SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V11I11P119

Volume 11 Issue 11, 228-243, November 2024

Original Article

Variable Kernel Feature Fusion and Transfer Learning for Pap Smear Image-Based Cervical Cancer Classification

Priya S A[1], V Mary Amala Bai[2]

[1]*Department of Computer Applications, Noorul Islam Centre for Higher Education, Kumaracoil, India.*
[2]*Department of Information Technology, Noorul Islam Centre for Higher Education, Kumaracoil, India.*

[1]*Corresponding Author : priya.s.a@outlook.com*

Received: 21 September 2024 Revised: 20 October 2024 Accepted: 18 November 2024 Published: 03 December 2024

Abstract *- Cervical cancer, a malignant tumour that forms in the cervix, significantly contributes to cancer-related mortality among women globally, making early diagnosis crucial for effective treatment. Pap smear images, which are microscopic images of cervical cells, are commonly used for the detection of abnormal cells that may lead to cervical cancer. This study introduces a novel classification approach, the Variable Kernel Feature Fusion-CNN (VKFF-CNN), which improves classification performance by fusing multi-scale features using convolutional layers with 3x3, 4x4, and 5x5 kernels. This architecture captures a diverse set of features, enhancing the ability of the model to accurately classify cervical cells. With an average accuracy of 98.03%, precision of 97.83%, recall of 97.11%, and an F1 score of 98.23%, the VKFF-CNN exhibited outstanding outcomes on the Herlev Pap Smear dataset. These results demonstrate that VKFF-CNN significantly outperforms traditional machine learning models. The model's confusion matrix indicated fewer misclassifications, underscoring its robustness and effectiveness. Including batch normalization and the softmax activation function further enhanced the model's stability and accurate classification. Overall, VKFF-CNN presents a promising advancement for automated cervical cancer screening, providing highly accurate and reliable detection.*

Keywords *- Cervical Cancer, Herlev Pap Smear Dataset, Variable Kernel Feature Fusion-CNN, Multi-Scale Feature Extraction, Softmax activation.*

1. Introduction

Cervical cancer arises from the tissue of the cervical region, located at the bottom level of the uterus. Infected uterine cells undergo proliferation and development with aberrant cell cycles. Cervical abnormalities can be detected by testing precancerous lesions with several screening techniques. Prolonged manifestation of Human Papillomavirus (HPV) infection results in the development of cervical cancer through the formation of malignant lesions. As early-stage cervical cancer is a slow-growing malignancy, women do not experience any symptoms until it progresses to invasive cancer and metastasizes to other organs in the body.

Due to the extended premalignant phase of the disease, early onset diagnosis is entirely controllable and curable [1]. Figure 1 displays the various stages of cervical cancer. According to the World Health Organisation (WHO), cervical cancer is the fourth most common cancer among women globally, with an estimated 660,000 instances being registered in 2022 [2]. In that particular year, almost 94% of the 350,000 deaths caused by cervical cancer took place in countries characterized by low- and middle-income levels. The regions with the greatest rates of cervical cancer incidence and mortality include Central America, Sub-Saharan Africa (SSA), and South-East Asia. This results from inadequate awareness of this condition and limited access to healthcare services. Conversely, developed countries have established techniques that enable precise and effective screening tools, enabling early detection and treatment of precancerous lesions [3]. Early detection and treatment of premalignant tumours are broadly acknowledged to effectively prevent the spread of cancer in around 90% of cervical cancer patients. Therefore, the early identification of cervical cancer is crucial.

Cellular testing, particularly the Pap smear test, is a widely recognized and easily accessible screening method for identifying cervical cancer. Advances in medical technology facilitate the identification of diseases by analyzing abnormalities in the structure of cervical cells, such as the shape and color of the nucleus and the cytoplasm shown in the image [4]. The images obtained from a Pap smear can reveal details regarding the presence of cervical cancer [5]. The manual classification of Pap smear cell films presents significant challenges due to the subtle visual differences between cell groupings, particularly in the cell nucleus size, which can appear indistinguishable across categories. This makes the task of identifying tumours through manual analysis prone to errors, potentially leading to misdiagnosis

and delayed treatment [6]. Additionally, manual screening is time-consuming, lacks the comprehensive accuracy required, and is vulnerable to human bias and subjective errors [7]. The variability in interpretation among cytologists can further exacerbate inconsistencies in detecting abnormalities. These limitations highlight the critical need for developing a more precise and cost-effective screening approach to ensure early and precise identification of cervical cancer, eventually enhancing patient results.

Artificial Intelligence (AI) has demonstrated promising application value in the diagnosis of a variety of diseases over the past years, such as skin malignancies [8], retinal disorders [9], and the imaging detection of tumours [10]. AI systems can autonomously process images, extract features, and analyze data using advanced algorithms, which is especially beneficial for the timely identification and diagnostics of cervical cancer. Modern Computer-Aided Diagnosis (CAD) systems leverage Deep Learning (DL) techniques to effectively extract relevant features from medical images, offering superior performance compared to traditional methods [11]. Convolutional Neural Networks (CNNs), a prominent DL architecture, have demonstrated impressive results in numerous medical domains because of their capacity to learn and generalize complex patterns from large datasets [12].

However, CNN models require extensive labeled data to avoid overfitting and ensure robust generalization. Transfer Learning (TL) is employed to address this challenge, where a pre-trained CNN framework, originally trained on a large dataset like ImageNet, is adapted for cervical cell image classification [13]. This approach helps transfer learned features to the target domain, improving classification performance while mitigating the need for large volumes of labeled data. The main goal of this research is to implement TL using a pre-trained CNN model and develop a novel feature fusion CNN architecture that integrates features from multiple convolutional kernels, improving the accuracy and efficiency of cervical cancer diagnosis. The proposed research offers the following main contributions:

- To introduce a Variable Kernel Feature Fusion-CNN (VKFF-CNN) that integrates features from multiple convolutional kernels to enhance feature extraction and classification accuracy.
- To utilize transfer learning by adapting a pre-trained CNN model to improve feature extraction and classification, addressing the challenge of limited labeled data.
- To demonstrate the VKFF-CNN's potential as a promising tool for automated cervical cancer screening, offering significant advancements in accuracy and efficiency over existing classification methods.

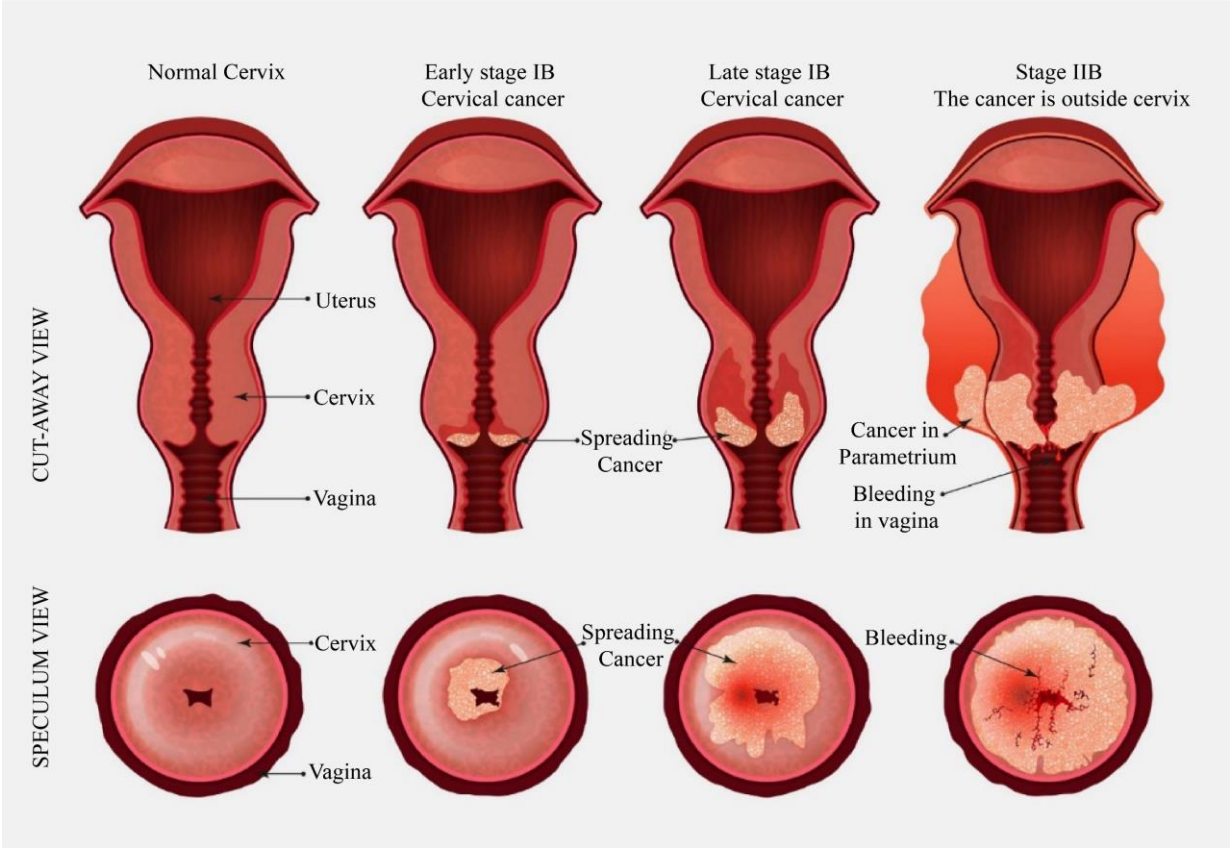

Fig. 1 Different stages of cervical cancer

The structure of the paper is organized as follows: Section 2 examines the existing research relevant to the proposed research, highlighting current methodologies and their limitations. Section 3 details the proposed methodology, comprising the design and implementation of the VKFF-CNN model and the application of transfer learning. Section 4 highlights and investigates the experiments' findings, analyzing the proposed approach's performance and effectiveness. Section 5 of the study provides a concise summary of the findings, analyzes their implications, and proposes possible avenues for future research.

2. Related Works

Using uterine cervix images and transfer learning methods, Hanife Göker (2024) [14] proposed an image processing-based method for identifying cervical cancer. The research employed a publicly accessible dataset including 917 images, utilizing Gaussian filtering and histogram equalization to improve and denoise the images. This study employed 10-fold cross-validation to compare several TL architectures. VGG19 has been identified as the most efficiently performing model. The study emphasized that integrating preprocessing techniques with the optimal performance of VGG19 yielded a precise and effective detection method.

Emmanuel Ahishakiye et al. (2024) [15] introduced two kernel-based methods, namely a Deep Gaussian Process (DGP) and an optimized Support Vector Machine (SVM) model for the automated classification of cervical cancer. They utilized a dataset of liquid-based cytology Pap smear images obtained from a total of 460 participants. This dataset encompassed images representing four different categories of cervical abnormalities. During preprocessing, the dataset was subjected to debris removal, image enhancement, segmentation, and data augmentation algorithms. A comparative analysis revealed that the SVM model had superior performance overall but was prone to overfitting, whereas the DGP model demonstrated greater robustness. One of the study's primary limitations was the inadequate availability of computing resources, which hindered model training progress.

Using deep learning models, Sher Lyn Tan et al. (2024) [16] achieved automatic detection of cervical cancer without needing segmentation or distinctive features. The researchers employed transfer learning with pre-trained CNN models to classify Pap smear images directly into seven distinct categories. They assessed 13 models using the Herlev dataset. The DenseNet-201 model exhibited superior effectiveness with regard to both accuracy and efficiency. The investigation revealed that whereas pre-trained CNN models had strong performance for specific categories, the overall classification performance displayed variability. DenseNet models, namely DenseNet-201, performed exceptionally well in handling multi-class classification problems, highlighting the efficacy of TL with pre-trained models in handling sparse data and resolving class imbalance constraints.

The CerviFormer model, developed by Bhaswati Singha Deo et al. (2024) [17], uses Transformers to handle large-scale Pap smear images effectively while minimizing architectural limitations. The proposed model utilized cross-attention and self-attention methods to combine input data into a condensed latent space, resulting in excellent performance on two datasets. Three-state classification on the Sipakmed dataset yielded a classification accuracy of 93.70%, while two-state classification on the Herlev dataset achieved a classification accuracy of 94.57%.

The challenge of cervical cancer diagnosis was addressed by Lenis Wong et al. (2023) [18] by developing an artificial intelligence system designed to analyze liquid-based Pap smears according to the Bethesda classification. To overcome the limitation of the Papanicolaou technique, which frequently fails to detect lesions because of sampling problems, the study employed a dataset of Pap smear images that was augmented to incorporate 2,676 images. They utilized ResNet50V2 and ResNet101V2 models within the deep learning and transfer learning methodologies framework. Integrated with an automated detection system, the ResNet50V2 model exhibited superior performance.

Omneya Attallah (2023) [19] examined the categorization of cervical cancer by implementing a CAD system named CerCan.Net on Pap smear images. The proposed model employed three lightweight CNNs to decrease the intricacy of classification while improving accuracy. CerCan·Net incorporated deep features collected from the final three layers of each CNN instead of depending on features from a single layer. The study illustrated the advantages of combining data from several CNNs and showed how feature selection could improve classification performance. Validation of CerCan·Net's performance was conducted using benchmark datasets, demonstrating its capacity to assist cytopathologists in addressing common challenges encountered in regular Pap smear diagnosis.

Sai Chandana et al. (2023) [20] emphasized the necessity of efficient cervical cancer screening, pointing out that timely removal of diseased tissues could greatly decrease death rates. Utilizing a transfer learning-based SE-ResNet152 model modified by the Deer Hunting Optimization (DHO) algorithm, the work presented a DL method for multi-class cervical cancer classification using Pap smear images. A cost-sensitive loss function was employed in the method to resolve the imbalance in the dataset, which consisted of 8,838 images divided into 11 classes. Although the study attained excellent classification performance, it also revealed constraints such as the high dimensionality of concatenated data and the complex nature of transfer learning models. The hybrid approach introduced by Madhura Kalbhor et al. (2023) [21] integrates

DL architectures with Machine Learning (ML) algorithms and a fuzzy min-max neural network to extract features and classify Pap smear images. They employed pre-trained DL models while analyzing the Herlev and Sipakmed datasets. Within the Sipakmed dataset, the fine-tuned ResNet-50 model demonstrated the highest classification accuracy of 95.33%, surpassing the performance of AlexNet. The hybrid technique suggested in this study aims to improve the accuracy and efficiency of classifying cervical cytology images.

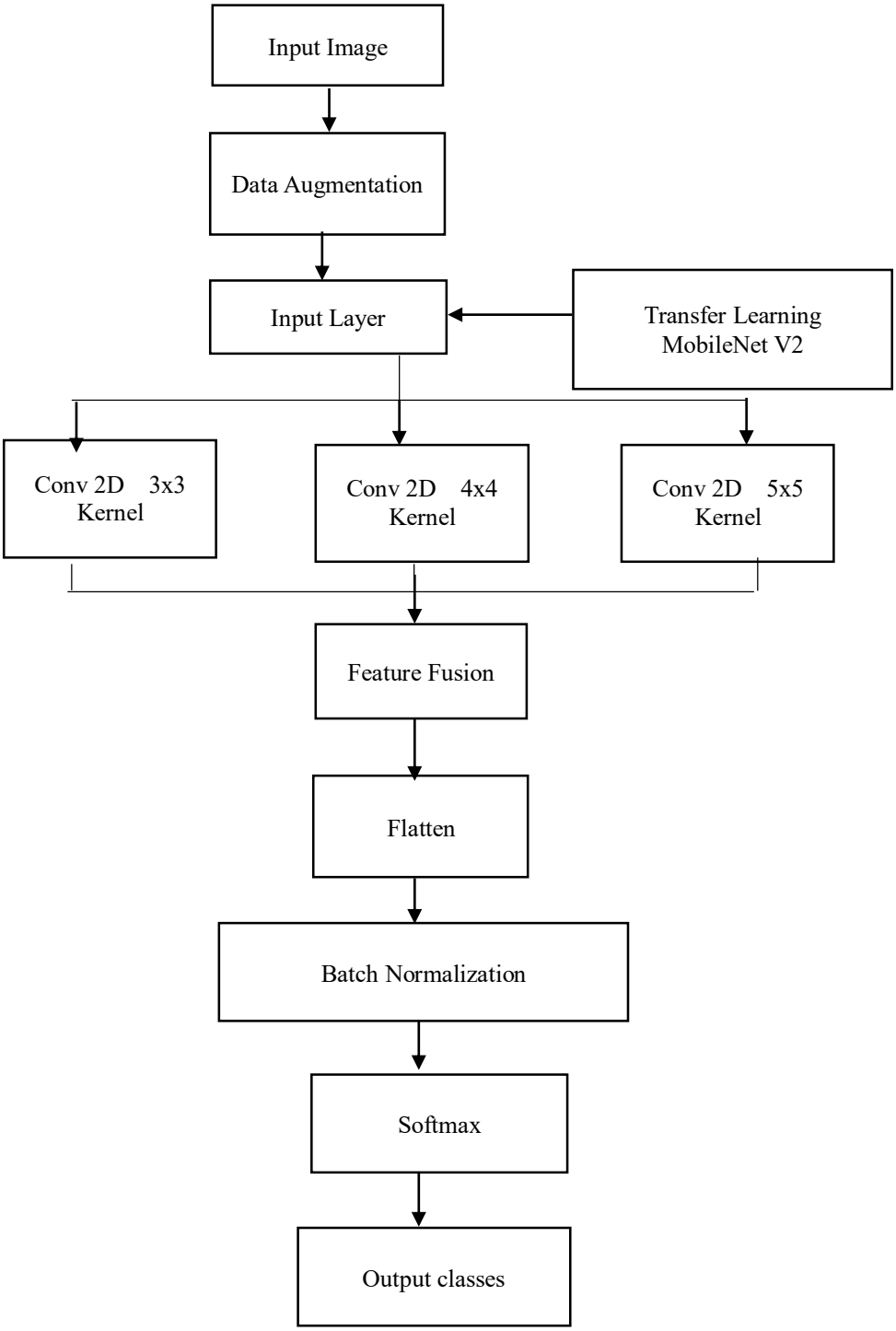

Fig. 2 Framework of proposed cervical cancer classifier

Using the SIPaKMeD dataset, which consists of 4045 isolated Pap smear cells, Aya Haraz et al. (2023) [22] developed an ML-based model to categorize cervical cancer cells from Pap smears into five major cell categories. The study utilized a rigorous preprocessing pipeline and a robust feature extractor, resulting in exceptional classification outcomes. Out of the applied classifiers, the SVM achieved the highest accuracy of 0.968, surpassing the accuracy of 0.958 for the Neural Network (NN) and 0.941 for the K-Nearest Neighbor (KNN). Habtemariam et al. (2022) [23] designed a resilient system for the automated categorization of cervix types and cervical cancer using deep learning methodologies. They gathered 915 histopathology and 4,005 colposcopy images from public sources and nearby medical facilities.

A MobileNetv2-YOLOv3 model was employed to extract the Region of Interest (ROI) from images of the cervix. These extracted ROIs were further categorized using the EfficientNetB0 model. The method attained a mean Average Precision (mAP) of 99.89% for detecting ROIs and classification accuracies of 96.84% for cervix type and 94.5% for cervical cancer. Litjens et al. [24] provided a detailed study of applications, highlighting the effectiveness of feature fusion techniques across different medical imaging domains. These studies underscore the potential of these advanced techniques to enhance automated diagnostic systems in healthcare.

According to the literature, although classical approaches for classifying cervical cancer laid an adequate foundation, deep learning methods, particularly CNNs, have significantly evolved the field. TL has shown efficacy in addressing the difficulties related to a scarcity of labeled data and computing limitations, whereas feature fusion has shown considerable promise in improving model performance by combining several features. The significance and effectiveness of these techniques are underscored by their extensive application in several medical imaging applications. This work aims to enhance the accuracy and effectiveness of cervical cancer detection by implementing and evaluating a TL and feature fusion CNN method designed in particular for classifying the Herlev dataset.

3. Materials and Methods

The proposed methodology introduces a novel method for classifying cervical cancer by leveraging TL and a VKFF-CNN. TL is employed to tackle the issue of insufficient annotated data and high computational demands by utilizing pre-trained models to enhance feature extraction capabilities. VKFF-CNN further refines this process by integrating features extracted from multiple convolutional kernels, allowing the model to capture both low-level and high-level patterns within cervical cell images.

This multi-faceted feature extraction enables a more nuanced understanding of the images, which is crucial for accurate classification.The methodology involves preprocessing and augmenting images from the Herlev dataset before feeding them into the VKFF-CNN classifier. In this study, we allocated 80% of the images for training and used the remaining 20% to test the model's performance. The classifier's efficiency is then determined using metrics such as accuracy, precision, recall, and F1 score. This approach is chosen for effectively combining diverse features and leveraging existing knowledge from pre-trained models, significantly improving classification accuracy and robustness in detecting cervical cancer. The process flow of the VKFF-CNN classification scheme is shown in Figure 2.

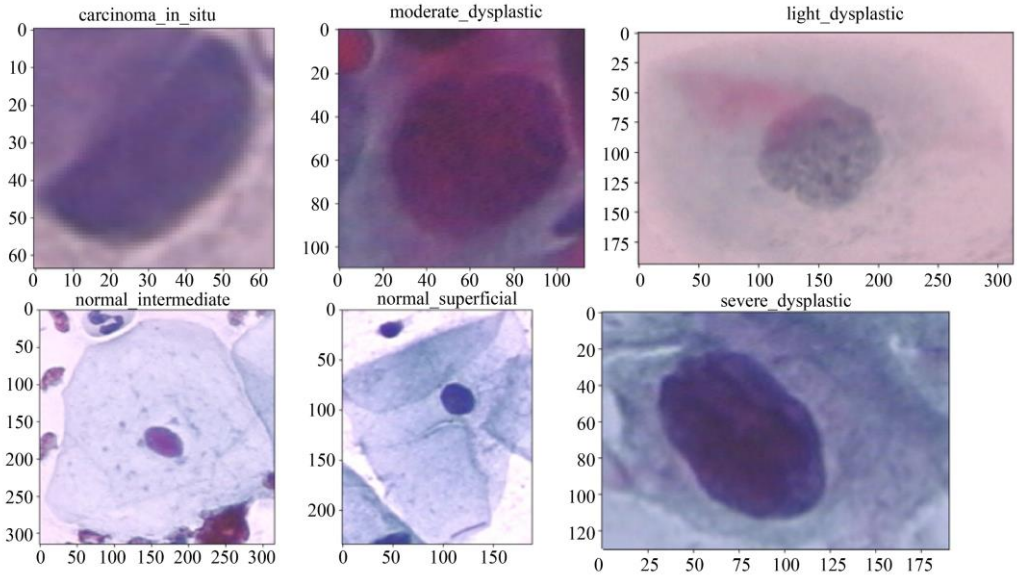

Fig. 3 Sample images from the dataset

3.1. Dataset

The Herlev Pap smear database utilized in this study was developed by Herlev University Hospital, which is publicly accessible online [25]. The database comprises 917 cell images organized into six major groups. The superficial and intermediate squamous epithelia are normal cells, while the other groups are malignant. The cellular types vary from normal to abnormal, with carcinoma in situ representing the most advanced pathological condition. Figure 3 depicts images from the dataset highlighting the three dysplastic cell stages. Unlike severe dysplasia, moderate dysplastic cells tend to regress without progressing to malignancy. Dysplastic cells frequently feature larger, darker nuclei that cluster together, whereas severe dysplasia is defined by enlarged, granular, and irregularly shaped nuclei [26].

These traits form the basis for dividing cervical cells into six categories for analysis. Figure 4 displays the distribution of different image categories. Figure 5 shows the distribution of normal and malignant cells present in the dataset.

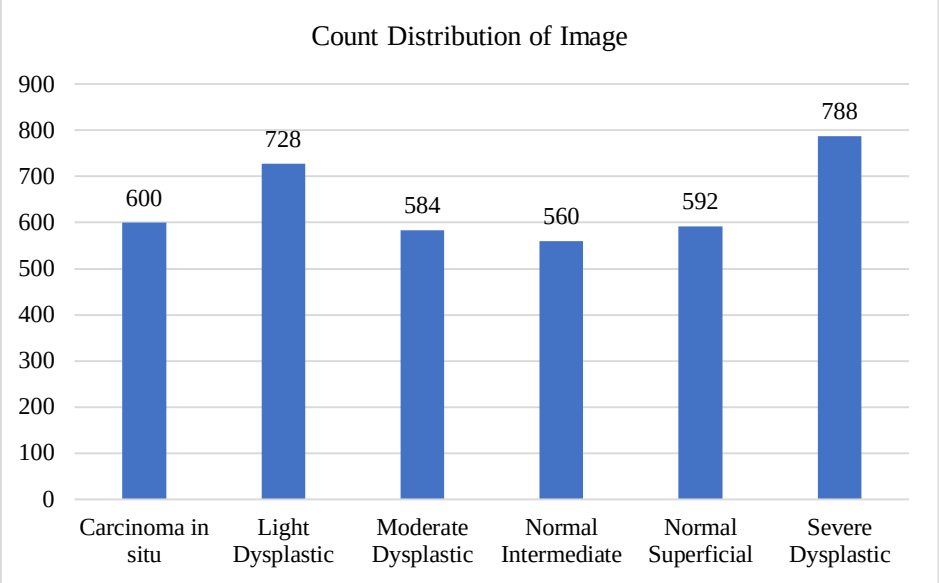

Fig. 4 Count distribution of images

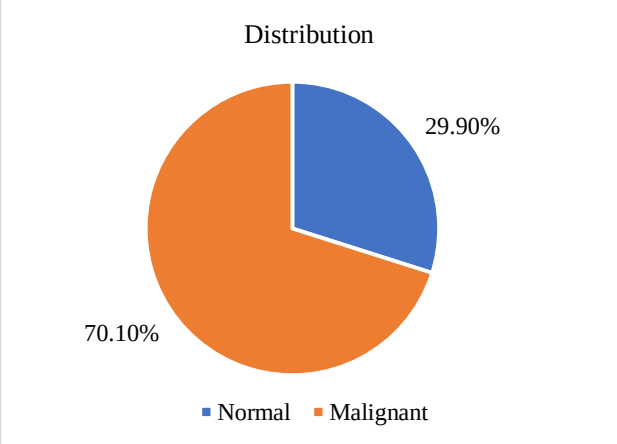

Fig. 5 Distribution of normal vs malignant cells

3.2. Preprocessing and Augmentation

Preprocessing refers to a series of techniques strategically designed to enhance the quality of images and assure consistency across the dataset. First, in order to maintain consistency and increase computational performance, images are resized and normalized. A data augmentation technique was employed utilizing the ImageDataGenerator class in Keras due to the dataset's limited image size and imbalanced distribution of images across classes [27]. The augmentation process involves using a series of modifications to artificially expand the dataset and improve the generalization capabilities of the model. The aforementioned transformations include geometric operations such as rotation, scaling, and flipping, which allow the model to preserve its structural integrity against spatial distortions. For instance, rotating images by random angles helps the model become invariant to cell orientation while scaling and zooming mimic different cell sizes and magnifications.

Flipping, both horizontally and vertically, introduces additional variations in cell appearance, and brightness and contrast adjustments help the model adapt to different staining and lighting conditions. Adding noise and applying elastic transformations make the model more robust to image acquisition imperfections and biological tissues' elastic nature. By employing preprocessing and augmentation techniques, the model is trained on data of exceptional quality and diversity, improving its accuracy and reliability in classifying cervical cancer from Pap smear images.

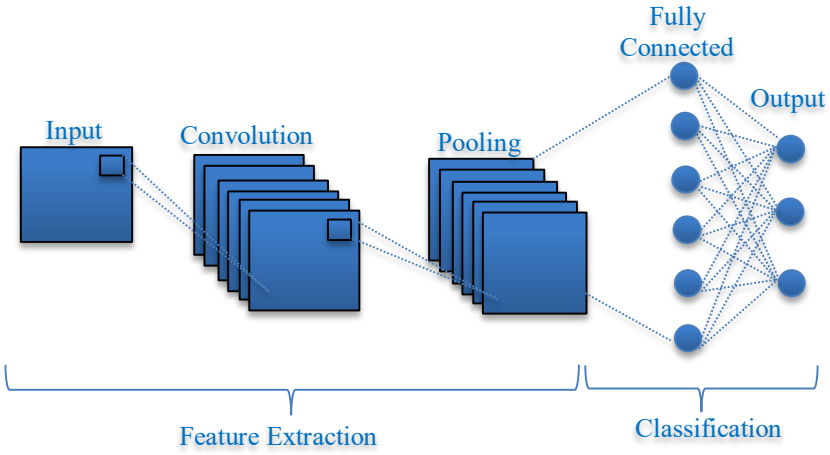

Fig. 6 Basic architecture of CNN

3.3. Model Development

3.3.1. Convolutional Neural Network

CNNs are a specific category of DL models suitable for analyzing imaging data [28]. The architecture of a CNN typically consists of several key layers, as shown in Figure 6. CNNs employ convolutional layers to automatically and effectively acquire spatial information hierarchies from input images.

The fundamental component of a CNN is the convolutional layer, which employs a set of learnable filters to the input image. Mathematically, the convolution operation for a single output channel is expressed as in Equation (1).

$$y_{i,j} = \sum_{m=1}^{M} \sum_{n=1}^{N} w_{m,n} \cdot x_{i+m,j+n} + b \quad (1)$$

Where $y_{i,j}$ is the output value at position (i,j), $w_{m,n}$ represents the filter weights, $x_{i+m,j+n}$ denotes the input image values, b is the bias term, and M and N represents the filter dimensions.

Following the convolutional layers, CNNs often incorporate pooling layers, which down-sample the spatial dimensions of the feature maps to lessen computational complexity and extract dominant features. For instance, max pooling is expressed as in Equation (2).

$$y_{i,j} = max_{m,n}\left(x_{i+m,j+n}\right) \quad (2)$$

Where the maximum value is taken from a local region of size $m \times n$ in the input feature map.

The last stages of a CNN often involve fully connected (dense) layers, which aggregate the features learned by the convolutional and pooling layers to generate class scores. If f is the output from the previous layer, the output of a fully connected layer is calculated according to Equation (3).

$$y_k = \sum_{i=1}^{N} w_{i,k} \cdot f_i + b_k \quad (3)$$

3.3.2. Transfer Learning and MobileNet V2

Transfer learning is a potent method in DL that utilizes the knowledge acquired from a previously trained model on a similar task to enhance performance on an entirely novel but interconnected task. In general, the process consists of two primary phases [29].

Initially, a base model, commonly known as the pre-trained model, is trained on an extensive dataset for a specific goal. This model learns to extract a rich set of features broadly applicable across several applications. In the second stage, the pre-trained model's learned features are assigned to an alternate target network and then fine-tuned on a more limited, task-specific dataset. This fine-tuning process involves modifying the base network's architecture, typically by adding task-specific layers or replacing some of its components, to modify it to the current task being performed.

For instance, the proposed study's base network is MobileNetV2, a lightweight deep-learning architecture designed for embedded and mobile devices [30]. Initial training of MobileNetV2 is conducted on a vast dataset, such as ImageNet, to acquire hierarchical features that effectively capture a diverse array of visual patterns. MobileNetV2 employs depthwise separable convolutions. In this module, a compressed representation with low dimensions is initially expanded to a higher dimension and then processed by a lightweight depth-wise convolution, as shown in Figure 7.

This step filters the data using one filter per input channel, enabling efficient feature extraction. The high-dimensional features are then remapped back to a low-dimensional space through a linear convolution, ensuring that the complexity of the network is kept manageable while still capturing important information.

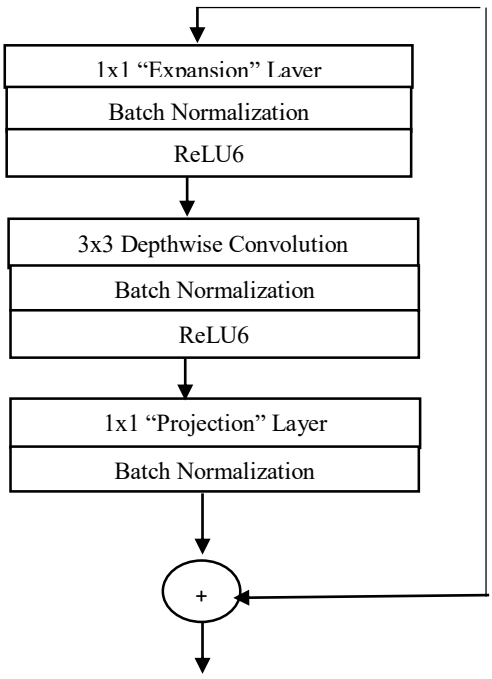

Fig. 7 MobileNetV2 with inverted residuals

One of the standout features of MobileNetV2 is the introduction of inverted residuals with linear bottlenecks. These inverted residuals consist of a lightweight depth-wise separable convolution followed by a linear layer. Using the linear bottleneck at the end of the residual block ensures that feature information is preserved without introducing additional non-linearities, further improving the network's overall efficiency.

In traditional convolutional layers, the computation of output is defined as in Eq. (4).

$$y_i = (W * X) + b \tag{4}$$

where W represents the weights, X is the input feature map. MobileNetV2 replaces this traditional structure with depth-wise and pointwise convolution. Depth-wise convolution (Y_d) assigns one filter to each input channel (depth), as shown in Figure 8, whereas the pointwise convolution (Y_p) utilizes a 1x1 convolution to merge these channels, as shown in Equations (5) and (6).

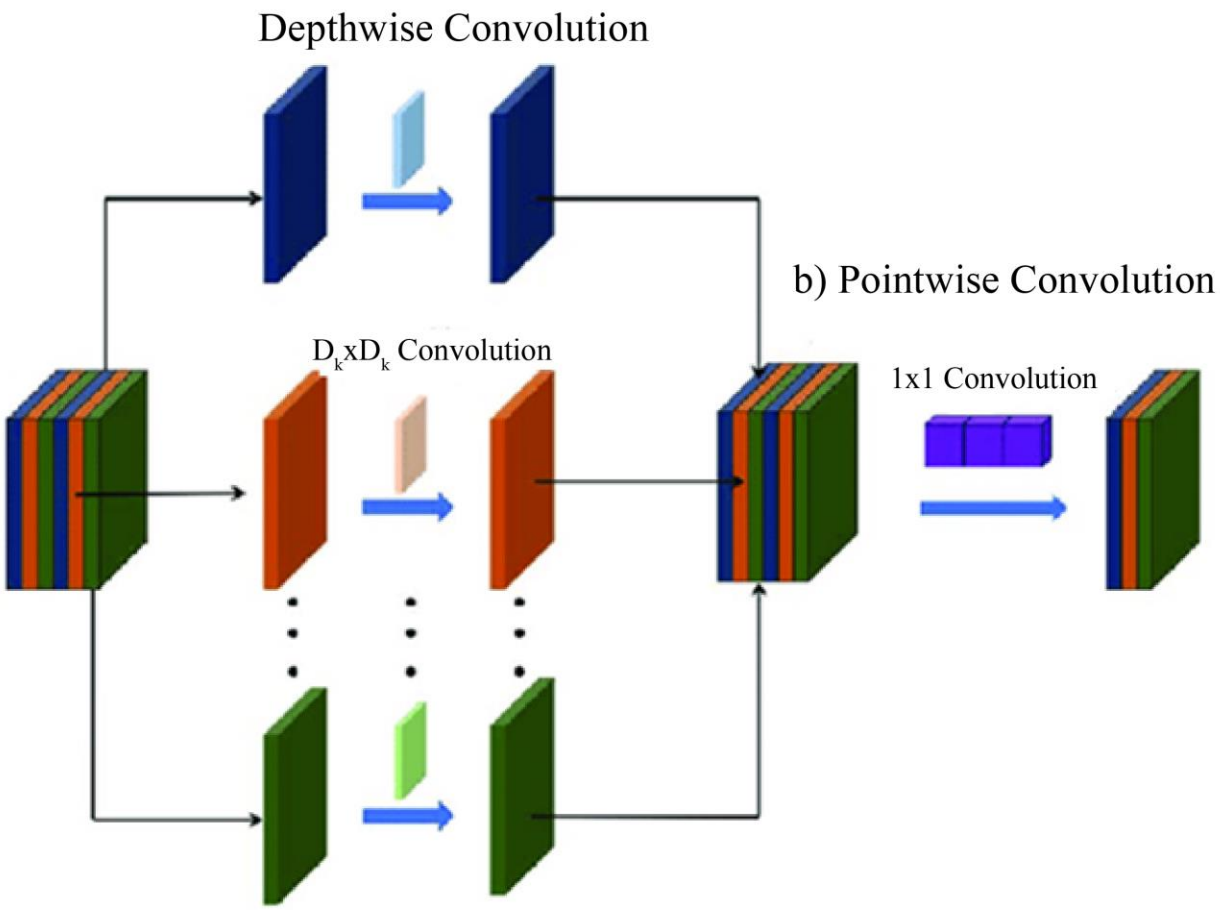

Fig. 8 Depth-wise separable convolution

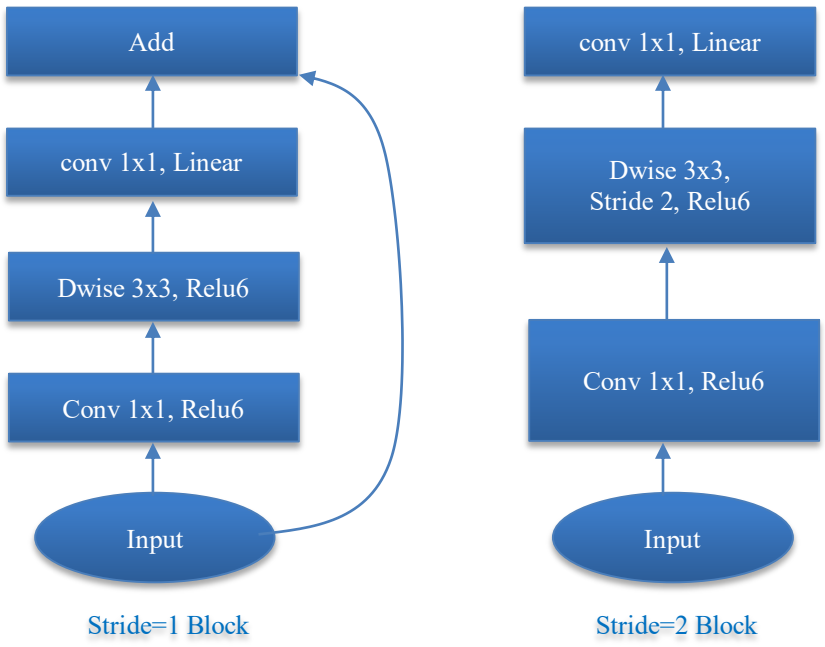

Fig. 9 MobileNet V2 Architecture

$$Y_d = W_d * X \quad (5)$$

$$Y_p = W_p * Y_d \quad (6)$$

This reduces the computational cost from $O(D_k^2.M.N)$ to $O(D_k^2.M + M.N)$. Kernel dimension D_k, input channel counts M and N are user-defined. MobileNetV2 utilizes two distinct types of bottleneck blocks: one stridden at 1 and the other stridden at 2, as shown in Figure 9. In the network, the block with stride 1 preserves the spatial dimensions of the input. In contrast, the block with stride 2 decreases the spatial dimensions by downsampling, controlling the resolution of feature maps as they progress.

The transfer learning process is most effective when the features learned from the base task are relevant and applicable to the target task, thus allowing the model to leverage previously acquired knowledge. In the context of cervical cancer classification, fine-tuning MobileNetV2 involves adjusting its pre-trained weights on the Herlev dataset and adapting it to the specific features of cervical cell images. This approach benefits from the rich feature representations learned from the vast and diverse ImageNet dataset, which includes numerous classes and image types. The fine-tuning process is mathematically represented by updating the weights W using backpropagation as shown in Eq. (7), where η is the learning rate, and L is the loss function.

$$W_{new} = W_{old} - \eta \frac{\partial L}{\partial W} \quad (7)$$

MobileNetV2's small model size facilitates its use in scenarios where computational resources and storage are limited. Applying TL with MobileNetV2 can significantly improve classification accuracy, especially when combined with data augmentation techniques. The pre-trained model provides a strong starting point, enabling the model to converge quickly even with a relatively small number of training images from the Herlev dataset. Leveraging TL with MobileNetV2 in cervical cancer classification can enhance early detection and diagnosis, potentially improving patient outcomes and contributing to more effective screening programs.

3.3.3. Proposed Variable Kernel Feature Fusion-CNN

Introducing a novel VKFF-CNN for cervical cancer classification offers significant improvements in capturing diverse and detailed features from the Herlev Pap Smear dataset. This innovative approach involves performing convolution operations using multiple kernel sizes, specifically 3x3, 4x4, and 5x5, in parallel. The VKFF-CNN architecture leverages the strength of these variable kernel sizes to learn and fuse features across different resolutions. The 3x3 kernel is adept at capturing fine details and local textures, while the larger 4x4 and 5x5 kernels can capture broader spatial patterns and structures. By combining these features, the model gains the ability to recognize patterns at multiple scales, improving its ability to differentiate between normal and abnormal cervical cells. This parallel convolutional strategy, described in Eq. (8), enhances the network's ability to generalize across a wide range of cervical cell image variations.

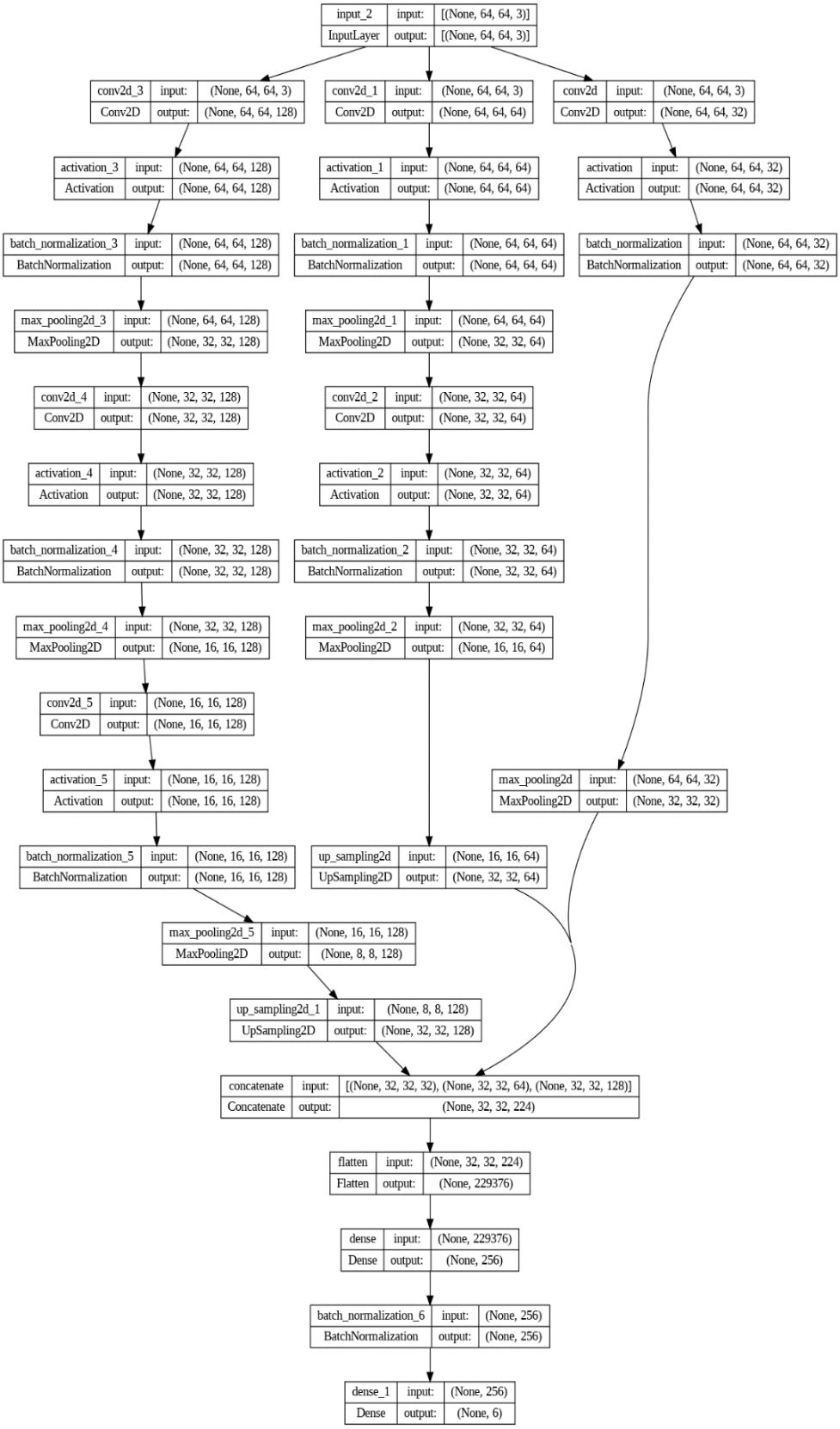

Fig. 10 Proposed VKFF-CNN Model

$$X_{3,3} = Conv_{3,3}(X)$$

$$X_{4,4} = Conv_{4,4}(X)$$

$$X_{5,5} = Conv_{5,5}(X) \quad (8)$$

The convolution operation for each kernel size is denoted as $Conv_{k,k}$, where k represents the dimensions of the convolutional filter. An integrated feature map is obtained by concatenating the outputs of parallel convolutional layers along the channel dimension, therefore incorporating information from different kernel sizes. This fusion mechanism, as represented in Equation (9), enhances the network's ability to capture both intricate details and extensive contextual information.

$$X_{fused} = Concat(X_{3,3}, X_{4,4}, X_{5,5}) \quad (9)$$

Where the concatenation operation, denoted as $Concat$, combines the feature maps from the 33×3, 4×4 and 5×5 convolutions into a single comprehensive feature map. Given that each convolutional layer outputs C' channels, the concatenated feature map will have $3 \times C'$ channels in total. This results in a fused feature map X_{fused} with dimensions $H \times W \times (3C')$, where H and W denote the height and width of the feature map, accordingly.

After the feature maps are concatenated, the next steps involve flattening, batch normalization, and the application of a softmax activation function. The flattening layer converts the multi-dimensional feature maps into a 1-D vector, making the data compatible with fully connected layers, which are essential for classification tasks. Once the data is flattened, batch normalization is applied to standardize the input to the next layer by normalizing activations within mini-batches. This process reduces internal covariate shift by subtracting the mean and dividing by the standard deviation of activations within each mini-batch, ensuring that the inputs to subsequent layers are on a consistent scale throughout training. This step also improves the stability and convergence speed of the neural network.

Two trainable parameters, scale (γ) and shift (β), are introduced by batch normalization for each feature map. These parameters allow the network to adjust the normalized values adaptively, enabling more flexible learning of the optimal data representation. By scaling and shifting the activations, the network gains more control over how features are represented and adjusted during training. Finally, the softmax activation function is employed to the output logits, transforming them into class probabilities. This final step allows the VKFF-CNN to output the likelihood of each class, facilitating accurate classification and decision-making. The VKFF-CNN model depicted in Figure 10 effectively combines multi-scale feature learning with robust and stable classification capabilities.

3.4. Hardware and Software Setup

The study utilized a powerful computing system, including an Intel Core i7 processor, NVIDIA GeForce GTX 1080Ti GPU, and 32GB of RAM, to implement the VKFF-CNN model for cervical cancer classification. Model development was carried out using the Keras library and executed in Python. Additionally, Google Colab was utilized for model experimentation and testing. Hyperparameters were carefully selected to optimize the model's performance. The specifics of these hyperparameter settings are detailed in Table 1, highlighting their importance in configuring the model for effective training and accurate classification.

Table 1. Hyperparameter specifications

Hyperparameters	Specifications
Optimizer	Adam
Epochs	100
Learning Rate	0.00001
Loss Function	Categorical Cross-Entropy
Activation Function	ReLU, Softmax

4. Results and Discussion

Accuracy plots illustrate the proportion of correctly classified instances among the total instances in each training epoch. An increasing trend in the accuracy plot typically indicates that the model improves its ability to make correct predictions. Initially, the training accuracy rapidly increases, reaching around 90% within the first 10 epochs, as shown in Figure 11, indicating the model's ability to quickly learn features from the training data.

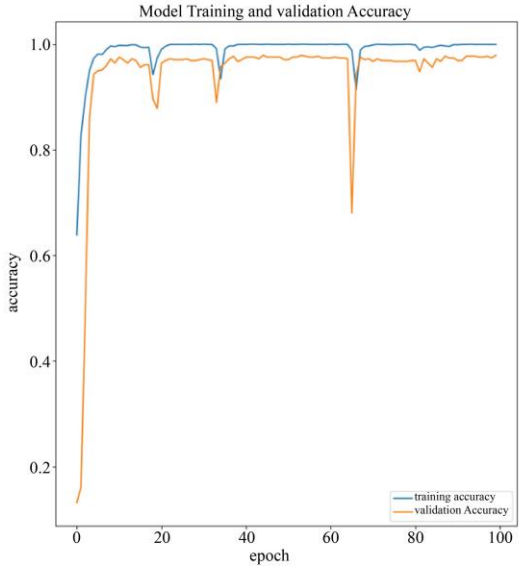

Fig. 11 Accuracy plot

Although there are minor variations in the training accuracy during the training process, it eventually converges near 98%, showcasing the model's high effectiveness on the training data. Validation accuracy increases steadily at the beginning but experiences several sharp drops around epoch

65, indicating potential overfitting or fluctuations in how the model generalizes to unseen data. Nonetheless, the validation accuracy stabilizes above 95%, suggesting the overall effectiveness of the feature fusion approach and transfer learning model in classifying cervical cancer images. However, the drop in validation accuracy could indicate sensitivity to certain subsets of validation data, potentially linked to noise or hard-to-classify samples. Loss plots depict the error or cost associated with the model's predictions. A declining trend in the loss plot implies that the model is learning effectively, as the error between predictions and true values is reducing. The training loss starts high, around 1.5, but quickly decreases to near zero by epoch 20, suggesting that the model is learning effectively and minimizing errors in the training data. A minor increase in training loss after epoch 50 indicates a slight disturbance in the optimization process but stabilizes soon after.

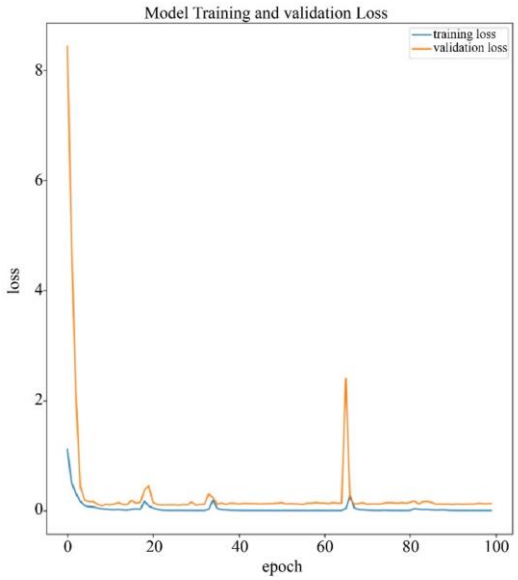

Fig. 12 Loss plot

In contrast, the validation loss begins much higher, exceeding 8 at epoch 4, as shown in Figure 12. This significant discrepancy between training and validation loss suggests initial difficulty generalizing to unseen data. As training progresses, the validation loss decreases but remains higher than the training loss, which may indicate overfitting. Notably, there is a spike in validation loss at epoch 65, corresponding to the dip in validation accuracy. Despite this, the validation loss eventually decreases, stabilizing at a low level by epoch 80. These plots suggest that the VKFF-CNN learned efficiently and avoided overfitting, maintaining a balance between training accuracy and generalization. The four main metrics are used to comprehensively assess the effectiveness of the proposed framework. These metrics, derived from the notions of False Negative (FN), False Positive (FP), True Positive (TP), and True Negative (TN), are crucial for evaluating the performance of the model. The mathematical expressions for these performance parameters are provided in Eq. (10), (11), (12), and (13).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{10}$$

$$Precision = \frac{TP}{TP+FP} \tag{11}$$

$$Recall = \frac{TP}{TP+FN} \tag{12}$$

$$F1-score = 2 \times \frac{precision \times Recall}{Precision+Recall} \tag{13}$$

The VKFF-CNN demonstrated outstanding performance over multiple training epochs, with 98.03% average accuracy, indicating a high level of precision in differentiating between normal and abnormal cervical cells. The VKFF-CNN model attained mean precision of 97.83%, recall of 97.11%, and F1 score of 98.23%, as shown in Figure 13. High values across all these metrics demonstrate the model's resilience and efficacy in accurately categorizing cervical cell images.

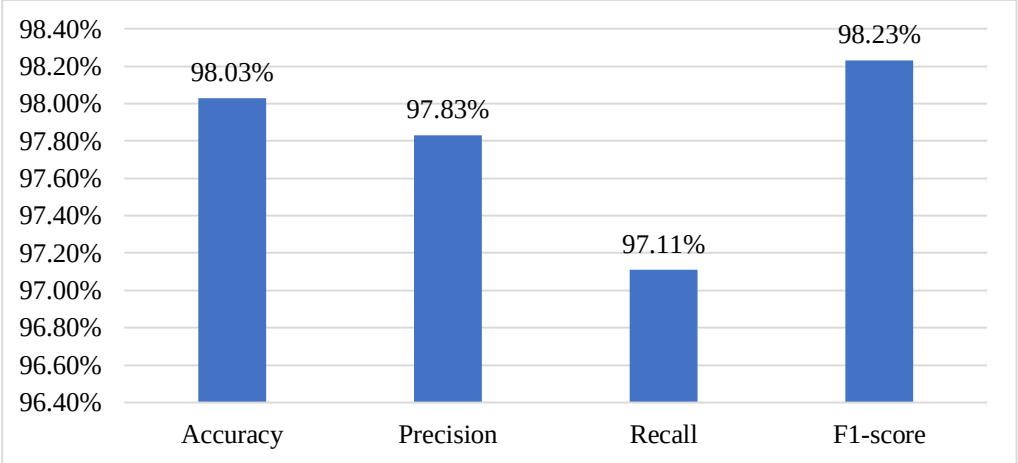

Fig. 13 Performance analysis of the proposed framework

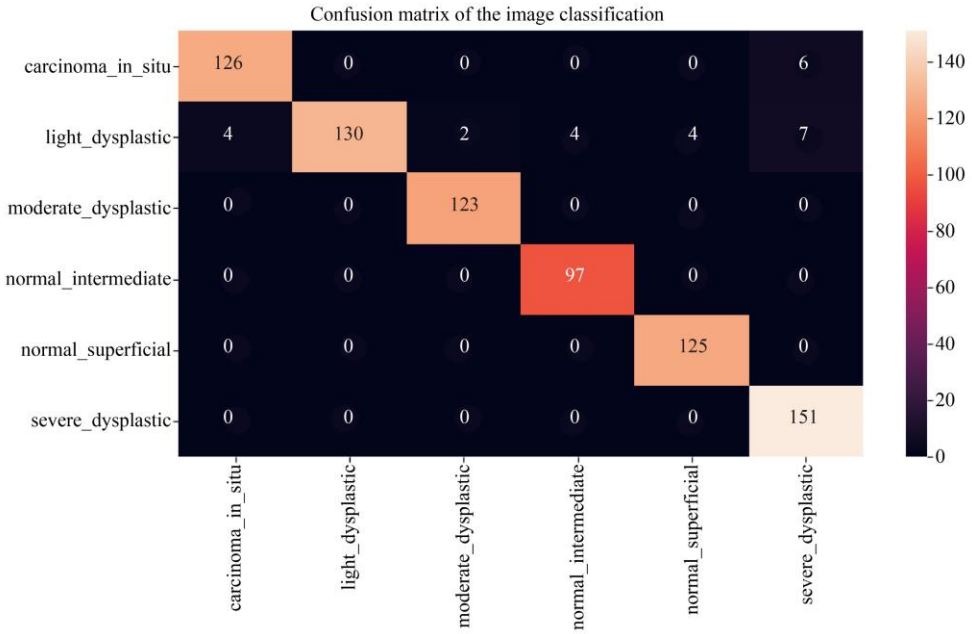

Fig. 14 Confusion matrix

The confusion matrix illustrated in Figure 14 further elucidates that the VKFF-CNN has a high true positive rate across all classes, with very few misclassifications. Low counts of FP and FN emphasize recall and precision, confirming that it rarely confuses one class for another. Achieving this level of accuracy is essential in medical diagnostics, where the consequences of misclassification can have serious implications for patient outcomes.

The analysis of predicted images showcases the model's ability to correctly classify cervical cell images. Visual inspection of these images reveals that the VKFF-CNN effectively distinguishes between normal and abnormal cells, accurately identifying various classes of cervical cells, as illustrated in Figure 15. The correct classification of these images underlines the model's potential utility in practical medical environments.

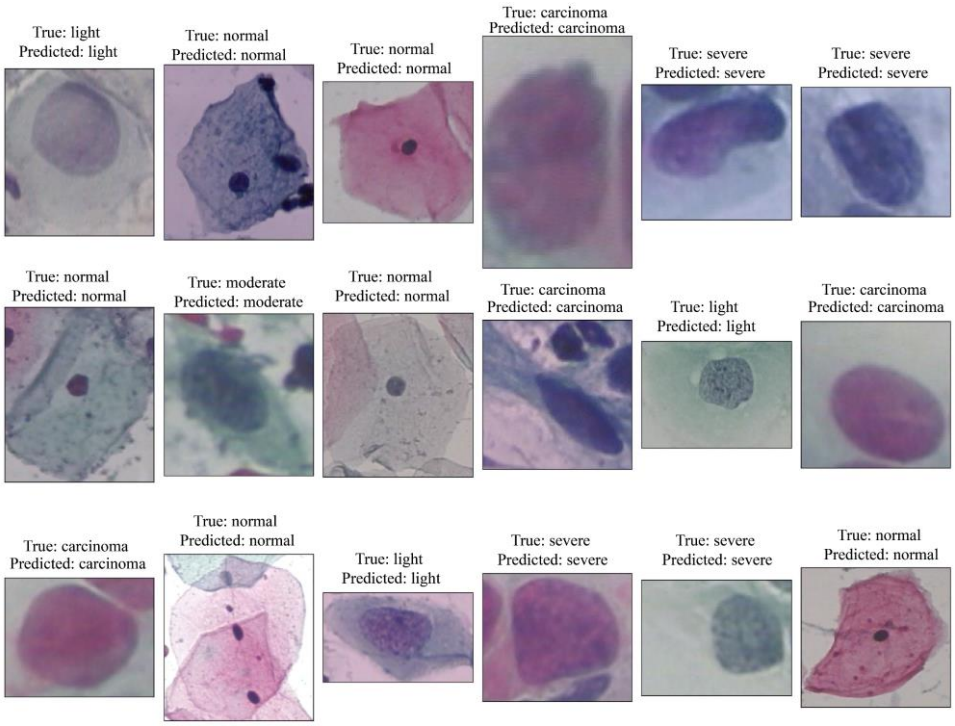

Fig. 15 Predictions of proposed VKFF-CNN classifier

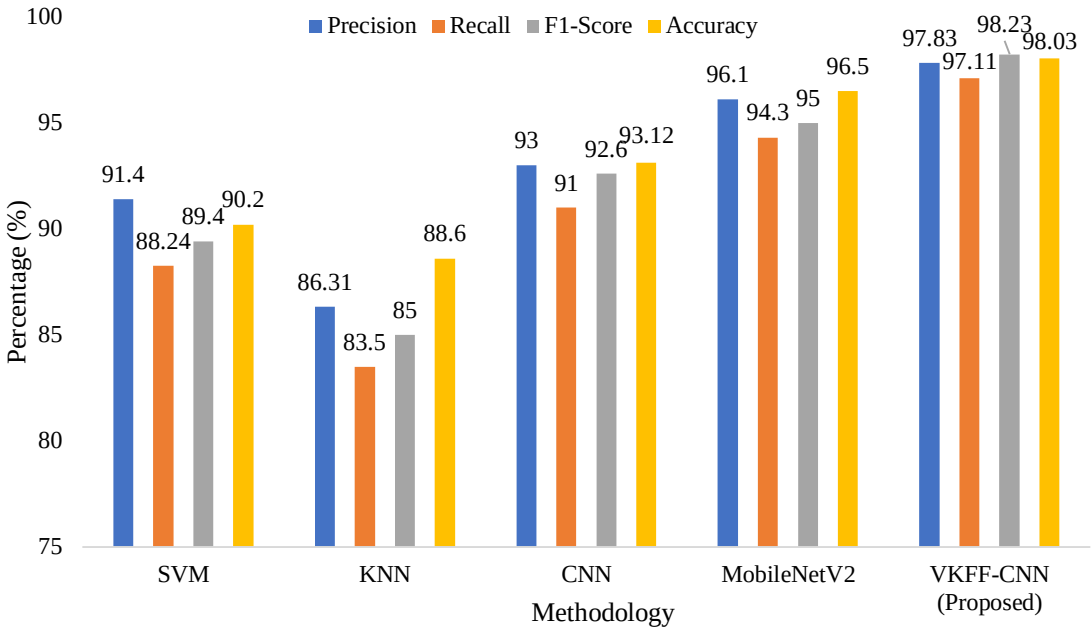

Fig. 16 Performance comparison

The performance of VKFF-CNN significantly outperforms SVM [31] and KNN [32], as well as conventional CNN and the pre-trained MobileNetV2 model, as depicted in Figure 16. The VKFF-CNN attained an average accuracy of 98.03%, surpassing the SVM and KNN, which typically exhibit lower accuracy due to their limited capacity to identify complex trends in high-dimensional data. Precision (97.83%) and recall (97.11%) also exceed those of the SVM and KNN, which often struggle with imbalanced datasets and intricate feature spaces. Compared to a standard CNN, the VKFF-CNN's innovative use of multiple kernel sizes enhances feature extraction, leading to a higher F1 score of 98.23% versus the conventional CNN's typically lower performance metrics. When benchmarked against MobileNetV2, the VKFF-CNN shows a slight edge in accuracy and precision because of specialized architecture that captures multi-scale features more effectively. MobileNetV2, while efficient and powerful, does not incorporate the same level of feature fusion, resulting in a marginally lower classification performance. The VKFF-CNN's superior performance across multiple metrics highlights its robustness and effectiveness in cervical cancer classification.

By integrating advanced CNN architecture with TL techniques, the VKFF-CNN leverages pre-trained models' capabilities, further enhancing its performance on the Herlev Pap Smear dataset. VKFF-CNN demonstrates significant promise for enhancing cervical cancer classification. Its high accuracy and robust performance metrics underscore its potential to support clinical decision-making and improve early detection of cervical cancer.

5. Conclusion

A malignancy originating from the cells lining the cervix, cervical cancer, represents a substantial health risk due to its ability for early detection and treatment through screening methods. This study introduced the VKFF-CNN architecture for cervical cancer categorization using the Herlev Pap Smear dataset. This model employs multi-scale feature extraction by incorporating convolutional layers with 3x3, 4x4, and 5x5 kernels, enabling it to capture diverse features and enhance classification performance. The VKFF-CNN demonstrated outstanding performance measures, achieving an average accuracy of 98.03%, precision of 97.83%, recall of 97.11%, and an F1 score of 98.23%. These outcomes significantly surpass those of SVM, KNN, traditional CNNs, and the pre-trained MobileNetV2 model. The robustness and accuracy of the VKFF-CNN were further validated through confusion matrix analysis and visual inspection of predicted images. Incorporating batch normalization and softmax activation functions contributed to the model's stability and the precision of its probabilistic classifications. These results indicate that VKFF-CNN is a robust automated cervical cancer screening tool, showing considerable advancements over existing methodologies. Future research may explore further optimizations of the VKFF-CNN model, the application of additional data augmentation techniques, and the validation of its efficacy on larger datasets to ensure its effectiveness in real-world clinical settings.

Acknowledgements

The author expresses profound appreciation to the supervisor for providing guidance and unwavering support throughout the course of this study.

References

[1] P.B. Shanthi, K.S. Hareesha, and Ranjini Kudva, "Automated Detection and Classification of Cervical Cancer using Pap Smear Microscopic Images: A Comprehensive Review and Future Perspectives," *Engineered Science*, vol. 19, pp. 20-41, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[2] Cervical Cancer, World Health Organization, 2024. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cervical-cancer

[3] Aabid Shariff et al., "Automated Image Analysis for High-Content Screening and Analysis," *SLAS DISCOVERY: Advancing the Science of Drug Discovery*, vol. 15, no. 7, pp. 726-734, 2010. [CrossRef] [Google Scholar] [Publisher Link]

[4] Ewert Bengtsson, and Patrik Malm, "Screening for Cervical Cancer using Automated Analysis of PAP-Smears," *Computational and Mathematical Methods in Medicine*, vol. 2014, no. 1, pp. 1-12, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[5] Aslı Gençtav, Selim Aksoy, and Sevgen Önder, "Unsupervised Segmentation and Classification of Cervical Cell Images," *Pattern Recognition*, vol. 45, no. 12, pp. 4151-4168, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[6] Jeovany Martínez-Mesa et al., "Inequalities in Pap Smear Screening for Cervical Cancer in Brazil," *Preventive Medicine*, vol. 57, no. 4, pp. 366-371, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[7] Chuin-Mu Wang et al., "New Methods for Image De-Noising and Edge Enhancement in Cervical Smear Images Segmentation," *International Journal of Computer, Consumer and Control*, vol. 2, no. 1, pp. 1-9, 2013. [Google Scholar]

[8] Vasuja Devi Midasala et al., "MFEUsLNet: Skin Cancer Detection and Classification using Integrated AI with Multilevel Feature Extraction-Based Unsupervised Learning," *Engineering Science and Technology, an International Journal*, vol. 51, pp. 1-17, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[9] Uday Pratap Singh Parmar et al., "Artificial Intelligence (AI) for Early Diagnosis of Retinal Diseases," *Medicina*, vol. 60, no. 4, pp. 1-15, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[10] Arohee Sinha, and Tarun Kumar, "Enhancing Medical Diagnostics: Integrating AI for Precise Brain Tumour Detection," *Procedia Computer Science*, vol. 235, pp. 456-467, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[11] I. Onur Sigirci, Abdulkadir Albayrak, and Gokhan Bilgin, "Detection of Mitotic Cells in Breast Cancer Histopathological Images Using Deep Versus Handcrafted Features," *Multimedia Tools and Applications*, vol. 81, pp. 13179-13202, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[12] Muhammad Fazal Ijaz, and Marcin Woźniak, "Recent Advances in Deep Learning and Medical Imaging for Cancer Treatment," *Cancers*, vol. 16, no. 4, pp. 1-5, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[13] Mohammad Zafer Khaliki, and Muhammet Sinan Başarslan, "Brain Tumor Detection from Images and Comparison with Transfer Learning Methods and 3-Layer CNN," *Scientific Reports*, vol. 14, no. 1, pp. 1-10, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[14] Hanife Göker, "Detection of Cervical Cancer from Uterine Cervix Images Using Transfer Learning Architectures," *Eskişehir Technical University Journal of Science and Technology A-Applied Sciences and Engineering*, vol. 25, no. 2, pp. 222-239, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[15] Emmanuel Ahishakiye, and Fredrick Kanobe, "Optimizing Cervical Cancer Classification through Transfer Learning and Kernel Methods: Analyzing the Performance of Deep Gaussian Processes and Support Vector Machines on Pap Smear Image Data," *Research Square*, pp. 1-15, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[16] Sher Lyn Tan et al., "Cervical Cancer Classification from Pap Smear Images Using Deep Convolutional Neural Network Models," *Interdisciplinary Sciences: Computational Life Sciences*, vol. 16, pp. 16-38, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[17] Bhaswati Singha Deo et al., "CerviFormer: A Pap Smear-Based Cervical Cancer Classification Method Using Cross-Attention and Latent Transformer," *International Journal of Imaging Systems and Technology*, vol. 34, no. 2, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[18] Lenis Wong et al., "Deep Learning and Transfer Learning Methods to Effectively Diagnose Cervical Cancer from Liquid-Based Cytology Pap Smear Images," *International Journal of Online and Biomedical Engineering*, vol. 19, no. 4, pp. 77-93, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[19] Omneya Attallah, "CerCan· Net: Cervical Cancer Classification Model via Multi-Layer Feature Ensembles of Lightweight CNNs and Transfer Learning. Expert Systems with Applications, vol. 229, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[20] Krishna Prasad Battula, and B. Sai Chandana, "Multi-Class Cervical Cancer Classification Using Transfer Learning-Based Optimized SE-ResNet152 Model in Pap Smear Whole Slide Images," *International Journal of Electrical and Computer Engineering Systems*, vol. 14, no. 6, pp. 613-623, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[21] Madhura Kalbhor et al., "Hybridization of Deep Learning Pre-Trained Models with Machine Learning Classifiers and Fuzzy Min–Max Neural Network for Cervical Cancer Diagnosis," *Diagnostics*, vol. 13, no. 7, pp. 1-16, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[22] Aya Haraz et al., "Deep Learning-Based Technique for the Perception of the Cervical Cancer," *Mansoura Engineering Journal*, vol. 48, no. 3, pp. 1-9, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[23] Lidiya Wubshet Habtemariam, Elbetel Taye Zewde, and Gizeaddis Lamesgin Simegn, "Cervix Type and Cervical Cancer Classification System Using Deep Learning Techniques," *Medical Devices: Evidence and Research*, vol. 15, pp. 163-176, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[24] Geert Litjens et al., "A Survey on Deep Learning in Medical Image Analysis," *Medical Image Analysis*, vol. 42, pp. 60-88, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[25] Yannis Marinakis, Georgios Dounias, and Jan Jantzen, "Pap Smear Diagnosis Using a Hybrid Intelligent Scheme Focusing on Genetic Algorithm Based Feature Selection and Nearest Neighbor Classification," *Computers in Biology and Medicine*, vol. 39, no. 1, pp. 69-78, 2009. [CrossRef] [Google Scholar] [Publisher Link]

[26] Anant R. Bhatt, Amit Ganatra, and Ketan Kotecha, "Cervical Cancer Detection in Pap Smear Whole Slide Images Using ConvNet with Transfer Learning and Progressive Resizing," *PeerJ Computer Science*, vol. 7, pp. 1-18, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[27] tf.keras.preprocessing.image.ImageDataGenerator, TensorFlow. [Online]. Available: https://www.tensorflow.org/api_docs/python/tf/keras/preprocessing/image/ImageDataGenerator

[28] Shadman Sakib et al., "An Overview of Convolutional Neural Network: Its Architecture and Applications," *Computer Science and Mathematics*, pp. 1-6, 2019. [Google Scholar] [Publisher Link]

[29] Sinno Jialin Pan, and Qiang Yang, "A Survey on Transfer Learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345-1359, 2010. [CrossRef] [Google Scholar] [Publisher Link]

[30] Mark Sandler et al., "MobileNetV2: Inverted Residuals and Linear Bottlenecks," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4510-4520, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[31] B. Ashok, and P. Aruna, "Comparison of Feature Selection Methods for Diagnosis of Cervical Cancer Using SVM Classifier," *International Journal of Engineering Research and Applications*, vol. 6, no. 1, pp. 94-99, 2016. [Google Scholar] [Publisher Link]

[32] L. Lavanya Devi, and P. Thirumurugan, "Cervical Cancer Classification from Pap Smear Images Using Modified Fuzzy C Means, PCA, and KNN," *IETE Journal of Research*, vol. 68, no. 3, pp. 1591-1598, 2021. [CrossRef] [Google Scholar] [Publisher Link]