CerConvNet: Cervical Cancer Cells Prediction Using Convolutional Neural Networks

Pallavi M, Smitha Patil, Madhusudhan M V, Smitha S Reddy, Vaishnavi K
Presidency University, Bangalore
E-mail: pallavim@presidencyuniversity.in, smithapatil@presidencyuniversity.in, madhusudhanmv@presidencyuniversity.in, smithasreddy13@gmail.com, vaishnavikosgi6772@gmail.com

Keywords: deep learning, CNN, SIPaKMeD, cervix.

Received: March 14, 2024

Cervix cancer is a distinct form of cancer occurring in women, originating in the cells of the cervix, which is the region of the uterus connecting to the vagina. About 90% of cases of cervix cancer are related to human papillomavirus (HPV) infection. The mortality rate in developed nations has decreased because of routine HPV testing for women. The absence of reasonably priced healthcare facilities, however, continues to make it difficult for developing countries to offer inexpensive remedies. Therefore, developing an accurate algorithm for cervical cancer prediction is necessary to identify women who are at risk of developing this condition. Architectures of Deep Learning have been employed in recent years to construct accurate models for the prediction of cervical cancer. This study offers a unique, straightforward transfer learning framework: ResNet50, DenseNet201, EfficientNetb1 and InceptionResNetV2, to classify cervical images using the SIPaKMeD dataset and different performance measures are gathered and examined. Still, the recommended Densenet201outperformed the most advanced methods. We obtained an average accuracy of 98.78% with CNN models which is the highest compared over the existing models. Resnet50 achieved even better results after augmentation with an accuracy of 99.51% and Precision, Recall, F1-score of 0.99. As a result, the findings support our approach to providing low-cost first-level screening.

Povzetek: Raziskava predstavlja CerConvNet, okvir za napovedovanje raka maternične vratu z uporabo konvolucijskih nevronskih mrež. Model DenseNet201 je dosegel najvišjo točnost, izboljšano z bogatenjem podatkov.

1 Introduction

In recent years, deep learning (DL), a sector of Artificial Intelligence (AI), has experienced rapid and substantial growth. The field of science is currently focusing on deep learning (DL) because of its many benefits, including high performance, diverse applications, great generalisation capability, and versatility. In view of the processing of huge amount of medical data and the advancement of computing capacity have created an interest in this domain. Utilizing deep learning could expedite and enhance the diagnostic process, leading to more precise and personalized treatment approaches. The study of computer-aided medical picture analysis holds great potential for bettering patient results [1]

As of 2020, cervical cancer ranks as the fourth most prevalent cancer among women, one in 168 Canadian women is predicted to get cervical cancer in their lives, and one in 478 Canadian women passes away from the disease. Cervical cancer is most widespread and fatal in countries with lower and middle incomes. The disparities emerge because of social and economic elements, along with restricted availability of nationwide HPV vaccination, cervical screening, and treatment resources, highlighting notable inequalities. The cause for cervical cancer is infection with high-risk variants of human papillomavirus (HPV). These disparities stem from social and economic factors, along with restricted availability of national HPV vaccination, cervical screening, and treatment resources, highlighting substantial inequities. The timely identification and prompt treatment are crucial elements in successfully curing cervical cancer. The cofactors that may increase the risk are tobacco smoking, immune suppression, multi-parity, sexually transmitted infections and poor diet. Lack of regular screening also increases the risk of Cervical cancer. The common symptoms include irregular or heavy bleeding, bleeding after intercourse, unpleasant odor in vaginal discharge and discomfort during sexual activity. Cervix cancer can be diagnosed by history or recto vaginal examination. The size of the tumour and the spread of the disease determine the cancer stage. In the approaching decades, countries globally are working towards accelerating the elimination of cervical cancer, aiming to achieve a specific set of three objectives by 2030.

2 Related work

Using DL algorithms, several previous research have suggested methods for detecting and classifying CC. Wanli Liu, Chen Li, Ning Xu et al. suggested a deep learning based image classification model known as CVM-Cervix to perform cervix cell classification [1]. CVM-Cervix is evaluated using a merged dataset that combines CRIC and SIPaKMeD datasets, constituting 11 classes. The pre-processed data is fed into the Convolutional Neural Networks and Virtualisation Technology modules for feature extraction. The features

extracted are then supplied into the MLP module for fusion and classification. Additionally, a streamlined post-processing technique is used to improve the model. The performance of CVM-Cervix is compared with 22 models (18 CNN models, 4 VT models). Among the 18 CNN models, high accuracy is shown by DenseNet169(88.99%) and DeiT (87.35%) has the highest accuracy among the VT models. However, CVM-Cervix surpassed both with an accuracy of 91.72%.

Anurag Tripathi et al. presents a deep learning approach on the SIPAKMED dataset using the ResNET-152 Architecture. This project involves classifying images into five groups. Normal cells and abnormal cells are the two subclasses that are created from the five classes. Transfer learning is implemented. The dataset was trained with 4 models, namely ResNet50, ResNet152, VGG16 and VGG19. All models were adjusted using identical hyper-parameter conFigureurations. The fine-tuning of model weights was performed over 50 scans over the entire dataset, considering a batch-size of 10, input image resolution set to 224×224. ResNet-152 yielded the highest accuracy of 94.89%, followed by VGG19(94.38%), ResNet-50(93.87%) and VGG16(92.85).

Swati Shinde, Madhura Kalbhor et al. presented an approach for prediction [2]. It introduces a deep learning framework DeepCyto, developed using pre-trained CNN models and ANN model. This frame work was tested on 3 cervical cancer datasets (Herlev, Sipakmed and LBC). In workflow 1, the feature vectors extracted from Principal Component analysis (PCA) are fed into 3 Machine learning classifiers (SVM, RF and FCN), the classifier's prediction is then passes into the voter. The maximum voting process's result determines the image's anticipated class. In workflow 2, the deep features are extracted using 4 CNN pre-trained models (ResNet50, XceptionNet, VGG16, VGG19), these features are then passed into a fully connected ANN. The pre-trained models' feature fusion vector is used to train the model, and the Artificial Neural Network's (ANN) output layer estimates the probability outputs that are used to make subsequent predictions. When the two workflows were compared, workflow two showed better performance. Achieving accuracy of 97-100% on different datasets.

Wasswa William, Andrew Ware et al. proposes a framework of automated cervical cancer diagnosis and categorisation. [3]. The suggested tool includes scene segmentation through the utilization of categoriser named WEKA region of interest and utilizes a method of systematically eliminating one item at a time to discard unwanted debris. Fuzzy C-means technique is used for classification, and simulated annealing is integrated with a wrapper filter for feature selection. The findings indicated that the approach surpasses numerous existing algorithms in terms of sensitivity is 99.28%, specificity of 97.47%, and accuracy is about 98.88% when tested on the Herlev dataset. The suggested system examines an entire pap smear slide in 3 minutes, contrasting with the 5–10 minutes required for manual analysis per slide.

Alquran, H, Alsalatie et al. proposes Cervical Net, a novel DL structure with distinguished group convolutional layers, utilizing depth-wise separable convolutions and grouped convolutions to extract depth features for improved accuracy in cervical cancer classification [4]. The model comprises stages such as image acquisition, image enhancement, extracting and choosing features, combining features, and carrying out classification. ShuffleNet V2 addresses the challenge of computational complexity in the model by using depth-wise convolutions and 1x1 tiny convolution kernels, resulting in a smaller model size without compromising accuracy. The extracted features from Cervix Net are input to different ML algorithms (SVM, RF, ANN Naïve Bayes, KNN) to determine the optimal classifier accuracy. These classifiers were verified on the SIPaKMeD datasets. The system attained its greater accuracy for 5 classes with 99.1 percent using the Support Vector Machine, while the Naïve-Bayes'classifier yielded lesser results, not surpassing eighty-five percent of accuracy.

Taranjit Kaur, Tapan Kumar Gandh et al. suggests the classification of brain images employing Convolutional Neural-Networks (CNN) and the application of transfer learning techniques [5]. Utilized CNN architectures such as Alexnet, ResNet-50, GoogLenet, VGG16, ResNet-101, VGG19, Inception-V3, and InceptionResNet-V2. It makes use of transfer learning - to increase the efficiency of the model, it makes use of the initial layers of pre-trained model by replacing last layers of the network. The model is evaluated on the images taken from the Harvard repository, Figure share repository, and Fortis Memorial Research Institute. The evaluation metrics used are sensitivity, specificity, and accuracy. The results showed AlexNet model yielded the most favorable results for all three datasets, with VGG-16 closely trailing behind. Employing pre-trained Deep-Convolutional-Neural-Network (DCNN) architectures with knowledge transfer technique for classification presents several advantages. Firstly, it automates the entire classification process. Secondly, it removes the traditional processes of noise filtering, outlining Regions of Interest (ROI), and conducting feature extraction and selection. Thirdly, it eliminates biases both between different observers and within the same observer, ensuring reproducible predictions from the pre-trained DCNN models.

Pin Wang, Lirui Wang et al. The suggested approach employs the Mean-shift clustering algorithm for identifying Regions of Interest (ROI). Subsequently, a versatile mathematical morphological operation is employed to separate nuclei that may be overlapping [6]. The project involves two steps, the first step is to accurately segment the cell nuclei present in the Pap smear images. This is achieved through Mean-Shift Clustering Algorithm and Mathematical Morphology. After segmenting the cell nuclei, the next step is to categorise samples of images based on the characteristics of the nuclei. Features derived from shape, texture, and Gabor are extracted from the segmented nuclei for classification. The proposed segmentation and classification methods are tested on a database of 362 cervical Pap smear images. The proposed system demonstrates high effectiveness in segmenting cell nuclei, with a sensitivity of approximately

94.25% and specificity of about 93.45% and an accuracy of more than 96%.

Md Mamunur Rahaman, Chen Li, Xiangchen Wu et al. proposed a model which includes, gathering image data, Image data cleansing, ROI of image extraction and categorisation of pictures using deep learning techniques, creating feature vectors and classifiers [7]. Three kinds of feature learnings, namely unsupervised (Autoencoders, Sparse and topological, etc), Supervised (ResNet, Fully CNN, VGG, Patch based, etc), Hybrid (Unsupervised pre-trained + CNN, CNN+ graph-based method, etc) are used for segmentation tasks. Classification using deep learning involves both representation learning techniques (such as Deep CNN, VGG16, AlexNet, ResNet, etc.) and cross representation learning techniques (including CNN + transfer learning, AlexNet + SVM, CNN + feature concatenation, etc.) [8]. CNN is acknowledged for its exceptional performance in both segmentation and classification tasks. It later explains a wide variety of CNN architectures namely AlexNet, ResNet, VGGNet, Inception, and mentions the available public and private cervical cytology datasets.

Pin Wang, En Fan, Peng Wang et al. contrasts and examines older machine learning and deep learning techniques for image categorisation, employing the SVM and CNN algorithms for comparison. The experiment is done on the MNIST hand-written digital picture dataset and COREL1000 picture set [9]. The article illustrates a substantial MNIST dataset, Support Vector Machine achieves a precision of 0.88, whereas CNN achieves a greater precision of 98%. In contrast, with a smaller COREL1000 dataset, SVM achieves a precision of 86%, while CNN's precision is 83%. The inference drawn is that conventional machine learning excels with smaller datasets, whereas deep learning frameworks show improved recognition accuracy with huge samples of data.

Tawsifur Rahman, Muhammad E et al. proposed system was trained and tested using dataset chest X-ray pneumonia, wherein image data consists of five thousand two hundred and forty-seven chest X-ray pictures and this was later increased using data augmentation [10]. Three classification tasks are performed, Normal vs. Pneumonia: Distinguishing between normal X-rays and those showing signs of pneumonia. Bacterial vs. Viral Pneumonia: Identifying whether the pneumonia is caused by bacterial or viral infection. The outcomes demonstrated that DenseNet201 performs better than the three other unique deep CNN networks. It resulted with 98%, 93.3% and 95% of accuracy for the 3 classification schemes, respectively.

Table I: Summary table of the existing methodologies.

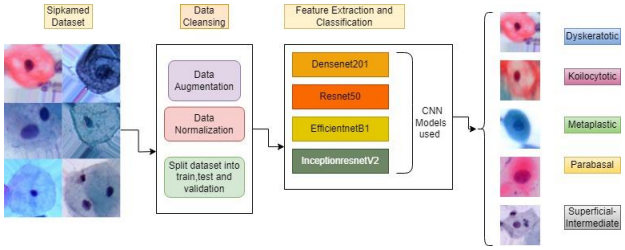
Reference	**Key Findings**	**Results**	**Limitations**
Wanli Liu et al. [1]	Hybrid approach with CNN and Visual Transformer for feature extractions, and Multilayer Perceptron for feature fusion. Tested on combined dataset of CRIC and SIPaKMeD.	Precision, accuracy and F1 of 91.8%, 91.7% and 91.7%.	Difficulty in distinguishing similar cell nuclei due to image ambiguity.
Anurag Tripathi et al. [2]	ResNet-152 used on SIPaKMeD dataset with transfer learning.	Classification accuracy of 94.9% achieved.	Need for dataset diversity.
Swati Shinde et al. [3]	Multiclass datasets converted into binary class (normal and abnormal). Workflow 1: PCA feature extraction with ensemble voting for classification. Workflow 2: Pre-trained with CNN with ANN for feature fusion and classification.	Accuracy of 97%, 99% and 100% for 7, 5 and 4-class classification.	Potential reduction in diagnostic accuracy due to class aggregation.
Wasswa William et al. [4]	Trainable Weka Segmentation; Simulated Annealing with Wrapper Filter feature selection; Fuzzy C-means classification.	A sensitivity of 99% and accuracy of 98% achieved.	Tested on small datasets. (917, 497, 60)
Alquran et al. [5]	New Feature Fusion method using Shuffle Net and Cervical Net. PCA used for feature extraction and dimensionality reduction.	Classification accuracy of 99.1% achieved.	Tested on Small dataset.(4049 images)

Taranjit Kaur et al. [6]	Use of pre-trained DCNN models(AlexNet, ResNet50, VGG16, GoogleNet, ResNet101, VGG19, InceptionV3, & InceptionResNetV2) with transfer learning for brain image classification.	AlexNet outperformed all other models.	Trained on small sample datasets.(50, 74, 160)
Pin Wang et al. [7]	Segmentation using Mean-shift clustering and classification based on nuclei features.	94.2% sensitivity and 93.4% specificity with 96% accuracy achieved.	Tested on small dataset. (362 images)
Pin Wang et al. [9]	Comparison of traditional machine learning(SVM) and deep learning(CNN) algorithm.	Machine learning performs better with small sample datasets and deep learning better with large datasets.	-
Mousser et al. [39]	Representation learning was done by transfer-learning techniques and categorisation of labels achieved through Multilayer-perceptron.	ResNet50 outperformed; Accuracy of 89.2% with five-fold cross-validation achieved.	Lower accuracy compared to expected standards.
Pal et al. [40]	Transfer learning models(backbone) with Deep metric learning(DML) for image feature discrimination; KNN classifier.	Contrastive, N-Pair embedding, and batch hard achieved mean-5 precision scores of 91.1%, 89.58%, and 85.14%, respectively.	Possible lack of generalizability during training. Class imbalance.
Md Mamunur Rahaman et al. [35]	Hybrid deep feature fusion(VGG16, VGG19, ResNet50, XceptionNet) technique for classification;	Accuracy of 99.1%(5-class) and 90.3%(7-class).	Performance degrades as class size increases.
Kalbhor et al. [41]	Discrete cosine transformer(DCT) and Haar transformer for feature extraction; Machine learning models for classification.	Classification accuracy of 81.1% achieved.	Lower accuracy; Low resolution images.
Chandana et al. [42]	SE-ResNet152 for classification; Deer Hunting Optimization for hyper-parameter tuning.	Precision, recall and F1-Score of 98.8%, 97.8% and 98.6%.	High dimensionality of concatenated features.
Li et al. [43]	Weekly supervised MHCRF based model for classification; Feature extraction of color, texture, and SOTA deep learning features.	Overall classification accuracy of 77.3% achieved.	Tested on small dataset(600 samples)

The related work section provides an extensive review of existing methodologies. The above summary table I compares the key findings, accuracy, and limitations of the reviewed studies.

3 Methodology

The following section includes the detailed explanation of the proposed methodology starting from pre-processing till classification.



Figureure 3.1: Proposed methodology framework.

Above Figureure 3.1 represents the framework of the proposed methodology. Each section of the framework is explained in detail as below.

3.1 Dataset description

To evaluate the proposed transfer learning methods, we utilized the Sipakmed Pap Smear dataset, a publicly available cervical cancer image dataset [11-14]. The Sipakmed dataset is a five-class dataset comprising a total of 4049 individual images featuring isolated cells were manually cropped from a collection of nine hundred and sixty-six cluster cell images. The cell classes are categorized into superficial-intermediate cells, parabasal cells, koilocytotic cells, Dysketarotic cells and Metaplastic cells. Dataset split ratio information is as follows: 80:5:15 for training, testing and validation respectively.

3.2 Data cleansing

The dataset's limited number of images were insufficient for effective model training, posing a risk of overfitting. We used data augmentation approaches to solve this issue, expanding the sample size using fundamental augmentation techniques. The transformation technique used are: rotation, height shift, wide shift, shear, zoom, horizontal flip, and fill mode. Table II below gives the detailed number of images class wise before and after augmentation.

Table II: Dataset description

Class no.	Class name	Number of original images	Number of images after augmentation
1	Dyskeratotic	813	1,602
2	Koilocytotic	825	1,610
3	Metaplastic	793	1,555
4	Parabasal	787	1,553
5	Superficial-Intermediate	831	1,630

Data augmentation techniques are basically used here to expand the amount and diversity of a sipkamed standard dataset as shown in Figure 3.2, hence improving the performance and robustness of CNN models. Geometric tranformations utilized here are: Rotation with a degree of 20, width_shift_range, height_shift_range, shear_range, zoom_range equal to 0.2, Flipping of images is done horizontally to help the model learn to recognize images from different perspectives and reduce the generalization errors which can be witnessed in the section 4.2.

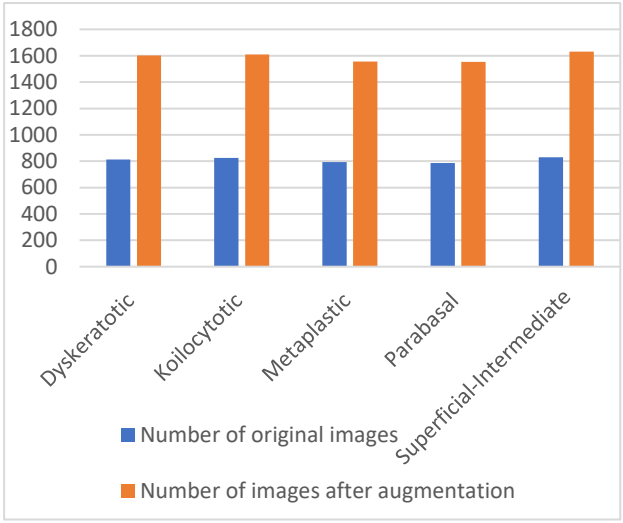


Figure 3.2 Bar graph representing number of images before and after augmentation.

3.3 Representation learning and classification

Convolutional neural networks (CNN) are artificial neural networks that specialize in detecting and understanding patterns, making them helpful for image analysis [15-19]. CNN consists of multiple layers, each with its own set of functions to perform on the input data. The architecture of the convolutional neural network comprises various layers, including convolutional, Rectified Linear Unit (ReLu) [39], and down-sampling layers with max strategy.

The design is influenced by AlexNet, as outlined in the next section and onward. It comprises of six layers: Conv-2D, Re-Lu, down-sampling, and a dense layer. The input layer stores raw data, the convolutional layer calculates the dot-product of patches of imagery and kernel filters, and an activation function is applied by the activation function layer to each portion of the output of the convolutional layer.

The layer of down sampling improves the efficiency of memory output from the preceding layers, resulting in decreased computational costs. The Dense layer receives input from its preceding layer and produces the calculated one-dimensional array of logit values. Techniques for detecting and objects' racking involve extracting features from images and videos, primarily applied for security purposes.

Enhancing training performance can be achieved by incorporating extra layers, like dropouts. The layer called

drop-out is specifically initiated during the period of training. Specifically, in the forward pass (input to the function). The dropout layer selectively eliminates a specified number of neurons and retains those that persist. Only the non-dropped neurons are updated during the backward pass. This dropout mechanism serves as a regularization technique. By allowing the model to acquire robust, neuron-independent characteristics, it prevents the learning of features that are overly dependent on specific neurons, thus mitigating over fitting during the training phase.

CNN, or Convolutional Neural Networks [20], has a significant impact in classification of images. Its approach is based on the use of numerous tiny kernels to determine the characteristics in every layer. CNN is made up of a collection of layers, which can extract various features from the input images. Deep feature extraction from SIPaKMeD datasets ‘of various classes images of original size 224x224x3 were accomplished by extracting features from pre-trained CNN models. Resnet50, Densenet201, EfficientnetB1, InceptionresnetV2 are the various CNN architectures used in this article.

Base ResNet50, Densenet201, Efficientnetb2, Inceptionnetv2 models are customised as described in the following sections. Hyper-parameters used in each of these architectures are detailed as follows: Standard sipkamed dataset is splitted into the training instances ratio of 80% and validation instances of 20%. Learning rate of 0.001 and adam optimizer are used for reducing the loss during training, dropout rate of 20% and L2 regularizers are utilized to overcome over fitting. Around 100 iterations were used to scan the entire dataset with the batch size of 32 to ease the training process. By choosing these hyper-parameters values, we have achieved good results compared to the existing work.

3.3.1 Resnet50

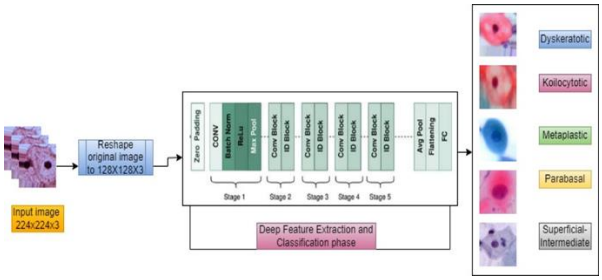


Figure 3.3 Base Resnet50 architecture.

ResNet-50 typically takes an input image of size 224x224 pixels. A standard convolutional layer with 7x7 filters and 64 output channels. Followed by batch normalization and ReLU activation. The core building blocks of ResNet are residual blocks [16]. ResNet-50 consists of 16 residual blocks grouped into four stages. Each stage contains a different number of residual blocks with varying numbers of filters. The ResNet50 [21-24] architecture comprises 50 trainable layers, with 48 of them being convolutional layers, in addition to a solitary layer for both average pooling and max-pooling. Spatial pooling is applied at different stages of the network to reduce spatial dimensions. Global Average Pooling (GAP) is used before making the layer dense. The final network layer incorporates a fully-connected layer that employs softmax activation to classify five classes in this context.

The number of nodes in the last layer aligns with the number of classes involved in the classification task. Diverging from its forerunners, ResNet50's shortcut connections deviate from the usual two layers and instead bypass three layers. The Figureure 3.3 illustrates the ResNet50 architecture, highlighting the point Res-Nets support two types of shortcut- connections. If the dimensions of both inputs and outputs are similar [25-28], straight-arrow connections are utilized. In cases where there is an increase in these dimensions, the alternative connections marked with dotted lines are activated. Parameter details of Resnet50 architecture are listed below in Table III.

Table III: Parameter details of Resnet50 architecture.

Type of the layer	Input shape	Output shape	Number of parameters
Image	224 x 224 x 3	128 x 128 x 3	0
Base model	128 x 128 x 3	4 x 4 x 2048	23,534,592
Global average pooling	4 x 4 x 2048	2048	0
Dense	2048	1024	2,098,176
Dropout	1024	1024	0
Dense (output layer)	1024	5	5,125
Total parameters			25,637,893

3.3.2 Densenet201

Initial input image size, convolution layers and their filter sizes in Densenet201 are similar to Resnet50. DenseNet concatenates feature maps from previous layers, propagating them to subsequent layers and connecting them to newly created feature maps. DenseNet offers benefits such as feature reuse and less issues with exploding or vanishing gradients [29].

The structure of DenseNet is composed of an input layer, three dense blocks, layers of transition, and a down-sampling with average strategy globally. Layers of the transition consist of a batch-normalization layer, a convolution layer of size 1×1, and a window size of 2×2 average down-sampling layer including step size of 2 as shown in Figure 3.3. Unlike conventional pooling algorithms, global average pooling (GAP) reduces a feature map from w × w × c to 1 × 1 × c, effectively condensing the entire slice into a single digit. The final dense-layer, is modified due to the original design of the last FCL to identify five categories [30-31]. Parameter details of Densenet201 architecture are listed below in Table IV.

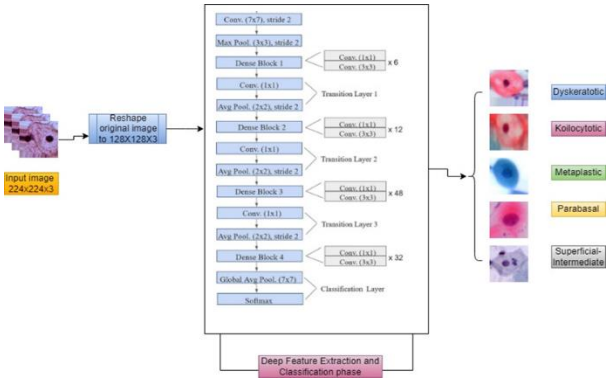


Figure 3.4 Base Densenet201 architecture.

Table IV: Parameter details of Densenet201 architecture.

Type of the layer	**Input shape**	**Output shape**	**Number of parameters**
image	224 x 224 x 3	128 x 128 x 3	0
Base model	128 x 128 x 3	4 x 4 x 1920	18,092,928
Global average pooling	4 x 4 x 1920	1920	0
Dense	1920	1024	1,967,104
Dropout	1024	1024	0
Dense (output layer)	1024	5	5,125
Total parameters			20,065,157

3.3.3 EfficientnetB1

EfficientNetB1 belongs to the EfficientNet series of convolutional neural networks (CNNs) formulated to achieve superior accuracy and efficiency in contrast to conventional CNN designs. The architecture of EfficientNetB1 is derived from a compound scaling approach, which uniformly adjusts the network's width, depth, and resolution. [32]. This allows the model to be more efficient in terms of computational resources while maintaining high accuracy. The "B1" in EfficientNetB1 denotes the specific scaling coefficients used for the width, depth, and resolution. The Efficientnet-B1 architecture's up-sampling network is made up of decoder blocks. Each block consists of up-sampling of window size 2x2 convolution 2D of the output from the previous layer, with a step size of two. This result is then concatenated [32] with the representation learning maps from the section of the encoder. Before proceeding to the next decoder [32] block, the combined tensor is passed through two convolution layers with Re-LU activation and Batch- Normalization as shown in Figure 3.4. The architecture's last layer is a soft-max convolution with the same number of channels as the output classes, which is five, and the output image size is similar as the input. Parameter details of EfficientnetB1 architecture are listed below in Table V.

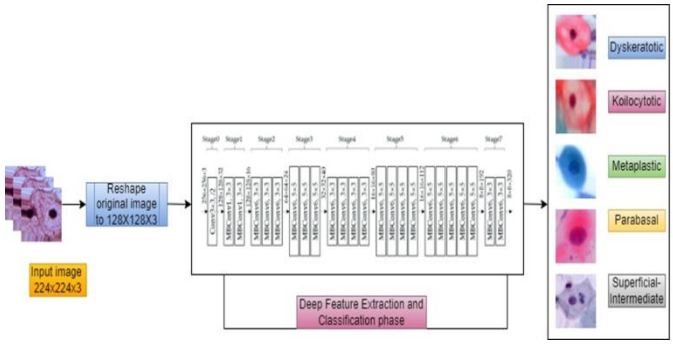


Figure 3.5 Base EfficientnetB1 Architecture [18].

Table V: Parameter details of EfficientnetB1 architecture

Type of the layer	**Input shape**	**Output shape**	**Number of parameters**
image	224 x 224 x 3	128 x 128 x 3	0
Base model	128 x 128 x 3	4 x 4 x 1280	6,513,184
Global average pooling	4 x 4 x 1280	1280	0
Dense	1280	1024	1,311,744
Dropout	1024	1024	0
Dense (output layer)	1024	5	5,125
Total parameters			7,830,053

3.3.4 InceptionresnetV2

The Inception-ResNet-v2 model merges the Inception structure with Residual connections. The fundamental characteristics [33] of the Inception-ResNet-v2 architecture include:

Inception Blocks (Inception modules): The network employs Inception blocks, comprising several parallel convolutional branches with varying kernel sizes. These branches capture features at various scales. Inception modules help the network learn diverse and rich representations. Residual Connections: In addition to Inception modules, Inception-ResNet-v2 incorporates residual connections. Residual connections aid in addressing the vanishing gradient problem and enable the training of extremely deep networks. These connections involve adding the input of a layer to its output, creating a shortcut connection. Reduction Blocks: Similar to the original Inception architecture, Inception-ResNet-v2 includes reduction blocks that simultaneously reduce the spatial dimensions of the input attribute maps and increase the number of channels. Top of Form This helps in

reducing computational complexity and extracting high-level features. Stem Block: Inception-ResNet-v2 has a stem block at the beginning of the network, which processes the input image and extracts initial features. The stem block typically includes a blend of layers of pooling and convolution, as demonstrated in Figureure 3.5. Auxiliary Classifiers: Inception-ResNet-v2 includes auxiliary classifiers during training to provide additional gradients for the earlier layers. This helps with the training of very deep networks. The Inception-Resnet [31] block combines different sized convolutional filters using residual connections. Using residual connections prevents degradation from deep structures and reduces training time. Parameter details of InceptionResnetV2 architecture are listed below in Table VI.

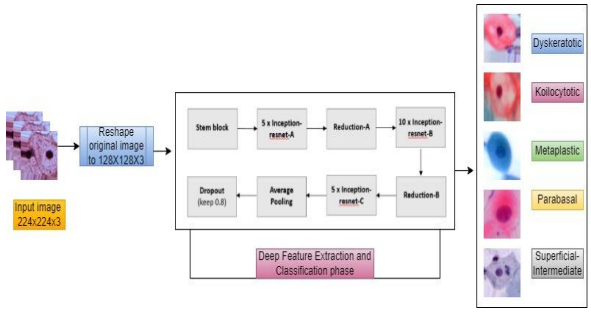


Figure 3.6 Base InceptionResnetV2 architecture.

Table VI: Parameter details of InceptionResnetV2 architecture.

Type of the layer	Input shape	Output shape	Number of parameters
image	224 x 224 x 3	128 x 128 x 3	0
Base model	128 x 128 x 3	2 x 2 x 1536	54,276,192
Global average pooling	2 x 2x 1536	1536	0
Dense	1536	1024	1,573,888
Dropout	1024	1024	0
Dense (output layer)	1024	5	5,125
Total parameters			55,855,205

4 Results and discussions

The Sipakmed dataset consists of five classes and includes 4049 individual images of isolated cells, manually extracted from 966 clustered cell images. The cell classes are categorized into superficial-intermediate cells, parabasal cells, koilocytotic cells, Dysketarotic cells and Metaplastic cells. Partitioning the SIPaKMeD dataset involves allocating 80% for training, 5% for testing, and the remaining 15% for validation. Following evaluation metrics have been used in the article for the categorization of Cervix cancer cells:

Accuracy, defined as the percentage of successfully forecasted rows to total rows based on the five categories supplied in the dataset. Precision determines how many anticipated outcomes are actually the same class label. Recall assesses the model's capacity to accurately identify all relevant instances within a dataset; F1-score is the balance of precision and recall, offering a single measure that optimizes the model's accuracy and recall. These metrics are calculated using the subsequent formulas as specified from equations 4.1 through 4.4:

$$\text{Accuracy}=\frac{(TP + TN)}{(TP + TN + FP + FN)} \quad \text{--------}\rightarrow \quad 4.1$$

$$\text{Precision}=\frac{TP}{TP+FP} \quad \text{---------}\rightarrow \quad 4.2$$

$$\text{Recall}=\frac{TP}{TP+FN} \quad \text{---------}\rightarrow \quad 4.3$$

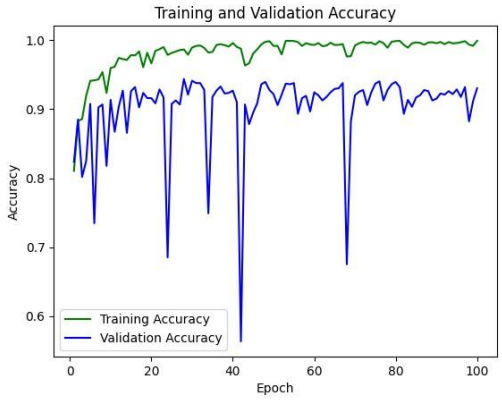
$$\text{F1-Score}=\frac{2*Precision*Recall}{Precision+Recall} \quad \text{---------}\rightarrow \quad 4.4$$

While accuracy is a common metric, it may not be sufficient in cases of imbalanced datasets. As SIPaKMeD dataset is balanced, accuracy can be considered as the most relevant metric in our research. The Error Matrix is a tabular representation that presents the numbers of correctly classified cancer cells and misclassified cancerous cells. This helps understand the types of errors made by the model.

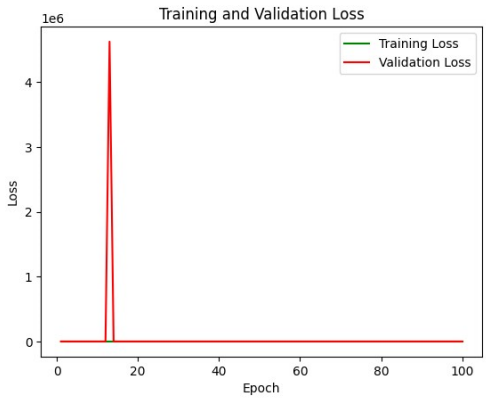
4.1 SIPaKMeD dataset classification before augmentation

Figure 4.1, 4.3 4.5 and 4.7 shows the accuracy and loss graphs of four CNN architectures plotted against training and validation data. Comparing training and validation accuracy helps assess whether the system generalizes well to new data or if there's over fitting, which is overcome by adding L2 regularisers in all the four models. The Loss graph represents the decrease in the model's loss on the training set over 100 epochs. A decreasing training loss indicates that the model is learning and adjusting its weights to minimize the error as and when epochs are increased. Monitoring validation loss helps identify overfitting. If the training loss keeps decreasing while the validation loss rises, it indicates that the model may be memorizing the training data and demonstrating suboptimal performance when presented with new, unseen data. Hence, along with the training dataset, model's loss and accuracy on a separate validation set are also depicted. Among Resnet50, Densenet201, Efficientnetb1 and InceptionResnetv2 models, Densenet201 achieved the highest accuracy of 96.74.

The Confusion matrix of the proposed methodology is represented in Figure 4.2,4.4,4.6 and 4.8. It offers an overview of the forecasts of the network in contrast to the real ground truth for five distinct classes. The matrix is particularly useful for understanding the types and frequencies of errors these four CNN models make [34].



(a) Resnet-50 Accuracy



(b) Resnet-50 Loss

Figure 4.1 Accuracy (a) and Loss (b) graphs of Training and Validation datasets of SIPaKMeD images for Resnet50 architecture.

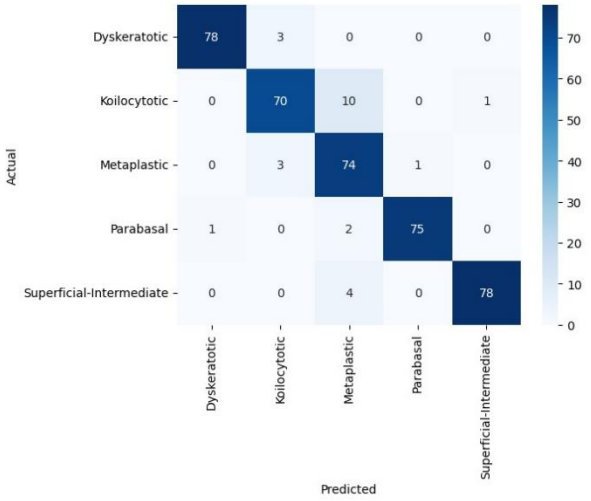
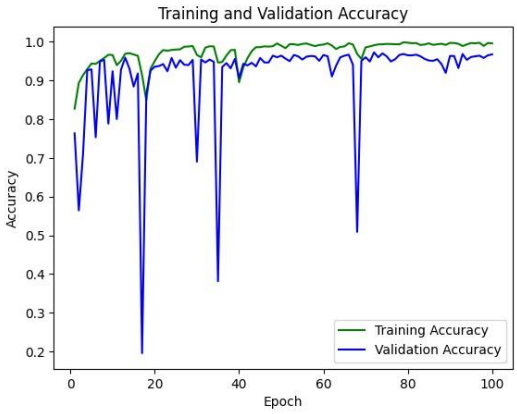
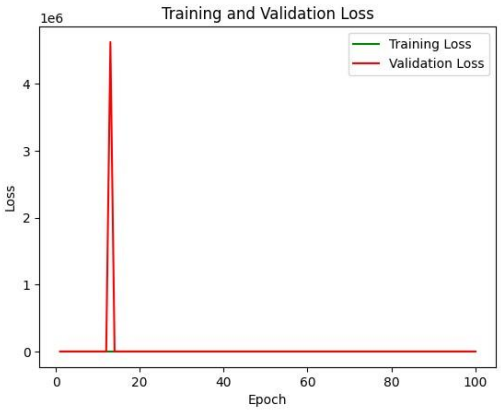


Figure 4.2 Confusion matrix obtained for Resnet50.



(a) Densenet-201 Accuracy



(b) Densenet-201 Loss

Figure 4.3 Accuracy (a) and Loss (b) graphs of Training and Validation datasets of SIPaKMeD images for Densenet201 architecture

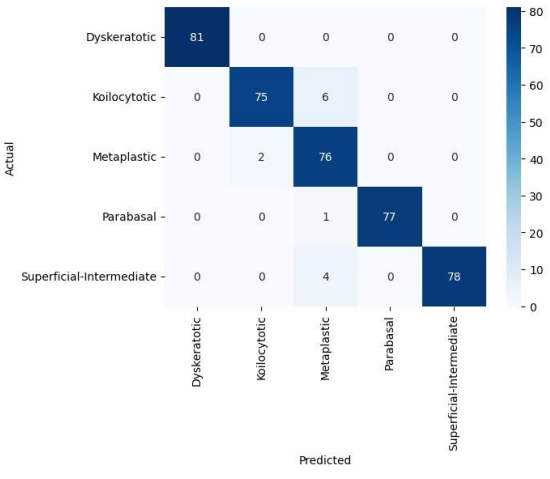
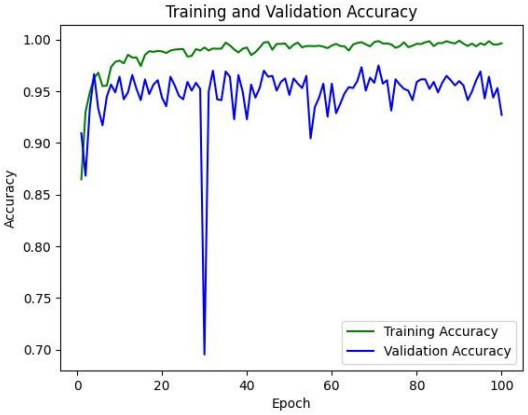
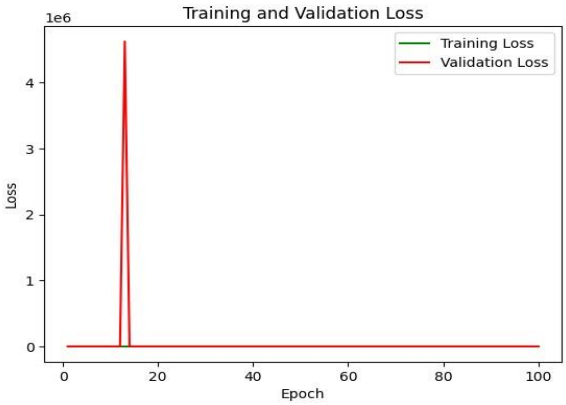


Figure 4.4 Confusion matrix obtained for Densenet201



a) Efficientnet-b1 Accuracy



(b) Efficientnet-b1 Loss

Figure 4.5 Accuracy (a) and Loss (b) graphs of Training and Validation datasets of SIPaKMeD images for EfficientnetB1 architecture

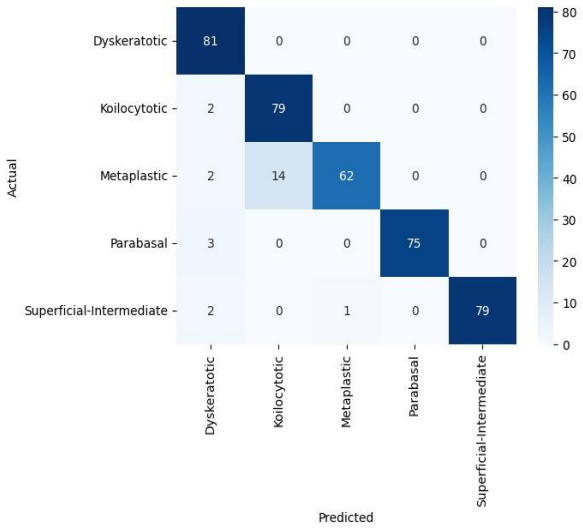
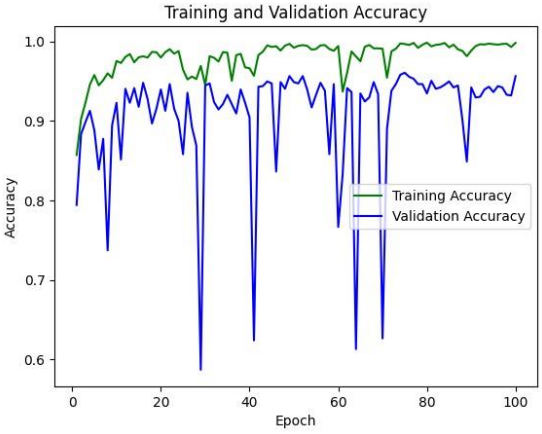
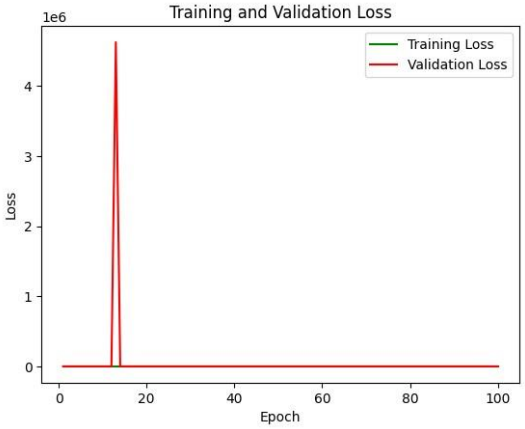


Figure 4.6 Confusion matrix obtained for EfficientnetB1



(a) Inception-Resnetv2 Accuracy



(b) Inception-Resnetv2 Loss

Figure 4.7 Accuracy (a) and Loss (b) graphs of Training and Validation datasets of SIPaKMeD images for InceptionResnetV2 architecture.

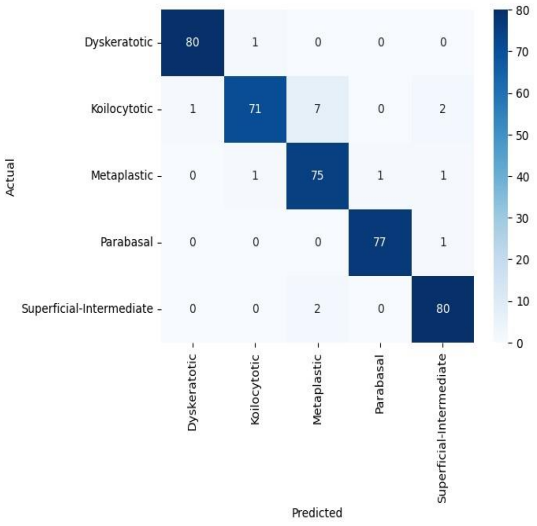


Figure 4.8 Confusion matrix obtained for InceptionResnetV2

4.2 SIPaKMeD dataset classification after augmentation

Different techniques for augmenting data including rotation, height shift, width shift, shear, zoom, horizontal flipping, and fill mode are used on SIPaKMeD dataset which increased number of images from 4049 to 7950. This has increased accuracy as described below. An improved accuracy and reduction in the loss of all the CNN models after augmentation are described in the Figureures 4.10, 4.12, 4.14 and 4.16.

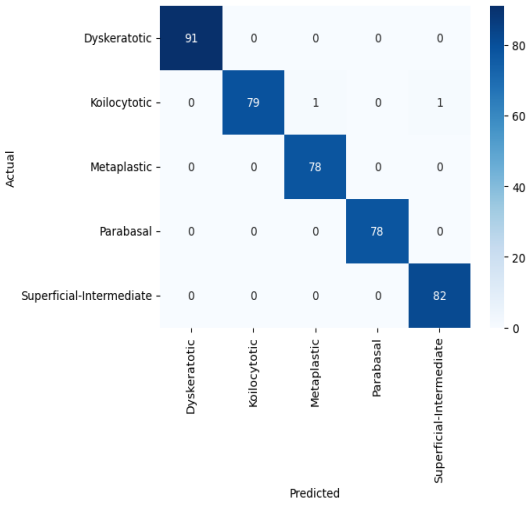
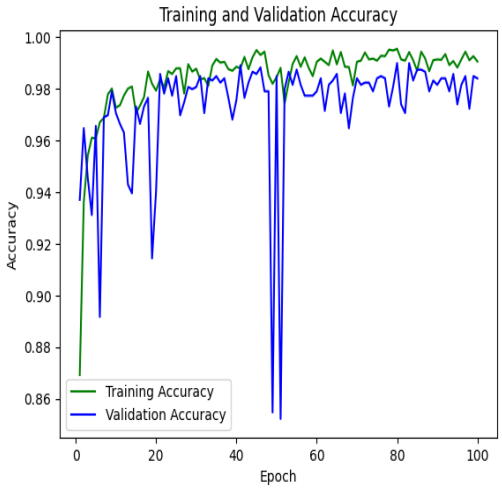
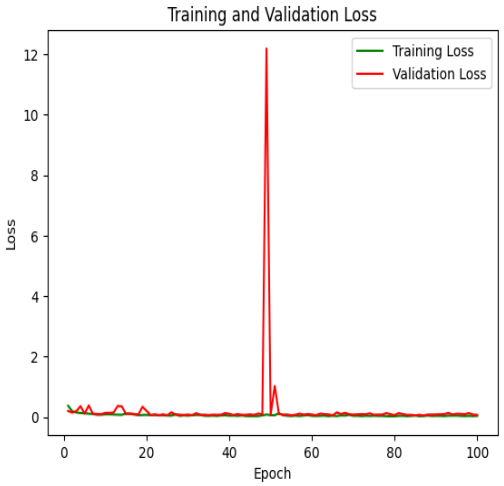


Figure 4.9 Confusion matrix obtained for Resnet50.



(a) Resnet-50 Accuracy



(b) Resnet-50 Loss

Figure 4.10 Accuracy (a) and Loss (b) graphs of Training and Validation datasets of SIPaKMeD images for Resnet50 architecture.

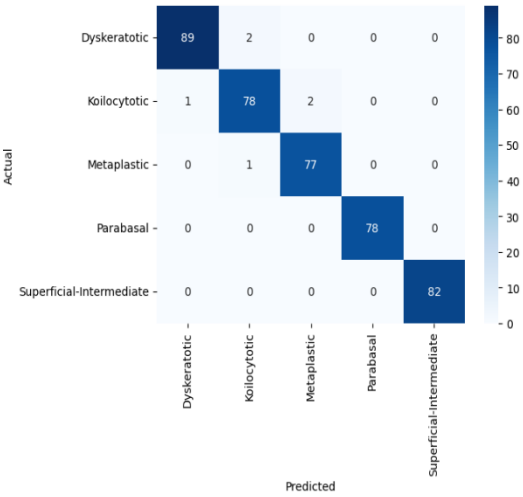
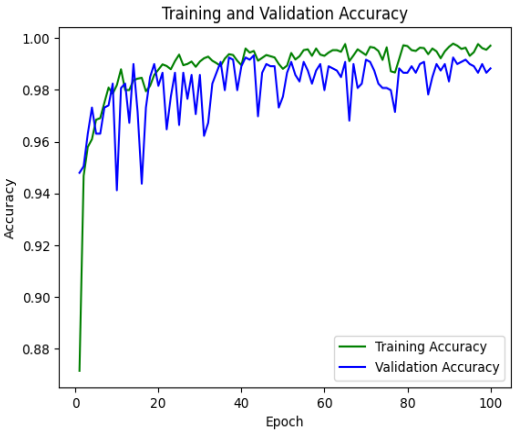
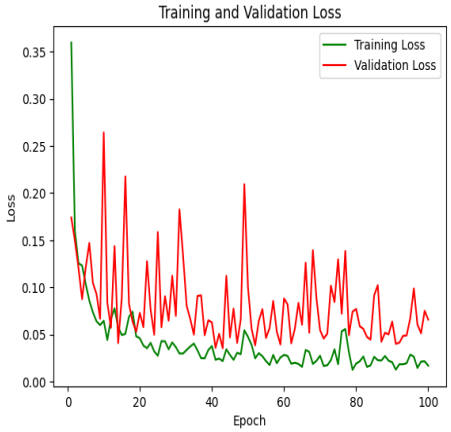


Figure 4.11 Confusion matrix obtained for Densenet201.



(a) Densenet-201 Accuracy



(b) Densenet-201 Loss

Figure 4.12 Accuracy and Loss graphs of Training and Validation datasets of SIPaKMeD images for Densenet201 architecture.

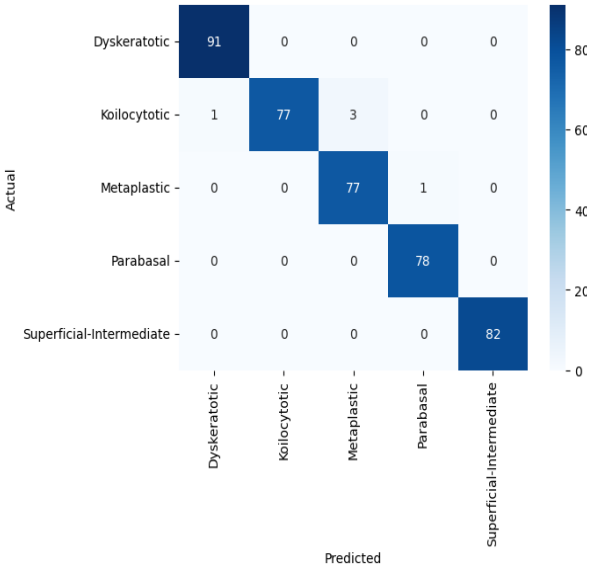
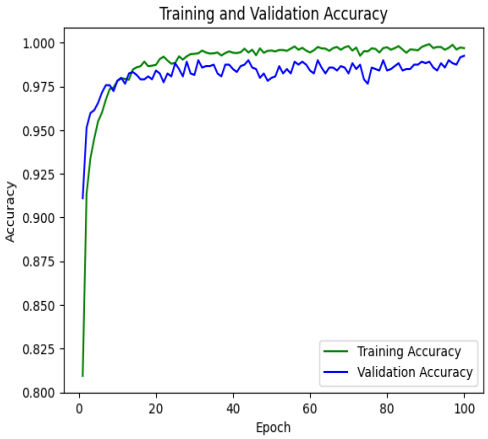
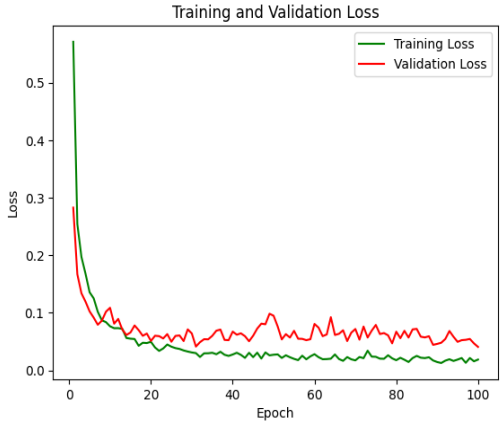


Figure 4.13 Confusion matrix obtained for Efficientnetb1.



(a) Efficientnet-b1 Accuracy



(b) Efficientnet-b1 Loss

Figure 4.14 Accuracy and Loss graphs of Training and Validation datasets of SIPaKMeD images for Efficientnetb1 architecture.

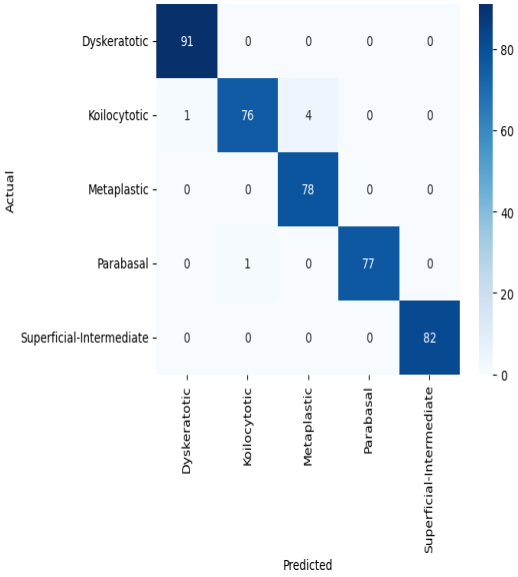
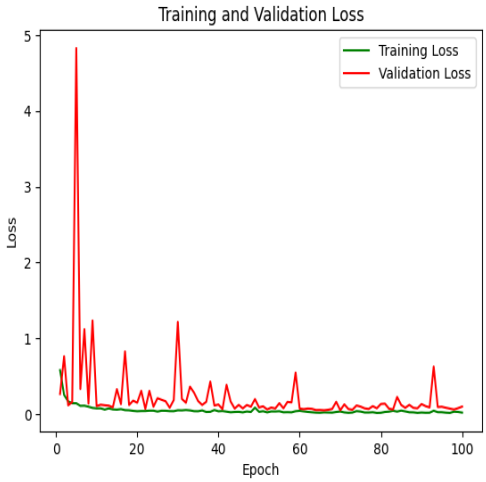


Figure 4.15 Confusion matrix obtained for InceptionResnetv2

The confusion matrix after the augmentation are shown in the Figure 4.9, 4.11,4.13 and 4.15.

(a) Inception-Resnetv2 Accuracy



(b) Inception-Resnetv2Loss

Figure 4.16 Accuracy and Loss graphs of Training and Validation datasets of SIPaKMeD images for InceptionResnetv2 architecture.

Table VII: Performance comparison between existing and proposed models.

Reference	**Model**	**Accuracy**	**Precision**
[35]	VGG16 VGG19 XceptionNet	98.27% 96.43% 65.77%	0.98 0.96 0.75
[36]	Bagging ensemble classifier	94.09%	
[37]	DenseNet-161	98.96%	
[3]	Voting ensemble (RF, SVM, FCNN)	88.95%	0.99
[2]	Resnet- 152	94.89%	
Proposed models	EfficientNet -b1 InceptionNet-Resnet-v2 DenseNet-201 Resnet -50	98.78% 98.53% 98.53% 99.51%	0.99 0.99 0.99 1.00

CNN architectures used in this work are compared against other ML and DL techniques as stated in table VII. The detailed metrics evaluation for the following proposed CNN [38] models are listed in the below table VIII. Resnet50 outperformed compared to the other three models.

Table VIII: Performance comparison of proposed models.

Sl No	**CNN model**	**Class labels**	**Accuracy**	**Precision**	**Recall**	**F1-Score**
1	Resnet -50	Dysketar otic	1.00	1.00	1.00	1.00
		koilocyto tic	1.00	1.00	0.98	0.99
		Metaplast ic	1.00	0.99	1.00	0.99
		parabasal	1.00	1.00	1.00	1.00
		superficia l-intermedi ate	1.00	0.99	1.00	0.99
2	InceptionNet-Resnet-v2	Dysketar otic	0.99	0.99	1.00	0.99
		koilocyto tic	0.99	0.99	0.94	0.96
		Metaplast ic	0.99	0.95	1.00	0.97
		parabasal	0.99	1.00	0.99	0.99
		superficia l-intermedi ate	0.99	1.00	1.00	1.00
3	EfficientNet -b1	Dysketar otic	0.99	0.99	1.00	0.99
		koilocyto tic	0.99	1.00	0.95	0.97
		Metaplast ic	0.99	0.96	0.99	0.97
		parabasal	0.99	0.99	1.00	0.99
		superficia l-intermedi ate	0.99	1.00	1.00	1.00
4	DenseNet-201	Dysketar otic	0.99	0.99	0.98	0.98
		koilocyto tic	0.99	0.96	0.96	0.96
		Metaplast ic	0.99	0.97	0.99	0.98
		parabasal	0.99	1.00	1.00	1.00
		superficia l-intermedi ate	0.99	1.00	1.00	1.00

5 Conclusion

This paper presents the multiclass classification of cervical cancer dataset named SIPaKMeD of five different classes using various CNN architectures namely Resnet50, Densenet201, Efficientnetb1, Inceptionresnetv2 and obtained an average accuracy of 95% on test data. Among these techniques, Densenet201 outperformed with the highest accuracy of 96.74%. After performing data augmentation with various techniques on the SIPaKMeD dataset, average accuracy has been increased to 98.78% with a great reduction in the loss and misclassification. Resnet50 won the race with the highest accuracy of 99.5% after augmentation. Moreover,

additional datasets can be incorporated for conducting classification analysis of cervical cancer. Additionally, deep learning architectures can facilitate stage-wise predictions in this context.

References

[1] Wanli Liu, Chen Li, Ning Xu, Tao Jiang, Md Mamunur Rahaman, Hongzan Sun, Xiangchen Wu, Weiming Hu, Haoyuan Chen, Changhao Sun, Yudong Yao, Marcin Grzegorzek, "CVM-Cervix: A hybrid cervical Pp-smear image classification framework using CNN, visual transformer and multilayer perceptron" – 2022. https://doi.org/10.1016/j.patcog.2022.108829

[2] Anurag Tripathi, Aditya Arora, Anupama Bhan, "Classification of cervical cancer using Deep Learning Algorithm" -2021. http://dx.doi.org/10.1109/ICICT50816.2021.9358570

[3] Swati Shinde, Madhura Kalbhor, Pankaj Wajire, "DeepCyto: a hybrid framework for cervical cancer classification by using deep feature fusion of cytology images" – 2022. http://dx.doi.org/10.3934/mbe.2022301

[4] Wasswa William, Andrew Ware, Annabella Habinka Basaza-Ejiri, Johnes Obungoloch, "A pap-smear analysis tool (PAT) for detection of cervical cancer from pap-smear images" – 2019. https://doi.org/10.1186/s12938-019-0634-5

[5] Alquran, H.; Alsalatie, M.; Mustafa, W.A.; Abdi, R.A.; Ismail, A.R. "Cervical Net: A Novel Cervical Cancer Classification Using Feature Fusion"- 2022. https://doi.org/10.3390%2Fbioengineering9100578

[6] Taranjit Kaur, Tapan Kumar Gandh, "Deep convolutional neural networks with transfer learning for automated brain image classification" – 2020. https://doi.org/10.1007/s00138-020-01069-2

[7] Pin Wang, Lirui Wang, Yongming Li, Qi Song, Shanshan Lv, Xianling Hu, "Automatic cell nuclei segmentation and classification of cervical Pap smear images" – 2018. https://doi.org/10.1016/j.bspc.2018.09.008

[8] Md Mamunur Rahaman, Chen Li, Xiangchen Wu, Yudong Yao, Zhijie Hu, Tao Jiang, Xiaoyan Li, Shouliang Qi, "A Survey for Cervical Cytopathology Image Analysis Using Deep Learning" – 2020. https://doi.org/10.1109/ACCESS.2020.2983186

[9] Pin Wang, En Fan, Peng Wang, "Comparative Analysis of Image Classification Algorithms Based on Traditional Machine Learning and Deep Learning" – 2020. https://doi.org/10.1016/j.patrec.2020.07.042

[10] Tawsifur Rahman, Muhammad E. H. Chowdhury, Amith Khandakar, Khandaker R. Islam, Khandaker F. Islam, Zaid B. Mahbub, Muhammad A. Kadir, Saad Kashem, "Transfer Learning with Deep Convolutional Neural Network (CNN) for Pneumonia Detection Using Chest X-ray" – 2020. http://dx.doi.org/10.3390/app10093233

[11] Kong, Shien Nie, et al. "Cervical Cancer Identification using Deep Learning Approaches." 2022 IEEE 8th International Conference on Computing, Engineering and Design (ICCED). IEEE, 2022. https://doi.org/10.1109/ICCED56140.2022.10010544.

[12] Gorantla, Rohan, et al. "Cervical cancer diagnosis using cervixnet-a deep learning approach." 2019 IEEE 19th international conference on bioinformatics and bioengineering (BIBE). IEEE, 2019. https://doi.org/10.1109/BIBE.2019.00078

[13] Hemalatha, K., and V. Vetriselvi. "Deep Learning based Classification of Cervical Cancer using Transfer Learning." 2022 International Conference on Electronic Systems and Intelligent Computing (ICESIC). IEEE,2022. https://doi.org/10.1109/ICESIC53714.2022.9783560

[14] Behnam Neyshabhr, Hanie Sedghi, Chiyun Zhang, "What is being transferred in transfer learning?" – 2020. https://doi.org/10.48550/arXiv.2008.11687

[15] Nanni, L., Ghidoni, S., Brahnam, S., "Ensemble of convolutional neural networks for bioimage classification"2021. https://doi.org/10.1016/j.aci.2018.06.002

[16] https://www.educba.com/keras-resnet50/.

[17] Jaiswal, Aayush, Neha Gianchandani, Dilbag Singh, Vijay Kumar, and Manjit Kaur. "Classification of the COVID-19 infected patients using DenseNet201 based deep transfer learning." Journal of Biomolecular Structure and Dynamics 39, no. 15 (2021): 5682-5689. http://dx.doi.org/10.1080/07391102.2020.1788642

[18] Jie, Yongshi, Xianhua Ji, Anzhi Yue, Jingbo Chen, Yupeng Deng, Jing Chen, and Yi Zhang. "Combined multi-layer feature fusion and edge detection method for distributed photovoltaic power station identification." Energies 13, no. 24 (2020): 6742. https://doi.org/10.3390/en13246742

[19] Krishna Adithya, Venkatesh, Bryan M. Williams, Silvester Czanner, Srinivasan Kavitha, David S. Friedman, Colin E. Willoughby, Rengaraj Venkatesh, and Gabriela Czanner. "EffUnet-SpaGen: an efficient and spatial generative approach to glaucoma detection." Journal of Imaging 7, no. 6 (2021): 92. http://dx.doi.org/10.3390/jimaging7060092

[20] Nguyen, Long D., Dongyun Lin, Zhiping Lin, and Jiuwen Cao. "Deep CNNs for microscopic image classification by exploiting transfer learning and feature concatenation." In 2018 IEEE international symposium on circuits and systems (ISCAS), pp. 1-5. IEEE, 2018. https://doi.org/10.1109/ISCAS.2018.8351550

[21] Salama, Gerges M., Asmaa Mohamed, and Mahmoud Khaled Abd-Ellah. "COVID-19 classification based on a deep learning and

machine learning fusion technique using chest CT images." Neural Computing and Applications (2023):1-19. http://dx.doi.org/10.1007/s00521-023-09346-7.

[22] Jogin, Manjunath, M. S. Madhulika, G. D. Divya, R. K. Meghana, and S. Apoorva. "Feature extraction using convolution neural networks (CNN) and deep learning." In 2018 3rd IEEE international conference on recent trends in electronics, information & communication technology (RTEICT), pp. 2319-2323. IEEE, 2018. https://doi.org/10.1109/RTEICT42901.2018.9012507.

[23] M. E. Plissiti, P. Dimitrakopoulos, G. Sfikas, C. Nikou, O. Krikoni, A. Charchanti, Sipakmed: A new dataset for feature and image based classification of normal and pathological cervical cells in Pap smear images, in IEEE International Conference on Image Processing (ICIP) 2018, (2018), 3144–3148. https://doi.org/10.1109/ICIP.2018.8451588

[24] SIPakMed Database, 2022. Available from: https://www.cs.uoi.gr/~marina/sipakmed.html.

[25] Guo lili, Ding shifei, "Deep learning research progress [J]", 2015.

[26] Akshay Mangawati, Mohana, Mohammed Leesan, H. V. Ravish Aradhya, "Object Tracking Algorithms for video survilllance applications" International conference on communication and signal processing (ICCSP), India, 2018, pp. 0676-0680. https://doi.org/10.1109/ICCSP.2018.8524260

[27] Apoorva Raghunandan, Mohana, Pakala Raghav and H. V. Ravish Aradhya, "Object Detection Algorithms for video survilllance applications" International conference on communication and signal processing (ICCSP), India, 2018, pp. 0570-0575. https://doi.org/10.1109/ICCSP.2018.8524461

[28] Kibriya, Hareem, and Rashid Amin. "A residual network-based framework for COVID-19 detection from CXR images." Neural Computing and Applications 35, no. 11 (2023): 8505-8516. https://doi.org/10.1007/s00521-022-08127-y

[29] Wang, Shui-Hua, and Yu-Dong Zhang. "DenseNet-201-based deep neural network with composite learning factor and precomputation for multiple sclerosis classification." ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) 16, no. 2s (2020): 1-19. https://doi.org/10.1145/3341095

[30] Krishna Adithya, Venkatesh, Bryan M. Williams, Silvester Czanner, Srinivasan Kavitha, David S. Friedman, Colin E. Willoughby, Rengaraj Venkatesh, and Gabriela Czanner. "EffUnet-SpaGen: an efficient and spatial generative approach to glaucoma detection." Journal of Imaging 7, no. 6 (2021): 92. https://doi.org/10.3390/jimaging7060092

[31] Nguyen, Long D., Dongyun Lin, Zhiping Lin, and Jiuwen Cao. "Deep CNNs for microscopic image classification by exploiting transfer learning and feature concatenation." In 2018 IEEE international symposium on circuits and systems (ISCAS), pp. 1-5. IEEE, 2018. https://doi.org/10.1109/ISCAS.2018.8351550

[32] Tan, Mingxing, and Quoc Le. "Efficientnet: Rethinking model scaling for convolutional neural networks." In International conference on machine learning, pp. 6105-6114. PMLR, 2019. https://doi.org/10.48550/arXiv.1905.11946.

[33] Szegedy, Christian, Sergey Ioffe, Vincent Vanhoucke, and Alexander Alemi. "Inception-v4, inception-resnet and the impact of residual connections on learning." In Proceedings of the AAAI conference on artificial intelligence, vol. 31, no. 1. 2017. http://dx.doi.org/10.1609/aaai.v31i1.11231

[34] Madhusudhan, M. V., V. Udayarani, and C. Hegde. "Finger vein based authentication using deep learning techniques." Int. J. Recent Technol. Eng 8, no. 5 (2020): 5403-5408. http://dx.doi.org/10.1088/1757-899X/1105/1/012032

[35] Md Mamunur Rahaman, Chen Li, Yudong Yao, Frank Kulwa, Xiangchen Wu, Xiaoyan Li, Qian Wang, "DeepCervix: A deep learning-based framework for the classification of cervical cells using hybrid deep feature fusion techniques" – 2021. https://doi.org/10.1016/j.compbiomed.2021.104649

[36] Kyi Pyar Win, Yuttana Kitjaidure, Kazuhiko Hamamoto, Thet Myo Aung, "Computer-Assisted Screening for Cervical Cancer Using Digital Image Processing of Pap Smear Images" – 2020.

[37] Muhammed Talo, "Diagnostic Classification of Cervical Cell Images from Pap Smear Slides"- 2019. http://dx.doi.org/10.33793/acperpro.02.03.116

[38] Shruthi, U., and V. Nagaveni. "TomSevNet: a hybrid CNN model for accurate tomato disease identification with severity level assessment." Neural Computing and Applications 36, no. 10 (2024): 5165-5181. http://dx.doi.org/10.1007/s00521-023-09351-w

[39] Mousser, Wafa, and Salima Ouadfel. "Deep feature extraction for pap-smear image classification: A comparative study." In Proceedings of the 2019 5th International Conference on Computer and Technology Applications, pp. 6-10. 2019. http://dx.doi.org/10.1145/3323933.3324060

[40] Pal, Anabik, Zhiyun Xue, Brian Befano, Ana Cecilia Rodriguez, L. Rodney Long, Mark Schiffman, and Sameer Antani. "Deep metric learning for cervical image classification." IEEE Access 9 (2021): 53266-53275. https://doi.org/10.1109/ACCESS.2021.3069346

[41] Kalbhor, Madhura, Swati Vijay Shinde, and Hemant Jude. "Cervical cancer diagnosis based on cytology pap smear image classification using fractional coefficient and machine learning classifiers." TELKOMNIKA (Telecommunication Computing Electronics and Control) 20, no. 5 (2022): 1091-1102. http://doi.org/10.12928/telkomnika.v20i5.22440

[42] Chandana, B. Sai. "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 14, no. 6 (2023): 613-623. http://dx.doi.org/10.32985/ijeces.14.6.1

[43] Li, Chen, Hao Chen, Le Zhang, Ning Xu, Dan Xue, Zhijie Hu, He Ma, and Hongzan Sun. "Cervical histopathology image classification using multilayer hidden conditional random fields and weakly supervised learning." Ieee Access 7 (2019): 90378-90397. https://doi.org/10.1109/ACCESS.2019.2924467