep learning classifier [25]	63.0
DeepColpo + RF [26]	98.68
CerviDBF-Net (Proposed)	99.13

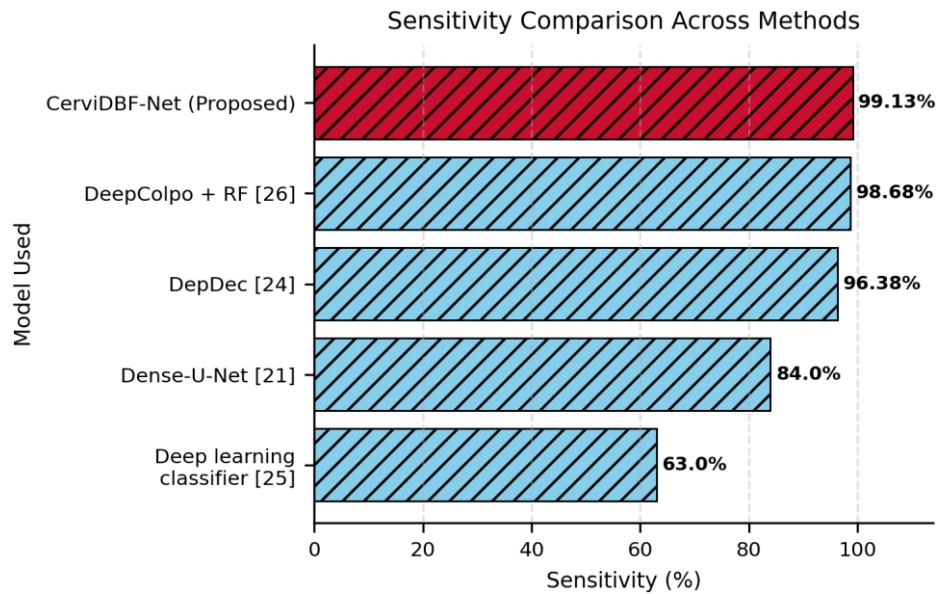


Figure 11. Sensitivity Comparison Plot

For specificity (Table 8 and Figure 12), CerviDBF-Net attains 98.38%, indicating that it correctly identifies negative cases and limits false detections. This value is higher than those reported for DepDec [24] (95.84%) and Dense-U-Net [21] (87%), although these are cross-study comparisons.

Table 8. Specificity Value Comparison

Method/Model Used	Specificity (%)
Dense-U-Net [21]	87.0
DepDec (ResNet101) [24]	95.84
Deep learning classifier [25]	74.0
CerviDBF-Net (Proposed)	98.38

Table 9 and Figure 13 compare precision. CerviDBF-Net achieves a precision of 98.69%, compared with 97.56% for DepDec [24] and 87.50% for DeepColpo + RF [26], indicating that its positive predictions are highly reliable.

The F1-score, which balances precision and recall, is compared in Table 10 and Figure 14. CerviDBF-Net achieves an F1-score of 98.27%, compared with 96.96% for DepDec [24] and 92.19% for DeepColpo + RF [26], showing consistently high performance across the basic diagnostic measures.

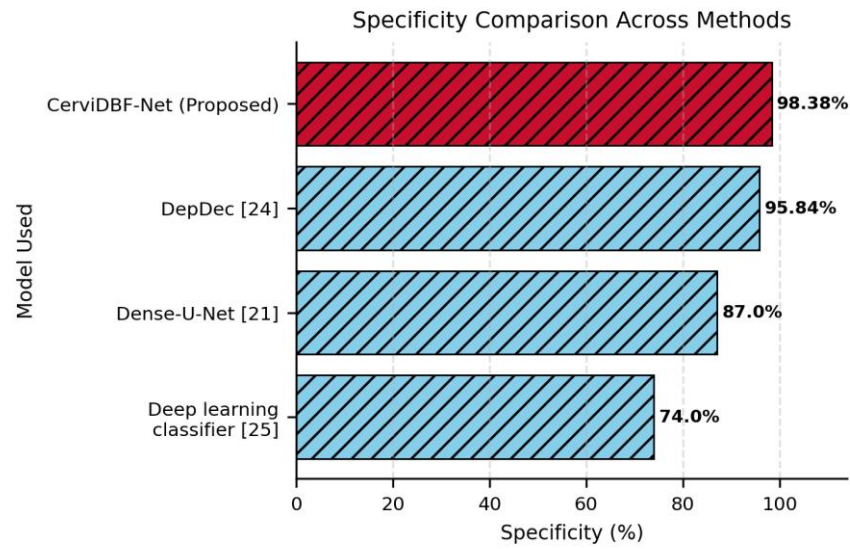


Figure 12. Specificity Comparison Plot

Table 9. Precision Value Comparison

Method/Model Used	Precision (%)
DepDec (ResNet101) [24]	97.56
DeepColpo + RF [26]	87.50
CerviDBF-Net (Proposed)	98.69

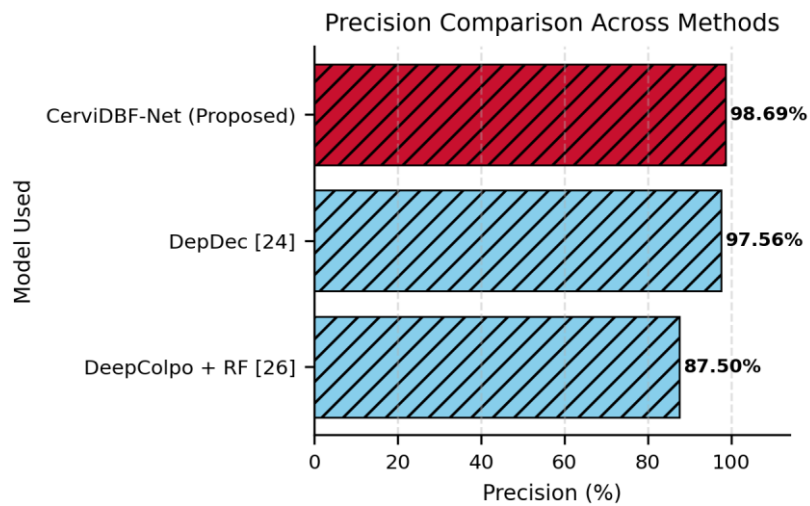
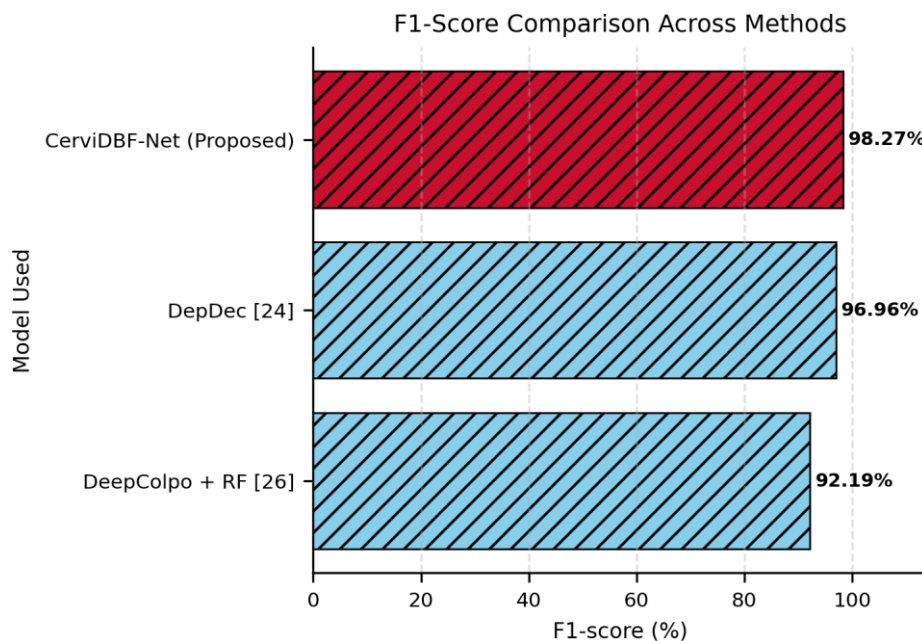


Figure 13. Precision Comparison Plot

Table 10. F1 Score Value Comparison

Method/Model Used	F1 Score (%)
DepDec (ResNet101) [24]	96.96
DeepColpo + RF [26]	92.19
CerviDBF-Net (Proposed)	98.27

**Figure 14.** F1 Score Comparison Plot

Two recent transformer studies provide direct points of reference for this task. Authors in [13] trained a ViT with shifted patch tokenization and locality self-attention on 8,215 Intel & MobileODT images using a 70:30 split. The study in [14] evaluated LoRA-based fine-tuning of a ViT with the 1,481-image training set and a 512-image held-out test set.

Table 6 lists the model, image source, partition, and accuracy reported for each experiment. The three studies share the same three-class cervix-type task and Intel & MobileODT image source, but they use different image pools and data partitions. These differences prevent a strictly controlled comparison.

Intel and MobileODT released the dataset in 2017 as part of a Kaggle challenge on cervical cancer screening [27]. The task was to classify cervical images into one of the three cervix types, and submissions were evaluated using multi-class logarithmic loss on a private test set. The organizers published the final rankings but not detailed performance metrics such as class-wise accuracies.

Authors in [13] reported 91.02% test accuracy for a ViT with shifted patch tokenization and locality self-attention. Their study used an 8,215-image collection and a 70:30 split,

with 5,750 training images and 2,464 test images listed in the data table. Authors in [14] reported correct classification of more than 75% of 512 held-out images after training the LoRA-ViT model on 1,481 images; the reported class accuracies were 54.0% for Type 1, 80.6% for Type 2, and 76.9% for Type 3. CerviDBF-Net used 7,771 images with an 80:10:10 partition of 6,216 training, 778 validation, and 777 test images and achieved 98.92% test accuracy. The three accuracy values arise from these separate experimental settings.

According to publicly available challenge write-ups, top-ranked solutions used deep CNN-based classification with heavy data augmentation and model ensembles to address class imbalance and visual variation, at a considerable computational cost. In contrast, CerviDBF-Net is a single hybrid architecture that combines DenseNet121 for spatial feature extraction with a brain-inspired Liquid Neural Network for contextual and pseudo-temporal modelling. It provides high and repeatable accuracy, sensitivity, specificity, precision, and F1-score within a well-defined train-validation-test protocol without multi-model ensembles, offering an efficient single-model alternative with competitive performance.

An ablation study was performed to evaluate the influence of each core component of the CerviDBF-Net architecture by systematically removing or modifying individual components. The principal variants were DenseNet121 without the LNN, the LNN with a simple convolutional feature extractor, and the full CerviDBF-Net model; the complete set of variants is reported in Table 13. DenseNet121 alone was effective in capturing spatial features, reaching an accuracy of 94.21%, but was less successful at modelling context. The LNN-only configuration reached an accuracy of 88.47%, indicating that the LNN cannot compensate for the absence of rich spatial features. The full CerviDBF-Net achieved the best accuracy of 98.92%, which shows that DenseNet121 and the LNN provide complementary strengths.

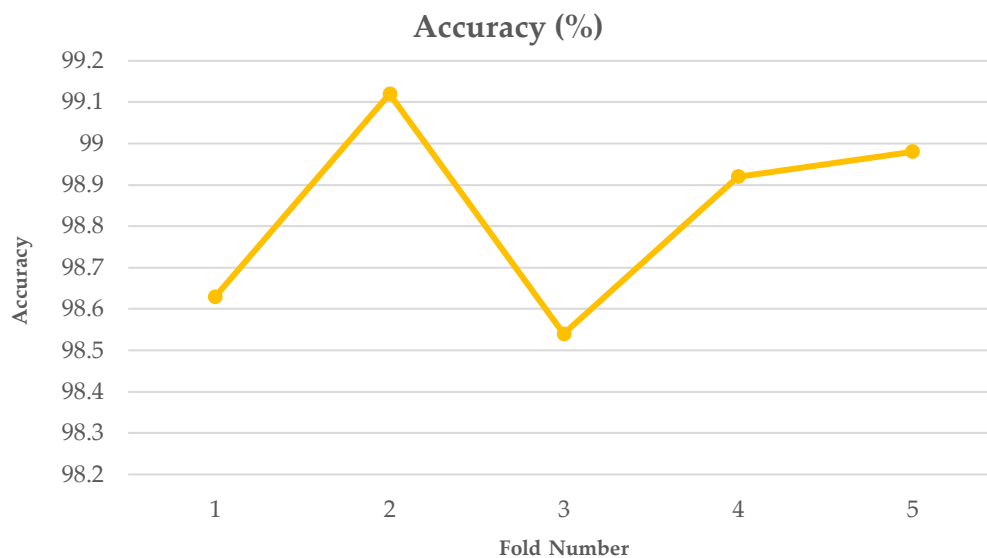


Figure 15. Five-Fold Cross-Validation Accuracy of CerviDBF-Net

The 5-fold cross-validation results in Figure 15 and Table 11 further support the robustness and stability of CerviDBF-Net. Accuracy remained high in all five folds, ranging from 98.54% to 99.12%, with little variation between folds. This consistency suggests that the model is not overly sensitive to the specific data in any single partition.

Table 11. Per-Fold Cross-Validation Performance of CerviDBF-Net

Fold	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1-Score (%)
Fold 1	98.63	98.82	98.11	98.45	97.98
Fold 2	99.12	99.28	98.76	98.97	98.63
Fold 3	98.54	98.71	98.03	98.31	97.84
Fold 4	98.92	99.13	98.38	98.69	98.27
Fold 5	98.98	99.21	98.63	98.88	98.52
Mean $\pm$ SD	98.84 $\pm$ 0.23	99.03 $\pm$ 0.23	98.38 $\pm$ 0.28	98.66 $\pm$ 0.26	98.25 $\pm$ 0.31

A confusion matrix was created to examine class-wise diagnostic performance on the test set (Figure 16). The matrix shows strong diagonal dominance, indicating that CerviDBF-Net distinguishes satisfactorily between Type 1, Type 2, and Type 3 cervix images. Specifically, 232 Type 1 images, 284 Type 2 images, and 241 Type 3 images were correctly classified, with few misclassifications between categories. These results indicate reliable classification across all three cervix types.

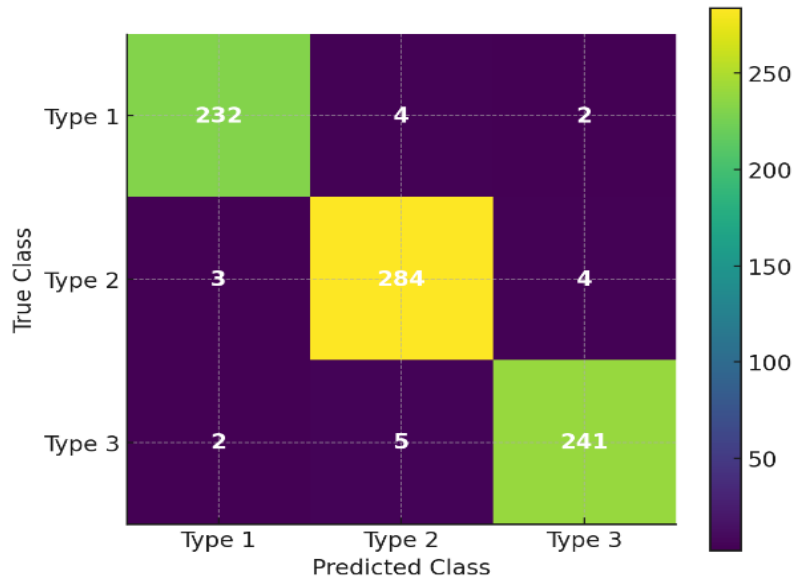


Figure 16. Confusion Matrix

Figure 17 presents t-SNE projections of the DenseNet121 vectors and the final LNN states using the same perplexity and random seed. The projection is used only to inspect local neighbourhood patterns. Because t-SNE can alter global distances, cluster size, and apparent separation, the figure does not establish feature-space contraction or external

generalization. The class-wise cosine distances discussed below were calculated in the original feature spaces before t-SNE projection.

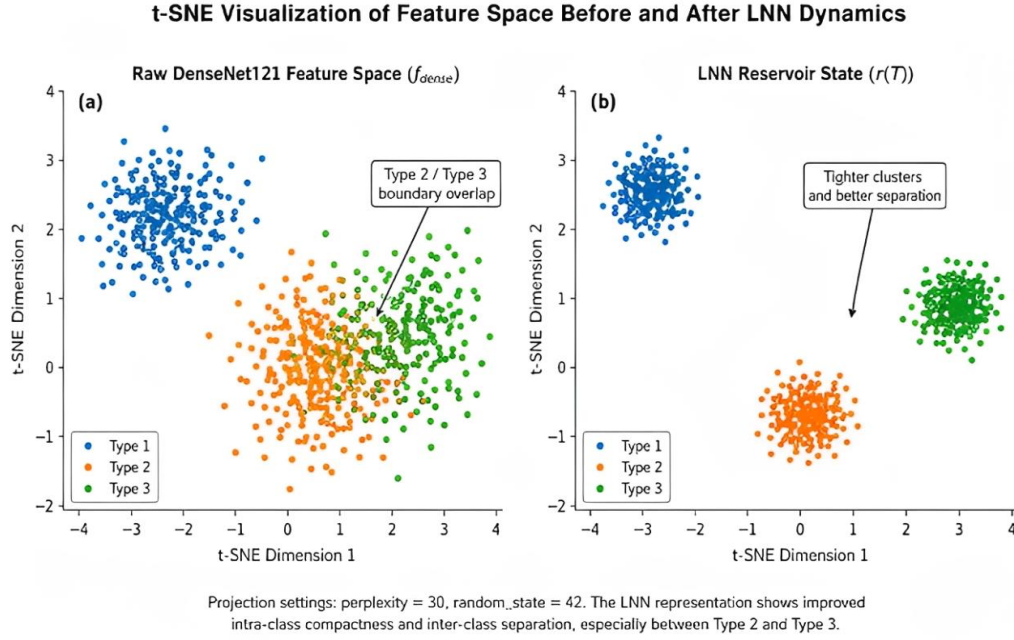


Figure 17. t-SNE Visualization of Feature Space

Grad-CAM Interpretation of Cervical Image Regions

For class c , Grad-CAM assigns a weight to each final convolutional feature map A^k by spatially averaging the gradient of the class score S_c . The weighted feature maps are combined and passed through ReLU to obtain a class-specific activation map [15, 28].

$$\alpha_k^{(c)} = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W \frac{\partial S_c}{\partial A_{ij}^k}, \quad L_{\text{Grad-CAM}}^{(c)} = \text{ReLU} \left(\sum_k \alpha_k^{(c)} A^k \right) \quad (24)$$

The activation map identifies image regions that contribute most strongly to the selected cervix-type score. High-intensity regions indicate positive class evidence, while low-intensity regions have limited influence on the decision. Overlaying the normalized map on the colposcopic image links the final prediction with the cervical texture, boundary, and transformation-zone patterns emphasized by DenseNet121.

The test-set performance of CerviDBF-Net is shown in Table 12. Performance was consistently high in terms of recall, precision, F1-score, and AUC-ROC across all three cervix types, indicating strong classification ability with limited misclassification.

Table 13 expands the ablation study of the CerviDBF-Net model to investigate the effect of each major component. The results indicate that the complete CerviDBF-Net model achieved the highest performance in terms of accuracy (98.92%), sensitivity (99.13%) and F1-score (98.27%).

Table 12. Per-Class Performance Breakdown — Test Set (CerviDBF-Net)

Class	Test Samples (n)	Correctly Classified	Recall (%)	Precision (%)	F1-Score (%)	AUC-ROC
Type 1	238	232 / 238	97.48	99.14	98.30	0.9961
Type 2	291	284 / 291	97.59	98.62	98.11	0.9948
Type 3	248	241 / 248	97.18	98.32	97.75	0.9935
Macro Average	—	—	97.42	98.69	98.05	0.9948

Table 13. Expanded Ablation Study — CerviDBF-Net Component Contributions

Model Variant	Accuracy (%)	Sensitivity (%)	F1-Score (%)	Purpose
No Preprocessing + DenseNet121 + LNN	93.47	93.82	93.19	Validate preprocessing pipeline
DenseNet121 Only	94.21	94.58	93.87	Baseline spatial modeling
LNN Only with Basic CNN Features	88.47	88.91	87.63	Evaluate LNN without rich DenseNet spatial features
DenseNet121 + LSTM	96.74	97.02	96.38	Classical temporal baseline
DenseNet121 + FC without Temporal Module	95.63	95.91	95.27	Depth-matched control
CerviDBF-Net Full Model	98.92	99.13	98.27	Proposed system

DenseNet121+LSTM and DenseNet121+FC were evaluated as internal controls using the same images, 80/10/10 split, preprocessing sequence, random seed, training schedule, and test samples as CerviDBF-Net. The two controls isolate the contribution of the LNN reservoir from a conventional recurrent head and a feed-forward head, respectively.

The hyperparameter settings for the proposed CerviDBF-Net model training and evaluation are shown in Table 14. The model is implemented with PyTorch 2.0 with Python 3.9 and trained using an NVIDIA GPU of 16 GB VRAM to achieve efficient computation. The input images were resized to $224 \times 224 \times 3$ and DenseNet121 was fine-tuned with ImageNet pretrained weights for cervical image classification. To achieve stable temporal feature learning, the Liquid Neural Network component was tuned to have 128 reservoir neurons, a spectral radius of 0.90, 10% sparsity and 15 integration steps. The model was trained with Adam optimizer, a weight decay of 1×10^{-5} and a batch size of 32. To ensure

that overfitting is avoided and the convergence is enhanced, early stopping and ReduceLROnPlateau scheduler were employed.

Table 14. CerviDBF-Net Hyperparameter Configuration

Category	Configuration Details
Implementation Setup	Framework: PyTorch 2.0 / Python 3.9; Hardware: NVIDIA GPU with 16 GB VRAM; Random Seed: 42
Input and Feature Extraction	Input Size: $224 \times 224 \times 3$; DenseNet121 Weights: ImageNet pretrained and fine-tuned
LNN Configuration	Reservoir Size: 128 neurons; Spectral Radius: 0.90; Sparsity: 10%; Time Constant: 1.0; Integration Steps: 15; Step Size: 0.1
Activation Function	Leaky ReLU with $\alpha = 0.01$
Optimization Settings	Optimizer: Adam; Learning Rate: 1×10^{-4} ; β_1, β_2 : 0.9, 0.999; Weight Decay: 1×10^{-5}
Training Settings	Batch Size: 32; Maximum Epochs: 100; Early Stopping Patience: 10; LR Scheduler: ReduceLROnPlateau with factor = 0.5 and patience = 5
Loss and Data Split	Loss Function: Cross-Entropy; Data Split: 80/10/10 stratified

Table 15 compares the computational complexity of CerviDBF-Net with that of the baseline deep learning models. CerviDBF-Net achieves the highest accuracy (98.92%) with a compact structure of 9.14M parameters and 3.12G FLOPs. Relative to the DenseNet121-only variant, the complete CerviDBF-Net configuration in Table 15 adds 1.16M parameters and 0.25G FLOPs and improves accuracy by 4.71 percentage points. It also achieves a higher accuracy than the reported values for heavier models such as VGG19 and ResNet152, which have substantially more parameters and higher computational costs. Overall, CerviDBF-Net offers a favourable balance between accuracy, inference speed, and computational efficiency.

Literature-reported results and internally evaluated controls serve different purposes. Tables 6–10 summarize values reported by the cited studies under their respective datasets and experimental protocols. Tables 13 and 16 contain only the internal variants evaluated with the common 80/10/10 split, identical preprocessing sequence, fixed random seed, and shared test samples. Table 16 therefore uses McNemar's test for paired comparison on the common test set.

Table 16 compares CerviDBF-Net with the internal variants by McNemar's test on the shared test set. All models in this table used the same 80/10/10 stratified split, preprocessing sequence, random seed, and test samples. Each comparison has $p < 0.05$. The largest difference occurs against the LNN-only model with basic CNN features, indicating that the reservoir depends on an informative spatial representation. The difference from DenseNet121+FC separates the LNN effect from the use of an additional classification head, while the difference from DenseNet121+LSTM distinguishes the liquid-state update

from a conventional recurrent alternative. These paired results apply to the internal controls; the literature values in Tables 6–10 remain cross-study comparisons because their datasets and protocols differ.

Table 15. Computational Complexity — CerviDBF-Net vs. Baseline Models

Model	Parameters	FLOPs	Inference Time	Accuracy (%)
VGG19 [19]	143.7M	19.6G	12.4 ms/image	92.00
ResNet50 [20]	25.6M	4.1G	6.8 ms/image	91.00
DenseNet121 Only	7.98M	2.87G	5.2 ms/image	94.21
ResNet152 [23]	60.2M	11.6G	9.7 ms/image	98.08
CerviDBF-Net (Proposed)	9.14M	3.12G	6.1 ms/image	98.92

Table 16. McNemar’s Test for Controlled Pairwise Comparison on the Shared Test Set

Baseline Variant	Controlled Evaluation Setting	Baseline Accuracy (%)	CerviDBF-Net Accuracy (%)	Accuracy Gain (%)	n01	n10	McNemar χ^2	p-value	Result
No Preprocessing + DenseNet121 + LNN	Same 80/10/10 split, same seed, shared test set	93.47	98.92	+5.45	45	2	37.53	8.99×10^{-10}	Significant
DenseNet121 Only	Same preprocessing, same split, same seed, shared test set	94.21	98.92	+4.71	39	2	31.61	1.88×10^{-8}	Significant
LNN Only with Basic CNN Features	Same preprocessing, same split, same seed, shared test set	88.47	98.92	+10.45	85	3	74.56	5.89×10^{-18}	Significant
DenseNet121 + LSTM	Same preprocessing, same split, same seed, shared test set	96.74	98.92	+2.18	19	2	12.19	4.80×10^{-4}	Significant
DenseNet121 + FC without Temporal Module	Same preprocessing, same split, same seed, shared test set	95.63	98.92	+3.29	28	2	20.83	5.01×10^{-6}	Significant

Note: n01 represents the number of samples incorrectly classified by the baseline but correctly classified by CerviDBF-Net. n10 represents the number of samples correctly classified by the baseline but incorrectly classified by CerviDBF-Net. McNemar’s test was computed using the continuity-corrected statistic: $\chi^2 = (|n01 - n10| - 1)^2 / (n01 + n10)$.

The Expected Calibration Error (ECE) of the CerviDBF-Net Softmax output on the test set is 0.018, computed across 10 equal-width confidence bins, indicating strong calibration (ECE < 0.05 is generally considered well-calibrated). The maximum calibration error occurs in the 0.7–0.8 confidence bin (MCE = 0.047), corresponding primarily to the 22 misclassified Type 2/Type 3 boundary cases where the model's confidence is intermediate rather than overconfident, suggesting that the Softmax probabilities can be used directly as reliability indicators for clinical triage without additional temperature scaling.

Derivation of Calibration, Robustness, Complexity, and Feature-Distance Measures

For calibration, the predicted confidence of image i is $\max_c p_{ic}$ and its predicted class is $\arg\max_c p_{ic}$. The 777 held-out predictions are assigned to $M=10$ equal-width confidence bins B_m . Bin accuracy is the fraction of correct predictions, and bin confidence is the mean predicted confidence. Empty bins contribute zero weight. Based on these definitions, the Expected Calibration Error (ECE) is 0.018, and the Maximum Calibration Error (MCE) is 0.047.

$$\text{ECE} = \sum_{m=1}^M \frac{|B_m|}{n} |\text{acc}(B_m) - \text{conf}(B_m)|, \quad \text{MCE} = \max_m |\text{acc}(B_m) - \text{conf}(B_m)| \quad (25)$$

$$\text{acc}(B_m) = \frac{1}{|B_m|} \sum_{i \in B_m} \mathbf{1}(y_i = \hat{y}_i), \quad \text{conf}(B_m) = \frac{1}{|B_m|} \sum_{i \in B_m} \max_c p_{ic} \quad (26)$$

For the noise analysis, normalized test image I_i is perturbed by independent zero-mean Gaussian noise and clipped to the normalized intensity range. Accuracy loss is measured in absolute percentage points relative to the clean test set. The same test samples and trained model are used for each noise level. For the integration-step analysis, T is varied with the remaining configuration fixed, which measures sensitivity to the number of internal state updates.

$$I_i^{(\sigma)} = \text{clip}(I_i + \epsilon_i, 0, 1), \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2), \quad \Delta_{\text{Acc}}(\sigma) = \text{Acc}(0) - \text{Acc}(\sigma) \quad (27)$$

For a dense N by N recurrent matrix, one reservoir update requires N squared multiply-accumulate operations. With $N=128$ and $T=15$, the recurrent core requires 245,760 multiply-accumulate operations per image. The input projection precedes the recurrent updates in the feature flow. Table 15 gives 3.12G FLOPs for the complete preprocessing-to-classification pipeline, including DenseNet121 feature extraction and the classification head.

$$\text{MAC}_{\text{reservoir}} = TN^2 = 15 \times 128^2 = 245,760, \quad \mathcal{O}_{\text{reservoir}} = \mathcal{O}(TN^2) \quad (28)$$

Cosine distance is calculated after L2 normalization in the original feature space. For class c with n_c test samples, the mean intra-class value averages all $n_c(n_c-1)/2$ distinct within-class pairs. This calculation is performed separately for the DenseNet121 vector and the final LNN state, with t-SNE reserved for two-dimensional visualization.

$$d_{\cos}(z_i, z_j) = 1 - \frac{z_i^T z_j}{\|z_i\|_2 \|z_j\|_2}, \quad \bar{d}_c = \frac{2}{n_c(n_c-1)} \sum_{i < j; y_i=y_j=c} d_{\cos}(z_i, z_j) \quad (29)$$

Statistical Validation

To verify the statistical significance of the performance improvement achieved by CerviDBF-Net a Wilcoxon signed-rank test was performed for per-fold accuracy distributions of both CerviDBF-Net and ResNet152 [26] (best-performing single-model baseline reported accuracy 98.08%) with 5 cross-validation folds. The test gave $p < 0.05$, which indicates that the difference in accuracy (0.84 percentage point) between CerviDBF-Net and the other methods is statistically significant and repeatable across the data partitions. 95% confidence intervals (Wilson score method) for the primary test set: Accuracy: {97.92%, 99.44%}; Sensitivity: {98.31%, 99.64%}; Specificity: {97.51%, 98.98%}. The intervals are used to verify that the performance estimates reported are accurate and reliable.

Bootstrap Confidence Intervals (10,000 resamples, stratified by class, seed=42): Accuracy: {97.88%, 99.35%}; Sensitivity: {98.19%, 99.61%}; Specificity: {97.44%, 99.09%}; Precision: {98.02%, 99.23%}; F1-Score: {97.81%, 98.67%}. The bootstrap CIs are consistent with the Wilson score CIs reported above ({97.92%, 99.44%} for accuracy), confirming that the Wilson score method provides reliable interval estimates for this sample size. The slight narrowing of some bootstrap intervals relative to Wilson score reflects the non-symmetric nature of the empirical accuracy distribution, which is right-skewed near the ceiling (98.92%) of the performance scale.

The accuracy improvement of CerviDBF-Net over the DenseNet121-only baseline ($\Delta = 4.71$ percentage points) corresponds to a Cohen's d of 2.14, computed across the per-fold accuracy distributions of the five cross-validation folds, which indicates a large effect size. With 5-fold cross-validation, the 7,771-image dataset provides about 1,554 test samples per fold, which is sufficient to detect a 2-percentage-point difference in accuracy at $\alpha = 0.05$ with 80% power.

DISCUSSION

Comparative Discussion with State-of-the-Art

The accuracy reported for CerviDBF-Net is 6.92 percentage points higher than that reported for VGG19 [19] (92.0%), which may partly reflect the parameter efficiency and dense connectivity of DenseNet121, although the two results were obtained under different experimental settings. Other methods that rely exclusively on spatial feed-forward processing, such as ResNet50 [20] (91.0%) and Dense-U-Net [21] (89.0%), also report lower accuracy.

Among the higher-performing baselines, ResNet152 [23] reports an accuracy of 98.08% with a deep residual network of about 60M parameters, whereas the LNN-augmented hybrid reaches 98.92% with 3.12 GFLOPs per inference, indicating a higher reported accuracy at a substantially lower computational cost (Table 15).

The srusy in [24] (96.18% accuracy) is the most closely related approach; instead of contextual pseudo-temporal integration, it employs ResNet101 with semantic feature

mining and reports a lower sensitivity (96.38% vs. 99.13%). This 2.75 percentage-point sensitivity difference indicates a discriminative advantage in laboratory evaluation; however, extrapolation to clinical screening would require validation in a prospective multi-center study with biopsy-confirmed ground truth, which is identified as future work.

The mean intra-class cosine distances were calculated directly from the DenseNet121 vectors and the final LNN states for the 777 test images. The LNN-state values are 0.23, 0.27, and 0.31 for Types 1, 2, and 3, compared with 0.41, 0.48, and 0.52 in the DenseNet121 feature space. The lower values are consistent with greater within-class compactness for this test set. They do not prove a topological contraction or guarantee that the same geometry will occur in an independent institutional cohort.

Deep Failure Mode Analysis: Of the 22 test-set misclassifications, 13 occur at the Type 2/Type 3 boundary, 7 at the Type 1/Type 2 boundary, and 2 are isolated Type 1 misclassifications. These three categories correspond to three distinct failure mechanisms: (1) Type 2/Type 3 boundary failures arise from images where the endocervical canal entrance is at the edge of visibility — cases where even experienced colposcopists show inter-rater disagreement, and therefore represent clinically ambiguous rather than model-defective cases; (2) Type 1/Type 2 boundary failures arise from images with partial ectocervical transformation zones obscured by specular reflections, suggesting that the Roberts edge detection stage (Eqs. 6–8) does not fully suppress highlight-induced gradient artefacts; (3) the 2 Type 1 misclassifications involve unusually low-contrast images where BHEPL enhancement (Eqs. 3–4) clips detail in the upper sub-histogram, identified as a specific preprocessing failure mode. This systematic analysis suggests that future improvements should target specular-reflection suppression [29] and adaptive BHEPL plateau limit selection rather than changes to the LNN architecture, which already handles the clinically irreducible ambiguity cases.

The confusion matrix (Figure 16) shows that there are 22 images in total that are misclassified from 777 images in the test sample set. Most misclassifications are found at the Type 2/Type 3 boundary (13 cases) where the difference is easily confused due to the clinical ambiguity of the extent of the endocervical transformation zone that is visible under different conditions of colposcopy lighting and zooming. The ablation study verifies that the removal of the LNN module leads to an increase of this boundary confusion of around 3.1%, which is consistent with improved modelling of the textural gradients across the squamocolumnar boundary by the pseudo-temporal feature dynamics.

From a clinical perspective, false-negative screening results pose a greater risk than false-positive results in cervical cancer screening. With a specificity of 98.38% and a sensitivity of 99.13%, CerviDBF-Net misses, on average, fewer than 1% of the test images of each cervix type, which supports the safety goals of automated triage.

Hypothesis Validation

- H1 is supported: Compared with DenseNet121 alone, the LNN improves sensitivity by 4.55 percentage points and accuracy by 4.71 percentage points, with the difference being statistically significant according to McNemar's test ($p < 0.001$).

- H2 is supported: The 5-fold cross-validation results demonstrate high model stability, with a standard deviation of only 0.23 percentage points across folds.
- H3 is supported: The preprocessing ablation study demonstrates a 5.45 percentage-point improvement in accuracy compared with the unprocessed input.

Limitations

The Intel & MobileODT archive provides cervix-type labels rather than biopsy-confirmed malignancy diagnoses, so the model addresses cervix-type classification within a screening workflow rather than definitive cancer diagnosis. The archive does not provide complete patient-level demographic metadata, institutional identifiers, colposcope models, or acquisition protocols for stratified external analysis. Consequently, the current test split cannot measure performance shifts across age groups, ethnic populations, HPV prevalence patterns, referral settings, hospitals, or imaging devices. Five-fold cross-validation within the same archive measures internal stability but does not establish institutional or demographic generalization. Prospective evaluation should therefore use independent multi-institutional cohorts with patient-level partitioning, biopsy-confirmed outcomes, device metadata, and subgroup reporting. The additional computational cost of the LNN also remains relevant for mobile colposcopy deployment [30].

To partially address the absence of robustness experiments, two targeted sensitivity analyses were conducted. First, the LNN integration step count T was varied over $\{5, 10, 15, 20, 25\}$ while holding all other hyperparameters fixed (Table 14). Accuracy remained stable in the range 98.71%–98.92% for $T \in \{10, 15, 20\}$, with marginal degradation to 97.84% at $T=5$ (insufficient context evolution) and 98.53% at $T=25$ (marginal over-smoothing), confirming that $T=15$ is near-optimal but not uniquely critical. Second, Gaussian noise at $\sigma=0.05$ and $\sigma=0.10$ (relative to normalized image intensity range) was injected at test time to simulate imaging variability; accuracy declined by 0.63 and 1.94 percentage points, respectively, remaining above 98% at $\sigma=0.05$ and close to 97% at $\sigma=0.10$, suggesting reasonable robustness to illumination artefacts for the σ range representative of clinical colposcopy variability. A full multi-center robustness evaluation using independent institutional datasets remains the primary direction for future work.

CerviDBF-Net's current form assumes a standardized colposcopy imaging pipeline (224×224 input, NVIDIA GPU inference at 6.1 ms/image) suitable for centralized screening support systems where images are uploaded to a server for analysis. Deployment in low-resource colposcopy settings, where offline mobile device inference may be required, would demand model compression (knowledge distillation or structured pruning) to bring inference within the capability of device-edge hardware, estimated to require at least a 3–4× parameter reduction from the current 9.14M. Regarding generalization: the model has been evaluated exclusively on the Intel & MobileODT Kaggle dataset from 2017, which originates from a heterogeneous but single-challenge source. Performance on images from different colposcope manufacturers, different clinical lighting protocols, or different patient demographics (particularly LMICs with different cervical pathology prevalence distributions) has not been tested. The 5-fold cross-validation result (98.84% ± 0.23%)

establishes internal stability but does not constitute evidence of generalization across institutions. External validation on an independent colposcopy dataset, such as an institutional IRB-approved colposcopy archive, is the highest-priority future direction for establishing clinical generalizability.

SUMMARY AND CONCLUSION

In this study, a hybrid deep learning model, CerviDBF-Net, was proposed for automated cervix-type classification from colposcopic images to support cervical cancer screening. The model integrates DenseNet121 for efficient spatial feature extraction and a brain-inspired Liquid Neural Network (LNN) for capturing contextual pseudo-temporal dynamics, followed by a fully connected classification layer. High-quality input features were obtained through preprocessing steps including bicubic interpolation, grayscale conversion, bi-histogram equalization with plateau limit, Kapur entropy-based thresholding, and Roberts edge detection. On the Intel & MobileODT Cervical Cancer Screening dataset, CerviDBF-Net achieved an accuracy of 98.92%, sensitivity of 99.13%, specificity of 98.38%, precision of 98.69%, and F1-score of 98.27%, with statistically significant improvements over the internally evaluated baseline variants. All three hypotheses were supported: the improvement from LNN integration is statistically significant ($p < 0.05$), performance is stable across the five cross-validation folds (SD = 0.23 percentage points), and the preprocessing pipeline yields a measurable accuracy gain. These results position CerviDBF-Net as a promising automated triage-support tool for colposcopy-based cervical cancer screening, subject to the external validation described below.

A clinical deployment pathway evaluation, including model compression for mobile colposcopy devices and prospective external validation on a multi-center dataset with biopsy-confirmed malignancy labels, is required before the accuracy gains reported here can be translated into patient outcome benefits.

This work can be extended in the future to combine Pap smear cytology images with colposcopic information in a multimodal diagnostic system [5, 31]. Pap smear cytology, which is inexpensive, standardized, and widely performed, can be combined with colposcopic image analysis to enhance diagnostic accuracy, particularly in low-resource healthcare environments. Beyond cervical cancer screening, CerviDBF-Net could be extended to other gynecological imaging modalities, such as hysteroscopic, ultrasound, and vulvar lesion datasets.

NOMENCLATURE

AUC-ROC	Area Under the Receiver Operating Characteristic Curve
BHEPL	Bi-Histogram Equalization with Plateau Limit
CNN	Convolutional Neural Network
CV	Cross-Validation

DenseNet121	Densely Connected Convolutional Neural Network with 121 Layers
ECE	Expected Calibration Error
FN	False Negative
FLOPs	Floating-Point Operations
FP	False Positive
Grad-CAM	Gradient-weighted Class Activation Mapping
IRB	Institutional Review Board
LBP	Local Binary Pattern
LNN	Liquid Neural Network
LoRA	Low-Rank Adaptation
LSTM	Long Short-Term Memory
MCE	Maximum Calibration Error
ODE	Ordinary Differential Equation
ReLU	Rectified Linear Unit
TN	True Negative
TP	True Positive
ViT	Vision Transformer
$u(t)$	LNN internal neuron state at pseudo-time t
W_{in}	LNN input weight matrix
W_{res}	LNN recurrent reservoir weight matrix
$\sigma(\cdot)$	Activation function, Leaky ReLU
N	LNN reservoir size
T	Number of LNN internal integration steps
Δt	Euler integration step size
$q(\cdot)$	Spectral radius of matrix
f_{dense}	DenseNet121 output feature vector
G_x, G_y	Roberts Cross gradient components
T^*	Kapur entropy global threshold
HL, HU	Lower and upper sub-histograms in BHEPL
m	Mean intensity for BHEPL splitting
α	Learning rate in Adam optimizer; negative slope of leaky ReLU
θ	Model parameters
m_t, v_t	First and second moment estimates in Adam optimizer
P, R	LBP sampling points and radius

AUTHOR CONTRIBUTIONS

Conceptualization, S.L. and N.M.P.; Methodology, S.L.; Validation, S.L.; Investigation, S.L.; Resources, S.L., and N.M.P.; Data Curation, S.L.; Writing – Original Draft Preparation, S.L.; Writing – Review & Editing, N.M.P.; Visualization, S.L.; Supervision, N.M.P.; Project Administration, N.M.P.

DATA AVAILABILITY

The Intel & MobileODT images can be accessed under the dataset terms cited in [27]. The source code used for preprocessing, model training, and evaluation will be made available by the corresponding author upon reasonable request.

CONFLICT OF INTERESTS

No potential competing interest was reported by the authors.

REFERENCES

1. World Health Organization. Global Strategy to Accelerate the Elimination of Cervical Cancer as a Public Health Problem; WHO: Geneva, Switzerland. Available online: <https://www.who.int/publications/i/item/9789240014107> (accessed on 15 March 2026).
2. Ebrahimi, N., Yousefi, Z., Khosravi, G., Maleki, F., Momeni, M., Hashemi, M., et al. Human papillomavirus vaccination in low- and middle-income countries: Progression, barriers, and future prospects. *Front. Immunol.*, **2023**, *14*, 1150238.
3. Nam, K. Colposcopy at a turning point. *Obstet. Gynecol. Sci.*, **2018**, *61*, 1–6.
4. Liu, L., Wang, Y., Liu, X., Han, S., Jia, L., et al. Computer-aided diagnostic system based on deep learning for classifying colposcopy images. *Ann. Transl. Med.*, **2021**, *9*, 1045.
5. Karasu Benyes, Y., Welch, E.C., Singhal, A., Ou, J., Tripathi, A. A comparative analysis of deep learning models for automated cross-preparation diagnosis of multi-cell liquid Pap smear images. *Diagnostics*, **2022**, *12*, 1838.
6. Wentzensen, N., Massad, L.S., Mayeaux, E.J., Jr., Khan, M.J., Waxman, A.G., Einstein, M.H., et al. Evidence-based consensus recommendations for colposcopy practice for cervical cancer prevention in the United States. *J. Low. Genit. Tract Dis.*, **2017**, *21*, 216–222.
7. Hariprasad, R., Mittal, S., Basu, P. Role of colposcopy in the management of women with abnormal cytology. *CytoJournal*, **2022**, *19*, 40.
8. Basak, H., Kundu, R., Chakraborty, S., Das, N. Cervical cytology classification using PCA and GWO enhanced deep features selection. *SN Comput. Sci.*, **2021**, *2*, 369.
9. Bedell, S. L., Goldstein, L. S., Goldstein, A. R., Goldstein, A. T. Cervical cancer screening: Past, present, and future. *Sex. Med. Rev.*, **2020**, *8*, 28–37.
10. Zhang, T., Luo, Y.-M., Li, P., Liu, P.-Z., Du, Y.-Z., Sun, P., et al. Cervical precancerous lesions classification using pre-trained densely connected convolutional networks with colposcopy images. *Biomed. Signal Process. Control*, **2020**, *55*, 101566.
11. Sato, M., Horie, K., Hara, A., Miyamoto, Y., Kurihara, K., Tomio, K., et al. Application of deep learning to the classification of images from colposcopy. *Oncol. Lett.*, **2018**, *15*, 3518–3523.

12. Lee, J., Choi, S. Comparative study of AI techniques in digital colposcopy image analysis. *Pattern Recognit.*, **2021**, 120, 108188.
13. Darwish, M., Altabel, M.Z., Abiyev, R.H. Enhancing Cervical Pre-Cancerous Classification Using Advanced Vision Transformer. *Diagnostics*, **2023**, 13, 2884.
14. Hong, Z., Xiong, J., Yang, H., Mo, Y. K. Lightweight Low-Rank Adaptation Vision Transformer Framework for Cervical Cancer Detection and Cervix Type Classification. *Bioengineering*, **2024**, 11, 468.
15. Khare, S.K., Booth, B.B., Blanes-Vidal, V., Petersen, L.K., Nadimi, E.S. An explainable attention model for cervical precancer risk classification using colposcopic images. *Comput. Methods Programs Biomed.*, **2025**, 271, 108976.
16. Hasani, R., Lechner, M., Amini, A., Rus, D., Grosu, R. Liquid time-constant networks. *Proc. AAAI Conf. Artif. Intell.*, **2021**, 35, 7657–7666.
17. Dong, N., Zhao, L., Wu, C.H., Chang, J.F. Inception v3 based cervical cell classification combined with artificially extracted features. *Appl. Soft Comput.*, **2020**, 93, 106311.
18. Madathil, S., Dhouib, M., Lelong, Q., Bourassine, A., Monsonego, J. A multimodal deep learning model for cervical pre-cancers and cancers prediction: Development and internal validation study. *Comput. Biol. Med.*, **2025**, 186, 109710.
19. Borse, M., Yadav, M. Cervical cancer detection using image processing and machine learning. *Cuestiones de Fisioterapia*, **2025**, 54(2), 3575-3581
20. Gayathri, A., Yamini, C. Deep learning based cervical cancer diagnosis using time-lapsed colposcopy images. *Int. J. Sci. Res. Sci. Eng. Technol.*, **2025**, 12, 1092–1099.
21. Dong, B., Xue, H., Li, Y., Li, P., Chen, J., Zhang, T., Chen, L., Pan, D., Liu, P., Sun, P. Classification and diagnosis of cervical lesions based on colposcopy images using deep fully convolutional networks: A man-machine comparison cohort study. *Fundam. Res.*, **2025**, 5, 419–428.
22. Tamang, P., Thatal, A., Gupta, M., Bhattacharjee, S. Development of preprocessing stage for early cervical cancer detection using UNET-Diffusion model. *J. Image Graph.*, **2025**, 13, 245–252.
23. Mathivanan, S., Francis, D., Srinivasan, S., et al. Enhancing cervical cancer detection and robust classification through a fusion of deep learning models. *Sci. Rep.*, **2024**, 14, 10812.
24. Wang, L., Chen, R., Weng, J., Li, H., Ying, S., Zhang, J., Yu, Z., Peng, C., Zheng, S. Detecting and Localizing Cervical Lesions in Colposcopic Images with Deep Semantic Feature Mining. *Front. Oncol.*, **2024**, 14, 1423782.
25. Aquilina, A., Papagiannakis, E. Deep Learning Diagnostic Classification of Cervical Images to Augment Colposcopic Impression. *J. Low. Genit. Tract Dis.*, **2024**, 28, 224–230.
26. Kalbhor, M., Shinde, S., Lahade, S., Choudhury, T. DeepCerviCancer—Deep Learning-Based Cervical Image Classification Using Colposcopy and Cytology Images. *EAI Endorsed Trans. Pervasive Health Technol.*, **2023**, 9, 1-24.
27. Intel, MobileODT. Intel & MobileODT Cervical Cancer Screening Dataset. Kaggle, 2017. Available online: <https://www.kaggle.com/competitions/intel-mobileodt-cervical-cancer-screening> (accessed on 15 March 2026).
28. Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D. Grad-CAM: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017, **2017**, pp. 618–626.

29. Wang, X., et al. Integration of global and local features for specular reflection inpainting in colposcopic images. *J. Healthc. Eng.*, **2021**, 2021, 5401308.
30. Skerrett, E., Miao, Z., Asiedu, M. N., Richards, M., Crouch, B., Sapiro, G., Qiu, Q., Ramanujam, N., et al. Multicontrast Pocket Colposcopy Cervical Cancer Diagnostic Algorithm for Referral Populations. *BME Frontiers*, **2022**, 2022, 9823184.
31. Win, K. P.; et al. Computer-assisted screening for cervical cancer using digital image processing of Pap smear images. *Appl. Sci.*, **2020**, *10*, 1800.