Journal of Artificial Intelligence and Computational Technology
Journal homepage: https://ojs.omgfzc.com/index.php/JAICT

Adaptive Multi-Scale Feature Extraction for Cervical Cancer Classification Using Dynamic Hierarchical Pooling

Abdalla Ibrahim Abdalla Musa[1,2] , Mahir Mohammed Sharif Adam[2,3]*

[1]Department of Computer Science, College of Computer, Qassim University, Buraydah, Saudi Arabia
[2]Computer Science Dept., Faculty of Computer Science and Information Technology, Omdurman Islamic University, Omdurman, Sudan,
[3]Computer and Self Development,Common First Year Unit,Prince Sattam bin Abdalziz University, Alkharj, Saudi Arabia

Emails: ab.musa@qu.edu.sa; mahiralsharif@oiu.edu.sd

Article Info

Article History:
Submitted/Received – 17- Dec, 2024
Revised in revised format –25- Jan - 2025
Accepted – 04-Feb-2025
Available -04-Feb-2025
Publication Date -01-Apr-2025

Keyword:
Cervical Cancer, Pap Smear, CNN, RNN, Dynamic Hierarchical Pooling, Hybrid Network, Feature Extraction

Cite this article:

ABSTRACT

Cervical cancer remains a leading cause of mortality among women worldwide, especially in low-resource settings where access to early screening and treatment is limited. Early detection through accurate and efficient diagnostic methods is critical for improving patient outcomes. This study proposes a novel method for classifying cervical cancer using Dynamic Hierarchical Pooling (DHP). To effectively capture multi-scale characteristics, DHP adaptively modifies the number of pyramid levels and pooling types according to the size of the input image. To be more precise, the module dynamically divides the feature maps into different spatial regions and applies various pooling operations to each region. This adaptive mechanism extracts fine-grained and coarse-grained features crucial for recognizing diverse patterns in cervical pap smear images. To facilitate efficient processing, feature maps are resized to a common size, regardless of the original image size. The Squeeze-and-Excitation (SE) attention module further enhances feature discrimination by dynamically updating attention weights, focusing on the most informative regions of the feature maps. Combining the strengths of Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), a hybrid architecture is employed to leverage local and global contextual information. Experimental results demonstrate the superior performance of the proposed method compared to state-of-the-art techniques, highlighting its potential for improving cervical cancer diagnosis.

1. INTRODUCTION

Cervical cancer is a major global public health issue, ranking as the fourth most common cancer among women and a leading cause of cancer-related deaths, especially in low- and middle-income countries (LMICs) (Ganguly, T., et. al., 2023). The World Health Organization (WHO) states that over 90% of cervical cancer deaths occur in LMICs due to limited access to early screening, timely diagnosis, and effective treatment options. Early detection of cervical cancer can significantly improve survival rates by allowing for prompt intervention and treatment during the precancerous or early stages of the disease. However, traditional diagnostic methods such as visual inspection with acetic acid (VIA), Pap smear testing, and HPV DNA testing are often difficult to access or too expensive in resource-limited settings (Tan, X., et. al., 2021, Xiang, Y., et. al., 2020). These challenges highlight the urgent need for automated, affordable, and scalable diagnostic solutions to address the increasing burden of cervical cancer worldwide. Advancements in medical imaging and artificial intelligence have opened up new possibilities for improving cervical cancer diagnosis. Deep learning techniques, especially CNNs have shown great potential in analyzing medical images because they can automatically extract and learn features from complex datasets (Feng, T., et. al., 2022, Li, Y.X., et. al., 2023). However, traditional CNN-based methods often struggle to capture important details and relationships in diverse datasets like cervical Pap smear images. Additionally, these methods may not perform well in real-world situations where variations in image resolution, quality, and staining protocols are common (Ghoneim, A., et. al., 2020).

To overcome these challenges, this study proposes a new framework for classifying cervical cancer using DHP. The DHP module is designed to adaptively capture both fine-grained and coarse-grained features by dynamically adjusting pyramid levels and pooling operations based on the size of the input image. Unlike fixed pooling methods, DHP divides feature maps into spatial regions and applies customized pooling strategies, allowing it to effectively handle images with different resolutions and sizes. This flexibility ensures that important diagnostic features such as cellular morphology and tissue organization are accurately preserved. This framework uses CNNs to extract spatial features and RNNs to capture temporal and sequential dependencies across regions of interest. By combining the strengths of both models, the hybrid CNN-RNN architecture enables the model to leverage local and global contextual information for improved classification performance. This architecture is designed to overcome limitations in traditional methods and improve diagnostic reliability in diverse and complex datasets.

By addressing the challenges of feature extraction, image variability, and contextual understanding, this study aims to develop a robust and adaptive diagnostic framework for cervical cancer classification, offering a scalable solution to enhance early detection and improve patient outcomes.

DOI: https://doi.org/ 10.70274/jaict.2025.2.1.40.
p- ISSN 34562-3478 e- ISSN 5443-1243

2. LITERATURE SURVEY

(Pacal, I., 2024) introduced MaxCerVixT, a lightweight Vision Transformer model tailored for cervical cancer detection in pap-smear images. By replacing MaxViT MBConv blocks with ConvNeXtv2 blocks, the model achieved 99.02% accuracy on SIPaKMeD and 99.48% on LBC datasets. Khan, A., et. al., 2023 developed CervixFormer, an advanced multi-scale Swin Transformer model for cervical cancer classification on whole-slide images (WSI). Incorporated a self-attention GAN for handling class imbalance. The framework included multi-scale ensemble learning for detailed cell identification, even on atypical squamous cells (ASC/ASCUS). (Yang, C., et. al., 2024) proposed SWINUNECCT, a novel model combining a bidirectional hash-based agent transformer with a multi-task learning approach for cervical cancer MRI segmentation and classification. SwinUNeCCt demonstrated superior performance on multiple metrics (HD, IoU, DSC) compared to existing 3D medical models, achieving the best results while balancing computational efficiency and model complexity. (Deo, B.S., et. al., 2024) proposed CerviFormer, a novel cross-attention-based Transformer model for accurate classification of cervical cancer in Pap smear images. This model effectively addressed the challenges posed by varying image sizes and complex patterns in Pap smears. By leveraging the power of Transformers, CerviFormer achieved impressive results on two publicly available datasets, Sipakmed and Herlev. (Abinaya, K., et. al., 2024) proposed a 3D-ViT-SE-KELM model, a novel deep-learning architecture designed for accurate cervical cancer classification. This model, which we have named 3D-ViT-SE-KELM for easy reference, combined 3D CNNs and ViTs to extract comprehensive features, with FPN integrating multi-level features. The 3D SE block refined feature importance, and KELM classified the input images.

Kang, (J., et. al., 2024) proposed CerviSegNet-DistillPlus, a deep-learning framework using knowledge distillation and model pruning to enhance cervical cancer cell detection and segmentation. Tested on the Cx22, DTU/HERLEV, and SIPaKMeD datasets, the model achieved top performance across all datasets. Mishra, (A.K., et. al., 2024) introduced automated Cervical Precancerous Lesion Classification using Quantum Invasive Weed Optimization with Deep Learning (CPLC-QIWODL) on biomedical pap smear images. The method utilized Gabor filtering for image preprocessing, SqueezeNet for feature extraction, and a Deep Variational Autoencoder for classification. (Yi, J., et. al., 2024) proposed Multi-scale Window Transformer (MWT) to improve cervical cytopathology image recognition 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. (Ding, B., et. al., 2024) proposed a method that leverages image and textual diagnostic knowledge from The Bethesda System (TBS) to improve abnormal cervical cell detection. A TBS Diagnostic Knowledge-based BERT (TDK-Bert) module extracted textual features from TBS and

DOI: https://doi.org/ 10.70274/jaict.2025.2.1.40.
p- ISSN 34562-3478 e- ISSN 5443-1243

aligned them with image features through an attention mechanism. Additionally, an Omnidirectional Dynamic Convolution-based Spatial Pyramid Pooling-Fast (ODC-SPPF) module improved the SPPF module for better feature extraction, especially for subtle cells and complex backgrounds. (Ganguly, T., et. al., 2023) proposed a self-attention-based ResNet model for cervical cancer detection using ICAR-WHO dataset images. Addressing dataset and labeling challenges, the model applied ResNet50 with CNNs and clustering via pseudo-labeling, enhancing detection accuracy.

(Feng, T., et. al., 2022) proposed the CT-YOLOv5 model to enhance cervical lesion detection by improving the YOLOv5s algorithm with transformers and a Convolutional Block Attention Module (CBAM). It used PANet and CBAM for refined feature extraction, the model achieved precision, recall, and mAP scores of 93.97%, 92.94%, and 92.8%, respectively. (Liu, Y., et. al., 2024) proposed an ensemble network combining three CNN architectures, DenseNet-169, VGG-19, and Xception with a SWIN transformer to improve cervical cytology image classification. The framework was tested on the SIPaKMeD and Mendeley LBC datasets, achieving 95.50% accuracy on SIPaKMeD and 98.65% on Mendeley LBC. (Liu, W., et. al., 2022) proposed CVM-Cervix, a hybrid framework that leveraged CNN and Visual Transformers (ViT) to extract both local and global features from cervical Pap smear images. These extracted features are then fused and processed by a Multilayer Perceptron (MLP) for accurate classification of cervical cells. (Chen, W., et. al., 2022) proposed a lightweight CNN architecture for efficient cervical cell classification. Using knowledge distillation, lightweight models like Xception, MobileNet, and MobileNetV2 achieved enhanced accuracy on the Herlev Pap smear dataset for 7-class classification tasks. The best-performing Xception model reached 72.25% accuracy, comparable to the more complex Inception-ResNetV2, but with only 40% of its size. (Fadlallah et al., 2024) presented a machine learning model organized into three key phases for data elicitation and preprocessing. The initial phase involves stakeholders and developers working together to assign initial weights to prioritization criteria. The second phase employs parallel ranking to enhance requirements prioritization, thereby shortening software implementation time through concurrent processes. The final phase applies a classification method using the Random Forest-driven MoSCoW Method (RF-MM).

(Talpur, D.B., et. al., 2024) proposed DeepCervixNet, an advanced deep learning model that leverages enhanced ResNet and DenseNet architectures with SE blocks and ensemble learning. By utilizing the Harlev dataset and applying image enhancement techniques, DeepCervixNet achieved remarkable accuracy in classifying cervical cells. In their study, (E.H.E. Yasin et al., 2024) utilized QGIS version 3.16.10 and ArcMap version 10.7 for image pre-processing. They began by downloading bands from four image scenes and saving them as individual .tiff files. These bands—specifically bands 2, 3, 4, and 8—were then merged sequentially using virtual raster creation, generating a false-color composite for visualization. A subset of this virtual raster was extracted and trimmed to match the full

DOI: https://doi.org/ 10.70274/jaict.2025.2.1.40.
p- ISSN 34562-3478 e- ISSN 5443-1243

extent of the study area, forming the basis for a training dataset for image classification. The process started with unsupervised classification to differentiate between various crop types in the study area, which helped reduce mixed pixels between classes and enhance classification accuracy. This was followed by supervised classification, where training signatures for predefined crop classes were created using polygons marked as Regions of Interest (ROIs).

The aim of this paper is to develop a robust and adaptive deep learning framework, Dynamic Hierarchical Pooling with Squeeze-and-Excitation Attention-guided Hybrid Network (DHP-SE-HN), to accurately classify cervical cancer from Pap smear images. The proposed method seeks to address limitations in existing diagnostic approaches by effectively capturing multi-scale features, enhancing feature discrimination, and leveraging local and global contextual information. The ultimate goal is to improve diagnostic accuracy and reliability for cervical cancer screening, particularly in resource-constrained clinical settings, thereby contributing to early detection and better patient outcomes. The key contributions are

- Adaptively adjusts pyramid levels and pooling operations based on the input image size, enabling the extraction of both fine-grained and coarse-grained features.
- Channel-wise attention mechanism to recalibrate feature importance dynamically, enhancing the focus on significant regions in Pap smear images.
- Provides a scalable and adaptive framework suitable for deployment in low-resource clinical environments, where traditional screening methods are limited.

This paper is organized into 6 sections. Section 1 introduces the research background, highlighting the limitations of traditional diagnostic methods and the need for scalable, automated solutions to address the challenges in cervical cancer diagnosis. Section 2 reviews related work in cervical cancer diagnosis and medical image classification. Section 3 provides a detailed explanation of the proposed methodology, including the DHP module, SE attention mechanism, and the hybrid CNN-RNN architecture. Section 4 outlines the experimental setup and presents the results, followed by an in-depth discussion. Finally, Section 5 concludes the paper and suggests potential directions for future research.

3. METHODOLOGY

The Dynamic Hierarchical Pooling with Squeeze-and-Excitation Attention-guided Hybrid Network (DHP-SE-HN) method introduces an innovative approach for cervical cancer classification from Pap smear images. This methodology begins with image preprocessing, where images are resized to a standardized resolution to ensure uniform input dimensions. The core of the model is the Dynamic Hierarchical Pooling (DHP) module, which adaptively adjusts the number of pyramid levels and pooling types based on the input image size. DHP divides the feature maps into multiple spatial regions and applies different pooling operations at each level, capturing both fine-grained and coarse-grained

DOI: https://doi.org/ 10.70274/jaict.2025.2.1.40.
p- ISSN 34562-3478 e- ISSN 5443-1243

features that are essential for recognizing subtle abnormalities in medical images. Next, the Squeeze-and-Excitation (SE) attention mechanism is applied to enhance the feature discrimination by recalibrating the importance of spatial regions, allowing the model to focus on the most informative areas of the image. The method then employs a hybrid architecture that combines Convolutional Neural Networks (CNNs) to extract rich local features and Recurrent Neural Networks (RNNs) to capture global context and sequential dependencies. This hybrid structure enables the model to effectively understand both local details and broader image contexts. The model is trained using a binary cross-entropy loss function, and its performance is evaluated on a dataset of Pap smear images. The DHP-SE-HN method demonstrates superior classification accuracy and robustness compared to existing techniques, providing a powerful tool for automated cervical cancer detection and enhancing diagnostic accuracy in clinical settings. Figure 1 illustrates the architecture of the Dynamic Hierarchical Pooling with Squeeze-and-Excitation Attention-guided Hybrid Network

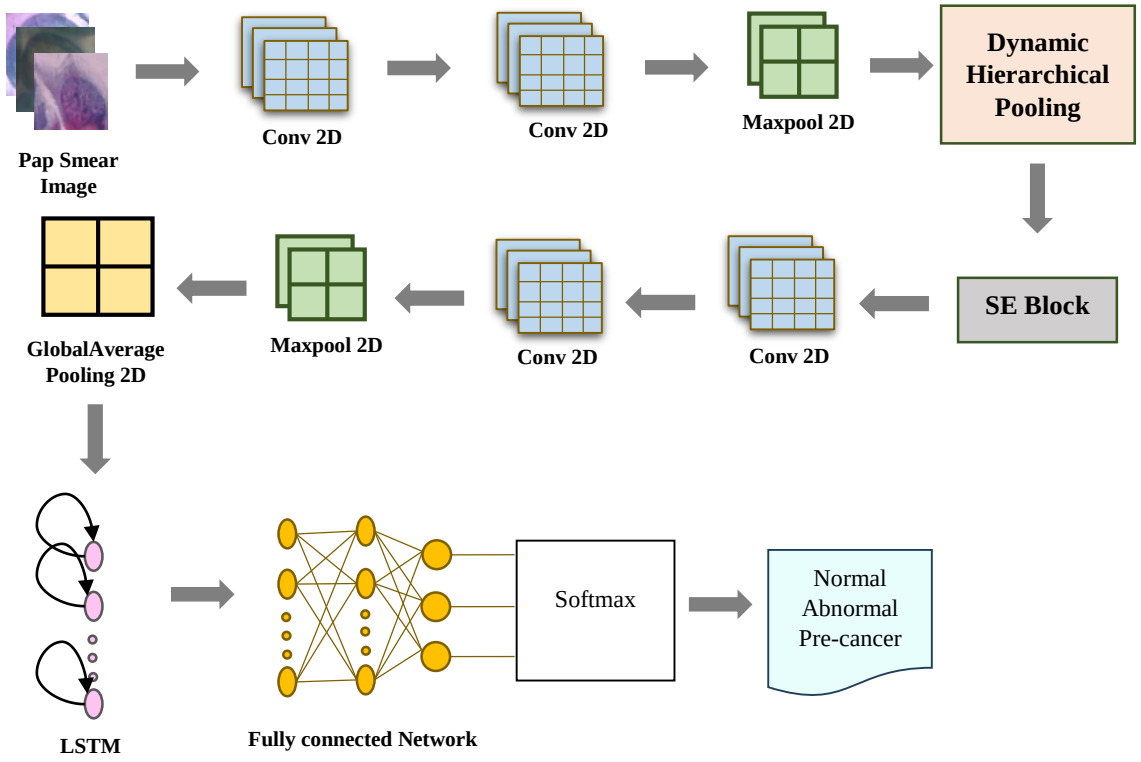


Figure 1: Architecture of the Dynamic Hierarchical Pooling with Squeeze-and-Excitation Attention-guided Hybrid Network

3.1 Preprocessing

3.1.1 Image Resizing

All Pap smear images are resized to a uniform dimension of 128x128 pixels to maintain dataset consistency and meet the neural network's input requirements. This standardization ensures efficient processing of each image by the model while preserving essential features necessary for accurate classification.

DOI: https://doi.org/ 10.70274/jaict.2025.2.1.40.
p- ISSN 34562-3478 e- ISSN 5443-1243

3.1.2 Contrast Enhancement (CLAHE)

Contrast Enhancement (CLAHE) is applied to resized Pap smear images to enhance contrast while preserving important details, particularly in regions with varying intensity distributions. Unlike traditional histogram equalization, CLAHE operates on localized image regions, dividing the image into small tiles and equalizing each independently. This localized approach prevents over-amplification of noise and ensures that subtle structures, such as cellular boundaries and nuclei, remain distinguishable. After equalizing the tiles, CLAHE blends the regions seamlessly to avoid artificial boundaries, producing an image with improved visibility of critical features. By emphasizing fine details in the resized 128x128 images, CLAHE enhances the model's ability to differentiate between normal and abnormal cells, ultimately improving classification accuracy.

3.1.3 Normalization

After enhancing the contrast, the pixel values of the images are normalized to a standard range, usually [0, 1]. This normalization is achieved by dividing each pixel value by 255, the maximum possible value in an 8-bit grayscale image. Normalizing in this way ensures that the neural network receives inputs with a consistent scale, helping to stabilize the training process, promote faster convergence, and ultimately enhance model performance.

3.2 Dynamic Hierarchical Pooling

Dynamic Hierarchical Pooling (DHP) is a pooling technique designed to capture multi-scale features by dynamically adjusting the number of pooling levels and operations based on the input feature map's size. Given an input feature map X with spatial dimensions H x W and C channels, DHP adaptively sets the number of pyramid levels L as a function of the smaller spatial dimension, min (H, W). Typically, L can be defined as:

$$L = min(max\_levels, log_2(min(H, W)) - 1) \qquad (1)$$

where max_levels is a predefined maximum number of levels, ensuring that pyramid levels remain feasible even for small inputs.

At each level l (where l = 0, 1, …, L-1), the feature map is divided into $2^l \times 2^l$ grid regions. Each region $R_{i,j}^l$ in this grid, where i, j $\epsilon \{0, 1, \dots, 2^l - 1\}$, has spatial dimensions:

$$\left(\frac{H}{2^l}, \frac{W}{2^l}\right) \qquad (2)$$

Within each region $R_{i,j}^l$, the pooling operation is adaptively chosen based on the variance of the feature values within that region. The variance $\sigma^2(R_{i,j}^l)$ is computed as:

$$\sigma^2(R_{i,j}^l) = \frac{1}{|R_{i,j}^l|} \sum_{(x,y)\in R_{i,j}^l} (F(x,y) - \mu_{R_{i,j}^l})^2 \qquad (3)$$

DOI: https://doi.org/ 10.70274/jaict.2025.2.1.40.
p- ISSN 34562-3478 e- ISSN 5443-1243

where $\mu_{R_{i,j}^l}$ is the mean value of the region. If the variance is high, Max Pooling is applied to emphasize dominant features otherwise, Average Pooling is used to aggregate smoother patterns. The pooled feature for a region is given by:

$$P_{i,j}^l = f_{pool}\left(R_{i,j}^l\right) \quad (4)$$

where f_{pool} is the pooling function selected based on $\sigma^2\left(R_{i,j}^l\right)$. The resulting pooled features from all regions at each level are then concatenated across levels to form a comprehensive feature representation P, that captures multi-scale information across spatial hierarchies:

$$P = concat\left(\left\{P_{i,j}^l | l = 0,1,\dots,L-1\right\}\right) \quad (5)$$

This dynamic configuration enables DHP to adaptively capture features at multiple resolutions, making it more effective for tasks requiring spatially-aware representations, like object detection and semantic segmentation.

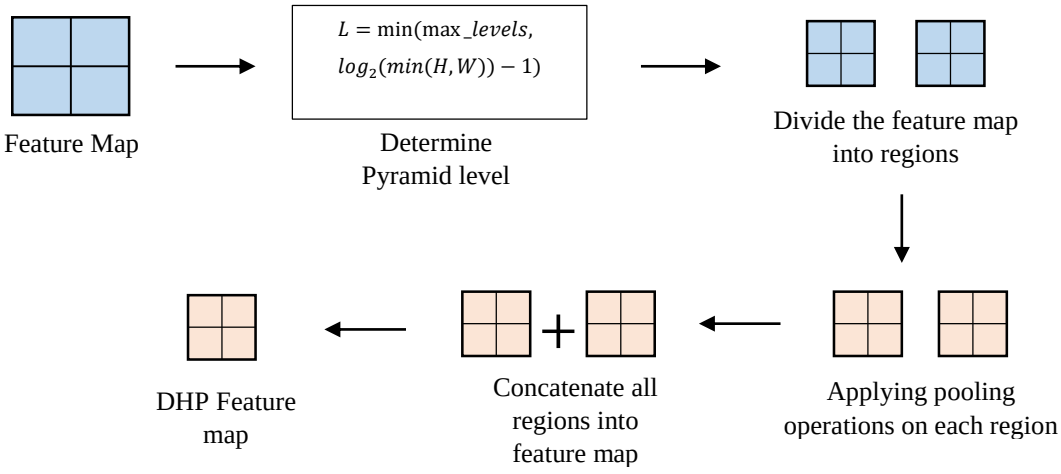


Figure 2: Working mechanism of DHP

Figure 2 illustrates the working mechanism of DHP showcasing its ability to adaptively capture multi-scale features this dynamic approach enhances feature representation for improved classification accuracy.

Algorithm: DHP ()

Adaptively split feature maps into hierarchical regions.

For each pyramid level *l*:

Divide the feature map into $2^l \times 2^l$ regions.

Apply pooling (e.g., max or average) on each region:

$$P_{(i,j)}^l = f_{pool} R_{(i,j)}^l$$

Concatenate pooled features across levels:

$$P = concat(\{P_{(i,j)}^l | l = 0,1,\dots,L-1\})$$

3.3 SE Attention Mechanism

DOI: https://doi.org/ 10.70274/jaict.2025.2.1.40.
p- ISSN 34562-3478 e- ISSN 5443-1243

The SE block is a channel-wise attention mechanism designed to recalibrate the feature maps by explicitly modeling the interdependencies between channels. Given an input feature map $X \in R^{H \times W \times C}$, where H and W are the spatial dimensions (height and width), and C is the number of channels, the SE block operates in two main steps: squeeze and excitation.

3.3.1 Squeeze

The global spatial information is captured by applying global average pooling (GAP) across the spatial dimensions of the input. This results in a channel descriptor, a vector $z \in R^C$, which represents the average value of each channel:

$$z_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} X_{ijc} \ for\ each\ channel\ c = 1,2, \dots, C \qquad (6)$$

Algorithm: SE ()

---

1. Squeeze:

Apply Global Average Pooling (GAP) to extract channel-wise descriptors:

$$S = GAP(P)$$

2. Excitation:

Pass through two fully connected layers with ReLU and sigmoid:

$$E_1 = Dense\left(S, \frac{C}{r}, ReLU\right), E_2 = Dense(E_1, C, Sigmoid)$$

3. Recalibrate:

Multiply attention weights E_2 with feature maps:

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

3.3.2 Excitation

The squeezed vector z is then passed through two fully connected (FC) layers, typically with a reduction ratio r, where the first FC layer reduces the dimensionality from C C/r, and the second restores it back to C. These layers generate a channel-wise attention vector $s \in R^C$ using a sigmoid activation:

$$s = \sigma\big(W_2 \delta(W_1 z)\big) \qquad (7)$$

where σ is the sigmoid function, δ is the ReLU activation, and W_1, W_2 are learned weight matrices.

3.3.3 Recalibration

The channel-wise attention vector s is then reshaped and multiplied with the original input feature map X, scaling the importance of each channel according to the learned attention:

$$X' = X \cdot s \qquad (8)$$

where s is broadcasted across the spatial dimensions. The output X' is the recalibrated feature map with enhanced channel-wise attention.

DOI: https://doi.org/ 10.70274/jaict.2025.2.1.40.
p- ISSN 34562-3478 e- ISSN 5443-1243

This mechanism allows the network to focus on important features and suppress irrelevant ones, improving the representational power and performance of the model.

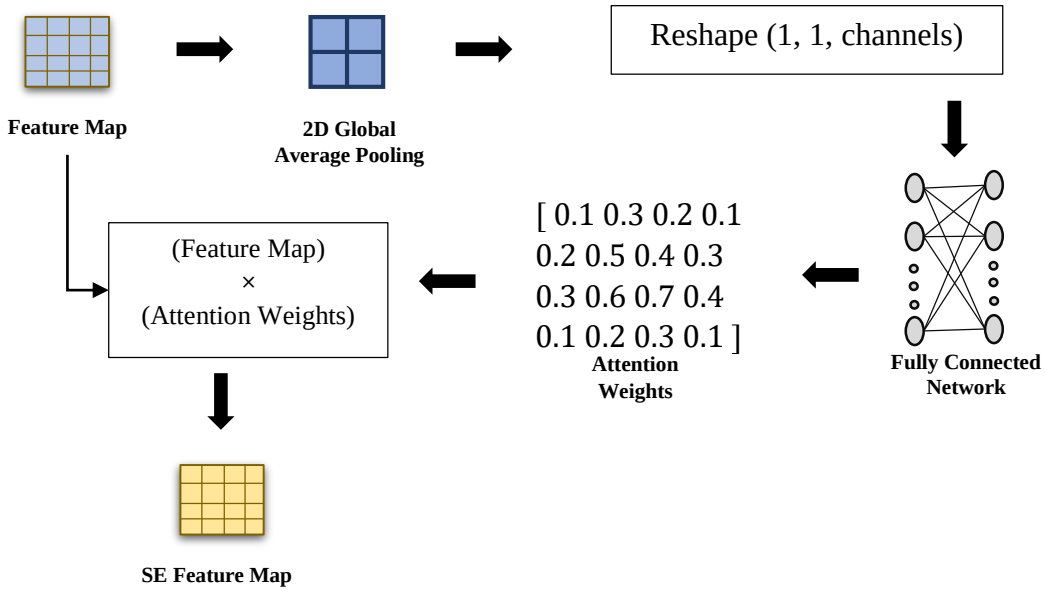


Figure 3: Working mechanism of Squeeze-and-Excitation Attention

Figure 3 illustrates the working mechanism of the Squeeze-and-Excitation Attention, which enhances feature representations by adaptively recalibrating channel-wise information. This process improves the network's ability to focus on critical regions, enhancing performance in tasks requiring fine-grained feature discrimination, such as cervical cancer classification.

3.4 DHP-SE-HN

The DHP-SE-HN is a sophisticated deep learning architecture for Pap smear image classification, combining multi-scale feature extraction and attention mechanisms to enhance performance. Starting with two convolutional layers (64 filters, 3x3, ReLU) followed by max pooling, the model extracts initial features from the input image. DHP then adaptively partitions these feature maps, applying multi-level pooling to capture diverse spatial scales. A SE block follows, applying channel-wise attention to recalibrate feature importance by leveraging global average pooling and two fully connected layers with ReLU and sigmoid activations. Further feature refinement occurs through two additional convolutional layers (128 filters, 3x3, ReLU) and another max-pooling layer. The resulting feature map undergoes global average pooling to reduce spatial dimensions before entering an LSTM layer (128 units), which captures temporal dependencies in the feature sequence. Finally, fully connected layers (256 and 128 neurons, ReLU) lead to the output layer consists of a dense layer with 3 neurons using softmax activation to classify into three classes: non-cervical cancer, pre-cervical cancer and cervical cancer. The model is optimized using the Adam optimizer and categorical cross-entropy loss, making it robust in learning both fine-grained and high-level patterns for pap smear image analysis.

DOI: https://doi.org/ 10.70274/jaict.2025.2.1.40.
p- ISSN 34562-3478 e- ISSN 5443-1243

Algorithm- DHP-SE-HN ()

Input: Pap smear image I

Output: Predicted class probabilities

1. Preprocessing

1.1 Image Resizing:

Resize image I to a fixed size 128x128:

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

1.2 Contrast Enhancement (CLAHE):

Enhance local contrast using Adaptive Histogram Equalization (CLAHE):

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

1.3 Normalization:

Normalize pixel values to the range [0, 1]:

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

$$I_{pre} = I_{normalized}$$

2. Model Architecture

2.1 Initial Convolutional Layers:

Apply two convolutional layers to extract initial features:

$$H_{conv1} = Conv2D(I_{pre}, W_1, b_1) + b_1$$

$$H_{relu1} = ReLU(H_{conv1})$$

Apply a second convolutional layer:

$$H_{conv2} = Conv2D(H_{relu1}, W_2, b_2) + b_2$$

$$H_{relu2} = ReLU(H_{conv2})$$

2.2 Max-Pooling:

Reduce spatial dimensions:

$$H_{pool1} = MaxPooling(H_{relu2})$$

2.3 Dynamic Hierarchical Pooling (DHP):

function DHP ()

2.4 Squeeze-and-Excitation (SE):

function SE ()

2.5 Additional Convolutional Layers:

Apply two additional convolutional layers for further feature refinement:

$$H_{conv3} = Conv2D(H_{SE}, W_3, b_3) + b_3$$

$$H_{relu3} = ReLU(H_{conv3})$$

$$H_{conv4} = Conv2D(H_{relu3}, W_4, b_4) + b_4$$

$$H_{relu4} = ReLU(H_{conv4})$$

2.6 Second Max-Pooling:

DOI: https://doi.org/ 10.70274/jaict.2025.2.1.40.
p- ISSN 34562-3478 e- ISSN 5443-1243

Further reduce spatial dimensions:

$$H_{pool2} = MaxPooling(H_{relu4})$$

2.7 Global Average Pooling (GAP):

Extract a fixed-size feature vector:

$$H_{flat} = GAP(H_{pool2})$$

2.8 Recurrent Layer (LSTM):

Reshape features and apply LSTM (128 units) to capture sequential dependencies:

$$H_{LSTM} = LSTM(Reshape(H_{flat}(1, -1)))$$

2.9 Dense Layers:

Apply dense layers for feature aggregation:

$$H_{dense1} = Dense(H_{LSTM}, 256, ReLU)$$
$$H_{dense2} = Dense(H_{dense1}, 128, ReLU)$$

2.10 Output Layer:

Apply final dense layer for classification into three classes:

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

Obtain Predicted class probabilities.

3. Training

3.1 Loss Function:

Use categorical cross-entropy loss:

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

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

3.2 Optimization:

Update model parameters using the 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 over a specified number of epochs or until convergence.

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 accuracy, precision, recall, and F1 score.

DOI: https://doi.org/ 10.70274/jaict.2025.2.1.40.
p- ISSN 34562-3478 e- ISSN 5443-1243

3.5 Cervical Cancer Classification

Cervical cancer classification is performed using a hybrid deep learning architecture that leverages dynamic hierarchical pooling and squeeze-and-excitation attention mechanisms. The process begins with preprocessing of Pap smear images, where each image is resized to a uniform size (128x128 pixels), enhanced using contrast adjustment and normalized to ensure consistent input features. The model then applies a series of convolutional layers to extract low-level features, followed by dynamic hierarchical pooling to capture multi-scale spatial patterns. The SE block further refines these features by recalibrating the importance of different channels, allowing the model to focus on the most relevant patterns. After additional convolutional layers and max-pooling operations, global average pooling reduces the spatial dimensions of the feature map into a compact vector. This vector is processed through an LSTM layer to capture sequential dependencies and temporal patterns, followed by dense layers to aggregate the features and produce final class predictions. The output layer, using s oftmax activation, classifies the images into three categories: non-cervical cancer, pre-cervical cancer and cervical cancer. The model is optimized using the Adam optimizer and trained with categorical cross-entropy loss to maximize classification accuracy. This architecture enables the model to effectively differentiate between abnormal and normal cellular patterns in Pap smear images, providing reliable predictions for cervical cancer diagnosis.

4. EXPERIMENTAL RESULTS AND ANALYSIS

4.1 Data Description

The dataset integrates three cytology image collections for classifying Pap smear and liquid-based cytology (LBC) samples. The Herlev Pap Smear Dataset includes 917 single-cell images divided into 7 classes, with 3 normal classes (intermediate squamous epithelial, columnar epithelial, and superficial squamous epithelial) totaling 242 images and 4 abnormal classes (various dysplasias and carcinoma) totaling 675 images. The Mendeley LBC Dataset contains 963 whole slide images across 4 classes. Of these, 613 images are normal, and the remaining 350 images represent abnormal conditions, including low- and high-grade squamous intraepithelial lesions (LSIL and HSIL) and squamous cell carcinoma (SCC). The SIPaKMeD Pap Smear Dataset features 4,049 images of isolated cells grouped into 5 classes. Two classes (superficial and intermediate, and parabasal) are normal, comprising 1,618 images. Two classes (koilocytotic and dyskeratotic) are abnormal, with 1,638 images. The fifth class (metaplastic), categorized as benign, contains 793 images. The combined dataset comprises 5,929 images, categorized into 2,473 normal (non-cervical cancer), 2,663 abnormal (cervical cancer), and 793 benign images (potentially precancerous). This comprehensive resource provides a strong foundation for advancing classification models in cytology.

4.2 Experimental Setup

DOI: https://doi.org/ 10.70274/jaict.2025.2.1.40.
p- ISSN 34562-3478 e- ISSN 5443-1243

The research was conducted on a workstation with an Intel i5-4300U 1.90 GHz processor, 8 GB of RAM, and a 64-bit Windows 10 operating system. Python code was developed using the Anaconda IDE, leveraging TensorFlow and Keras for model implementation.

4.3 Result and Performance Analysis

The Dynamic Hierarchical Pooling with Squeeze-and-Excitation Attention-guided Hybrid Network (DHP-SE-HN) incorporates preprocessing steps, including resizing images to 128x128 pixels, contrast enhancement using CLAHE, and normalization to stabilize the training process. The core of the model is the DHP module which adaptively captures multi-scale features by adjusting pooling levels and operations based on the input feature map size, enabling both fine-grained and coarse-grained feature extraction. To further enhance feature discrimination, the SE attention mechanism recalibrates channel-wise feature importance focusing on diagnostically significant regions. The hybrid architecture combines CNNs for local spatial feature extraction and RNNs for capturing sequential dependencies and global context. Trained with the Adam optimizer and categorical cross-entropy loss, the model classifies images into three categories: non-cervical cancer, pre-cervical cancer, and cervical cancer. The proposed model is being compared with existing models, including our previous work, SE-AG-HN as well as the CerviFormer model proposed by Deo, B.S., et al., 2024 and the 3D-ViT-SE-KELM model proposed by Abinaya, K., et al., 2024.

In our previous work, SE-AG-HN, a deep learning architecture was proposed to classify cervical cancer from cytology images. The architecture integrates three key modules, a CNN module for feature extraction, an S&E module for feature recalibration and an RNN module for final classification. To prepare the images for analysis, preprocessing techniques like resizing, contrast enhancement using Adaptive Histogram Equalization and normalization were employed. The CNN module captures spatial features, while the SE module enhances the model's focus on relevant features. On the other hand the RNN module captures sequential dependencies in the data. The model was trained using the Adam optimizer with categorical cross-entropy as the loss function and accuracy as the evaluation metric. The model achieved accurate classification into three categories: cervical cancer, pre-cancerous lesions, and non-cancerous conditions. Deo, B.S., et. al., 2024 proposed CerviFormer, a novel cross-attention-based Transformer model for accurate classification of cervical cancer in Pap smear images. This model effectively addressed the challenges posed by varying image sizes and complex patterns in Pap smears. By leveraging the power of Transformers, CerviFormer achieved impressive results on two publicly available datasets, Sipakmed and Herlev. Abinaya, K., et. al., 2024 proposed 3D-ViT-SE-KELM model, a novel deep learning architecture designed for accurate cervical cancer classification. This model, which we have named 3D-ViT-SE-KELM for easy reference, combined 3D CNNs and ViTs to extract comprehensive features, with FPN integrating multi-level features. The

DOI: https://doi.org/ 10.70274/jaict.2025.2.1.40.
p- ISSN 34562-3478 e- ISSN 5443-1243

3D SE block refined feature importance, and KELM classified the input images. Table 1 shows a comparison of actual and predicted result for different types of input images.

Table 1: Input Image with Actual Result and Predicted Result

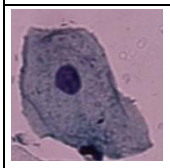
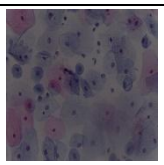
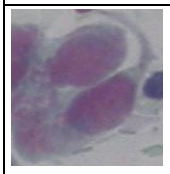
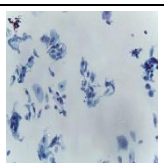
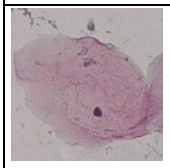
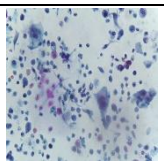
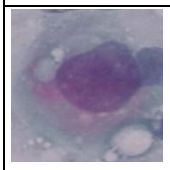
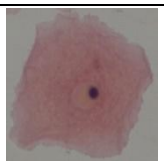
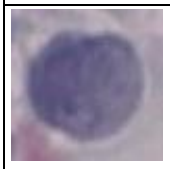
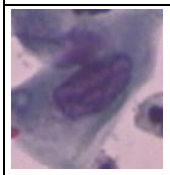
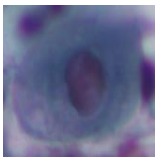
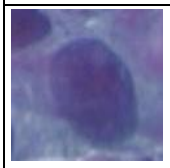
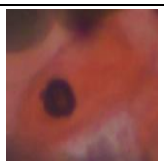
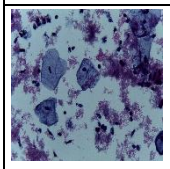
Input Image	Actual Result	Predicted Result	Input Image	Actual Result	Predicted Result
	Intermediate Squamous Epithelial - Normal	Normal		Low Grade Squamous Intraepithelial Lesion (LSIL) - Abnormal	Abnormal
	Columnar Epithelial - Normal	Normal		High Grade Squamous Intraepithelial Lesion (HSIL) - Abnormal	Abnormal
	Superficial Squamous Epithelial - Normal	Normal		Squamous Cell Carcinoma (SCC) - Abnormal	Abnormal
	Mild Squamous Non-Keratinizing Dysplasia - Normal	Normal		Superficial Intermediate - Normal	Normal
	Squamous Cell Carcinoma in-situ Intermediate - Abnormal	Abnormal		Parabasal - Normal	Normal
	Moderate Squamous Non-Keratinizing Dysplasia - Abnormal	Abnormal		Koilocytotic - Abnormal	Abnormal
	Severe Squamous Non-Keratinizing Dysplasia - Abnormal	Abnormal		Dyskeratotic - Abnormal	Abnormal
	Negative for Intraepithelial Malignancy - Normal	Normal		Metaplastic - Benign	Benign

Table 2 presents the training and testing performance of a model over 10 epochs. The training loss decreases while the training accuracy increases, indicating the model's improvement on training

DOI: https://doi.org/ 10.70274/jaict.2025.2.1.40.
p- ISSN 34562-3478 e- ISSN 5443-1243

data. Table 3 compares the performance of the proposed work DHP-SE-HN with the existing works, SE-AG-HN, CerviFormer and 3D-VIT-SE-KELM across multiple validation sets. DHP-SE-HN model outperforms the existing models in sensitivity, as indicated by its higher True Positive Rate (TPR) and lower False Negative Rate (FNR) across multiple validation sets.

Table 2: Training and testing metrics of the proposed model

Epochs	Training Loss	Testing Loss	Training accuracy	Testing accuracy
1	0.217	0.222	0.966	0.968
2	0.204	0.194	0.957	0.976
3	0.184	0.127	0.989	0.974
4	0.173	0.219	0.976	0.983
5	0.211	0.201	0.982	0.967
6	0.221	0.134	0.967	0.989
7	0.214	0.213	0.961	0.967
8	0.167	0.251	0.978	0.978
9	0.216	0.189	0.975	0.96
10	0.174	0.229	0.988	0.989
Overall	**0.1981**	**0.1979**	**0.9739**	**0.9751**

Table 4 compares the performance of DHP-SE-HN, SE-AG-HN, CerviFormer, and 3D-VIT-SE-KELM across multiple validation sets. DHP-SE-HN consistently outperforms the other models in terms of sensitivity, indicating its effectiveness in identifying true positive cases, while maintaining a balance with other metrics like specificity, accuracy, precision, and F1-score. This makes DHP-SE-HN a strong contender for applications where high sensitivity is crucial.

Table 3: Comparative analysis of performance measures obtained from proposed work DHP-SE-HN and existing works, SE-AG-HN, CerviFormer and 3D-VIT-SE-KELM

Validation Sets	DHP-SE-HN				SE-AG-HN				CerviFormer				3D-ViT-SE-KELM			
	TPR	FPR	FNR	TNR	TPR	FPR	FNR	TNR	TPR	FPR	FNR	TNR	TPR	FPR	FNR	TNR
1	0.987	0.046	0.013	0.954	0.945	0.048	0.017	0.944	0.867	0.113	0.133	0.887	0.824	0.188	0.176	0.812
2	0.966	0.063	0.034	0.937	0.934	0.025	0.002	0.925	0.881	0.082	0.119	0.918	0.839	0.166	0.161	0.834
3	0.979	0.038	0.021	0.962	0.962	0.038	0.032	0.942	0.901	0.093	0.099	0.907	0.798	0.151	0.202	0.849
4	0.968	0.062	0.032	0.938	0.978	0.019	0.015	0.937	0.927	0.103	0.073	0.897	0.811	0.209	0.189	0.791
5	0.987	0.051	0.013	0.949	0.981	0.025	0.021	0.949	0.891	0.137	0.109	0.863	0.784	0.212	0.216	0.788

DOI: https://doi.org/ 10.70274/jaict.2025.2.1.40.
p- ISSN 34562-3478 e- ISSN 5443-1243

6	0.976	0.049	0.024	0.951	0.984	0.008	0.027	0.945	0.912	0.135	0.088	0.865	0.769	0.246	0.231	0.754
7	0.989	0.031	0.011	0.969	0.959	0.028	0.011	0.937	0.893	0.129	0.107	0.871	0.813	0.199	0.187	0.801
Overall	0.979	0.049	0.021	0.951	0.963	0.027	0.018	0.940	0.896	0.113	0.104	0.887	0.805	0.196	0.195	0.804

DOI: https://doi.org/ 10.70274/jaict.2025.2.1.40.
p- ISSN 34562-3478 e- ISSN 5443-1243

Table 4: Comparative analysis of validation sets of proposed work DHP-SE-HN and existing works, SE-AG-HN, CerviFormer and 3D-VIT-SE-KELM

Validation Sets	DHP-SE-HN					SE-AG-HN					CerviFormer					3D-ViT-SE-KELM				
	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	0.987	0.954	0.967	0.981	0.958	0.945	0.944	0.945	0.972	0.958	0.867	0.887	0.887	0.861	0.873	0.824	0.812	0.814	0.833	0.804
2	0.966	0.937	0.973	0.967	0.952	0.934	0.925	0.929	0.971	0.952	0.881	0.918	0.937	0.892	0.928	0.839	0.834	0.832	0.798	0.79
3	0.979	0.962	0.959	0.953	0.961	0.962	0.942	0.952	0.953	0.961	0.901	0.907	0.891	0.927	0.87	0.798	0.829	0.828	0.821	0.774
4	0.968	0.938	0.956	0.949	0.956	0.978	0.937	0.958	0.934	0.956	0.927	0.897	0.882	0.843	0.901	0.811	0.791	0.791	0.848	0.819
5	0.987	0.949	0.964	0.977	0.955	0.981	0.949	0.965	0.93	0.955	0.891	0.863	0.906	0.888	0.912	0.784	0.788	0.782	0.781	0.27
6	0.976	0.951	0.988	0.961	0.968	0.975	0.933	0.944	0.962	0.968	0.912	0.865	0.852	0.909	0.898	0.769	0.754	0.779	0.777	0.839
7	0.989	0.969	0.978	0.973	0.945	0.966	0.924	0.969	0.944	0.945	0.893	0.871	0.879	0.911	0.876	0.813	0.801	0.806	0.824	0.781
Overall	0.979	0.951	0.969	0.966	0.956	0.963	0.936	0.952	0.953	0.956	0.896	0.887	0.891	0.890	0.894	0.805	0.801	0.805	0.812	0.725

Figure 4 illustrates the training and testing accuracy of a DHP-SE-HN model over 10 epochs. The model initially learns quickly, leading to a sharp increase in both training and testing accuracy. Figure 5 illustrates the training and testing loss of DHP-SE-HN model over 10 epochs. The model exhibits promising performance, with the training loss steadily decreasing over the epochs, indicating effective learning. The testing loss also shows a decreasing trend, suggesting good generalization to unseen data.

DOI: https://doi.org/ 10.70274/jaict.2025.2.1.40.
p- ISSN 34562-3478 e- ISSN 5443-1243

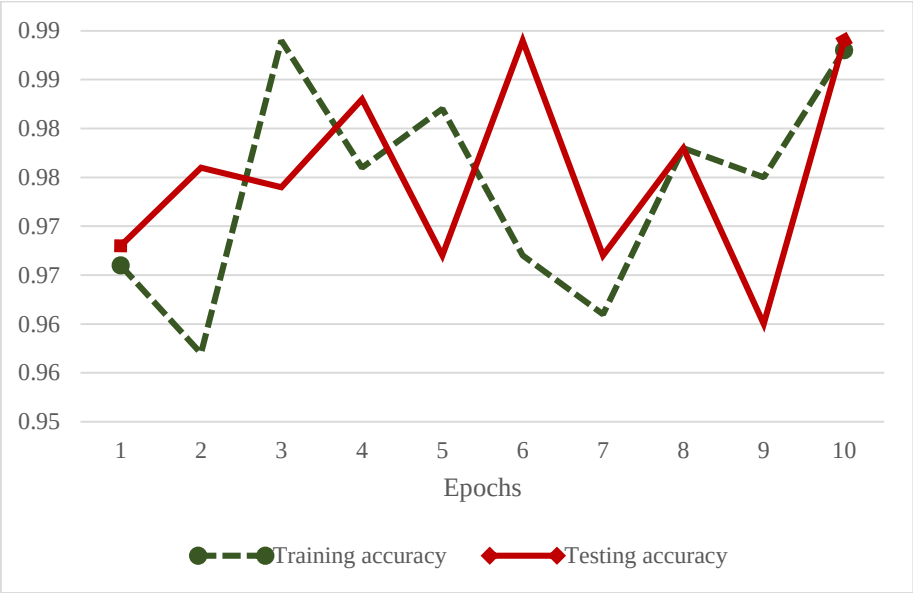


Figure 4: Graphical representation of Training and testing accuracy of DHP-SE-HN

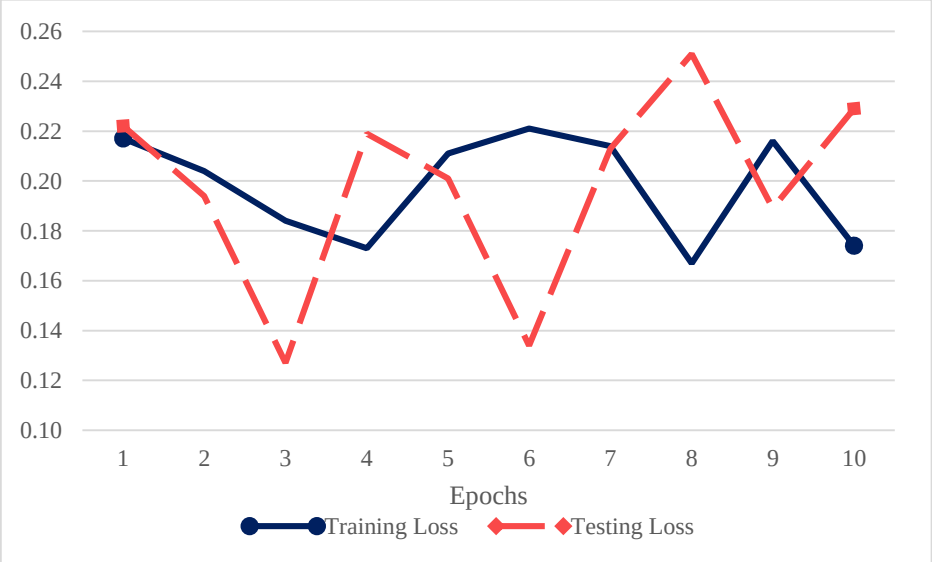


Figure 5: Graphical representation of Training and testing loss of the DHP-SE-HN

Figure 6 compares the performance of four models (DHP-SE-HN, SE-AG-HN, CerviFormer, and 3D-VIT-SE-KELM) across five metrics (Sensitivity, Specificity, Accuracy, Precision, and F1-Score). DHP-SE-HN exhibit the highest accuracy, indicating their ability to make correct predictions consistently. Overall, DHP-SE-HN shows better performance in sensitivity, accuracy, and F1-Score, indicating their ability to identify positive cases and overall performance.

DOI: https://doi.org/ 10.70274/jaict.2025.2.1.40.
p- ISSN 34562-3478 e- ISSN 5443-1243

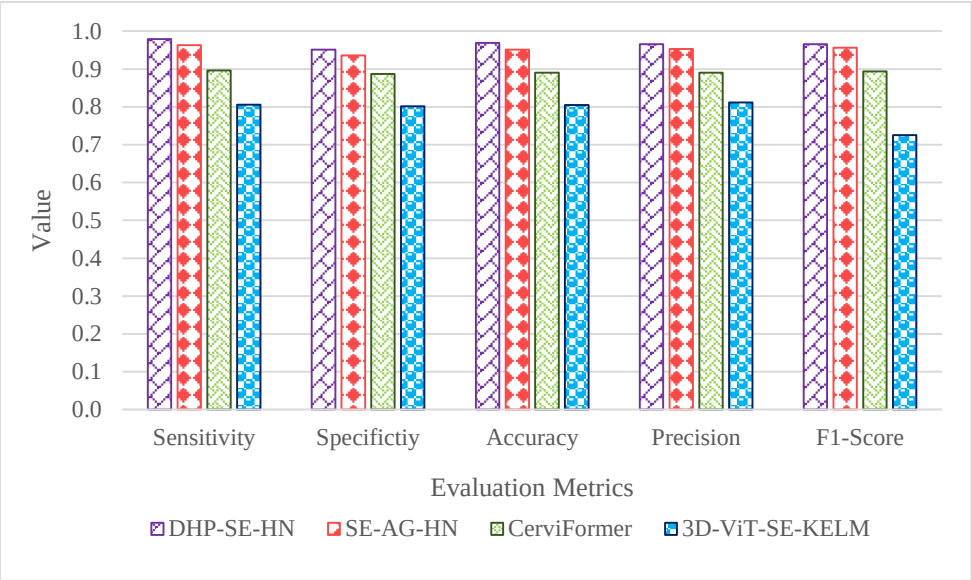


Figure 6: Graphical representation of performance obtained from proposed work DHP-SE-HN and existing works, SE-AG-HN, CerviFormer and 3D-VIT-SE-KELM

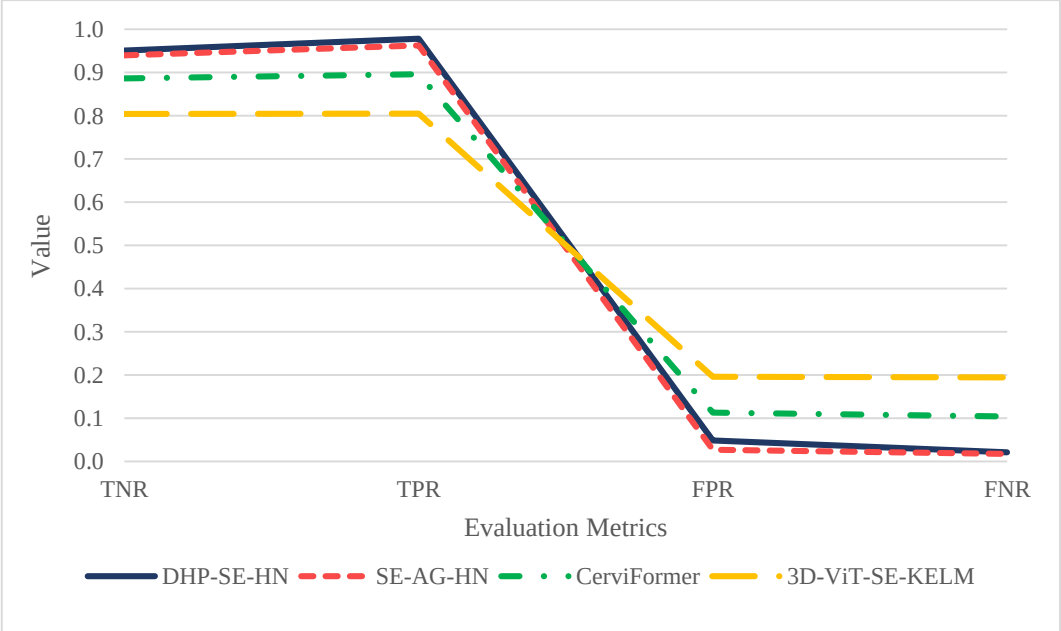


Figure 7: Graphical representation of TPR, FPR, FNR, TNR obtained from proposed work DHP-SE-HN and existing works, SE-AG-HN, CerviFormer and 3D-VIT-SE-KELM

Figure 7 compares the performance of four models (DHP-SE-HN, SE-AG-HN, CerviFormer, and 3D-VIT-SE-KELM) across four metrics: True Positive Rate (TPR), False Positive Rate (FPR), False Negative Rate (FNR), and True Negative Rate (TNR). DHP-SE-HN outperforms the other models in terms of sensitivity, as evidenced by its higher True Positive Rate (TPR) and lower False Negative Rate (FNR). This indicates its superior ability to correctly identify positive cases. While it may have a slightly higher False Positive Rate (FPR), its overall performance, particularly in terms of sensitivity, makes it a strong contender for applications where accurate identification of positive cases is crucial.

DOI: https://doi.org/ 10.70274/jaict.2025.2.1.40.
p- ISSN 34562-3478 e- ISSN 5443-1243

4.4 Discussion

The DHP-SE-HN demonstrates superior performance compared to SE-AG-HN, CerviFormer, and 3D-ViT-SE-KELM in cervical cancer classification from Pap smear images. DHP-SE-HN excels in sensitivity and TPR, indicating its superior ability to correctly identify cancerous cases, which is crucial for early diagnosis. It also achieves a high F1-score of 0.956, highlighting its balance between precision and recall. While it shows slightly higher FPR, its FNR is significantly lower, ensuring fewer missed positive cases. The model outperforms the others in terms of accuracy, precision, and specificity, with CerviFormer and 3D-ViT-SE-KELM showing marginally better TNR. The flexibility of the DHP module and the feature enhancement provided by the SE mechanism contribute to the model's ability to handle variations in image quality and resolution, making it adaptable to real-world clinical environments. Overall, DHP-SE-HN's robust performance across multiple metrics and its scalability position it as a highly effective tool for cervical cancer detection, particularly in resource-constrained settings. Future work will focus on optimizing the model's precision and exploring its use in other medical image domains.

5. CONCLUSION

DHP-SE-HN excels by combining DHP for adaptive multi-scale feature extraction and SE attention to focus on diagnostically important regions. The hybrid CNN-RNN architecture enables the model to capture both local spatial features and global contextual dependencies, resulting in superior performance across multiple evaluation metrics. The model outperforms existing methods in terms of metrics making it well-suited for early detection of cervical cancer, where minimizing missed diagnoses is critical. Its ability to handle variations in image quality, resolution, and staining protocols ensures that it is adaptable to real-world clinical settings, including low-resource environments. A unique strength of the proposed framework lies in the combination of fine-grained and coarse-grained feature extraction via dynamic pooling and the enhanced focus on relevant features through attention mechanisms. These features make DHP-SE-HN a robust tool for accurate classification. Future enhancements could focus on integrating semi-supervised learning or transfer learning to improve performance with limited labeled data. Expanding the model to differentiate between more granular categories of cervical cancer and incorporating interpretability features would increase its clinical value. Overall, DHP-SE-HN offers a scalable, reliable solution for cervical cancer screening, with strong potential for real-world application and broader medical imaging domains.

REFERENCES:

Abinaya, K. and Sivakumar, B., 2024. A Deep Learning-Based Approach for Cervical Cancer Classification Using 3D CNN and Vision Transformer. *Journal of imaging informatics in medicine*, *37*(1), pp.280-296.

DOI: https://doi.org/ 10.70274/jaict.2025.2.1.40.
p- ISSN 34562-3478 e- ISSN 5443-1243

Chen, W., Gao, L., Li, X. and Shen, W., 2022. Lightweight convolutional neural network with knowledge distillation for cervical cells classification. *Biomedical Signal Processing and Control*, *71*, p.103177.

Deo, B.S., Pal, M., Panigrahi, P.K. and 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), p.e23043.

Ding, B., Zhou, X., Li, C., Zhou, J. and He, Y., 2024. Fusion Diagnostic Knowledge and Foundation Model forAbnormal Cervical Cell Detection.

Feng, T., Ying, J., Yang, H., Li, F. and Li, H., 2022, September. Regional detection of cervical lesions based on self-attention mechanism and multi-scale feature enhancement. In *Chinese Intelligent Systems Conference* (pp. 182-190). Singapore: Springer Nature Singapore.

Ganguly, T., Singh, R.P. and Kumar, P., 2023, November. Self-attention based resnet model for cervical cancer detection. In *2023 Second International Conference on Informatics (ICI)* (pp. 1-6). IEEE.

Ghoneim, A., Muhammad, G. and Hossain, M.S., 2020. Cervical cancer classification using convolutional neural networks and extreme learning machines. *Future Generation Computer Systems*, *102*, pp.643-649.

Kang, J. and Li, N., 2024. CerviSegNet-DistillPlus: An Efficient Knowledge Distillation Model for Enhancing Early Detection of Cervical Cancer Pathology. *IEEE Access*, *12*, pp.85134-85149.

Khan, A., Han, S., Ilyas, N., Lee, Y.M. and Lee, B., 2023. CervixFormer: A Multi-scale swin transformer-Based cervical pap-Smear WSI classification framework. *Computer Methods and Programs in Biomedicine*, *240*, p.107718.

Li, Y.X., Chen, F., Shi, J.J., Huang, Y.L. and Wang, M., 2023. Convolutional neural networks for classifying cervical cancer types using histological images. *Journal of Digital Imaging*, *36*(2), pp.441-449.

Liu, W., Li, C., Xu, N., Jiang, T., Rahaman, M.M., Sun, H., Wu, X., Hu, W., Chen, H., Sun, C. and Yao, Y., 2022. CVM-Cervix: A hybrid cervical Pap-smear image classification framework using CNN, visual transformer and multilayer perceptron. *Pattern Recognition*, *130*, p.108829.

Liu, Y., Wang, Z., Dai, A. and Gu, W., 2024, May. A novel ensemble framework based on CNN models and Swin transformer for cervical cytology image classification. In *Seventh International Conference on Computer Graphics and Virtuality (ICCGV 2024)* (Vol. 13158, pp. 150-161). SPIE.

Mishra, A.K., Gupta, I.K., Diwan, T.D. and Srivastava, S., 2024. Cervical precancerous lesion classification using quantum invasive weed optimization with deep learning on biomedical pap smear images. *Expert Systems*, *41*(7), p.e13308.

Pacal, I., 2024. MaxCerVixT: A novel lightweight vision transformer-based Approach for precise cervical cancer detection. *Knowledge-Based Systems*, *289*, p.111482.

DOI: https://doi.org/ 10.70274/jaict.2025.2.1.40.
p- ISSN 34562-3478 e- ISSN 5443-1243

Fadlallah, K. I. A., Sharif, M. M., & Eldow, M. E. Y. (2024). Developing parallel requirements prioritization machine learning model integrating with MoSCoW method. Journal of Artificial Intelligence and Computational Technology, 1(1).

Talpur, D.B., Raza, A., Khowaja, A. and Shah, A., 2024. DeepCervixNet: An Advanced Deep Learning Approach for Cervical Cancer Classification in Pap Smear Images. *VAWKUM Transactions on Computer Sciences*, *12*(1), pp.136-148.

Yasin, E.H.E., Sharif, M.M.S., Yahia, M.Y.A., Othman, A.Y., Ibrahim, A.O., Kheiry, M.A. and Musa, M.2024. Crop type classification and spatial mapping in River Nile and Northern State, Sudan, using Sentinel-2 satellite data andfield observation. Journal of Degraded and Mining Lands Management 11(3):5997-6007, doi:10.15243/jdmlm.2024.113.5997.

(PDF) Crop type classification and spatial mapping in River Nile and Northern State, Sudan, using Sentinel-2 satellite data and field observation. Available from: https://www.researchgate.net/publication/379447835_Crop_type_classification_and_spatial_mapping_in_River_Nile_and_Northern_State_Sudan_using_Sentinel-2_satellite_data_and_field_observation [accessed Dec 13 2024].

Tan, X., Li, K., Zhang, J., Wang, W., Wu, B., Wu, J., Li, X. and Huang, X., 2021. Automatic model for cervical cancer screening based on convolutional neural network: a retrospective, multicohort, multicenter study. *Cancer cell international*, *21*, pp.1-10.

Xiang, Y., Sun, W., Pan, C., Yan, M., Yin, Z. and Liang, Y., 2020. A novel automation-assisted cervical cancer reading method based on convolutional neural networks. *Biocybernetics and Biomedical Engineering*, *40*(2), pp.611-623.

Yang, C., Tan, Z., Wang, Y., Bi, R., Shi, T., Yang, J., Huang, C., Jiang, P. and Fu, X., 2024. SwinUNeCCt: bidirectional hash-based agent transformer for cervical cancer MRI image multi-task learning. *Scientific Reports*, *14*(1), p.24621.

Yi, J., Liu, X., Cheng, S., Chen, L. and Zeng, S., 2024. Multi-scale window transformer for cervical cytopathology image recognition. *Computational and Structural Biotechnology Journal*, *24*, pp.314.

DOI: https://doi.org/ 10.70274/jaict.2025.2.1.40.
p- ISSN 34562-3478 e- ISSN 5443-1243