



Attention-Guided Hybrid Network for Cervical Cancer Classification

Abdalla Ibrahim Abdalla Musa^{1,2*}, Mahir Mohammed Sharif Adam^{2,3}

¹ Department of Computer Science, College of Computer, Qassim University, Buraydah 51452, Saudi Arabia

² Computer Science Department, Faculty of Computer Science and Information Technology, Omdurman Islamic University, Omdurman 382, Sudan

³ Computer and Self Development, Common First Year Unit, Prince Sattam bin Abdulaziz University, Alkharj 16273, Saudi Arabia

Corresponding Author Email: ab.musa@qu.edu.sa

Copyright: ©2025 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/isi.300116>

ABSTRACT

Received: 12 October 2024

Revised: 6 December 2024

Accepted: 14 January 2025

Available online: 25 January 2025

Keywords:

cervical cancer, deep learning, hybrid network, classification, pap smear, se attention

Cervical cancer remains a significant global health burden, necessitating accurate and efficient diagnostic tools. This paper proposes a novel deep learning architecture, the Squeeze-and-Excitation Attention-Guided Hybrid Network (SE-AG-HN), for the classification of cervical cancer from Pap smear images. The proposed method effectively addresses the challenges posed by image variability and subtle abnormalities by integrating Squeeze-and-Excitation (SE) attention and a hybrid convolutional neural network (CNN) - recurrent neural network (RNN) structure. The SE attention module recalibrates feature channels to enhance discriminative information, while the hybrid architecture leverages both local and global contextual features. Experimental results on a benchmark cervical cancer dataset demonstrate the superior performance of SE-AG-HN compared to state-of-the-art methods, highlighting its potential as a valuable tool for cervical cancer screening and diagnosis.

1. INTRODUCTION

Cervical cancer is a malignancy arising from the cervix and remains a significant global health concern, especially in developing regions. It arises primarily due to the persistent infection of high-risk Human Papillomavirus (HPV) types, such as 16 and 18. The progression of cervical cancer is often insidious, with early stages typically asymptomatic. As the disease advances, symptoms may manifest as abnormal vaginal bleeding, pelvic pain, and persistent vaginal discharge. Identifying and treating cervical cancer early significantly enhances the chances of successful treatment and better patient outcomes. Traditional screening methods, such as Pap smears and HPV testing, have been instrumental in reducing cervical cancer incidence. However, these methods show limitations, including subjectivity in Pap smear interpretation and the potential for false negatives. Additionally, the accessibility and affordability of these screening modalities remain challenges in many regions. Figure 1 illustrates the pathogenesis of cervical cancer, highlighting the role of HPV infection, risk factors, prevention methods, symptoms, and treatment options. This highlights the critical role of vaccination against HPV, regular screening tests, and early detection of cervical cancer in effectively preventing and managing this disease.

To address these limitations, there has been a growing interest in developing automated diagnostic systems based on deep learning. Existing deep learning models have achieved promising results in cervical cancer classification. However,

these models often struggle with the variability in image quality, subtle abnormalities, and the complex interplay of different image features.

This study proposes a novel deep learning architecture, the Squeeze-and-Excitation Attention-Guided Hybrid Network, to overcome these challenges. By integrating Squeeze-and-Excitation attention and a hybrid convolutional neural network recurrent neural network structure, our model effectively captures both local and global image features, enabling more accurate and robust classification of cervical cancer from Pap smear images. The SE attention module enhances the model's ability to focus on discriminative image regions, while the hybrid architecture leverages the strengths of both CNNs and RNNs for comprehensive feature extraction. By addressing the limitations of existing methods, the proposed SE-AG-HN model has the potential to improve cervical cancer screening and diagnosis, leading to earlier detection, reduced mortality rates, and improved patient outcomes.

2. LITERATURE SURVEY

Pacal and Kılıcarslan [1] developed a robust cervical cancer classification system using CNN and ViT models, tested on the SIPaKMeD pap-smear dataset. The model employed data augmentation and ensemble learning techniques, finding that ViT models outperformed CNN models in diagnostic accuracy suggesting potential clinical application for early and precise cancer identification. Kalbhor and Shinde [2] investigated

cervical cancer diagnosis using deep learning approaches using pre-trained models as feature extractors combined with machine learning algorithms, achieving 92.03% accuracy with ResNet-50. Additionally, they applied transfer learning, with Google Net fine-tuning yielding a classification accuracy of 96.01%. Kumari et al. [3] proposed an automated cervical cancer classification system using a Deep Neural Network (DNN) to address early-stage prediction challenges. The method involved four stages of pre-processing, outlier elimination, dimensionality reduction via Principal Component Analysis (PCA) and classification. The DNN classifier achieved effective performance in distinguishing normal from abnormal cervical cells. Younezzade et al. [4] proposed a cervical cancer detection model using convolutional neural networks and colposcopy images. It examined the impact of increasing the number of classes on model accuracy, which reached 99% during training but dropped to 43.11% during testing. Cheng et al. [5] explored the application of deep learning in cervical cancer image processing discussing the workflow of image acquisition, preprocessing, feature extraction, and target detection, with a focus on CNN, generative adversarial networks (GANs) and autoencoders. Talpur et al. [6] proposed DeepCervixNet, an advanced deep learning model for cervical cancer detection in Pap smear images. By enhancing ResNet101 and DenseNet169 with sequence and excitation blocks and using ensemble learning, the model achieved an accuracy of 99.89%. Bueno-Crespo et al. [7] developed a deep learning model for cervical cancer classification, combining CNNs with the Grad-CAM technique for explainability. The heatmap from Grad-CAM was merged with the original image, with a 10% intensity fusion proving most effective. This hybrid model achieved an accuracy of 94%, aiding pathologists by highlighting regions of interest for review.

Devaraj et al. [8] employed pre-trained models like ResNet50V2, InceptionV3, Xception to cervical smear images for cervical cancer prediction. This analysis validated with cross-validation, showed ResNet50V2 achieving the highest accuracy. Metrics used included accuracy, precision, sensitivity (or recall), and F1-score, demonstrating that deep learning can accurately classify cervical cancer, thereby enhancing early diagnosis without the need for invasive procedures. Mathivanan et al. [9] investigated the use of pre-trained deep neural networks such as AlexNet, InceptionV3, ResNet-101 and ResNet-152 for feature extraction in cervical cancer detection. Their methodology combined these features with various machine learning algorithms. ResNet-152 outperformed other models tested, achieving an accuracy of 98.08% on the SIPaKMeD dataset. This hybrid approach of DL and ML aims to enhance cervical cancer classification and detection efficiency. Tan et al. [10] developed deep learning models for automated cervical cancer detection without segmentation or custom features, using transfer learning with 13 pre-trained CNN models on Pap smear images. Performance evaluation was conducted on the Herlev dataset, in which DenseNet-201 achieved the best performance. Their approach demonstrated good accuracy and efficiency, requiring minimal computing time.

Tripathi et al. [11] investigated the classification of cervical cancer using deep learning algorithms. The deep learning method, ResNet-152 was applied to the SIPAKMED pap-smear image dataset and achieved a classification accuracy of 94.89%. Deo et al. [12] proposed CerviFormer, a cross-attention-based Transformer model, for cervical cancer

classification in pap smear images. The model effectively handled large-scale input data and achieved competitive results on two publicly available datasets. It showed potential improvement in early detection and treatment of cervical cancer. Jeysihri and Kowsigan [13] proposed attention-based model effectively segmented and classified cervical cancer in biomedical images. It combined Multiscale ResUNet++ with Fuzzy C-means Clustering for segmentation and Serial Cascaded Residual Attention with Long Short-Term Memory for classification. Hyperparameters were tuned using Hybrid Arithmetic Dolphin Swarm Optimization.

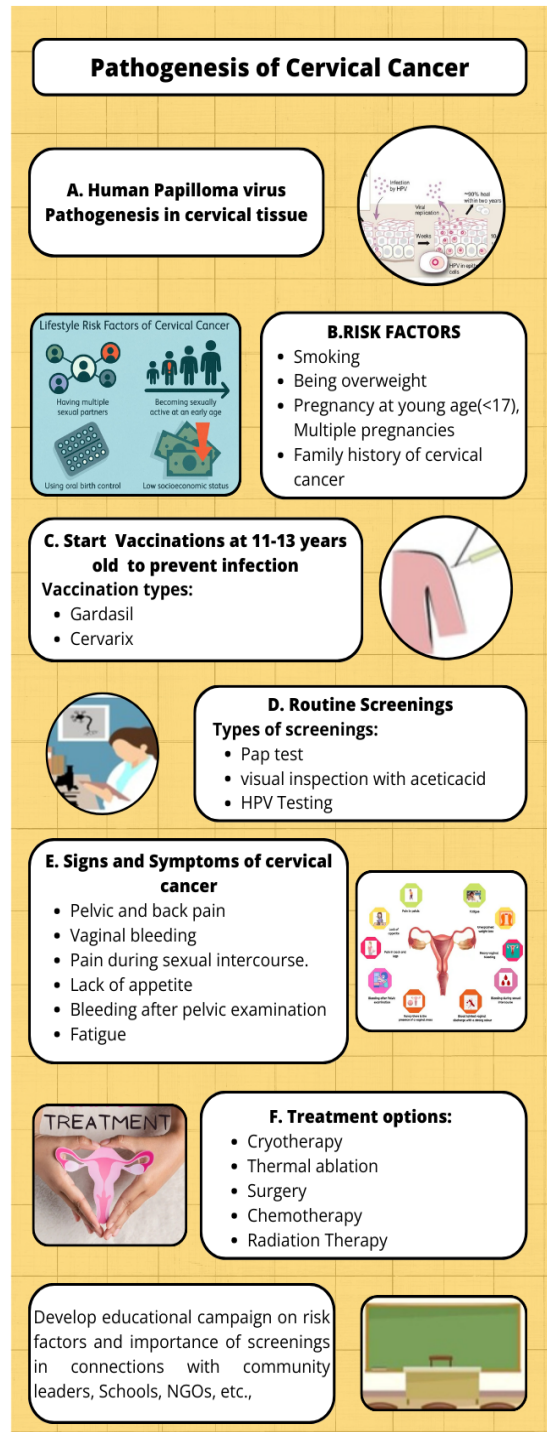


Figure 1. Pathogenesis of cervical cancer

Ganguly et al. [14] combined convolutional neural networks, clustering, and pseudo-labeling, effectively detected and

classified cervical cancer using images from the ICAR-WHO dataset. It addressed challenges related to limited labeling and dataset availability, demonstrating promising results for early diagnosis and treatment planning. Xia et al. [15] proposed SPFNet, a novel network structure for cervical cancer cell detection. It used different combination strategies and head components, and incorporated data preprocessing techniques. Experimental results demonstrated its superior performance in detecting cervical cancer cells, potentially reducing the workload of doctors and enhancing the accuracy of cervical cancer diagnosis. Ghoneim et al. [16] proposed CNN-ELM-based system which effectively detected and classified cervical cancer cells, achieving high accuracy rates on the Herlev database. It utilized deep-learned features extracted through transfer learning and fine-tuning, and benefitted from the efficiency of the Extreme Learning Machine (ELM) classifier. Fan et al. [17] proposed CAM-VT framework, combining Conjugated Attention Mechanism and Visual Transformer, effectively identified cervical cancer nest images. It outperformed other deep learning models and demonstrated strong performance in both ablation and extended experiments, highlighting its potential for practical clinical application in cervical cancer screening. Habtemariam et al. [18] proposed a deep learning-based system to classify cervix types and diagnose cervical cancer. A lightweight MobileNetV2-YOLOv3 model was used for region of interest (ROI) extraction, while EfficientNetB0 models were used for cervix type and cervical cancer classification. Tanimu et al. [19] employed a decision tree classifier to predict cervical cancer outcomes based on risk factors. Recursive Feature Elimination (RFE) and Least Absolute Shrinkage and Selection Operator (LASSO) were used to select the most significant features, while SMOTETomek addressed data imbalance and missing values. With an accuracy of 98.72% and a sensitivity of 100%, the model effectively predicted cervical cancer outcomes. Feng et al. [20] developed the CT-YOLOv5 model to enhance cervical lesion detection by improving the YOLOv5s algorithm with transformers and a Convolutional Block Attention Module (CBAM). Using PANet and CBAM for refined feature extraction, the model achieved precision, recall, and mAP scores of 93.97%, 92.94%, and 92.8%, respectively. CT-YOLOv5 outperformed models like SSD and YOLOv5, aiding in accurate identification of affected areas and disease severity in cervical images, thereby advancing cervical cancer detection. Yi et al. [21] introduced the Multi-scale Window Transformer (MWT) to improve cervical cytopathology image recognition, aiming to address the challenge of manually intensive cervical cancer screenings. The MWT incorporated multi-scale window multi-head self-attention (MW-MSA) to extract local and integrated cell features enhancing feature interaction without needing whole-image self-attention. By using convolutional feed-forward networks within a pyramid architecture, the model achieved efficient and accurate representation. Tested on large datasets with over 360,000 images across two- and four-category classifications, the MWT outperformed both general and specialized cytopathology classifiers.

The aim of this research is to develop and assess an innovative deep learning architecture, the Squeeze-and-Excitation Attention-Guided Hybrid Network, designed to enhance cervical cancer classification from Pap smear images. This study integrates a sophisticated Squeeze-and-Excitation attention mechanism with a hybrid convolutional neural network and recurrent neural network framework to tackle

challenges such as image variability and subtle abnormalities. By refining the model's ability to focus on critical features and capture complex patterns, the research aims to demonstrate that SE-AG-HN significantly improves classification accuracy over current methods, ultimately offering a more effective tool for cervical cancer screening and diagnosis. The key contributions are:

- SE attention mechanism recalibrates feature channels which improves the model's ability to focus on critical regions within Pap smear images and accurately distinguish between cervical cancer and other conditions.
- The CNN-RNN hybrid combines convolutional layers for spatial feature extraction with LSTM layers for sequential pattern analysis, improving classification accuracy and robustness.

This paper is organized into five sections. The Introduction defines cervical cancer, explores its symptoms and highlights the need for improved diagnostic tools. It introduces the Squeeze-and-Excitation Attention-Guided Hybrid Network, a novel deep learning architecture designed to enhance cervical cancer classification from Pap smear images by addressing challenges like image variability and subtle abnormalities. The Literature Review examines recent research and existing methods for cervical cancer detection using Pap smear images, discussing the limitations of current approaches and the need for advanced models. The Methodology section explains the Squeeze-and-Excitation attention module and the hybrid CNN-RNN structure. It covers preprocessing, normalization, and feature extraction techniques and describes the model training process, including loss functions and evaluation metrics. In the Results and Analysis section, experimental results are presented, demonstrating the model's superior performance compared to other methods. The analysis includes strengths, limitations, and performance metrics. Finally, the Conclusion summarizes the key findings, evaluates the model's effectiveness, and suggests areas for future research and potential improvements to further enhance diagnostic accuracy.

3. METHODOLOGY

The proposed SE-AG-HN method integrates a Squeeze-and-Excitation attention mechanism with a hybrid convolutional neural network and recurrent neural network architecture to enhance cervical cancer classification from Pap smear images. The model begins by applying a series of convolutional layers, interspersed with max-pooling, to extract hierarchical spatial features from the preprocessed input images. The SE attention mechanism is then utilized to refine these features by recalibrating the importance of each channel. The refined feature maps are subsequently processed by an LSTM layer, which captures sequential dependencies within the spatial representations. This is followed by fully connected layers that further transform the features, culminating in a final dense layer with a softmax activation function that produces a multiclass classification outcome, indicating three classes cervical cancer, pre cervical cancer and noncervical cancer. The model's training process employs the Adam optimization algorithm and uses categorical cross-entropy to measure the difference between the model's predictions and the actual labels.

Figure 2 depicts the architecture of the Squeeze-and-Excitation Attention-Guided Hybrid Network for classifying

Pap smear images into Normal, Abnormal, and Pre-Cancer categories. The diagram shows the SE-AG-HN architecture for cervical cancer classification using Pap smear images. It extracts features through convolutional layers, refines them with a Squeeze-and-Excitation block, and captures sequential dependencies using an LSTM layer. The final classification into Normal, Abnormal, and Pre-Cancer categories is done through fully connected layer, combining spatial and sequential information for improved accuracy.

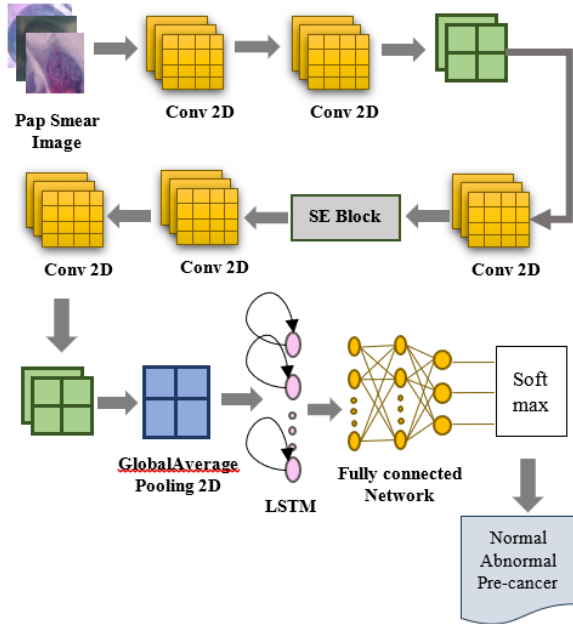


Figure 2. Architecture of the Squeeze-and-Excitation Attention-Guided Hybrid Network

3.1 Preprocessing

3.1.1 Image resizing

To ensure consistency across the dataset and compatibility with the input requirements of the neural network, all Pap smear images are resized to a uniform dimension of 128x128 pixels. This resizing standardizes the input size, allowing the model to process each image efficiently while maintaining the essential features required for accurate classification.

3.1.2 Adaptive Histogram Equalization (CLAHE)

After resizing the images, a technique called Adaptive Histogram Equalization (CLAHE) is used to improve the image contrast, making it easier to distinguish between different features. Unlike standard histogram equalization, CLAHE operates on small regions (tiles) within the image, adjusting the contrast locally to highlight subtle differences in tissue structures. This localized contrast enhancement is particularly effective in Pap smear images, where variations in cell morphology and texture are critical for identifying cancerous and non-cancerous regions. CLAHE also prevents over-amplification of noise by limiting contrast in homogeneous areas, ensuring that important features are emphasized without introducing artifacts.

3.1.3 Normalization

After contrast enhancement, the pixel values of the images are normalized to a common range, typically [0, 1]. This normalization process involves dividing the pixel values by 255, which is the maximum value in an 8-bit grayscale image.

Normalization helps to stabilize the training process by ensuring that the inputs to the neural network have a consistent scale, allowing for faster convergence and improved model performance.

3.2 SE attention mechanism

The SE attention block consists of two main operations: squeeze and excitation. During the squeeze phase, global average pooling compresses the spatial dimensions of the input feature maps into a channel descriptor vector of size C (the number of channels in the input tensor), where each value represents global information about its corresponding channel. This vector is reshaped to have dimensions $(1, 1, C)$ to match the input tensor format for further processing.

During the excitation phase, the reshaped vector is processed by two dense layers. The first dense layer reduces the dimensionality of the vector by a factor determined by the reduction ratio, typically set to 16. This reduces the vector to a size of $C // \text{reduction_ratio}$, allowing the network to learn a more compact representation of the channels. The ReLU activation function is employed to introduce non-linearity into the model, enabling it to learn more complex relationships within the data. The second fully connected layer in the model reshapes the data back to its original number of channels (C). Following this, a sigmoid activation function is applied, generating channel-wise attention weights. These weights, ranging from 0 to 1, indicate the significance of each channel in the overall image information.

Finally, these attention weights are multiplied element-wise with the original feature maps to recalibrate the channels, emphasizing the more informative ones and suppressing less relevant ones. This channel recalibration enhances the network's focus on critical features, which is particularly useful for tasks like cervical cancer classification where identifying fine-grained patterns in medical images is essential for accurate diagnosis.

The SE block improves feature map by recalibrating channel-wise weights which dynamically adjusts the importance of each channel. Given an input tensor X with dimensions (H, W, C) , where C represents the number of channels, H represent the height of the feature map and W be the width of the feature map. The SE block works in the following manner:

(1) Feature Map: Begin with the feature map obtained from a convolutional layer.

(2) Squeeze Operation:

- **Global Average Pooling:** Compute the channel-wise statistics by averaging over the spatial dimensions:

$$z_c = \frac{1}{H \times W} \sum_{h=1}^H \sum_{w=1}^W X_{h,w,c} \quad (1)$$

This results in a vector z of shape (C) , where each element z_c represents the global average of the feature map across spatial dimensions for each channel.

- **Reshape:** Reshape this vector z into the shape $(1, 1, C)$. This reshaping ensures that z can be used for channel-wise scaling. The reshaped tensor has a shape of $(1, 1, 128)$, which matches the dimensions of the feature map X for the subsequent operations.

(3) Excitation Operation:

- **Fully Connected Network:** Pass the reshaped vector z through a fully connected network to compute channel-wise

attention weights. This involves:

- **Dense Layer 1:** Apply a dense layer with weights W_1 and a ReLU activation function:

$$z' = \text{ReLU}(W_1 \cdot z) \quad (2)$$

- **Dense Layer 2:** Apply another dense layer with weights W_2 and a sigmoid activation function:

$$s = \sigma(W_2 \cdot z') \quad (3)$$

The result s is a vector of attention weights, indicating the importance of each channel.

- **Multiply with Feature Map:** Scaling the original feature map X by these attention weights s :

$$X_{SE} = X \cdot s \quad (4)$$

This produces the final SE-enhanced feature map X_{SE} , where each channel of the feature map has been recalibrated according to its importance.

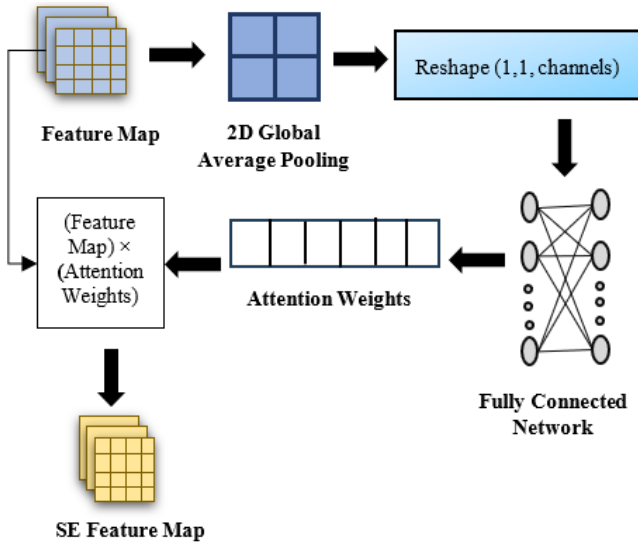


Figure 3. Squeeze-and-Excitation attention architecture

Figure 3 shows the Squeeze-and-Excitation Attention architecture processes the input feature map by applying global average pooling and reshaping in the Squeeze phase. In the Excitation phase, it uses fully connected networks to compute channel-wise attention weights and multiplies these weights with the original feature map to produce the final SE-enhanced feature map.

3.3 SE-AG-HN

The SE-AG-HN model is designed for multiclass cervical cancer classification by integrating convolutional neural networks (CNNs) with recurrent neural networks (RNNs) and fully connected network. SE attention mechanism in SE-AG-HN model significantly enhances performance by recalibrating feature maps through channel-wise attention. It allows the model to focus on most significant features while suppressing less relevant ones. This is especially beneficial in identifying subtle and localized patterns which is crucial for accurate classification. By dynamically adjusting the feature importance, the SE block ensures the model emphasizes key

discriminative features, thereby boosting accuracy and robustness.

The hybrid network architecture integrates CNNs and RNNs networks to leverage their complementary strengths. CNNs capture spatial features and texture patterns from image data, while RNNs effectively model sequential dependencies such as morphological changes across image regions. This enables the model to capture both fine grained local details and broader contextual relationships which are essential for reliable medical diagnoses. Inclusion of SE block further enhances this hybrid structure by refining feature selection, enabling better generalization to diverse and complex datasets. This integration of precise spatial and temporal features along with adaptive feature recalibration, makes SE-AG-HN model a more reliable and innovative approach for cervical cancer classification.

The model starts with five convolutional layers, using 32, 64, 128, 256, and 512 filters respectively, each with a kernel size of (3, 3) and ReLU activation functions. MaxPooling2D layers with a pool size of (2, 2) are applied after some convolutional layers to reduce spatial dimensions. Following the convolutional layers, a Squeeze-and-Excitation (SE) block recalibrates feature importance by emphasizing the most relevant features. After this, GlobalAveragePooling2D compresses the spatial dimensions into a single feature vector. This vector is then reshaped to (1, 512), preparing it for the LSTM layer. The LSTM layer with 64 units processes this reshaped vector to capture sequential dependencies and patterns. This is followed by two fully connected layers with 128 and 64 neurons, respectively, to further refine the features. The final output is a Dense layer with 3 neurons and a SoftMax activation function, providing probabilities for three classes (non-cervical cancer, pre-cervical cancer, and cervical cancer).

The CNN component extracts spatial features with multiple convolutional layers followed by max-pooling layers:

(1) **Convolutional Layers:** For a convolutional layer with k filters, kernel size (k_h, k_w) , and ReLU activation function:

$$X_{out} = \text{ReLU}(W * X + b) \quad (5)$$

where, W represents the filter weight matrix, $*$ signifies the convolution operation and b denotes the bias term.

(2) **Max-Pooling Layers:** Apply max-pooling with pool size (2, 2):

$$X_{pool}(i, j) = \max X(2i + p, 2j + q) \quad (6)$$

where, (p, q) are the indices within the pooling window.

After convolutional and SE block processing, global average pooling is applied:

$$z = \frac{1}{H \times W} \sum_{h=1}^H \sum_{w=1}^W X_{h,w} \quad (7)$$

Reshape z to fit the LSTM input shape (1, -1).

(3) **LSTM Layer:** The LSTM layer with 64 units processes the reshaped feature vector. The LSTM's output at time step t , denoted as h_t , is defined as follows:

$$h_t = \text{LSTM}(x_t, h_{t-1}) \quad (8)$$

where, x_t is the input at time t and h_{t-1} is the hidden state from the previous time step.

(4) Fully Connected Network: After the LSTM layer, the feature vector is passed through a fully connected network consisting of two dense layers before reaching the final output layer:

The initial dense layer comprises 128 neurons, each employing a ReLU activation function:

$$h_1 = \text{ReLU}(w_1 \cdot h + b_1) \quad (9)$$

The second dense layer has 64 neurons, also with a ReLU activation function:

$$h_2 = \text{ReLU}(w_2 \cdot h_1 + b_2) \quad (10)$$

where, w_1 and b_1 are the weights and biases for the first dense layer, w_2 and b_2 are for the second dense layer.

Table 1. Input and output shape for each layer in the proposed SE-AG-HN

Layer	Input Shape	Filters	Output Shape
Conv 2D	(128, 128, 3)	32	(128, 128, 32)
Conv 2D	(128, 128, 32)	64	(128, 128, 64)
Maxpooling 2D	(128, 128, 64)	-	(64, 64, 64)
Conv 2D	(64, 64, 64)	128	(64, 64, 128)
Globalaveragepooling 2D	(64, 64, 128)	-	(128)
Reshape	(128)	-	(1, 1, 128)
Dense	(1, 1, 128)	-	(1, 1, 8)
Dense	(1, 1, 8)	-	(1, 1, 128)
Multiply	(1, 1, 128) x (64, 64, 128)	-	(64, 64, 128)
Conv 2D	(64, 64, 128)	256	(64, 64, 256)
Conv 2D	(64, 64, 256)	512	(64, 64, 512)
Maxpooling 2D	(64, 64, 512)	-	(32, 32, 512)
Globalaveragepooling 2D	(32, 32, 512)	-	(512)
Reshape	(512)	-	(1, 512)
LSTM	(1, 512)	-	(64)
Dense	(64)	-	(128)
Dense	(128)	-	(64)
Dense	(64)	-	(3)

Output Layer: The final dense layer has 3 neurons corresponding to the three classes, with a softmax activation function to produce the multiclass output:

$$\hat{y} = \text{softmax}(w_3 \cdot h_2 + b_3) \quad (11)$$

where, w_3 and b_3 are the weights and biases of the output layer, and \hat{y} represents the predicted probability distribution across the three classes. The model is configured for training with the Adam optimization algorithm and uses categorical cross-entropy to measure the discrepancy between predicted and actual class probabilities:

$$\text{Loss} = -\sum_{c=1}^C y_c \log(\hat{y}_c) \quad (12)$$

where, y_c is the true label and \hat{y}_c is the predicted probability for class c . Accuracy is used as the evaluation metric to measure the model's performance. Table 1 represents the input and output shape for each layer in the proposed hybrid network (SE-AG-HN).

Algorithm – SE-AG-HN ()

1. Preprocessing

Input: Pap smear image I

Output: Preprocessed image I_{pre}

1.1. Image Resizing:

Resize image I to a fixed size (128, 128):

$$I_{resized} = \text{resize}(I, (128, 128))$$

1.2. Adaptive Histogram Equalization (CLAHE):

Apply CLAHE to enhance local contrast of $I_{resized}$:

$$I_{clahe} = \text{CLAHE}(I_{resized})$$

1.3. Normalization:

Normalize pixel values of I_{clahe} to the range [0, 1]:

$$I_{normalized}(x, y) = \frac{I_{clahe}(x, y)}{255}$$

Output Image: $I_{pre} = I_{normalized}$

2. Model Architecture

2.1. Convolutional Layers:

Apply 2D convolution with filters W and bias b :

$$H_{conv} = \text{Conv2D}(I_{pre}, W, b) + b$$

Apply ReLU activation function:

$$H_{relu} = \text{ReLU}(H_{conv})$$

2.2. Squeeze-and-Excitation (SE) Block:

Squeeze: Global average pooling:

$$S = \text{GlobalAveragePooling2D}(H_{relu})$$

Excitation:

$$E_1 = \text{Dense}(S, \frac{C}{R}, \text{ReLU})$$

$$E_2 = \text{Dense}(E_1, C, \text{Sigmoid})$$

Recalibrate: Multiply attention weights E_2 with feature maps:

$$H_{SE} = H_{relu} \times E_2$$

2.3. Max-Pooling:

Apply MaxPooling to reduce spatial dimensions:

$$H_{pool} = \text{MaxPooling}(H_{SE})$$

2.4. Flatten and Reshape for RNN Processing:

Flatten the output from the last layer:

$$H_{flat} = \text{GlobalAveragePooling2D}(H_{pool})$$

Reshape for RNN input:

$$H_{reshaped} = \text{Reshape}(H_{flat}, (1, -1))$$

2.5. Recurrent Layer (LSTM):

Apply LSTM to process the sequential data:

$$H_{LSTM} = \text{LSTM}(H_{reshaped}, 64)$$

2.6. Dense Layers:

Apply Dense layers:

$$\begin{aligned} H_{dense1} &= Dense(H_{LSTM}, 128, ReLU) \\ H_{dense2} &= Dense(H_{LSTM}, 64, ReLU) \end{aligned}$$

2.7. Output Layer:

Apply final Dense layer for classification:

$$H_{output} = Dense(H_{dense2}, 3, Softmax)$$

Output: Predicted class probabilities for Cervical Cancer, Non-Cervical Cancer, and Pre-Cervical Cancer.

3. Training

3.1. Loss Function:

Use categorical cross-entropy loss to measure the discrepancy between predicted probabilities \hat{y}_c and actual labels y :

$$Loss = - \sum_{c=1}^3 y_c \log(\hat{y}_c)$$

where, y_c is the true label for class c and \hat{y}_c is the predicted probability for class c .

3.2. Optimization:

Update model parameters using Adam optimizer:

$$\theta_{new} = \theta - \eta \cdot \nabla \theta L$$

where, η is the learning rate and $\nabla \theta L$ is the gradient of the loss function with respect to the parameters θ .

3.3. Iteration:

Repeat training for a specified number of epochs or until convergence criteria are met.

4. Evaluation

4.1. Prediction on Test Data:

Predict class probabilities for test images I_{test} :

$$\hat{y}_{test} = Predict(I_{test}, \theta)$$

4.2. Performance Assessment:

Assess model performance using metrics such as accuracy, precision, recall, and F1 score.

3.4 Cervical cancer classification

Cervical Cancer Classification from pap smear images involve meticulous preprocessing pipeline followed by the application of a hybrid convolutional neural network and recurrent neural network architecture. Initially, the raw Pap smear images are resized to a uniform dimension of 128x128 pixels to standardize input size across the dataset. Subsequently, Adaptive Histogram Equalization (CLAHE) is employed to enhance the local contrast of the images, improving the visibility of critical features such as cell boundaries and morphological details. This contrast-enhanced image is then normalized by scaling pixel values to the range [0, 1], ensuring consistent input data for the neural network.

Once preprocessed, the image is passed through the hybrid network. The CNN component extracts hierarchical spatial features through a series of convolutional and pooling layers, which are further refined using a Squeeze-and-Excitation (SE) block to emphasize significant features. The output from the CNN is then flattened and reshaped into a sequence suitable

for the LSTM layer of the RNN, which captures temporal dependencies and patterns across the feature dimensions. Subsequently, the fully connected network further refines the features, culminating in a final dense layer with a softmax activation function that outputs a probability distribution across the three classes. Through this hybrid approach, the network learns to classify each image based on its learned features and patterns, distinguishing between cervical cancer, non-cervical cancer and pre-cervical cancer with high accuracy.

4. EXPERIMENTAL RESULTS AND ANALYSIS

4.1 Dataset description

The dataset comprises three distinct collections of cytology images used for classifying Pap smear and liquid-based cytology (LBC) samples. The Herlev Pap Smear Dataset contains 917 single-cell images, categorized into 7 classes. Among these, 3 classes are normal, Class 1 (intermediate squamous epithelial) with 70 images, Class 2 (columnar epithelial) with 98 images, and Class 3 (superficial squamous epithelial) with 74 images. The 4 abnormal classes include Class 4 (mild squamous non-keratinizing dysplasia) with 182 images, Class 5 (squamous cell carcinoma in situ intermediate) with 150 images, Class 6 (moderate squamous non-keratinizing dysplasia) with 146 images, and Class 7 (severe squamous non-keratinizing dysplasia) with 197 images.

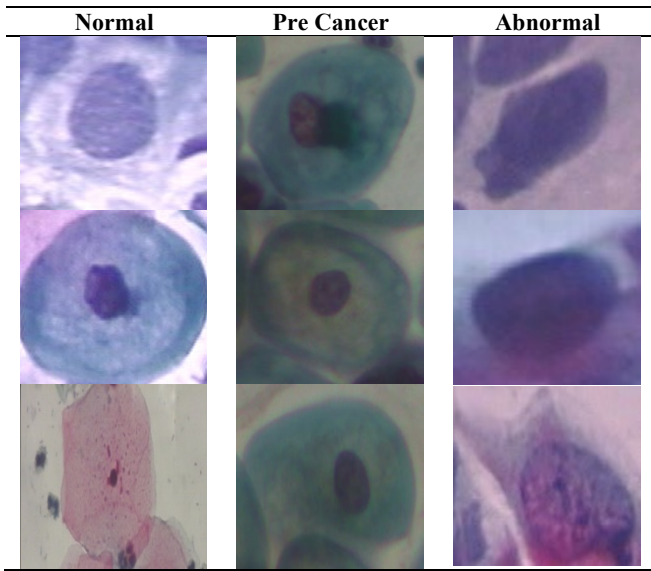
The Mendeley Liquid-Based Cytology (LBC) Dataset includes 963 whole slide images divided into 4 classes. Class 1, "negative for intraepithelial malignancy," is normal and consists of 613 images. Class 2, "low-grade squamous intraepithelial lesion (LSIL)," is abnormal with 163 images. Class 3, "high-grade squamous intraepithelial lesion (HSIL)," is abnormal and contains 113 images. Class 4, "squamous cell carcinoma (SCC)," is abnormal with 74 images. The SIPaKMeD Pap Smear Dataset features 4049 images of isolated cells extracted from 966 whole slide images, spanning 5 classes. Class 1, "superficial and intermediate," and Class 2, "parabasal," are normal with 831 and 787 images respectively. Class 3, "koilocytotic," and Class 4, "dyskeratotic," are abnormal with 825 and 813 images respectively. Finally, Class 5, "metaplastic," is categorized as benign and includes 793 images.

The combined dataset integrates images from the three distinct collections, offering a comprehensive resource for classifying Pap smear and liquid-based cytology samples into three main categories: normal, abnormal, and benign. These datasets include various image type such as single-cell images and whole slide images which is captured under different acquisition conditions. This provides an initial indication of the model's robustness. The aggregated dataset consists of 5,929 images, including 2473 non-cervical cancer images, 2663 cervical cancer images and 793 benign images, which represent non-cancerous but potentially precancerous conditions. This extensive dataset provides a robust basis for analyzing and improving classification models in cytology.

Table 2 shows sample images of three classes these datasets offer a wide range of examples for training and testing models using various types of cell images which helps in advancing diagnostic tools for early detection and classification of abnormal and normal cell types in cytology. The diversity in classes and image types supports the development of strong

models capable of handling real-world variations in cytological data.

Table 2. Sample images of the three classes



4.2 Experimental setup

The research was carried out on a workstation equipped with an Intel i5-4300U Processor clocked at 1.90 GHz, 8 GB of RAM, and a 64-bit operating system with x64-based architecture, running the Microsoft Windows 10 Pro operating system. Python implementation code was written using the Anaconda integrated development environment (IDE), with TensorFlow and Keras libraries utilized for model implementation.

4.3 Performance analysis

To ensure a thorough evaluation of the cervical cancer classification, a 7-fold cross-validation strategy is applied to the dataset of 5,929 images from three collections (Herlev Pap Smear, Mendeley Liquid-Based Cytology, and SIPaKMeD Pap Smear datasets). The dataset is split into seven folds, with each containing about 847 images. Stratified sampling is used to maintain balanced class representation across all folds. The SE-AG-HN model is trained and tested iteratively. Each fold serves as the test set once while the model is trained on the combined data of the remaining six folds totalling approximately 5,082 images. Performance metrics are assessed on the unseen fold during each iteration. This approach, involving the averaging of metrics across all folds, provides a more reliable assessment of the model's performance by accounting for the variability inherent in different data splits and offering a more robust evaluation of its ability to generalize to new, unseen data.

4.4 Result analysis

The proposed method utilizes a deep learning architecture called the Squeeze-and-Excitation Attention-Guided Hybrid Network to classify cervical cancer from cytology images. The model is designed to effectively handle the complexities of Pap smear and liquid-based cytology samples by integrating three key modules, CNN module for extracting feature maps, S&E module for recalibrating feature importance, and RNN

module for final classification. Preprocessing steps, including image resizing, contrast enhancement using Adaptive Histogram Equalization, and normalization, prepare the images for analysis. The CNN module captures spatial features, while the SE module enhances the model's focus on relevant features, and the RNN module captures sequential dependencies in the data. The model is compiled with the Adam optimizer, categorical cross entropy as the loss function, and accuracy as the evaluation metric. Training typically involves running the model for 20 epochs, though this adjusted based on performance. Learning rate provides balance between stability and convergence speed and it is set to 0.001. Table 3 shows the hyperparameters of the proposed SE-AG-HN model.

Table 3. Hyperparameters of proposed SE-AG-HN model

Number of Epochs	20
Batch size	32
Optimizer	Adam
Activation	Relu
Learning Rate	0.001
Loss	Categorical cross entropy

Table 4. Training and testing metrics of the proposed model

Epochs	Training Loss	Testing Loss	Training Accuracy	Testing Accuracy
1	0.215	0.235	0.956	0.945
2	0.176	0.215	0.969	0.952
3	0.238	0.21	0.951	0.954
4	0.245	0.268	0.95	0.938
5	0.253	0.205	0.949	0.955
6	0.221	0.168	0.953	0.96
7	0.125	0.213	0.93	0.952
8	0.182	0.275	0.968	0.934
9	0.205	0.192	0.958	0.965
10	0.174	0.203	0.965	0.956
Overall	0.2034	0.2184	0.9549	0.9511

Table 4 provides a summary of the training and testing metrics across 10 epochs for the proposed model. The training loss decreases from 0.215 in the first epoch to 0.174 in the tenth epoch, with an average of 0.2034, indicating that the model is effectively learning and reducing errors. The testing loss also shows a general decrease with some fluctuations, starting at 0.235 and ending at 0.203, averaging 0.2184, which suggests that the model is improving its ability to generalize to unseen data. Training accuracy increases from 95.6% in the first epoch to 96.5% in the tenth epoch, with an average of 95.49%, reflecting better performance on the training set over time. Similarly, testing accuracy rises from 94.5% in the first epoch to 95.6% in the tenth epoch, with an average of 95.11%, demonstrating improved performance on unseen data. The model also exhibits very high TPR at low FPR values for all classes, which indicates strong performance in identifying true positives while maintaining minimal false positives. As the FPR increases slightly ranging from 0 to 0.091 for class 0, 0.078 for class 1, and 0.038 for class 2, the TPR rises significantly, showing high sensitivity while keeping false positives low. The similarity in FPR and TPR patterns across the different classes suggests that the model performs consistently.

Performance of the proposed method is compared with three existing works, includes work of Pacal and Kılıncarslan [1], Habtemariam et al. [18] and Tanimu et al. [19]. Pacal and

Kılıcarslan [1] compared the performance of 40 CNN-based models and 20 ViT-based models on the SIPaKMeD pap-smear dataset. Data augmentation and ensemble learning were used to improve performance. They demonstrated ViT-based model and it required higher computational resources when comparing to CNN models. Habtemariam et al. [18] trained a MobileNetv2-YOLOv3 model for ROI extraction from cervix images and used pre-trained EfficientNetB0 models for cervix type and cervical cancer classification. This lightweight model

structure was advantageous for the deployment in resource constrained settings but had limitations in capturing subtle image variations due to its relatively shallow architecture. Tanimu et al. [19] employed a decision tree to predict cervical cancer outcomes, selecting important features using RFE and LASSO and addressing data imbalances with SMOTE Tomek. While this method achieves high accuracy in structured datasets, its reliance on hand-crafted features limited its adaptability to complex or unstructured image data.

Table 5. Comparative analysis of performance measures obtained by proposed work and the works of Pacal and Kılıcarslan [1], Habtemariam et al. [18] and Tanimu et al. [19]

Validation Sets	Proposed method				Pacal and Kılıcarslan [1]				Habtemariam et al. [18]				Tanimu et al. [19]			
	TPR	FPR	FNR	TNR	TPR	FPR	FNR	TNR	TPR	FPR	FNR	TNR	TPR	FPR	FNR	TNR
1	0.945	0.048	0.017	0.944	0.886	0.052	0.034	0.883	0.916	0.059	0.078	0.798	0.787	0.075	0.081	0.768
2	0.934	0.025	0.002	0.925	0.892	0.041	0.045	0.927	0.866	0.075	0.064	0.824	0.798	0.085	0.079	0.765
3	0.962	0.038	0.032	0.942	0.918	0.064	0.055	0.879	0.899	0.085	0.068	0.806	0.841	0.063	0.082	0.781
4	0.978	0.019	0.015	0.937	0.932	0.069	0.059	0.895	0.905	0.073	0.071	0.807	0.869	0.059	0.087	0.791
5	0.981	0.025	0.021	0.949	0.927	0.045	0.063	0.913	0.875	0.052	0.053	0.825	0.853	0.060	0.095	0.779
6	0.984	0.008	0.027	0.945	0.919	0.038	0.041	0.861	0.854	0.066	0.054	0.819	0.802	0.095	0.098	0.771
7	0.959	0.028	0.011	0.937	0.890	0.054	0.038	0.906	0.849	0.057	0.062	0.786	0.774	0.086	0.086	0.759
Overall	0.963	0.027	0.018	0.940	0.909	0.052	0.048	0.894	0.880	0.067	0.064	0.809	0.818	0.075	0.087	0.773

Table 6. Comparative analysis of validation sets among proposed work and existing works of Pacal and Kılıcarslan [1], Habtemariam et al. [18] and Tanimu et al. [19]

Validation Sets	Proposed method					Pacal and Kılıcarslan [1]					Habtemariam et al. [18]					Tanimu et al. [19]				
	Sensitivity	Specificity	Accuracy	Precision	F1 Score	Sensitivity	Specificity	Accuracy	Precision	F1 Score	Sensitivity	Specificity	Accuracy	Precision	F1 Score	Sensitivity	Specificity	Accuracy	Precision	F1 Score
1	0.945	0.944	0.945	0.972	0.958	0.957	0.946	0.889	0.906	0.946	0.856	0.761	0.857	0.902	0.873	0.789	0.755	0.876	0.842	0.748
2	0.934	0.925	0.929	0.970	0.952	0.936	0.912	0.892	0.892	0.938	0.882	0.740	0.862	0.854	0.884	0.749	0.721	0.856	0.851	0.782
3	0.962	0.942	0.952	0.959	0.961	0.896	0.897	0.946	0.901	0.933	0.892	0.798	0.891	0.849	0.908	0.812	0.644	0.791	0.802	0.849
4	0.978	0.937	0.958	0.934	0.956	0.928	0.861	0.918	0.926	0.928	0.887	0.892	0.945	0.824	0.914	0.795	0.679	0.768	0.846	0.828
5	0.981	0.949	0.965	0.930	0.955	0.883	0.923	0.867	0.877	0.908	0.919	0.876	0.856	0.878	0.902	0.842	0.713	0.881	0.799	0.816
6	0.975	0.933	0.944	0.962	0.968	0.978	0.886	0.901	0.923	0.927	0.911	0.857	0.932	0.861	0.884	0.837	0.694	0.777	0.846	0.799
7	0.966	0.924	0.969	0.944	0.945	0.961	0.908	0.899	0.905	0.918	0.882	0.864	0.911	0.841	0.862	0.806	0.782	0.798	0.789	0.808
Overall	0.963	0.936	0.952	0.953	0.956	0.934	0.905	0.902	0.904	0.923	0.890	0.827	0.893	0.858	0.890	0.804	0.713	0.821	0.825	0.804

Table 7. Comparative analysis of different categories among proposed work and existing works of Pacal and Kılıcarslan [1], Habtemariam et al. [18] and Tanimu et al. [19]

Classes	Proposed method					Pacal and Kılıcarslan [1]					Habtemariam et al. [18]					Tanimu et al. [19]				
	Sensitivity	Specificity	Accuracy	Precision	F1 Score	Sensitivity	Specificity	Accuracy	Precision	F1 Score	Sensitivity	Specificity	Accuracy	Precision	F1 Score	Sensitivity	Specificity	Accuracy	Precision	F1 Score
Normal	0.969	0.936	0.946	0.958	0.954	0.929	0.906	0.911	0.875	0.916	0.878	0.838	0.886	0.846	0.910	0.794	0.691	0.786	0.795	0.806
Abnormal	0.964	0.925	0.958	0.947	0.951	0.938	0.919	0.893	0.914	0.945	0.882	0.827	0.903	0.869	0.876	0.801	0.709	0.846	0.841	0.789
Pre-cancer	0.957	0.948	0.951	0.955	0.963	0.935	0.889	0.901	0.924	0.924	0.911	0.815	0.890	0.859	0.883	0.818	0.738	0.813	0.839	0.817
Overall	0.963	0.936	0.952	0.953	0.956	0.934	0.905	0.902	0.904	0.923	0.890	0.827	0.893	0.858	0.890	0.804	0.713	0.821	0.825	0.804

In contrast, proposed SE-AG-HN model integrates a hybrid CNN-RNN architecture with a Squeeze-and-Excitation (SE) block, enabling it to capture both spatial and sequential dependencies while adaptively focusing on critical image features. This capability provides a distinct advantage in handling subtle abnormalities and complex image variability inherent in cervical cancer datasets. The SE Block in our proposed SE-AG-HN model improved interpretability by automatically identifying and recalibrating the importance of

image features. By emphasizing critical features such as cell boundaries, tissue textures, or morphological structures. SE block allows the model to focus on the most informative regions in the image. This process provides insight into which features most significantly impact the model's classification decisions, thereby enhancing transparency and offering a clearer understanding of how the model arrives at its conclusions.

Table 5 shows the comparative analysis performance

measures obtained by the proposed work and the works of Pacal and Kılıcarslan [1], Habtemariam et al. [18] and Tanimu et al. [19], across five validation sets using key metrics such as True Positive Rate (TPR), False Positive Rate (FPR), False Negative Rate (FNR), and True Negative Rate (TNR). These results suggest that the proposed method offers a competitive advantage in classification accuracy and reliability, particularly in achieving higher TPR and TNR values, making it a robust choice overall compared to the other methods.

Table 6 shows the comparative analysis of validation sets among the proposed work and existing works of Pacal and Kılıcarslan [1], Habtemariam et al. [18] and Tanimu et al. [19]. The metrics assessed are Sensitivity, Specificity, Accuracy, Precision, and F1 Score. The proposed method consistently achieves high values in these metrics, with an overall accuracy of 94.98%, sensitivity of 96%, specificity of 93.94%, precision of 95.3%, and an F1 score of 95.64%. Table 7 shows the comparative analysis of different categories among the proposed work and existing works of Pacal and Kılıcarslan [1], Habtemariam et al. [18] and Tanimu et al. [19].

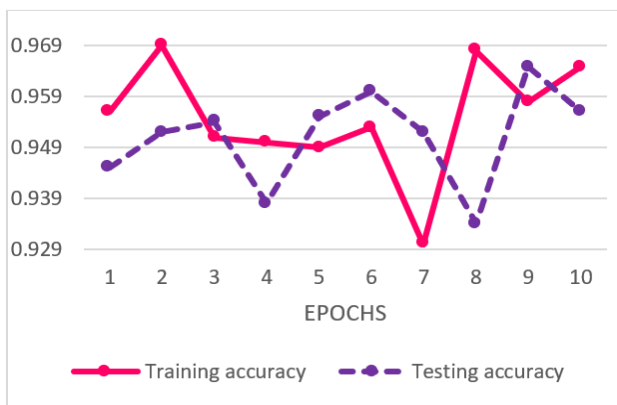


Figure 4. Graphical representation of training and testing accuracy of the proposed model

Figure 4 shows the training and testing accuracy over epochs. The pink line represents training accuracy, starting at 0.956, peaking at 0.969 in the second epoch, and ending at 0.965 in the tenth epoch. The violet line represents testing accuracy, starting lower at 0.945, dipping to 0.938 around the fourth epoch, and peaking at 0.965 in the ninth epoch. Both accuracies fluctuate but remain high, with the second and ninth epochs showing the highest values.

Figure 5 shows graphical representation of the training and testing loss of the proposed model over epochs. The blue line represents the training loss, which decreases from 0.215 in the first epoch to 0.125 by the seventh epoch and then slightly increases to 0.174 by the tenth epoch. The red line depicts the testing loss, starting at 0.235, reaching a minimum of 0.168 by the sixth epoch, peaking at 0.275 in the eighth epoch, and ending at 0.203 by the tenth epoch. Both losses generally decrease over time, with the testing loss showing more fluctuations.

Based on the on the specificity and sensitivity values shown in Figure 6, the proposed method consistently achieves the highest values. It also achieves a remarkable 95.2% accuracy compared with other methods indicating strong overall performance. Figure 7 shows the graphical representation of TPR, FPR, FNR, and TNR obtained by the proposed work and the works of Pacal and Kılıcarslan [1], Habtemariam et al. [18] and Tanimu et al. [19].

and Tanimu et al. [19]. Figure 8 shows the graphical representation of performance measures like sensitivity, specificity, accuracy, precision and f1 score obtained in different categories by the proposed work and the works of Pacal and Kılıcarslan [1], Habtemariam et al. [18] and Tanimu et al. [19].

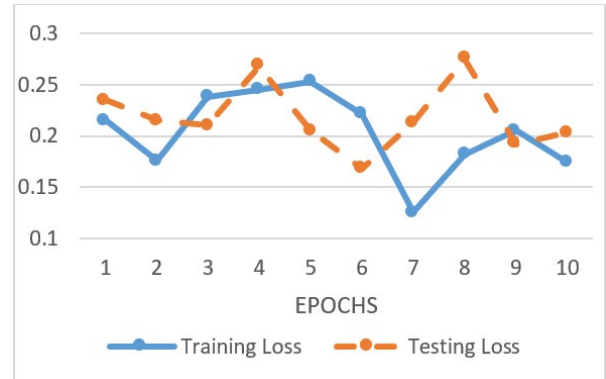


Figure 5. Graphical representation of training and testing loss of the proposed model

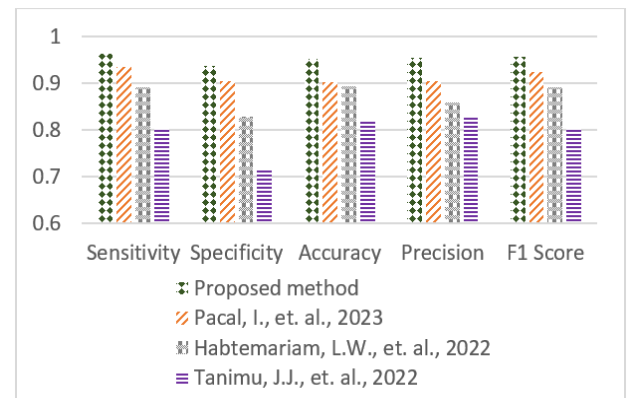


Figure 6. Graphical representation of performance obtained by the proposed work and the works of Pacal and Kılıcarslan [1], Habtemariam et al. [18] and Tanimu et al. [19]

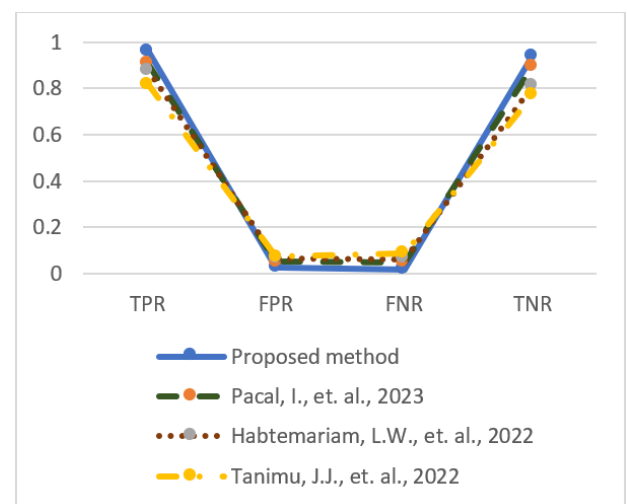


Figure 7. Graphical representation of TPR, FPR, FNR, TNR obtained by the proposed work and the works of Pacal and Kılıcarslan [1], Habtemariam et al. [18] and Tanimu et al. [19]

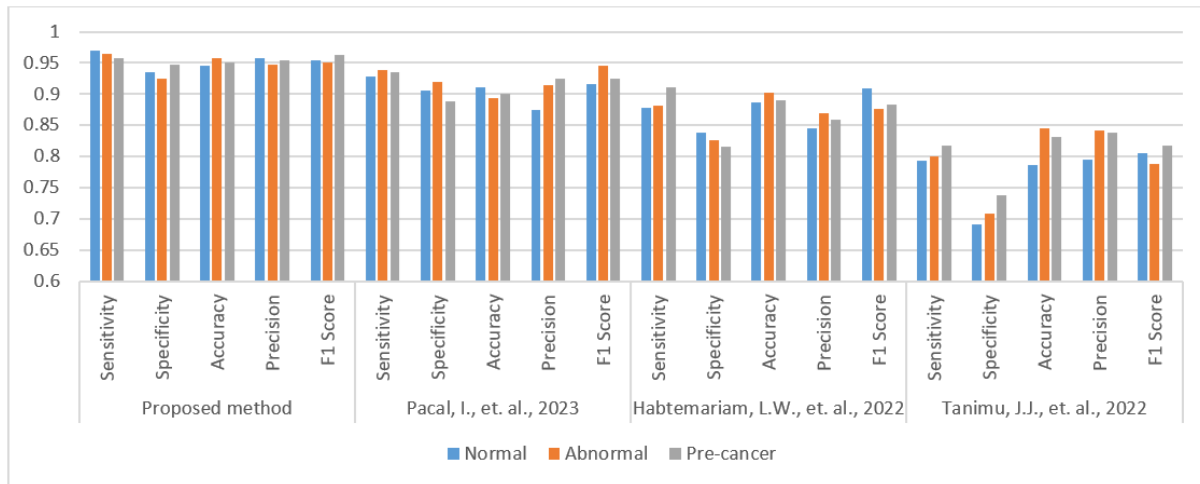


Figure 8. Graphical representation of performance obtained in different categories by the proposed work and the works of Pacal and Kılıcarslan [1], Habtemariam et al. [18] and Tanimu et al. [19]

The proposed method outperforms other across all key performance metrics. It achieves the highest TPR of 0.963, indicating superior effectiveness in identifying positive cases. Additionally, it has the lowest FPR of 0.027 and FNR of 0.018, demonstrating fewer errors in classifying cases. The method also boasts a higher TNR of 0.94, reflecting better performance in correctly identifying negative cases. Overall, these metrics highlight the proposed method's robustness and reliability, making it a more effective approach for accurate classification.

5. CONCLUSION

The proposed method, a Squeeze-and-Excitation Attention-Guided Hybrid Network effectively integrates a SE attention mechanism with a hybrid CNN and RNN architecture to enhance cervical cancer classification from Pap smear images. The SE attention mechanism recalibrates the importance of feature channels, allowing the model to focus on the most critical parts of the image, while the hybrid CNN-RNN architecture captures both spatial and sequential dependencies. This approach was applied to a Herlev Pap Smear Dataset, Mendeley Liquid-Based Cytology Dataset and SIPaKMeD Pap Smear Dataset, where it successfully identified patterns and features essential for distinguishing between cervical cancer, pre-cervical cancer, and non-cervical cancer.

Compared to traditional methods, which often rely on either CNN or RNN architectures alone, this hybrid approach influences the strengths of both. The CNN component excels at extracting local features, while the RNN captures temporal and sequential relationships, providing a comprehensive understanding of the data. The incorporation of the SE attention mechanism further enhances the model's ability to identify subtle yet crucial features, improving its classification accuracy. The experimental results demonstrated the effectiveness of this proposed method, with consistent improvements, showing a strong ability to correctly classify cervical cancer cases when compared to other methods. This would benefit the physicians to double their diagnostic accuracy.

The proposed SE-AG-HN model's performance depends significantly on high-quality input images, as noise, blur or poor contrast can adversely affect the classification accuracy.

Limited or homogeneous datasets may lead to overfitting and reduce the model's generalizability. While the model has been validated using Pap smear images, additional modifications may be needed to adapt it to other medical imaging modalities. Future work will focus on enhancing pre-processing techniques, increasing dataset diversity, integrating multimodal data, exploring advanced attention mechanisms, and optimizing the model's architecture for improved efficiency, broader applicability, and better generalization using transfer and few-shot learning strategies.

ACKNOWLEDGMENT

This study is supported via funding from Prince Sattam bin Abdulaziz University (Grant No.: PSAU/2024/R/1446).

REFERENCES

- [1] Pacal, I., Kılıcarslan, S. (2023). Deep learning-based approaches for robust classification of cervical cancer. *Neural Computing and Applications*, 35(25): 18813-18828. <https://doi.org/10.1007/s00521-023-08757-w>
- [2] Kalbhor, M.M., Shinde, S.V. (2023). Cervical cancer diagnosis using convolution neural network: Feature learning and transfer learning approaches. *Soft Computing*, 1-11. <https://doi.org/10.1007/s00500-023-08969-1>
- [3] Kumari, C.M., Bhavani, R., Padmashree, S., Priya, R. (2024). Automated cervical cancer classification using deep neural network classifier. *International Journal of Modeling, Simulation, and Scientific Computing*, 15(1): 2450008. <https://doi.org/10.1142/S1793962324500089>
- [4] Youneszade, N., Marjani, M., Shafiq, D.A. (2024). Exploring the impact of increasing the number of classes on the performance of cervical cancer detection models using deep learning and colposcopy. *Journal of Engineering Science and Technology*, 19(2): 629-647.
- [5] Cheng, C., Yang, Y., Qu, Y. (2024). Exploration of cervical cancer image processing technology based on deep learning. In *International Conference on Image, Signal Processing, and Pattern Recognition (ISPP 2024)*. SPIE, 13180: 255-263.

- <https://doi.org/10.1117/12.3033802>
- [6] Talpur, D.B., Raza, A., Khowaja, A., Shah, A. (2024). DeepCervixNet: An advanced deep learning approach for cervical cancer classification in pap smear images. *VAWKUM Transactions on Computer Sciences*, 12(1): 136-148. <https://doi.org/10.21015/vtcs.v12i1.1812>
- [7] Bueno-Crespo, A., Martínez-España, R., Morales-García, J., Ortíz-González, A., Imbernón, B., Martínez-Más, J., Rosique-Egea, D., Álvarez, M.A. (2024). Diagnosis of cervical cancer using a deep learning explainable fusion model. In *International Work-Conference on the Interplay Between Natural and Artificial Computation*. Springer: Cham, pp. 451-460. https://doi.org/10.1007/978-3-031-61137-7_42
- [8] Devaraj, S., Madian, N., Menagadevi, M., Remya, R. (2024). Deep learning approaches for analysing papsmear images to detect cervical cancer. *Wireless Personal Communications*, 1-18. <https://doi.org/10.1007/s11277-024-10986-8>
- [9] Mathivanan, S.K., Francis, D., Srinivasan, S., Khatavkar, V., P, K., Shah, M.A. (2024). Enhancing cervical cancer detection and robust classification through a fusion of deep learning models. *Scientific Reports*, 14(1): 10812. <https://doi.org/10.1038/s41598-024-61063-w>
- [10] Tan, S.L., Selvachandran, G., Ding, W., Paramesran, R., Kotecha, K. (2024). Cervical cancer classification from pap smear images using deep convolutional neural network models. *Interdisciplinary Sciences: Computational Life Sciences*, 16(1): 16-38. <https://doi.org/10.1007/s12539-023-00589-5>
- [11] Tripathi, A., Arora, A., Bhan, A. (2021). Classification of cervical cancer using Deep Learning Algorithm. In *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*, Madurai, India, pp. 1210-1218. <https://doi.org/10.1109/ICICCS51141.2021.9432382>
- [12] Deo, B.S., Pal, M., Panigrahi, P.K., Pradhan, A. (2024). CerviFormer: A pap smear-based cervical cancer classification method using cross-attention and latent transformer. *International Journal of Imaging Systems and Technology*, 34(2): e23043. <https://doi.org/10.1002/ima.23043>
- [13] Jeysari, J., Kowsigan, M. (2024). Multi-stage attention-based long short-term memory networks for cervical cancer segmentation and severity classification. *Iranian Journal of Science and Technology, Transactions of Electrical Engineering*, 48(1): 445-470. <https://doi.org/10.1007/s40998-023-00664-z>
- [14] Ganguly, T., Singh, R.P., Kumar, P. (2023). Self-attention based resnet model for cervical cancer detection. In *2023 Second International Conference on Informatics (ICI)*, Noida, India, pp. 1-6. <https://doi.org/10.1109/ICI60088.2023.10421309>
- [15] Xia, M., Zhang, G., Mu, C., Guan, B., Wang, M. (2020). Cervical cancer cell detection based on deep convolutional neural network. In *2020 39th Chinese Control Conference (CCC)*, Shenyang, China, pp. 6527-6532. <https://doi.org/10.23919/CCC50068.2020.9188454>
- [16] Ghoneim, A., Muhammad, G., Hossain, M.S. (2020). Cervical cancer classification using convolutional neural networks and extreme learning machines. *Future Generation Computer Systems*, 102: 643-649. <https://doi.org/10.1016/j.future.2019.09.015>
- [17] Fan, Z.Z., Wu, X.C., Li, C.Z., Chen, H.Y., Liu, W.L., Zheng, Y.C., Chen, J., Li, X.Y., Sun, H.Z., Jiang, T., Grzegorzec, M., Li, C. (2023). CAM-VT: A weakly supervised cervical cancer nest image identification approach using conjugated attention mechanism and visual transformer. *Computers in Biology and Medicine*, 162: 107070. <https://doi.org/10.1016/j.compbiomed.2023.107070>
- [18] Habtemariam, L.W., Zewde, E.T., Simegn, G.L. (2022). Cervix type and cervical cancer classification system using deep learning techniques. *Medical Devices: Evidence and Research*, 15: 163-176. <https://doi.org/10.2147/mder.s366303>
- [19] Tanimu, J.J., Hamada, M., Hassan, M., Kakudi, H., Abiodun, J.O. (2022). A machine learning method for classification of cervical cancer. *Electronics*, 11(3): 463. <https://doi.org/10.3390/electronics11030463>
- [20] Feng, T., Ying, J., Yang, H., Li, F., Li, H. (2022). Regional detection of cervical lesions based on self-attention mechanism and multi-scale feature enhancement. In *Proceedings of Chinese Intelligent Systems Conference*. Springer: Singapore, pp. 182-190. https://doi.org/10.1007/978-981-19-6203-5_18
- [21] Yi, J.X., Liu, X.L., Cheng, S.H., Chen, L., Zeng, S.Q. (2024). Multi-scale window transformer for cervical cytopathology image recognition. *Computational and Structural Biotechnology Journal*, 24: 314-321. <https://doi.org/10.1016/j.csbj.2024.04.028>