University of Thi-Qar Journal

ISSN (print): 2706- 6908, ISSN (online): 2706-6894

Vol.21 No.2 june 2026

Hybrid Deep Learning Framework for Cervical Cancer Cell Classification Using Discrete Wavelet Transform and EfficientNetV2-B3

Zahoor M. Aydam[1,*] and Baidaa Mutasher Rashed[2]

[1,*]Department of information technology, College of Computer Science and Mathematics, University of Thi-Qar, Thi-Qar, 64001, Iraq

[2]Department of Computers, Faculty of Computer Science & Mathematics, University of Thi-Qar, Nasiriyah 00964, Thi-Qar, Iraq.

***Corresponding Author Email: zahoor.mosad@utq.edu.iq**

ABSTRACT

Cervical cancer is among the most common and fatal cancers that affect women worldwide, timely and accurate diagnosis using cytopathology is essential in determining treatment options. Traditional cytological tests, including the Pap smear, involve manual processes that are laborious, time-consuming, and highly variable between different observers. In this paper, a novel dual-path hybrid deep learning architecture is proposed for automatic multi-class classification of cervical cancer cells based on the integration of Discrete Wavelet Transform (DWT) as frequency-domain features with deep spatial features extracted using the EfficientNetV2-B3 architecture. The proposed architecture extracts multi-scale wavelet features of dimension 256 in parallel with deep semantic features of dimension 1280. Feature vectors derived from both pathways are concatenated to create an integrated 1536-dimensional feature representation for multi-class classification of cervical cancer cells into five categories. The result is compare two experimental designs in the study: (1) EfficientNetV2-B3 alone with 97.04% accuracy and 0.9908 AUC score, and (2) hybrid Deep Wavelet + EfficientNetV2-B3 model with higher performance results (accuracy: 99.26%, precision: 99.28%, recall: 99.26%, F1-score: 99.27%, and AUC: 99.38%).

Keywords: Cervical Cancer Classification; Deep Learning; Discrete Wavelet Transform; EfficientNetV2-B3; Feature Fusion; Hybrid Architecture; Medical Image Analysis; Pap Smear.

1. INTRODUCTION

Cervical cancer is listed as one of the top four most frequent cancers affecting women, with nearly 604,000 new cases and 342,000 deaths annually, based on the statistics published by the World Health Organization (WHO) [1]. This type of cancer primarily stems from an infection caused by high-risk human papillomaviruses (HPVs), mainly HPV-16 and HPV-18, responsible for about 70% of cases. Despite the existence of preventive HPV vaccines and screening programs, this cancer continues

University of Thi-Qar Journal

ISSN (print): 2706- 6908, ISSN (online): 2706-6894

Vol.21 No.2 june 2026

to pose a serious problem, particularly for low- and middle-income countries with limited healthcare resources and expertise [2].

The Papanicolaou (Pap) test and liquid-based cytology serve as the core screening techniques that can detect pre-cancer lesions on the cervix and provide timely intervention, hence resulting in a reduction of cancer-related deaths. Routine cytological examination is the microscopic evaluation of the cervical cells by the cytopathologists using various criteria (e.g., Bethesda system or similar). Thus, it is a very labor intensive method and is accompanied with a high false negative rate (5–10%) along with a higher inter observer variability [3].

In recent years, the advent of deep learning (DL) has significantly advanced automated cervical cell analysis algorithms, and CNNs work effectively in extracting discriminative features directly from raw images [4]. But typical existing methods using CNN operates in the spatial domain and thus fails to exploit the textural and frequency domain features that are important in identifying certain pathological conditions from cytological images.

Which the wavelet transform provides a solid foundation for conducting multi-resolution analysis while simultaneously analyzing both spatial and frequency domains of the signal or image [5]. The discrete wavelet transform divides images into sub-bands containing different approximations and details in the wavelet domain. Thus, wavelet-based multi-resolution analysis allows extracting textural and structural features that are not covered by deep convolutional features from images. This research aims to investigate a novel way of integrating deep wavelet multi-resolution representations into deep convolutional features and explore its application to automated cervical cell classification.

The EfficientNet family of CNNs demonstrates impressive performance on classification tasks by applying a combination of compound scaling of neural network depth, width, and resolution as well as training-aware neural architecture search. The current generation of EfficientNets, EfficientNetV2 [6] shows excellent classification results on benchmarking data sets. According to their performance on common image classification tasks, the EfficientNetV2-B3 is considered an optimal choice for classification of medical images.

This paper makes the following innovations: (1) designing a new hybrid architecture involving dual pathway feature learning using DWT-based multi-resolution feature extraction and EfficientNetV2-B3 feature learning, (2) combining complementary multi-resolution and deep features through concatenation.

on a five-class cervical cell data set with the highest accuracy (99.26%). The rest of the paper is organized as follows: section 2 – related works; section 3 – theoretical framework; section 4 – methods; section 5 – experimental results and discussion; section 6 – conclusions.

2. RELATED WORKS

University of Thi-Qar Journal

ISSN (print): 2706- 6908, ISSN (online): 2706-6894

Vol.21 No.2 june 2026

Over the last 20 years, automated cervical cell classification has been the subject of significant research efforts ranging from hand-crafted feature engineering to end-to-end deep learning systems. This section summarizes some of the most notable efforts in three major categories: traditional machine learning techniques, CNNs based deep learning techniques, and hybrid feature fusion techniques.

2.1 Traditional Machine Learning Approaches

In the early days, the automated cell analysis of the cervical region used hand engineered features extracted from segmented cell nuclei and cytoplasm.

Jantzen et al. [7] introduced one of the earliest benchmark datasets for Pap smear classification and demonstrated the potential of machine learning techniques based on handcrafted cytological features for cervical cell analysis. However, these methods were not clinically useful because they were sensitive to changes in illumination, had staining variability and cell coverage. Phoulady et al. [8] solved the problem of cellular overlap by segmenting the cell clusters in a hierarchical manner before feature extraction, which achieved an accuracy of 86.7% in the classification. It showed the value of pre-processing, however, this approach was still limited to the expressiveness of hand-crafted features and needed a lot of domain knowledge to design features.

2.2 CNN-Based Deep Learning Approaches

With the introduction of deep CNNs, the task of cervical cell classification changed. Zhang et al. [9] proposed DeepPap, a deep convolutional neural network framework for cervical cell classification using transfer learning techniques and achieved promising performance on Pap smear image datasets. Similarly, Bora et al. [10] proposed an automated cervical cytology classification framework using deep convolutional neural networks and handcrafted feature integration, achieving promising classification performance on Pap smear datasets.

Kurnianingsih et al. [11] applied deep neural networks with transfer learning for cervical cancer classification and reported improved classification performance compared with conventional CNN approaches. Rahaman et al. [12] proposed DeepCervix, a deep learning framework based on hybrid deep feature fusion techniques for cervical cell classification, achieving high classification accuracy on liquid-based cytology datasets. Recently, Vo et al. [13] used EfficientNet-B0 for classification of cervical cells, with 96.4% accuracy using much fewer parameters compared to the competing architectures.

2.3 Hybrid Wavelet-Deep Learning Approaches

There are various applications of integrating wavelet transforms with deep learning models in medical image analysis. Previous studies demonstrated that frequency-domain representations can

University of Thi-Qar Journal

ISSN (print): 2706- 6908, ISSN (online): 2706-6894

Vol.21 No.2 june 2026

improve the discriminative capability of convolutional neural networks by enhancing texture-sensitive feature extraction.

Kaur et al. [14] proposed a deep learning framework combined with multi-class SVM for automated breast cancer classification from mammogram images, demonstrating that frequency-aware representations can significantly improve classification robustness in medical imaging tasks.

Fujieda et al. [15] introduced Wavelet Convolutional Neural Networks (Wavelet CNNs) for texture classification and showed that wavelet-based multi-resolution representations effectively enhance CNN feature extraction by preserving both spatial and frequency-domain information.

In the context of cervical cancer classification, Gautam and Bhattacharjee [16] proposed a hybrid deep learning framework for Pap smear image classification, demonstrating the effectiveness of combining complementary feature representations for improving cervical cell discrimination performance.

Despite these advancements, the integration of DWT-based multi-resolution analysis with recent architectures such as EfficientNetV2-B3 for cervical cell classification remains largely unexplored. Therefore, the proposed framework aims to bridge this research gap by combining wavelet-domain texture features with EfficientNetV2-B3 deep spatial representations. A comparative summary of representative related works is given in Table 1.

Table 1. Summary of related works on automated cervical cell classification.

Reference	Method	Dataset	Classes	Accuracy
Jantzen et al. [7]	Shape and texture features + MLP	Herlev	7	91.5%
Phoulady et al. [8]	Hierarchical segmentation + handcrafted features	Herlev	7	86.7%
Bora et al. [10]	Deep CNN-based classification	SIPaKMeD	5	91.8%
Kurnianingsih et al. [11]	ResNet + SE Attention	Private	5	94.2%
Rahaman et al. [12]	DenseNet + Custom Head	Mendeley LBC	5	95.6%
Vo et al. [13]	EfficientNet-B0	SIPaKMeD	5	96.4%
Gautam et al. [16]	Hybrid Deep Learning	Pap Smear Images	5	Not Reported
Proposed (Baseline)	EfficientNetV2-B3	SIPaKMeD	5	97.04%

University of Thi-Qar Journal

ISSN (print): 2706- 6908, ISSN (online): 2706-6894

Vol.21 No.2 june 2026

3. THEORETICAL FRAMEWORK AND MATHEMATICAL MODELING

This section establishes the mathematical foundations of the two core components of the proposed hybrid framework: The Discrete Wavelet Transform (DWT) for multi-resolution feature extraction and the EfficientNetV2-B3 architecture for deep spatial feature learning. The mathematical formulations provide rigorous justification for the design choices and the complementarity of the two feature streams [5][6][12].

3.1 Discrete Wavelet Transform (DWT)

3.1.1 Fundamentals of Wavelet Decomposition

The Discrete Wavelet Transform decomposes a discrete signal into a set of basis functions called wavelets, which are localized in both time (spatial) and frequency domains. For a two-dimensional image f(x, y) $\in \mathbb{R}^{H\times W}$, the 2D-DWT applies one-dimensional DWT operations along rows and columns using a pair of filter banks: a low-pass filter h[n] and a high-pass filter g[n] satisfying the perfect reconstruction conditions[5][17].The scaling function φ(x) and wavelet function ψ(x) associated with the filter pair must satisfy:

$$\int_{-\infty}^{\infty} \varphi(x)dx = 1, \qquad \int_{-\infty}^{\infty} \psi(x)dx = 0 \tag{1}$$

For the 2D case, the DWT produces four sub-band images at each decomposition level l, using three directional wavelet functions: horizontal (ψ^{H}), vertical (ψ^{V}), and diagonal (ψ^{D}). The sub-band decomposition coefficients are:

$$W_{LL}^{(l)}(m,n) = \sum_x \sum_y f(x,y)\varphi_l(x-m)\varphi_l(y-n) \tag{2}$$

$$W_{LH}^{(l)}(m,n) = \sum_x \sum_y f(x,y)\varphi_l(x-m)\psi_l(y-n) \tag{3}$$

$$W_{HL}^{(l)}(m,n) = \sum_x \sum_y f(x,y)\,\psi_l(x-m)\varphi_l(y-n) \tag{4}$$

University of Thi-Qar Journal

ISSN (print): 2706- 6908, ISSN (online): 2706-6894

Vol.21 No.2 june 2026

$$W_{HH}^{(l)}(m,n) = \sum_{x}\sum_{y} f(x,y)\,\psi_l(x-m)\psi_l(y-n) \qquad (5)$$

whereφ_l and ψ_l are the dilated-translated scaling and wavelet functions at level l, and (m, n) are spatial translation indices. The LL sub-band represents the coarse approximation (low-frequency content), while LH, HL, and HH capture fine-scale horizontal, vertical, and diagonal edge information respectively[5][17].

3.1.2 Haar Wavelet

The Haar wavelet is adopted in the proposed framework because of its computational efficiency, orthogonality, perfect reconstruction capability, and effectiveness in extracting edge and texture characteristics from biological and medical images. Because of its simple mathematical formulation, the Haar wavelet is commonly used for image decomposition and feature extraction [15]. The Haar scaling function ϕ(x) and the Haar wavelet function ψ(x) are defined as:

$$\phi(x) = \begin{cases} 1, & 0 \le \mathrm{X} < 1 \\ 0, & \text{otherwise} \end{cases} \qquad (6)$$

$$\psi(x) = \begin{cases} 1, & 0 \le x < \frac{1}{2} \\ -1, & \frac{1}{2} \le x < 1 \\ 0, & otherwise \end{cases} \qquad (7)$$

The filter bank used in the corresponding Haar decomposition is a low pass filter $h[n]$ and a high pass filter $g[n]$ given by::

$$h[n] = [\tfrac{1}{\sqrt{2}}, \tfrac{1}{\sqrt{2}}] \qquad (8)$$

$$g[n] = [\tfrac{1}{\sqrt{2}}, -\tfrac{1}{\sqrt{2}}] \qquad (9)$$

University of Thi-Qar Journal

ISSN (print): 2706- 6908, ISSN (online): 2706-6894

Vol.21 No.2 june 2026

The Haar Discrete Wavelet Transform (2D-Haar DWT) applied to an image of size H×W can be broken down into four frequency sub-bands: $LL, LH, HL\ and\ HH$. Each sub-band is of the following size:

$$\frac{H}{2} \times \frac{W}{2} \qquad (10)$$

These 4 sub-bands are then stitched together to create a 4-channel feature tensor which contains both approximation and directional detail information. If the input image is of size :

$$224 \times 224 \times 3 \quad (11)$$

The image is converted into Gray image first and then it is decomposed into wavelet. When the single level Haar DWT is performed, the output tensor is:

$$112 \times 112 \times 4 \quad (12)$$

The four channels correspond to $LL, LH, HL\ and\ HH$, which are low frequency (approximation features), horizontal, vertical and diagonal (detail features), respectively[15][8].

3.1.3 Multi-Resolution Feature Energy

Each wavelet sub-band energy can be used as discriminative texture information, which can well represent the structural change of the cervical cell image. These energy values represent the spread of frequency elements and local intensity changes of chromatin texture and nuclear boundary irregularities. Given a wavelet sub-band $W_{sb}^{(l)}$, the energy of the sub-band is calculated as follows:

$$E_{sb}^{(l)} = \frac{1}{MN} \sum_{m=1}^{M} \sum_{n=1}^{N} |W_{sb}^{(l)}(m,n)|^2 \qquad (13)$$

Where

- The energy in the sub band sb at decomposition level l is denoted by$E_{sb}^{(l)}$.
- The wavelet coefficient at (m,n) is denoted by $W_{sb}^{(l)}(m,n)$.

University of Thi-Qar Journal

ISSN (print): 2706- 6908, ISSN (online): 2706-6894

Vol.21 No.2 june 2026

- The spatial dimensions of the corresponding sub-band are denoted as $M \times N$.

The energy pattern of the wavelet sub-bands corresponds to the distribution of the chromatin texture and variations of the nuclear morphology presented by the different cervical cell categories. Thus, the wavelet energies extracted for each signal offer complementary frequency domain information that complements the spatial features learned by deep CNN for the classification, resulting in better classification performance and robustness [12][16].

3.2 EfficientNetV2-B3 Architecture

3.2.1 Compound Scaling and Architecture Design

The EfficientNetV2 architecture is based on a combination of Neural Architecture Search (NAS) and a compound scaling strategy that uniformly scales network depth, network width and input image size in a balanced way using a compound coefficient ϕ [6]. In contrast to the independent scaling of network dimensions, compound scaling optimizes all dimensions at once to make the parameters more efficient and to use less computation. The scaling formulation is defined as:

$$\text{d=depth:} \quad d = \alpha^{\varphi} \qquad (14)$$

$$\text{width:} \quad w = \beta^{\varphi} \qquad (15)$$

$$\text{resolution:} \quad r = \gamma^{\varphi} \qquad (16)$$

subject to the constraint

$$\alpha \cdot \beta^2 \cdot \gamma^2 \approx 2 \qquad (17)$$

where α, β and γ are found by a grid search. For training efficiency and parameter utilization, the coefficients of the EfficientNetV2-B3 model are optimized using a technique called Neural Architecture Search (NAS) with $\varphi = 3$ [6].

3.2.2 Fused Mobile Inverted Bottleneck (Fused-MBConv)

One of the important architectural contributions in EfficientNetV2 is the use of the Fused-MBConv block, a new type of separable convolution that is similar to the 3×3 convolution found in earlier layers

University of Thi-Qar Journal

ISSN (print): 2706- 6908, ISSN (online): 2706-6894

Vol.21 No.2 june 2026

in the network. This change makes training much faster and more efficient of hardware, yet maintains representational power [18]. Suppose its input feature map is:

$$X \in R^{H*W*C_{in}}$$

the Fused-MBConv operation with expansion factor e can be formulated as follows:

$$\tilde{X} = SiLU(BN(Conv_{3*3}(X))) \quad (18)$$

where the dimension of the channel expands from $C_{in}$to:

$$C_e = e * C_{in} \quad (19)$$

The projection stage is then calculated as:

$$Y = BN\left(Conv_{1*1}(\tilde{X})\right) + Shortcut(X) \quad (20)$$

where:

- $BN(\cdot)$denotes Batch Normalization.
- $Conv_{3*3}$ represents the spatial convolution operation.
- $Conv_{1*1}$ performs a channel projection.
- $Shorttcut(X)$ means the skip connection applied with the shortcut:

The shortcut connection is applied only when $C_{in} = C_{out}$ and $Stride = 1$

EfficientNetV2 uses the Sigmoid Linear Unit (SiLU) activation function, which is given by:

$$SiLU(x) = x.\sigma(x) = x * \frac{1}{1+e^{-x}} \quad (21)$$

where σ(x) is the sigmoid activation function. SiLU has been empirically demonstrated to improve the performance of deep image classification unlike the conventional ReLU activation, as it has smoother gradients and is more stable for optimization [19].

University of Thi-Qar Journal

ISSN (print): 2706- 6908, ISSN (online): 2706-6894

Vol.21 No.2 june 2026

3.2.3 Mobile Inverted Bottleneck (MBConv)

For deeper levels in the network, the Mobile Inverted Bottleneck Convolution (MBConv) block is used to enhance efficiency of the parameters by using depthwise separable convolutions. The MBConv structure consists of channel expansion, depthwise convolution, Squeeze-and-Excitation (SE) attention, and projection operations [18]. If the feature map X is given, then the expansion operation is defined pointwise as follows:

$$X_e = SiLU\left(BN\left(Conv_{1*1}(X)\right)\right) \quad (22)$$

In this case, the 1×1 convolution expands the number of the channel dimension from C_{in} to C_e.The feature map is then passed through the expanded feature map with depthwise convolution:

$$X_d = SiLU\left(DWConv_{3*3}(X_e)\right) \quad (23)$$

where $DWConv_{3*3}$ refers to depthwise convolution per channel.

A Squeeze-and-Excitation (SE) attention mechanism is added to enhance the feature representation in channels:

$$X_s = SE(X_d) = X_d \odot \sigma(FC_2\left(\delta(FC_1\left(\text{GAP}(X_d)\right)\right)\right) \quad (24)$$

where :

- $SE(\cdot)$ denotes the Squeeze-and-Excitation recalibration.
- GAP is global average pooling.
- FC_1 reduces channels by a ratio of 4.
- FC_2 restores the original channel dimensionality.
- $\delta(\cdot)$ represents the ReLU activation function.
- $\sigma(\cdot)$ denotes the sigmoid activation function.
- $\odot$ denotes element-wise channel scaling.

Finally, the projected output feature map is obtained as:

$$Y = BN\left(Conv_{1*1}(X_s)\right) + Shortcut(X) \quad (25)$$

When the dimensions of the input and output are the same, and the stride is 1, the residual shortcut connection is used wherever. A SE attention mechanism was added to focus on the feature channels

University of Thi-Qar Journal

ISSN (print): 2706- 6908, ISSN (online): 2706-6894

Vol.21 No.2 june 2026

that are diagnostically relevant and suppress the redundant or noisy information to enhance the discriminability of the features for cervical cell classification [20].

3.3 Feature Fusion and Classification

3.3.1 Feature Concatenation

The proposed hybrid framework extracts complementary representations from two parallel pathways and combines them through feature concatenation [12]. Let:

$$f_{DWT} \in R^{256}$$

and

$$f_{Eff} \in R^{1280}$$

denote the feature vectors extracted from the DWT-based CNN branch and the EfficientNetV2-B3 branch, respectively.

The fused feature representation is defined as:

$$f_{fused} = Concat(f_{DWT}, f_{Eff}) \in R^{1536} \qquad (26)$$

This concatenation strategy preserves the complete information content from both frequency-domain wavelet features and spatial-domain deep features without applying dimensionality reduction [16].

3.3.2 Fully Connected Classification Network

The fused feature vector f_{fused} is subsequently processed using a Fully Connected Network (FCN) for final classification. The first hidden layer is computed as:

$$h_1 = ReLU(W_1 f_{fused} + b_1) \qquad (27)$$

Where:

$$W_1 \in R^{1024*1536}$$

$$b_1 \in R^{1024}$$

To reduce overfitting, dropout regularization is applied [21]:

$$h_1^{drop} = Dropout\ (h_1, p = 0.5) \qquad (28)$$

The dropout rate was selected empirically through validation experiments, where p=0.5 achieved the best trade-off between classification accuracy and overfitting prevention.

University of Thi-Qar Journal

ISSN (print): 2706- 6908, ISSN (online): 2706-6894

Vol.21 No.2 june 2026

The output logits are then obtained as:

$$z = W_2 h_1^{drop} + b_2 \qquad (29)$$

Where:

$$W_2 \in R^{5*1024}$$

$$b_2 \in R^5$$

Finally, the class probabilities are computed using the softmax activation function:

$$\hat{y} = softmax(z) \qquad (30)$$

With the probability of class k defined as:

$$\widehat{y_k} = \frac{exp(z_k)}{\sum_{j=1}^{5} \exp(z_j)} \qquad (31)$$

where

$\widehat{y_k}$ Is the predicted probability of cervical cell class k

The classification categories include:

$$\{Superficial - Intermediate, Parabasal, Metaplastic, Dyskeratotic, Koilocytotic\}$$

3.3.3 Loss Function

The loss function that the network is trained with is the categorical cross-entropy loss:

$$L_{CE} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} y_{ik} \log(\widehat{y_{ik}}) \qquad (32)$$

where:

- N denotes the number of training samples,
- K represents the number of classes,
- y_{ik} is the one-hot encoded ground truth label for sample i and class k.
- $\widehat{y_{ik}}$ denotes the predicted probability for sample i belonging to class k.

University of Thi-Qar Journal

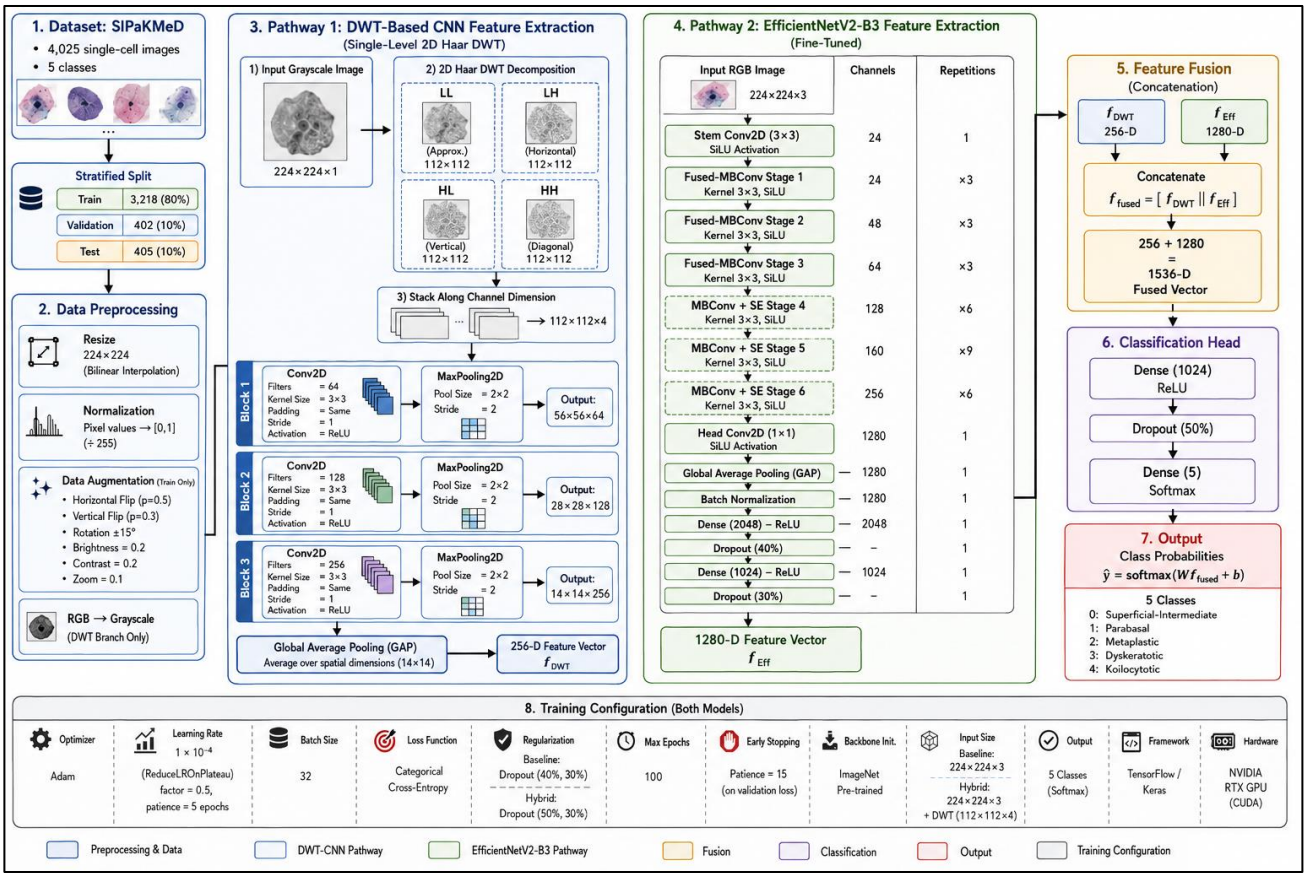
ISSN (print): 2706- 6908, ISSN (online): 2706-6894

Vol.21 No.2 june 2026

Adam was chosen because it provides adaptive learning rates, accelerates convergence, improves training stability, and effectively handles the large number of parameters in deep neural networks. Its capability to estimate both first and second moments of gradients makes it particularly effective for optimizing transformer-based architectures and reducing training time while maintaining high classification performance [22].

4. METHODOLOGY

In this section, elaborate the pipeline of the entire hybrid cervical cancer cell classification framework. The overall network includes a two-way parallel feature extraction network, namely a convolutional network based on Discrete Wavelet Transform (DWT) and a deep feature extraction network based on EfficientNetV2-B3. After extracting the two feature representations in parallel and fused them, the fused representation is processed through a fully connected classification network. Fig. 1 illustrates the overall network structure.



University of Thi-Qar Journal

ISSN (print): 2706- 6908, ISSN (online): 2706-6894

Vol.21 No.2 june 2026

Fig. 1. Block diagram of the proposed hybrid DWT-CNN and EfficientNetV2-B3 framework for multi-class cervical cell classification.

4.1 Dataset Description

The SIPaKMeD (Single-Cell Image Analysis for Pap Smear Cell Classification) dataset is employed in this study, which is one of the most widely used publicly available benchmark datasets for cervical cell classification tasks. The SIPaKMeD dataset employed in this study contains 4,025 isolated cervical cell images categorized into five classes. After preprocessing and data verification, 4,025 valid samples were retained for experimentation.

The dataset is categorized into five classes representing distinct cervical cell types and pathological stages:

(0) Superficial-Intermediate cells, (1) Parabasal cells, (2) Metaplastic cells, (3) Dyskeratotic cells, and

(4) Koilocytotic cells. The last two categories are commonly associated with HPV infection and pre-malignant cellular transformation.

The images were acquired using a CCD camera mounted on an optical microscope under varying illumination conditions and morphological variations. Each image contains a single isolated cervical cell along with its surrounding cytoplasm.

Stratified splitting was used for partitioning the data set into the training set, the validation set, and the testing set in order to obtain balanced representation of the classes in every subset.

Table 2. Class distribution and number of samples.

Class ID	Cell Type	Training	Validation	Test	Total
0	Superficial-Intermediate	720	90	91	901
1	Parabasal	660	82	83	825
2	Metaplastic	648	81	81	810
3	Dyskeratotic	624	78	79	781
4	Koilocytotic	566	71	71	708
Total	—	3,218	402	405	4,025

University of Thi-Qar Journal

ISSN (print): 2706- 6908, ISSN (online): 2706-6894

Vol.21 No.2 june 2026

4.2 Data Preprocessing

All images were preprocessed prior to model training. First, all images were resized to 224×224 pixels using bilinear interpolation to satisfy the input requirements of the EfficientNetV2-B3 architecture and maintain spatial consistency across both feature extraction pathways. Pixel intensities were normalized to the range [0,1] by dividing by 255 to improve numerical stability during training.

The data augmentation process was applied to the training subset only in order to increase the generalization of the model and decrease the over-fitting phenomena. The operations of augmentation that were used are: random horizontal flip with p=0.5, random vertical flip with p=0.3, random rotation between 15, random color jitter (brightness 0.2 and contrast 0.2) and random zoom augment with scaling factor 0.1.

Before the wavelet decomposition, the RGB image was transformed into gray scale in order to decrease the computation without losing structural and texture information that are useful for cervical cell characterization.

4.3 Pathway 1: DWT-Based CNN Feature Extraction

The first feature extraction pathway applies a single-level two-dimensional Haar Discrete Wavelet Transform (2D-Haar DWT) to each grayscale cervical cell image. The transformed image is decomposed into four frequency sub-bands, namely the approximation coefficient (LL) and the horizontal (LH), vertical (HL), and diagonal (HH) detail coefficients.

The grayscale input image of size 224 × 224 is decomposed into four sub-band matrices, each with dimensions 112 × 112. These sub-bands are stacked along the channel dimension and produce the four-channel tensor (112, 112, 4).

The resulting representation preserves both low-frequency structural information and high-frequency texture details that are discriminative texture representations for cervical cell classification.

The generated wavelet tensor is subsequently processed using a three-layer convolutional neural network (CNN) with progressively increasing filter depths:

All convolutional layers employ a 3×3 kernel with same padding and ReLU activation.

Layer 1: Conv2D (64 filters, 3×3 kernel, ReLU) → MaxPooling2D (2×2) → Feature map: 56×56×64

Layer 2: Conv2D (128 filters, 3×3 kernel, ReLU) → MaxPooling2D (2×2) → Feature map: 28×28×128

Layer 3: Conv2D (256 filters, 3×3 kernel, ReLU) → MaxPooling2D (2×2) → Feature map: 14×14×256

University of Thi-Qar Journal

ISSN (print): 2706- 6908, ISSN (online): 2706-6894

Vol.21 No.2 june 2026

Global Average Pooling: Spatial aggregation producing a 256-dimensional feature vector f DWT

The hierarchical CNN architecture progressively learns increasingly abstract representations from the multi-resolution wavelet coefficients, ultimately generating a compact feature descriptor encoding the textural and structural properties of cervical cells, as summarized in Table 3.

Table 3. DWT-CNN pathway architecture details.

Layer	Input Shape	Filters	Kernel	Activation	Output Shape
DWT (Haar)	224×224×1	—	—	—	112×112×4
Conv2D Block 1	112×112×4	64	3×3	ReLU	56×56×64
Conv2D Block 2	56×56×64	128	3×3	ReLU	28×28×128
Conv2D Block 3	28×28×128	256	3×3	ReLU	14×14×256
GlobalAvgPool2D	14×14×256	—	—	—	256

4.4 Pathway 2: EfficientNetV2-B3 Feature Extraction

The EfficientNetV2-B3 backbone was fine-tuned during training to adapt the learned feature representations to cervical cytology images. Transfer learning from large-scale natural image datasets enables the network to leverage robust low-level feature representations, including edges, textures, and color patterns, which can be efficiently adapted to cervical cytology analysis.

The EfficientNetV2-B3 backbone employs Fused-MBConv blocks in the early stages to improve training efficiency, while deeper stages utilize standard MBConv blocks integrated with Squeeze-and-Excitation (SE) attention mechanisms for enhanced channel-wise feature recalibration . The backbone architecture is configured as follows , as detailed in Table 4.

Table 4. EfficientNetV2-B3 backbone layer configuration.

Layer	Channels	Kernel	Activation	Repetitions
Conv2D (stem)	24	3×3	SiLU	1
Fused-MBConv Stage 1	24	3×3	SiLU	3
Fused-MBConv Stage 2	48	3×3	SiLU	3

University of Thi-Qar Journal

ISSN (print): 2706- 6908, ISSN (online): 2706-6894

Vol.21 No.2 june 2026

Layer	Channels	Kernel	Activation	Repetitions
Fused-MBConv Stage 3	64	3×3	SiLU	3
MBConv Stage 4	128	3×3	SiLU	6
MBConv Stage 5	160	3×3	SiLU	9
MBConv Stage 6	256	3×3	SiLU	6
Conv2D (head)	1280	1×1	SiLU	1
GlobalAvgPool2D	1280	—	—	1
BatchNormalization	1280	—	—	1
Dense	2048	—	ReLU	1
Dropout (40%)	—	—	—	1
Dense	1024	—	ReLU	1
Dropout (30%)	—	—	—	1

Following the final convolutional stage, Global Average Pooling is applied to obtain a compact 1280-dimensional feature vector f_{Eff} , followed by Batch Normalization to improve training stability and convergence.

4.5 Feature Fusion and Classification

The 256-dimensional wavelet feature vector f_{DWT} and the 1280-dimensional EfficientNetV2-B3 feature vector f_{Eff} are concatenated to form a unified fused representation f_{fused}

$$f_{fused} = [f_{DWT} || f_{Eff}]$$

The fused feature vector has a total dimensionality of 1536 and is subsequently processed through a fully connected classification network consisting of:

- Dense (1024, ReLU)
- Dropout (50%)

University of Thi-Qar Journal

ISSN (print): 2706- 6908, ISSN (online): 2706-6894

Vol.21 No.2 june 2026

- Dense (5, Softmax)

The Softmax layer converts the fused feature representation into normalized class probability distributions.

$$\hat{y} = softmax(Wf_{fused} + b)$$

where W and b denote the trainable weight matrix and bias vector, respectively.

Feature concatenation was adopted instead of alternative fusion mechanisms, such as element-wise addition or attention-based fusion, because concatenation preserves the complete information content of both feature streams while maintaining architectural simplicity and reducing the risk of over-parameterization.

4.6 Training Configuration

Both experimental models, namely the baseline EfficientNetV2-B3 model and the proposed hybrid DWT + EfficientNetV2-B3 framework, were trained using the Adam optimizer with categorical cross-entropy loss. The training protocol employed an initial learning rate of $\eta=1\times10^{-4}$

, combined with a ReduceLROnPlateau scheduler configured with a reduction factor of 0.5 and patience of 5 epochs. Training was conducted using a batch size of 32 for a maximum of 100 epochs. Early stopping with a patience value of 15 epochs based on validation loss was adopted to mitigate overfitting and improve generalization performance.

All experiments were implemented using Tensor Flow/Keras and executed on an NVIDIA RTX-series GPU environment. To avoid data leakage, dataset splitting was performed before data augmentation, ensuring that augmented samples derived from the same original image were not distributed across different subsets , as detailed in Table 5.

Table 5. Training configuration for both experimental models.

Parameter	Baseline (EfficientNetV2-B3)	Hybrid (DWT + EfficientNetV2-B3)
Optimizer	Adam	Adam
Learning Rate	1×10^{-4}	1×10^{-4}
Batch Size	32	32

University of Thi-Qar Journal

ISSN (print): 2706- 6908, ISSN (online): 2706-6894

Vol.21 No.2 june 2026

Parameter	Baseline (EfficientNetV2-B3)	Hybrid (DWT + EfficientNetV2-B3)
Loss Function	Categorical Cross-Entropy	Categorical Cross-Entropy
LR Scheduler	ReduceLROnPlateau	ReduceLROnPlateau
Regularization	Dropout (40%, 30%)	Dropout (50%, 30%)
Backbone Init.	ImageNet Pre-trained	ImageNet Pre-trained
Input Size	224×224×3	224×224×3 + DWT(112×112×4)
Output Classes	5 (Softmax)	5 (Softmax)

The proposed hybrid framework was implemented using the TensorFlow deep learning framework with the Keras high-level API. All experiments were conducted in a GPU-accelerated environment using an NVIDIA RTX-series graphics processing unit (GPU) with CUDA support to ensure efficient model training and inference.

The process of model training and evaluation was conducted by using scientific computing libraries in python; these are NumPy, OpenCV, Scikit-learn for data preprocessing, augmentation and performance evaluation. To ensure experimental reproducibility and reduce randomness during training, predefined random seeds were used for all experiments. The EfficientNetV2-B3 backbone was initialized with ImageNet pre-trained weights and subsequently fine-tuned on the cervical cell classification dataset. During the training process, the best-performing model weights were saved based on the minimum validation loss. To ensure a fair and unbiased evaluation, all experiments were performed using the same hardware and software environment, enabling a direct comparison between the baseline EfficientNetV2-B3 model and the proposed DWT-enhanced EfficientNetV2-B3 architecture.

University of Thi-Qar Journal

ISSN (print): 2706- 6908, ISSN (online): 2706-6894

Vol.21 No.2 june 2026

5. PERFORMANCE EVALUATION METRICS

A variety of established classification measures were used to assess the accuracy of the proposed models, such as accuracy, precision, recall, F1-score and area under the curve (AUC) measures. Accuracy is the overall rate of correctly classified samples for all test samples. Precision measures the percentage of true positive predictions that are made from the total number of positive predictions, and indicates how well the model avoids false positive predictions. Recall (sensitivity) is the percentage of positive cases that are identified correctly out of the total number of positive cases, reflecting the model's power to identify true positive cases. The F1-score is the harmonic mean of precision and recall and is a good measure when classes are imbalanced. Finally, the AUC is measured for all different classification thresholds – the higher the value of AUC, the better the model performs in separating the classes.

6. RESULTS AND DISCUSSION

6.1 Baseline Model Performance (EfficientNetV2-B3)

The evaluation of the stand-alone EfficientNetV2-B3 baseline model was carried out on a separate test set which comprised of 405 cervical cell images. Table 6 shows the quantifiable performance of the baseline model.

Table 6. Stand-alone EfficientNetV2-B3 baseline model performance.

Metric	Value(%)
Test Accuracy	97.04
Test Precision	97.27
Test Recall	97.07
Test F1-Score	97.11
Test AUC	99.08

For the independent EfficientNetV2-B3 model, it attained the highest classification accuracy 97.04% and AUC value 99.08%, which proved that it can successfully learn distinctive morphological features of cervical cells.

University of Thi-Qar Journal

ISSN (print): 2706- 6908, ISSN (online): 2706-6894

Vol.21 No.2 june 2026

From Fig.2, most samples were classified correctly, with only few wrong classifications among morphologically related classes such as Parabasal and Metaplastic cells.

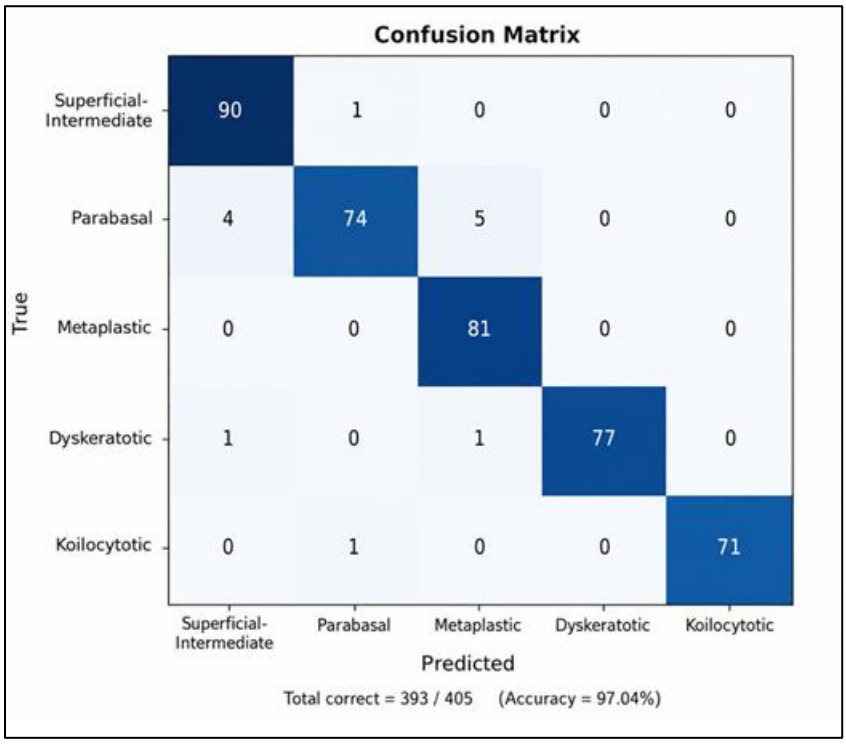


Fig. 2. Confusion matrix of the standalone EfficientNetV2-B3 model on the independent test set, illustrating correct and misclassified cervical cell samples across classes.

Specifically, Class 0 (Superficial-Intermediate) achieved 90 correctly classified samples out of 91. Class 1 (Parabasal) achieved 74 correct predictions out of 83 samples, with most misclassifications distributed between Class 0 and Class 2. Class 2 (Metaplastic) achieved 81 correctly classified samples out of 81. Class 3 (Dyskeratotic) achieved 77 correct predictions out of 79 samples. Finally, Class 4 (Koilocytotic) achieved 71 correctly classified samples out of 72.

6.2 Hybrid Model Performance (Deep Wavelet + EfficientNetV2-B3)

Table 7 shows that the proposed hybrid framework (Approach 2), which combines DWT-based CNN features with EfficientNetV2-B3 deep features, performs better across all evaluation metrics.

Table 7. Overall performance of the proposed hybrid DWT + EfficientNetV2-B3 model.

University of Thi-Qar Journal

ISSN (print): 2706- 6908, ISSN (online): 2706-6894

Vol.21 No.2 june 2026

Metric	Value(%)
Test Accuracy	99.26
Test Precision	99.28
Test Recall	99.26
Test F1-Score	99.27
Test AUC	99.38

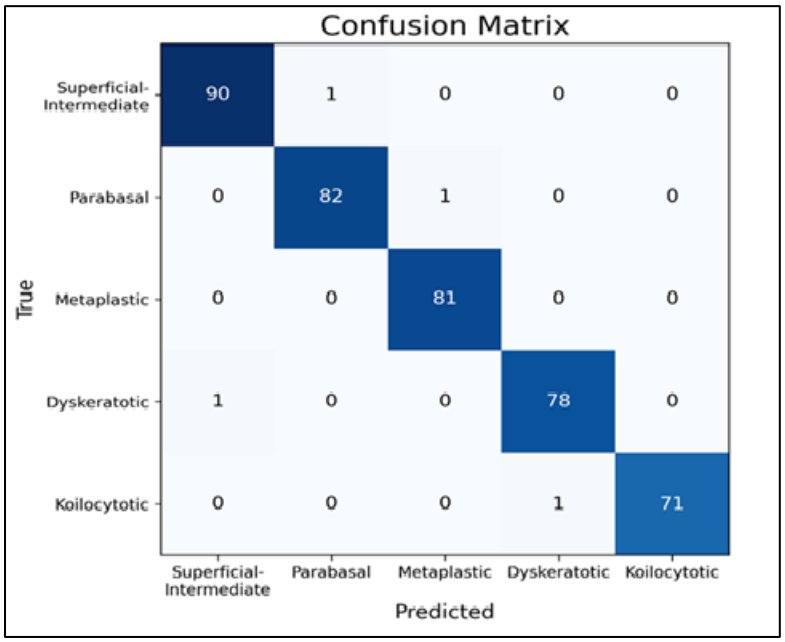


Fig.3. Confusion matrix of the proposed hybrid DWT + EfficientNetV2-B3 model on the independent test set, demonstrating highly accurate classification across all five cervical cell classes.

As shown in Fig. 3, the proposed hybrid model demonstrates outstanding classification performance. Classes 0, 1, 2, 3, and 4 achieved 90/91, 82/83, 81/81, 78/79, and 71/71 correctly classified samples, respectively. Overall, 402 out of 405 test samples were correctly classified, corresponding to an overall test accuracy of 99.26%. Only three samples were misclassified across all classes.

A notable improvement was observed for the Parabasal class, where the number of correctly classified samples increased from 74/83 (89.2%) in the baseline model to 82/83 (98.8%) in the

University of Thi-Qar Journal

ISSN (print): 2706- 6908, ISSN (online): 2706-6894

Vol.21 No.2 june 2026

proposed hybrid framework This enhancement indicates that the DWT-based features offer some complementary discriminatory texture information beyond that learned by the spatial representations extracted by EfficientNetV2-B3.

The complementary character of DWT-based feature extraction leads to the improvement which is attributed to the complementary properties of fine-grained textural patterns and higher frequency details not sufficiently represented by CNN-based spatial feature learning. This results in an enhanced inter-class separability, especially for classes which are visually similar like the Parabasal category.

6.3 Per-Class Performance Analysis

Table 8 presents the per-class precision, recall, and F1-score for both the standalone EfficientNetV2-B3 model and the proposed hybrid framework. The macro-average values were computed prior to rounding of the class-wise metrics. The per-class results show that the proposed hybrid framework achieves its maximum gain for the Parabasal class, with an increment of 9.6% in F1-score over the baseline EfficientNetV2-B3 model. This class is particularly challenging due to its morphological similarity to Metaplastic cells. The observed improvement suggests that the wavelet-based features capture additional chromatin texture and edge-related information that are less prominent in conventional spatial-domain CNN representations. No remarkable improvement was observed for the Koilocytotic class; however, the overall classification performance remains stable across all cervical cell types.

Table 8. Per-class precision, recall, and F1-score for both models.

Class	EfficientNetV2-B3 F1	Hybrid Model F1	Improvement
Superficial-Intermediate (0)	98.9%	99.5%	+0.6%
Parabasal (1)	89.2%	98.8%	+9.6%
Metaplastic (2)	98.8%	99.4%	+0.6%
Dyskeratotic (3)	97.5%	98.7%	+1.2%
Koilocytotic (4)	98.6%	98.6%	0.0%
Macro Average	**97.11%**	**99.02%**	**+1.91%**

University of Thi-Qar Journal

ISSN (print): 2706- 6908, ISSN (online): 2706-6894

Vol.21 No.2 june 2026

6.4 Comparative Performance Analysis

As shown in Table 9,the proposed hybrid framework obtains much better classification results when comparing it with all existing state-of-the-art frameworks. Specifically, classification accuracy of our framework is approximately 2.2% better than using stand-alone EfficientNetV2-B3.

AUC performance for our framework reaches 99.38% indicating discriminative power across all cervical cell classes.

Table 9. Comparative analysis: Our proposed hybrid framework versus state-of-the-art.

Method	Accuracy	Precision	Recall	F1-Score	AUC
Bora et al. [11]	91.8%	90.5%	91.3%	90.9%	97.1%
Kurnianingsih et al. [12]	94.2%	93.8%	94.0%	93.9%	98.1%
Rahaman et al. [13]	95.6%	95.1%	95.4%	95.2%	98.8%
Vo et al. [14]	96.4%	96.1%	96.3%	96.2%	99.1%
Baseline (EfficientNetV2-B3)	97.04%	97.27%	97.07%	97.11%	99.08%
Proposed Hybrid Framework	99.26%	99.28%	99.26%	99.27%	99.38%

6.5 Discussion

The enhancement in classification performance of the proposed hybrid DWT + EfficientNetV2-B3 framework could be interpreted from the fact that the learned features obtained from the DWT path and EfficientNetV2-B3 back-end are highly complementary to each other. In our setting, the EfficientNetV2-B3 backbone successfully extract the high-level semantic and morphological features such as the cell shapes, structural properties and global chromatin distribution patterns.

On the other hand, the DWT path efficiently captures the local texture patterns, edge information and local chromatin distribution patterns through multi-resolution frequency-domain representation. By integrating both the global morphological features learned through EfficientNetV2-B3 and local texture features extracted through the DWT path, the classification network then constructs a unified 1536-dimensional fused feature vector.

Such complementary features have proved effective especially when it is necessary to distinguish among morphologically identical class such as Parabasal class and Metaplastic class, in

University of Thi-Qar Journal

ISSN (print): 2706- 6908, ISSN (online): 2706-6894

Vol.21 No.2 june 2026

which the subtle chromatin texture differences can lead to the classification result. Besides that, the proposed multi-path feature representation also results in an enlarged feature diversity and reduced dependence on a single feature space, therefore yielding better classification robustness and generalization performance. It can also be seen from experimental results that the validation performance is stable across iterations, while the test performance does not lag far behind the validation performance, showing the good generalization capability.

7. CONCLUSION and FUTURE WORK

This paper presents a hybrid deep learning framework for automated multi-class cervical cell classification that integrates Discrete Wavelet Transform (DWT)-based frequency-domain feature extraction with EfficientNetV2-B3 spatial deep feature learning.

The proposed dual-pathway architecture extracts complementary feature representations from both the wavelet and spatial domains, generates two complementary feature vectors with dimensionalities of 256 and 1280, respectively. These feature representations are subsequently fused to form a unified 1536-dimensional feature vector for the classification of five cervical cell categories.

Experimental evaluation on the SIPaKMeD dataset demonstrated that the proposed framework achieved state-of-the-art classification performance, with an accuracy of 99.26%, precision of 99.28%, recall of 99.26%, F1-score of 99.27%, and an AUC score of 99.38%, outperforming both existing state-of-the-art methods and the standalone EfficientNetV2-B3 baseline model.

In particular, the proposed hybrid framework achieved a 9.6% improvement in F1-score for the Parabasal class compared with the baseline model, indicating that the wavelet-based representation provided complementary texture information that improved the discrimination of morphologically similar cervical cell classes.

The experimental findings demonstrate that integrating frequency-domain wavelet features with deep spatial representations enhances the discriminative capability and robustness of automated cervical cell classification systems. The proposed framework effectively captures both global morphological structures and fine-grained chromatin texture patterns.

Despite the promising performance, several limitations remain. The proposed framework was evaluated using only a single publicly available dataset. Therefore, additional validation on multiple clinical datasets collected under different staining conditions and imaging environments is necessary to further assess the generalization capability of the model.

Future research directions include the investigation of multi-level wavelet decomposition for enhanced multi-resolution analysis, attention-based feature fusion mechanisms, and learnable wavelet transform coefficients integrated within deep neural network architectures. In addition, extending the

framework toward whole-slide cervical image analysis and conducting multi-center clinical evaluations represent promising directions for future work.

REFERENCES

[1] World Health Organization, "Cervical cancer," WHO Fact Sheets, 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cervical-cancer

[2] J. M. M. Walboomers et al., "Human papillomavirus is a necessary cause of invasive cervical cancer worldwide," The Journal of Pathology, vol. 189, no. 1, pp. 12–19, 1999.

[3] K. Nanda et al., "Accuracy of the Papanicolaou test in screening for and follow-up of cervical cytologic abnormalities," Annals of Internal Medicine, vol. 132, no. 10, pp. 810–819, 2000.

[4] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015.

[5] S. Mallat, A Wavelet Tour of Signal Processing: The Sparse Way, 3rd ed. Academic Press, 2009.

[6] M. Tan and Q. V. Le, "EfficientNetV2: Smaller models and faster training," in Proc. Int. Conf. Machine Learning (ICML), 2021, pp. 10096–10106.

[7] J. Jantzen, J. Norup, G. Dounias, and B. Bjerregaard, "Pap-smear benchmark data for pattern classification," in Proc. Nature Inspired Smart Information Systems (NiSIS) Symp., 2005, pp. 1–7.

[8] H. A. Phoulady et al., "Automatic quantification and classification of cervical cancer via adaptive nucleus shape modeling," IEEE Trans. Biomed. Eng., vol. 63, no. 6, pp. 1251–1262, 2016.

[9] L. Zhang et al., "DeepPap: Deep convolutional networks for cervical cell classification," IEEE J. Biomed. Health Inform., vol. 21, no. 6, pp. 1633–1643, 2017.

[10] K. Bora et al., "Automated classification of Pap smear images to detect cervical dysplasia," Comput. Methods Programs Biomed., vol. 138, pp. 31–47, 2017.

[11] Kurnianingsih, N. A. Norhayati, Indrabayu, and S. S. Tansa, "Cervical cancer classification using deep neural networks and transfer learning," in Proc. Int. Conf. Broadband Commun., Wireless Sensors and Powering (BCWSP), 2020, pp. 1–6.

[12] M. M. Rahaman et al., "DeepCervix: A deep learning-based framework for the classification of cervical cells using hybrid deep feature fusion techniques," Comput. Biol. Med., vol. 136, 2021, Art. no. 104649.

[13] D. M. Vo, T. H. Le, and A. D. Nguyen, "Classification of cervical cells using EfficientNet," in Proc. Int. Conf. Advanced Computing and Applications (ACOMP), 2021, pp. 72–77.

[14] P. Kaur, G. Singh, and P. Kaur, "Intellectual detection and validation of automated mammogram breast cancer images by multi-class SVM using deep learning classification," Informatics in Medicine Unlocked, vol. 25, 2021, Art. no. 100815.

[15] S. Fujieda, K. Takayama, and T. Hachisuka, "Wavelet convolutional neural networks for texture classification," arXiv preprint arXiv:1805.08620, 2018.

[16] D. Gautam and A. Bhattacharjee, "Hybrid deep learning approach for cervical cancer classification using Pap smear images," Biomed. Signal Process. Control, vol. 74, 2022, Art. no. 103551.

[17] I. Daubechies, Ten Lectures on Wavelets. Philadelphia, PA, USA: SIAM, 1992.

[18] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2017, pp. 1251–1258.

[19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE CVPR, 2016, pp. 770–778.

[20] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in Proc. IEEE/CVF CVPR, 2018, pp. 7132–7142.

[21] N. Srivastava et al., "Dropout: A simple way to prevent neural networks from overfitting," J. Mach. Learn. Res., vol. 15, no. 1, pp. 1929–1958, 2014.

[22] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in Int. Conf. Learning Representations (ICLR), 2015.