EAI Endorsed Transactions
on Pervasive Health and Technology

Research Article EAI.EU

DeepCerviCancer - Deep Learning-Based Cervical Image Classification using Colposcopy and Cytology Images

Madhura Kalbhor[1], Dr. Swati Shinde[2,*], Sagar Lahade[3] and Dr. Tanupriya Choudhury[4]

[1] Pimpri Chinchwad College of Engineering, Sector -26, Pune, 411044, Maharashtra, India.
[2] Pimpri Chinchwad College of Engineering, Sector -26, Pune, 411044, Maharashtra, India.
[3] Pimpri Chinchwad College of Engineering, Sector -26, Pune, 411044, Maharashtra, India.
[4] University of Petroleum and Energy Studies, Dehradun, Uttarakhand, India.

Abstract

INTRODUCTION: Cervical cancer is a deadly malignancy in the cervix, affecting billions of women annually.

OBJECTIVES: To develop deep learning-based system for effective cervical cancer detection by combining colposcopy and cytology screening.

METHODS: It employs DeepColpo for colposcopy and DeepCyto+ for cytology images. The models are trained on multiple datasets, including the self-collected cervical cancer dataset named Malhari, IARC Visual Inspection with Acetic Acid (VIA) Image Bank, IARC Colposcopy Image Bank, and Liquid-based Cytology Pap smear dataset. The ensemble model combines DeepColpo and DeepCyto+, using machine learning algorithms.

RESULTS: The ensemble model achieves perfect recall, accuracy, F1 score, and precision on colposcopy and cytology images from the same patients.

CONCLUSION: By combining modalities for cervical cancer screening and conducting tests on colposcopy and cytology images from the same patients, the novel approach achieved flawless results

Keywords: Deep learning, KNN, SVM, LDA

Received on 20 June 2023, accepted on 24 September 2023, published on 03 October 2023

doi: 10.4108/eetpht.9.3473

[1]Corresponding author. Email: swati.shinde@pccoepune.org

1. Introduction

Cervical cancer is a term that describes an instance of malignancy that originates in the cervix, the bottom portion of the uterus that binds the vagina. Human papillomavirus is the primary causative agent [1,2,3]. The outer layer of the cervix is lined with squamous (thin, flat) cells, and these can transform into squamous cell carcinoma, which is one of the two elementary kinds of cervical cancer. Adenocarcinoma, on the other hand, develops in the glandular (column-shaped) cells lining the cervical canal [1]. Among female malignancies, cervical cancer has a high incidence rate. The prevalence of cervical cancer among women is a stark illustration of the effects of global health disparity. Cervical cancer affects over 5.6 billion women worldwide every year and it has a 90% fatality rate. The research projects an annual increase of over 300,000 deaths and over 500,000 new cases. And the study indicates that 85% of those deaths occurred in developing countries [4]. More than 700 women die every day from cervical cancer, and that number is expected to rise to an astounding 400,000 by the year 2030. More than 266,000 women lost their lives to cervical cancer in 2012. More than 311,000 women aged 20-39 lost their lives to cervical cancer last year. Each year, over 0.025 million females in Europe lose their lives to cervical cancer, and over 0.060 million new cases are identified [5]. Every year, 0.12 million Indian females are diagnosed with cervical cancer,

this constitutes 15.2% of all cancer-related deaths in the nation of India [6,7].

In most instances, the cancer of the cervix was detected too late because of a lack of awareness. Cervical abnormalities, which may eventually lead to cervical cancer, progress very gradually. Cervical cancer is difficult to detect early since there are few warning signs. Patients in developing countries are less likely to get regular tests due to a lack of awareness about the need of doing so. Contrary to expectations, high-income countries with well-established cervical screening programs had the lowest death rates [4,8,9]. With sufficient quality and quantity of screenings, the incidence of cervical cancer might be lowered by as much as 90% [10].

Cervical cancer screening and diagnosis are hindered by inadequate healthcare infrastructure in low and middle-income nations. Cervical cancer is curable if recognized early however, detection rates are low because of barriers including high testing costs and inadequate accessibility. To detect precancerous cells, the Pap smear test is presently the gold standard. The Pap smear test is important, but it takes a long time to acquire the results. Pathologists must analyze hundreds of cells on a single slide and label them all correctly [11]. The Pap smear test has been in use since the 1940s, yet it still has some serious drawbacks. Problems at any point in the examination process might introduce errors, increasing the likelihood of both false positives (in which a lesion is incorrectly identified) and false negatives (in which an existing lesion is not discovered). Cytological sample collection through lesion analysis is an example of these steps. The percentage of mistakes made during hand microscopic inspection of smears might reach 0.62 [12-15].

Cervical cancer cells are readily detectable and testable in their early stages thanks to screening techniques. Tests for early detection are discussed in the next paragraph. Cells from the cervix of the uterus are examined under a microscope in a procedure known as a Pap smear. A Pap smear involves the doctor inserting brushes within the cervix with the goal of collecting cells. In the laboratory, more cells are examined for issues. This test is very accurate and reliably differentiates between cervix cells that are normal and those that aren't. Human Papillomavirus (HPV) test: The infected cervix of the uterus is sampled for this test. Women over the age of 30 are encouraged to do this test, whereas younger women should get a Pap smear instead. Diagnosis is the process of identifying the ailment that is causing the symptoms a patient is experiencing. Colposcopy (a magnifying instrument) aids in the analysis of cervical cancer cells by a physician. During a diagnostic clinical technique called a punch biopsy, a tiny sample of skin and underlying tissue is taken away and scrutinized under a microscope. Using this technique, malignant cells in the study region may be inspected and analyzed. A curette, fashioned like a spoon, is used in endocervical curettage (ECC) to scrape away the cervix mucosa. The following tests may be performed if the infection found during the punch biopsy or ECC requires further investigation. Electrical wire loop: This procedure includes extracting a small tissue sample using a low-voltage electrical cable that is tenuous and stretchy. This is often carried out in the office under local anesthetic. Cone biopsy: A cone biopsy allows a trained healthcare professional to acquire cervical tissue from the interior layers of the cervix for analysis in the laboratory [16].

Rapid advancements in artificial intelligence and machine learning, as well as digital platforms, have led to unprecedented growth in the healthcare sector during the last decade [17]. Deep learning (DL) is one of the most promising disciplines because of the many possibilities it presents in the medical field. The use of DL has the potential to improve healthcare by allowing doctors to make more precise and timely diagnosis that may then inform more targeted treatment plans for each patient. Significant progress in medical image analysis for the detection and diagnosis of many forms of cancer has also been proven by DL, allowing for speedier and more effective treatment of patients. Improved patient care is the outcome of this technology's decreased dependence on physicians while still delivering timely and accurate findings to patients.

Multiple advances, including AI, have been put into practice out of a desire to improve clinical care's efficiency and effectiveness. The need to optimize and streamline clinical procedures has grown in importance in light of rising healthcare service demands and the massive amounts of data being created every day from several concurrent sources. Artificial intelligence's strength is in its ability to spot intricate patterns in pictures, which presents a unique chance to make the previously qualitative and subjective process of image interpretation into a measurable and easily repeatable one. Artificial intelligence has the potential to improve how diagnoses are made by doctors by gleaning hidden details from images. Medical imaging, genetics, pathology, electronic health records (EHRs), and social media might all benefit from the incorporation of artificial intelligence into more efficient diagnostic systems [18].

In this paper, the authors proposed an ensemble approach to improve the classification accuracy of abnormal and normal classes for colposcopy and pap screening images. The approach involved training two separate models, M1 and M2, on different datasets of colposcopy and Pap screening images. These models were then used in an ensemble approach where M1 took a colposcopy image of a patient and M2 took a Pap smear image of the same patient at the same time. The ensemble approach utilized the probabilities of abnormal and normal classes for both images, which were obtained by the models, and used them as features for machine learning approaches. This unique approach, using images of exactly the same patient for both colposcopy and pap screening, sets this research apart.

The following aspects distinguish this research from previous work, making it novel and valuable in terms of its contributions.

- The integration of colposcopic and cytology images into an AI system for cervical cancer detection marks a remarkable milestone in the field. This approach offers a more comprehensive and rigorous method for identifying potential cancer cases in patients. By leveraging multiple imaging modalities, the system can produce more accurate and reliable results.

- The Models were trained using numerous datasets, which resulted in a more comprehensive knowledge base and better performance when detecting cancer in unseen pictures.

- Real-world datasets from a hospital in Assam, India were collected and utilized to account for biological or physiological adaptations, which we named as Malhari. This approach ensured that the system was trained on relevant data that would lead to a more accurate and effective cervical cancer detection model.

- The ensemble model was subjected to a performance evaluation using colposcopy screening and cytology screening images sourced from the same patients, representing a real-world use case scenario. This approach deviates from the norm of using random images of the same category, as it offers a more reliable and relevant evaluation of the AI system's efficacy in detecting cervical cancer.

- According to the study's findings, the ensemble approach employed in this research achieved a perfect score of 100% for accuracy, recall, f1 score, and precision.

- Given the significant threat that cervical cancer poses to developing countries, where the availability of doctors is often limited, the development of AI solutions has become crucial in addressing this issue. To this end, the dataset used in this study, collected from real patients, has been made available for further research.

The subsequent article is divided into many sections. The data sets that were used are described in the "Datasets" section. The "Proposed System" section explains how the researchers implemented the ensemble approach. In the next part, it describes the results and analyses. Recent studies and advancements in this field are summarized in the "related work" section.

2. Related Work

Table 1 shows the summaries of the different papers studied and analyzed
.

Ref.	Methodology	Description	Advantage	Disadvantage	Result
[1]	The cervical cancer cell nucleus is segmented using the gradient force model and the balloon force model, using approaches to pre-process such as edge mapping with a double threshold for eliminating noise from edges. They use two parametric deformable models to strike	The contributors make available an approach for the automated computerized identification of cervical cell nuclei.	This approach's incorporation of both models permits finding a trade-off between effectiveness along with computational economy to be found that works satisfactorily in practice.	Conditional on both moles	Accuracy of 92%

.	an equilibrium between computational cost and precision.				
[4]	Cervix localization in the input images was performed using a faster R-CNN model. To enhance the effectiveness of the model, synthetic samples were generated with the assistance of GAN. Ultimately, the observed cervix was categorized by means of a classifier.	According to the authors present knowledge, the FSOD-GAN framework is the initial model that can execute multilevel, multiclass classification tasks in order to tell the difference between healthy and unhealthy cervical images and to divide illnesses according to their severity and kind.	Incorporating both the faster R-CNN and the GAN structures into a single deep-learning model is what makes the FSOD-GAN unique.	Synthetic sample utilization calls into doubt the reliability of the results.	Diagnosis success rate of 99%

[6]	MobileODT is an apparatus comprising a smartphone and other required components, envisioned for snapping photographs of women's cervical sections in remote settings. Following the photographs that have been taken, then they are uploaded to a website where a physician may examine them. The patient is informed of the findings and given instructions on how to proceed.	Platform designed to scale up distant cervical cancer screening efforts.	It may be dependent on battery	Depending on the accessibility and expertise of medical professionals, network assets	-
[19]	Preprocessed photos of the cervical region were analyzed, and characteristics were identified and fed into Efficient Net to determine whether the images were normal or infectious. The anomalous region was segmented using SegNet, and the dataset was split into training and testing sets. Cervical cancer was classified using the trained model.	In the study, researchers suggested a method that would use image processing to improve Pap smear scans and extract attributes that might be used to identify healthy and unhealthy cervical tissue.	1. Adopted a technique referred to as CLAHE aimed at enhancing the appearance of pictures. 2. Wavelet and GLCM image characteristics were employed to train an efficient net. 3. The SegNet approach is utilized on the abnormal cervical image to locate and segment the cancer zone. 4. Herlev dataset is used	1. Because the research relied on Pap smear pictures, it may be limited in its usefulness to identifying and monitoring other kinds of cervical cancer that may need alternative imaging techniques. 2. Required a massive quantity of data.	Accuracy was 98.29%, while sensitivity was 97.42%.

[20]	1. MobileNet, ResNet and shuffleNet used to extract features. 2. GLCM, DWT, GW were used.	1. Initially, the Pap smear pictures are created, and then deep learning and handmade characteristics are obtained in the next step. After that, the features are merged and reduced using principal component analysis before being clustered using support vector machines.	1. The elimination of the need for pre segmentation simplifies the procedure. 2. Added complexity by merging different CNN. 3. Images of pap smears were analyzed in the spatial and time-frequency domains to extract a number of textural characteristics.	1. The stated CAD has several shortcomings, which include being the case that it does not fine-tune the CNNs. 2. The proposed technique has not been employed outside of the context of Pap images.	100% accuracy
[21]	The authors commenced by designing three distinct foundation modules for gathering feature data from a variety of kernels of varying sizes. The cervical cell categorization model was then built by stacking a number of these simple modules.	Developed a deep CNN by stacking fundamental modules.	1. Create a model for classifying cervical cells that do not need segmentation and are very accurate. 2. Utilized datasets are Herlev and SIPaKMeD.	The proposed technique has not been employed outside of the context of cervical cell image categorization in this research.	Achieved a 95.628 percentile on the SIPaKMeD dataset for accuracy. In contrast, the Herlev dataset only produced a result of 92.717.
[22]	Relu, Prelu, Leaky Relu used	Three different implementations of the same design were used to investigate the effect of distinct activation functions on a residual neural network. In order to learn how the activation function affects the precision of the model.	ResNet was built with a consistent structure across trials so as to make the most of the influence of the activation function on the performance of the model.	One more general drawback of comparing the effect of activation functions on a ResNet model is that the results may not be transferable to other models or challenges.	Accuracy of 90% and 100% were achieved using Leaky-RELU and PRELU, respectively. However, when compared to their performance, ReLU fell short.
[23]	1. The network architecture of the CYENET model includes 15 layers for ConvNet operations, 12	In order to identify cervical cancer using colposcopy pictures, the researchers developed and deployed two	1. When trained, the CYENET model was 0.971 accurate, whereas the VGG_19 (TL) model was only 0.870 accurate. 2. In contrast to CYENET's steady validation methodology and smooth loss curve,	1. Both the dataset's size and the layer's depth have an effect on the model's efficiency. Due	1. The CYENET model outperforms the VGG19 (TL) model by 0.19 points in terms of classification accuracy, reaching 0.923 2. CYENET Accuracy

	stages for activation function application on the output, 5 layers for down sampling the input signal via max-pooling and 4 layers for local response normalization (LRN). 2. Only 12.76 percent of the 5679 photos were classified as Type 1, while 55.04 percent were classified as Type 2, and 33.08 percent were classified as Type 3. (Imbalanced data, so oversampling and augmentation used) 3. used parallel block to extract features 4. Relu function used	different ConvNet architectures: VGG19 (TL) and CYENET.	VGG19's is unstable. 3. Model is better than DenseNet-121, DenseNet-169, Colponet, SVM,Inception-Resnet-v2	to the interclass similarity issue, the deep architecture might have a negative impact on the overall model's classification performance.	92.30, Sensitivity 92.40, Specificity 96.20
[24]	A logistic regression model was developed using ResNet50's output probability, and clinical considerations were integrated into it.	1. 15,276 pictures were analyzed, representing 7,530 patients. 2. Used a colposcopist's assistance to map out the area around the suspected lesion before entering it into the model.	Researchers assessed the accuracy of diagnoses provided by experts with varied degrees of expertise to guarantee the usefulness of our proposed strategy. Case 1: NC vs. LSIL+ CAD model wins against the expert	Comparison with colposcopists is done Case 2: HSIL vs. HSIL+ CAD model fail against expert but in case of sensitivity model wins	1. ResNet alone isn't as effective as the model that incorporates clinical characteristics. 2. The overall model has 0.95 accuracy when comparing NC 1 to LSIL+ and a 0.90 accuracy when comparing HSIL to HSIL+.
[25]	1. A real-world dataset consisting of 230 photographs was used to evaluate the classification methods. There were 31.30% NILM pictures, 24.34% LSIL pictures,	Different approaches tested on pap smear multi-cell images like pre-trained approach, Ensemble approach, unique CNN model without auto encoders as well as with auto encoders	The implemented model (AE CNN) has very less trainable parameters (30,647) as compared to resNet-50 (23,587,712)	1. model is not effective that much to differentiate between HSIL and SSC because of similar features 2. resNet-50 beats all the other model like a pre-trained model,	1. Result on pre-trained models ResNet50, DenseNet121, ResNet101, ResNet153, and EfficientNetB0 are used for training and testing among them resnet-50 had highest accuracy of 99.2 2. Result on Ensemble Method (ResNet50, ResNet101, ResNet153, and

.	20.86% HSIL pictures, and 23.47% SCC pictures. 2. Augmentation is used to handle class imbalance problem 3. With approximately 0.060772 million parameters, the auto encoder-free model has a substantially smaller number of trainable parameters. This is around 99.78% less parameters than the typical amount seen in transfer learning models. 4. The number of parameters was lowered to 20,355 by using the AE CNN model. 5. The Sure Path liquid cytology Dataset developed by Hussain et al. used			Ensemble model, CNN and AE CNN but at the cost of high computation	EfficientNetB0) accuracy obtained 98% 3. Result on mode with auto encoders (AE CNN) accuracy obtained 3.90% 3. Result on mode with auto encoders (AE CNN) accuracy obtained 96.54%
[26]	The study participants used a shape-based iterative technique to locate nuclei and a marker-controlled watershed approach to separate overlapping cytoplasm in cell segmentation. The RF method was used to extract characteristics from segmented nuclei and cytoplasm. By integrating the outputs of, k nearest neighbor, LD, support vector machine, boosted trees, and bagged trees, a bagging ensemble classifier was employed to increase accuracy during classification.	The goal of the research was to develop a computer-based screening method for cervical cancer detection that makes use of digital image processing of Pap smear pictures.	1. Use of the Herlev and SIPaKMeD datasets 2. Both single-cell and multi-cell data were put through the tests.	As more classifiers are used, complexity increases.	Classification accuracy of 94.09% for five classes, and 98.27% for five
[27]	As a first step, the cervical area was acquired using enhanced Region-CNN. cls-net used for	The work covers the design and implementation of a whole ensemble deep learning model for the	The ASPP module is used to capture multi-scale features.	Cascade model used	Accuracy obtained 99.6

	feature extraction. Features were extracted using cls-net. After CNN determines if an image is normal or not, a second model is trained to categorize anomalous pictures into one of three groups: SSC, HSIL, or LSIL.	autonomous diagnosis of a whole slide image.			
[28]	Six ML classifiers were employed in constructing the voting classifier and performance is compared with DNN	Prognosis of cancer based on the patient's medical characteristics	Superior performance compared to DNN	1. The precision of DNNs is susceptible to variations in model design. 2. Additional processing power is a prerequisite	The voting classifier has a sensitivity of 100%, with an accuracy of 97-99%.
[29]	Nuclei are detected and classified as abnormal or normal using Mask R-CNN.	Research effectively proved the effectiveness of Mask R-CNN for pap smear histology slides for cervical cancer screening	Revolutionary new method	Since it is a brand-new experiment, there is no standard against which to judge it.	Exceptional accuracy (89.8%) and high sensitivity (72.5%)
[30]	The ensemble method was developed with the use of neural network architectures such as Inception Net, which is XceptionNet, MobileNet, and EfficientNetB6, 5, 4, 3, 2, 1, and 0.	Ensemble model built	1. Accuracy is better than individual models 2. Drop out, cross-validation used	more complex, required more resources	Recall, precision, accuracy 96%
[31]	Combination of adaptive pruning and transfer learning used	Considering the intent of categorizing pap cervical cancer screening pictures, an adaptive pruning deep TL model, PsiNet-TAP is highlighted.	Novel model is implemented to handle limited dataset	Tested Inadequate size of the picture dataset	accuracy> 98

3. Dataset:

The section that follows offers a detailed summary of all the datasets utilized in this investigation. The datasets were carefully chosen for their relevance to the study questions and aims, and they came from reliable sources to assure their quality and trustworthiness.

3.1. IARC Colposcopy Image Dataset (Dataset-I)

The IARC Colposcopy Image Dataset is a collection of 1313 abnormal and normal high-quality cervical cancer screening images from two datasets, the IARC Visual Inspection with Acetic Acid (VIA) Image Bank and the IARC Colposcopy Image Bank. In the following manuscript, this dataset is referred to as D1.

3.1.1. The IARC Visual Inspection with Acetic Acid (VIA) Image Bank

This dataset is a valuable resource for cervical cancer screening and research [32]. The database includes over 408 cervical images with various attributes, including the visibility and location of the transformation zone, features of the Acetowhite region and its existence, and the results of VIA screening. These attributes provide important information for healthcare professionals to identify cervical abnormalities and determine the appropriate management. The acetowhite area, in particular, is a significant indicator of cervical abnormalities and can vary in color, margin, surface, location, and size. The database comprises data regarding the nature of the conducted tests, including pre- or post-application of Lugol's iodine same for acetic acid, as well as qualification for ablative treatment when relevant. Moreover, the images are annotated with histological results, spanning from normal to different levels of cervical cancer. Figure 1 shows the folder of the IARC image bank VIA datasets.

IARCImageBankVIA (Folder Structure)
├──
Case 001
│ ├── AFC0.jpg
│ └── AFC1.jpg
├── Case 002
│ ├── AJL0.jpg
│ └── AJL1.jpg
.
.
.

├── Case 185
│ ├── ADF0.jpg
│ └── ADF1.jpg
├── Case 186
│ ├── ABK0.jpg
│ └── ABK1.jpg
├── Cases - Images.xlsx
├── Cases Meta data.xlsx

Figure 1. IARC image bank VIA dataset's folder structure.

3.1.2. The IARC Colposcopy Image Bank

This dataset is a valuable resource for researchers and medical professionals studying cervical cancer [33]. This dataset contains colposcopy screening images along with metadata files. Metadata for colposcopy images included in the Image Bank dataset in , including information about the patient, imaging equipment, and image characteristics. The images are annotated by expert colposcopists, providing a standardized and reliable dataset for research and clinical applications. The availability of this dataset can contribute to the development of improved screening methods, diagnostic tools, and treatment options for cervical cancer, ultimately leading to better outcomes for patients. Researchers can use this metadata to analyze the images and identify patterns or features that may be useful for diagnosis, developing smart deep learning screening solutions and treatment planning. Figure 2 shows the folder of the IARC image bank colpo datasets.

The cases metadata file contains following attributes

1. HPV (human papillomavirus) Status: is a sexually transmitted virus that is known to cause nearly all cases of cervical cancer. The virus can infect the cells of the cervix, causing abnormal changes that may lead to cancer over time. Therefore, HPV testing is a crucial part of cervical cancer screening and detection.
2. Provisional diagnosis: Based on the results of these tests, a healthcare provider may make a provisional diagnosis of cervical cancer, which may include information such as the type of cancer, the stage of cancer, and the degree of malignancy.
3. Histopathology: Histopathology refers to the examination of tissue samples under a microscope to detect abnormalities, such as cancer or precancerous changes.
4. Type: elaborate image category a) After Lugol's iodine b) After normal saline c) With green filter d) After acetic acid.

IARCImageBankColpo
├── Case 001
│ ├── AAAW1.jpg
│ ├── AAAW2.jpg
│ ├── AAAW3.jpg
│ └── AAAW4.jpg
├── Case 002
│ ├── AADE0.jpg
│ ├── AADE1.jpg

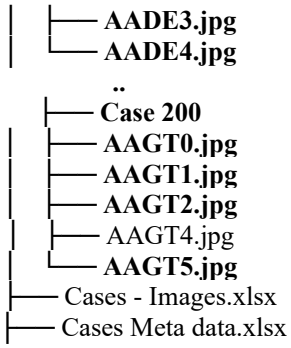

Figure 2. IARC image bank colpo dataset's folder structure.

Image Separation Logic

1. IARC Image Bank VIA

   Based on the observations noted in the VIA column of the cases metadata file, if diagnosis in the VIA column is negative then that image is considered as normal image else abnormal.

2. IARC Image bank colposcopy dataset

   Based on the observations noted in the Provisional Diagnosis column of the cases metadata file, certain diagnoses have been identified as normal, and as a result, the corresponding images have been categorized as normal.

   The following observations are considered as normal diagnosis and the corresponding image is considered as normal image or else abnormal.

   1) Type 1 squamocolumnar junction, normal
   2) Type 2 squamocolumnar junction, normal
   3) Type 3 squamocolumnar junction, normal
   4) Type 2 squamocolumnar junction, normal cervix with atrophic change
   5) Type 3 squamocolumnar junction, normal cervix with atrophic change
   6) Type 1 squamocolumnar junction, normal with pregnancy-induced changes
   7) Type 1 squamocolumnar junction, normal with ectropion
   8) Type1 squamocolumnar junction, normal with ectropion
   9) Type 1 squamocolumnar junction, normal with evidence of candida infection
   10) Type 1 squamocolumnar junction, normal with evidence of trichomoniasis
   11) Tuberculosis ulcer of cervix healed
   12) Type 1 squamocolumnar junction, normal with endocervical polyp
   13) Type 1 squamocolumnar junction, SPI
   14) Type 1 squamocolumnar junction, SPI
   15) Type 1 squamocolumnar junction, congenital transformation zone

   Note- All separation is done automatically by using a Python script in Google colab environment using the GPU. Figure 3 shows the overview of the D1 dataset, and Table 2 shows the cell distribution of the IARC Dataset.

D1 Dataset statistics

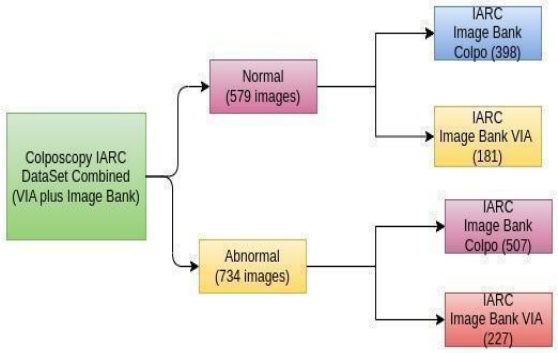

Figure 3. D1 Dataset Overview

Table 2. Categories and Counts of the IARC Image Bank VIA and IARC Image Bank Colpo Datasets, including the number of normal and abnormal images in each dataset [32][33].

Sr. No.	Dataset Name	Category	Count	Total - Dataset D1

.

				Normal	Abnormal
1	IARC Image Bank VIA (Total Images 408)	Normal	181		
				579	734
		Abnormal	227		
2	IARC Image Bank Colpo (Total Images 905)	Normal	398		
		Abnormal	507		

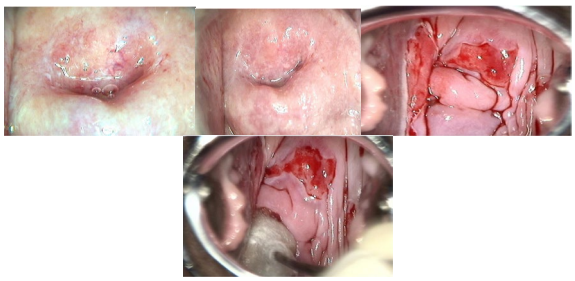

Figure 4. Sample images from the D1 Dataset.

The above Figure 4 displays a subset of a dataset, wherein the first two images represent negative cases, and the last two images represent positive cases of colposcopy screening pictures.

3.2. Liquid based-cytology Pap smear dataset (Dataset-II)

The Gauhati Medical College and Hospital's LBC (Liquid-Based Cytology) image collection is included in the dataset [34]. A cone-shaped brush is used to extract the target sample from the transformation zone, and the sample is then stored in a container with additive fluid to remove detritus. The samples are then layered, sedimented, and centrifuged at 2500 rpm for 5 minutes, then stained with hematoxylin and eosin before being prepared on slides. Each slide was photographed at the smear level using a Leica microscope so that cellular features could be identified. The 10 best images from each slide, together with the patient's medical history, were uploaded to an Excel file. Based on the patient's account and subsequent examination by a pathologist, the pictures were categorized as NILM, LSIL, HSIL, and SSC. The pictures are 2048 pixels wide by 1536 pixels high. Table 3 shows the Image distribution of the dataset. Figures 5 and 6 show the sample images from the LBC dataset in normal and abnormal classes. In the following manuscript, this dataset is referred to as D2.

Table 3. Categories and Counts of the Liquid based-cytology Pap smear dataset (D2) Dataset, including the number of normal and abnormal images in dataset [34].

Image Type	Category	Count	Total - Dataset D2	
			Normal	Abnormal
NILM	Normal	613		

LSIL	Abnormal	113		
HSIL	Abnormal	163	613	350
SSC	Abnormal	74		

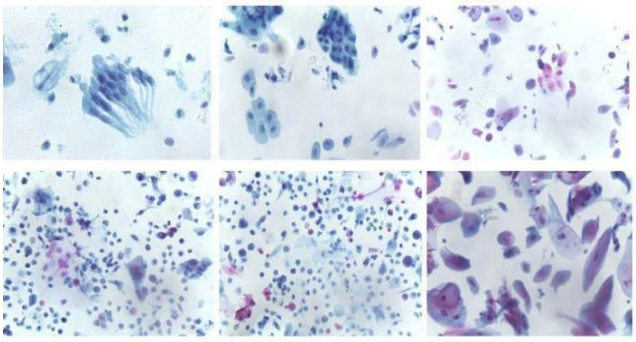

Figure 5. Sample Images of the HSIL and SCC categories from Liquid based-cytology Pap smear dataset (D2) [34].

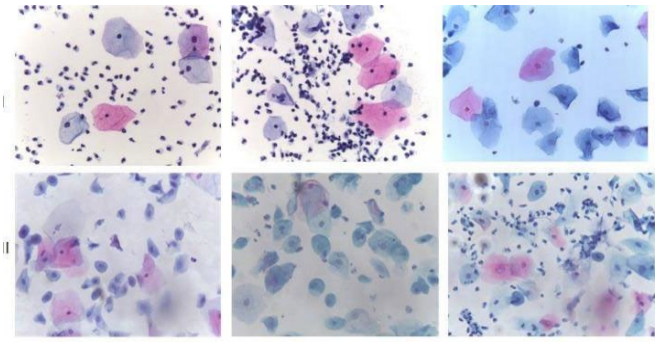

Figure 6. Sample Images of the NILM and LSIL categories from Liquid based-cytology Pap smear dataset (D2) [34].

3.3. Malhari dataset (Dataset-III)

We have collected the LBC and Colposcopy images of the same patient from Asam Hospital in India and named it as Malhari dataset. Patients permitted the hospital to release their data for research and development purposes under a strict confidentiality agreement. Important data from the dataset will be utilized in the study, and all necessary precautions will be taken to protect the privacy of the patient's information.

The dataset includes information from 32 patients, with each patient having four images captured from colposcopy, as well as 10 image patches obtained from a single Pap test image (In most of the cases). To our knowledge, no prior study has utilized both colposcopy and Pap smear images from the same set of patients, making this dataset a unique resource for research and analysis in this field. Table 4 shows the overview of the Malhari dataset. In the following manuscript, this dataset is referred to as D3.

Table 4. Categories and Counts of the (D3) Dataset, including the number of normal and abnormal images in dataset.

Type	Patients	Abnormal	Normal	Total
Colposcopy	32	81	53	134
Pap Smear		160	158	318
Total		241	211	452

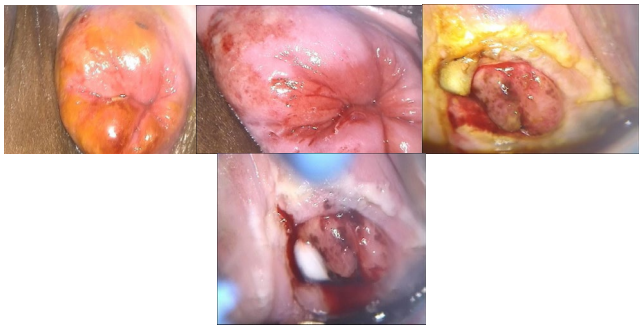

Figure 7. sample screening images from colposcopy portion of dataset D3.

The above Figure 7 displays a subset of a dataset, where the first two images represent negative cases, and the last two images represent positive cases of colposcopy screening pictures.

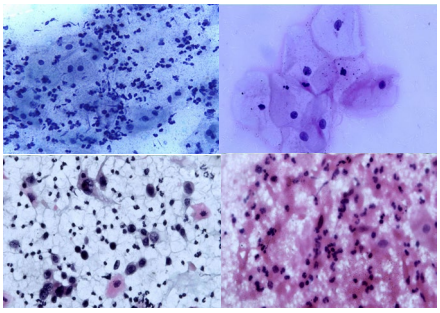

Figure 8. sample screening images from cytology portion of dataset D3.

The above Figure 8 displays a subset of a dataset, wherein the first two images represent negative cases, and the last two images represent positive cases of a pap smear.

4. Proposed Method

This section explains how the final ensemble model was implemented to predict whether a particular patient has

cervical cancer or not, using both colposcopy and Pap screening images of the same patient. This paper refers to the DeepColpo model as M1, which has been developed to classify colposcopy images into normal or abnormal categories with high accuracy. Similarly, the DeepCyto+ model is referred to as M2, and has been created to accurately classify cytology images into normal or abnormal categories. The section is divided into various sub-sections, where Subsection 4.1 explains the architecture of model M1, Subsection 4.2 describes the architecture of model M2, Subsection 4.3 describes the training process of models, and Subsection 4.4 will explain the ensemble approach using machine learning. Starting with the architecture of individual models and working our way up to the final ensemble model, this section strives to explain everything in as much detail as possible.

4.1. Architecture of M1

The M1 architecture has been designed to accurately classify abnormal and normal images obtained through colposcopy screening tests. To enhance its ability, the architecture has undergone training using diverse datasets, in order to acquire a robust knowledge base and improve its sensitivity in detecting abnormalities. This approach ensures that the M1 architecture can effectively differentiate between normal and abnormal colposcopy images, thereby aiding in the early detection and diagnosis of potential health issues. Figure 9 shows the architecture of model M1.

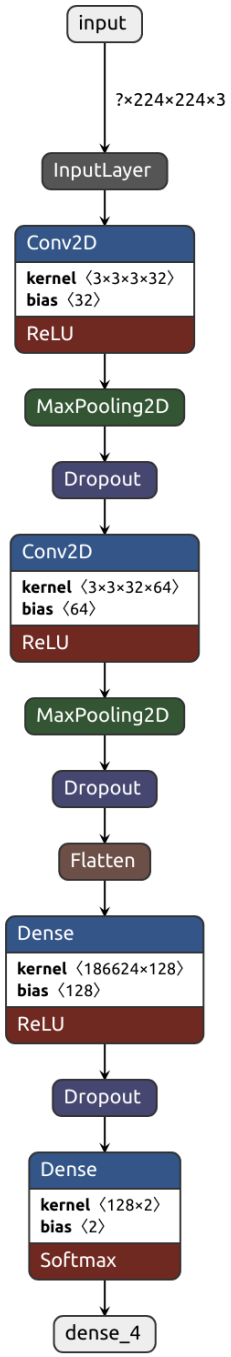

Figure 9. Architecture of Model M1.

The model (M1) comprises an input layer with a shape of (None, 224, 224, and 3) to accept input images with a size of 224x224 and 3 color channels (RGB). A 2D convolutional layer (Conv2D) is then applied, using 32 filters and a kernel size of 3x3, followed by batch normalization and an activation function to introduce non-linearity. The output's spatial dimensions are then cut in half through a down-sampling layer with a pool size of 2x2. Over fitting may be avoided by including a dropout layer with a rate of 0.25,

which eliminates 25% of the input units at random during training. Subsequently, a flatten layer, fully connected layers, and a classification layer with a softmax function are appended to the model (M1). The model consists of 12 layers. The total number of parameters for this model is 100,936,450, out of which 100,935,874 are trainable parameters, and the remaining 576 are non-trainable parameters. The evaluation of model M1's performance was carried out using different activation functions, as explained in the later part of the paper.

The performance of the M1 architecture has been found to be superior to that of other well-known architectures, despite its comparatively simple design and lower number of layers. This is a notable achievement, given the complexity of the dataset involved in classifying abnormal and normal colposcopy images. By achieving better results with a simpler architecture, the M1 model is a testament to the efficacy of its design and training approach.

4.2. Architecture of M2

The M2 model has been specifically developed and trained to classify abnormal and normal Pap smear screening images with a focus on achieving robust classification performance. To achieve this, the model has been trained using various datasets, including D2 and D3. Handling biological or physiological adaptations is crucial to ensuring accurate classification of images. For this reason, data from Asam Hospital has been collected under a confidentiality agreement with patients. As this dataset is derived from the Indian region, it represents a more realistic real-world scenario, accounting for the impact of regional environments and eating habits on internal body features, which are commonly referred to as "biological or physiological adaptations." Figure 10 shows the architecture of model M2.

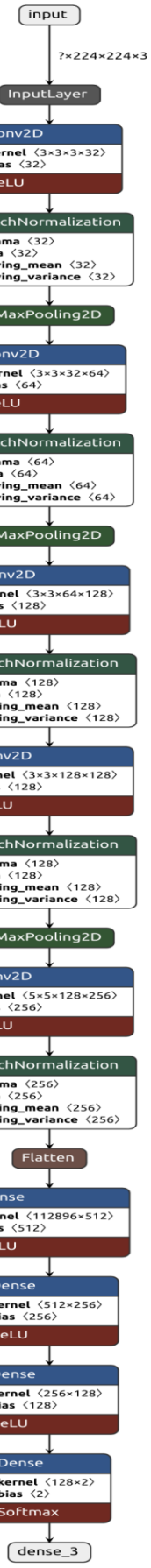

Figure 10. Architecture of Model M2.

.

Specifically, M2 is a deep neural network model trained to distinguish between abnormal and normal Pap smear images. It has many layers: five convolutional layers with ReLU as an activation function, batch normalization layers, and a maximum pooling layer. The initial step in feature extraction involves running the input image through a sequence of convolutional layers. Training stability is increased, and convergence is sped up with the aid of batch normalization layers, which standardize the activations of preceding layers. To do this, we employ max pooling layers to compress the feature maps' spatial dimensions while maintaining their essential features.

For the final classification, the convolutional layers' output is flattened and sent through numerous fully linked layers (dense layers). The last dense layer has a 2 softmax activation function neuron that produces a probability score that corresponds to the input image's class label.

Because of its enormous number of trainable parameters (59 million), M2 is able to learn complicated patterns and achieve high accuracy on tough image classification tasks, such as differentiating between abnormal and normal pap smear images.

The M2 model has achieved an impressive 100% accuracy in testing and very good accuracy in training, despite its simple architecture. This level of accuracy indicates that the model is capable of generalizing well and handling real-world scenarios. Consequently, it is highly suitable for use in real-world applications. The detailed training process of the M2 model will be explained in the next section.

4.3. Process of Model Training

This section explains how the models M1 and M2 were trained to be used in an ensemble approach, whereby the same patient's colposcopy and Pap images are passed through both models, and a prediction is generated. Details about this process will be explained as follows.

Throughout the training and testing phase, the data was divided into three sets. The training set was used to train the models, the validation set was used to fine-tune the models and prevent over fitting, and the testing set was used to evaluate the performance of the trained models.

To avoid bias in the selection of training, validation, and testing sets, the data was split in such a way that each set contained a representative sample of both abnormal and normal images.

Step 1: Training of Model M1 on Dataset D1

In the initial part of our research, authors trained Model M1 on Dataset D1 using several activation functions. The data has been divided into three separate sets, to facilitate the assessment and contrast of various models' efficacy during both the training and testing phases. With specified proportions, as shown in the table below. The findings of these tests are reported in the results section for future reference and comparison. The results section also discusses the training and performance of well-known CNN architectures tested on the same dataset (D1) with the same proportion using both transfer learning and new-from-scratch methodologies.

Table 5 shows the distribution of images across various sets during the training and testing of the model M1 on dataset D1.

Table 5. The distribution of images in Dataset D1 used in developing and evaluating Model M1.

	Training		Validation		Testing		Total
	Abnormal	Normal	Abnormal	Normal	Abnormal	Normal	
	512	405	147	116	75	58	1313
Total Images	917		263		133		
Percentage	70%		20%		10%		100%

Step 2: Training of Model M2 on Dataset D2

Dataset D2 was used for the training and evaluation phases of the M2 architecture. There were a total of 963 images in the dataset, including 35 abnormal and 61 normal examples for testing, 245 abnormal and 429 normal examples for training, and 70 abnormal and 123 normal examples for validating.

The dataset was split into three subsets with a 7:2:1 as shown in the following table. The results are outlined in the next section.

Table 6 shows the distribution of images across various sets during the training and testing of the model M2 on dataset D2.

Table 6. The distribution of images in Dataset D2 used in developing and evaluating Model M2.

	Training		Validation		Testing		Total
	Abnormal	Normal	Abnormal	Normal	Abnormal	Normal	
	245	429	70	123	35	61	963
Total Images	674		193		96		
Percentage	70%		20%		10%		100%

Step 3: Fine tuning of Model M1 and M2 Using Dataset D3
After training on the D1 and D2 datasets, Models M1 and M2 were fine-tuned on the D3 dataset, which included both colposcopy and pap smear images of the same patients. This dataset was deemed critical as it reflects real-world scenarios and can provide valuable insights into the performance of the models. Colposcopy images were utilized for Model M1's fine-tuning, whereas pap smear images were utilized for Model M2's. The dataset comprises images with varying grades of cervical intraepithelial neoplasia (CIN), where CIN1 is considered normal and CIN2 and CIN3 are categorized as abnormal. Images with NILM are regarded as normal for Pap smear tests, while any other classification is regarded as abnormal. This approach of fine-tuning on a diverse dataset with different types of images allowed the models to learn and generalize better on unseen data, potentially improving the models' diagnostic accuracy in identifying cervical lesions. Model M1 with the ELU activation function was chosen for further tuning because its performance on Dataset D1 was superior to that of other activation functions. Table 7 shows the distribution of images in dataset D3. Table 7 shows the distribution of images across various sets during the fine tuning and testing of the model M1 on dataset D3 colposcopy only.

Table 7. The distribution of images in Dataset D3 used in fine tuning and evaluating Model M1 on colposcopy image

	Training		Validation		Testing		Total
	Abnormal	Normal	Abnormal	Normal	Abnormal	Normal	
	47	41	5	6	5	6	110
Total Images	88		11		11		
Percentage	80%		10%		10%		100%

Table 8 shows the distribution of images across various sets during the fine tuning and testing of the model M2 on dataset D3 Pap smear only.

Table 8. The distribution of images in Dataset D3 used in fine-tuning and evaluating Model M2 on pap smear images.

	Training		Validation		Testing		Total
	Abnormal	Normal	Abnormal	Normal	Abnormal	Normal	
	104	104	13	13	13	13	260
Total Images	208		26		26		

Percentage	80%	10%	10%	100%

4.4. Ensemble Approach Using Machine Learning

To ensure a more realistic and clinically relevant evaluation of the ensemble approach, the testing process involves using screening images from the same patient for both colposcopy and pap smear screenings. This means that the same colposcopy and pap smear images from a particular patient are used in the evaluation, which helps to simulate a real-world scenario and better assess the performance of the ensemble model. The aim of the ensemble approach is to leverage the strengths of two models, M1 and M2, which are trained and tested on different datasets, to acquire robust knowledge for the classification of abnormal and normal colposcopy and pap smear screening images, respectively. By combining the outputs of both models, the ensemble approach seeks.

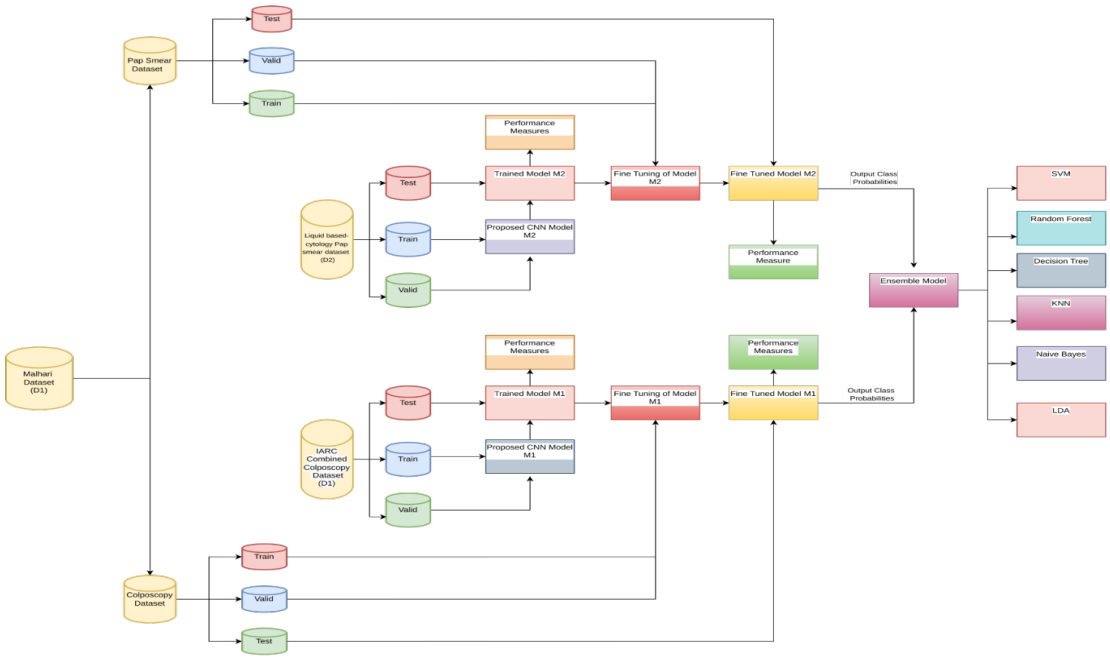

Figure 11. Architecture of Proposed Model

The accuracy and reliability of the classification results.

To combine the outputs of models M1 and M2, the probabilities of the output classes are taken and used as features that are then passed as input to various machine learning algorithms, such as SVM, random forest, decision tree, KNN, LDA, and naive Bayes. A certain portion of the data is used for training, and the rest is used for testing. The performance of the ensemble approach is evaluated in the result section, which provides insights into the accuracy and robustness of the model.

Figure 11 shows the architecture of the proposed method.

The dataset provided by D3 consists of records for 32 patients. However, for the purposes of this particular study, only data pertaining to 23 patients was utilized. Specifically, only those patients for whom both cytology and colposcopy images belonged to the same category were included in the analysis. For the purpose of this testing, pairs of identical patients' colposcopy and pap screening pictures are chosen from dataset D3, which was included in the testing set despite the fact that these photos have never been seen before. In order to create a greater number of pairings, various permutations and combinations are constructed. A total of 136 pairs were generated, of which 41 were used for testing and 95 for training. Table 9 shows the distribution of pairs used for ML algorithms.

Table 9. Distribution of pairs of colposcopy and Pap images.

.

			Training		Testing	
Normal Pair	Abnormal Pair	Total Pair	Normal Pair	Abnormal Pair	Normal Pair	Abnormal Pair
9	127	136	6	89	3	38
	Total		95		41	

5. Result and Analysis:

5.1. Result of Training Process

1. The outcome of the training process of Model M1 on Dataset D1 (Training Step 1)

Table 10 presents the performance metrics of model M1 trained and tested on dataset D1. It includes various metrics.

Table 10. Performance metrics of model M1 on dataset D1.

			Training		Testing			
Model	Activation Function used After Convo Layers	Output Layer Activation Function	Training Accuracy	Validation Accuracy	Accuracy	Precision	Recall	F1 Score
M1	swish	sigmoid	64.00	69.00	68.41	69.00	68.42	68.53
	elu	sigmoid	62.00	66.00	72.18	73.07	72.18	72.29
	tanh	sigmoid	60.00	58.00	59.40	62.67	59.40	59.01
	leaky relu	sigmoid	62.49	69.20	65.41	65.29	65.41	65.34
	relu	sigmoid	62.16	71.10	68.42	68.21	68.41	67.92

The M1 model using elu activation function and sigmoid activation function in the output layer achieved the highest accuracy, precision, recall, and F1 score among the activation function evaluated, with scores of 72.18%, 73.07%, 72.18% and 72.29% respectively. But which is still low to identify root cause of low accuracy the same dataset (D1) is used to train and test well performing architectures with and without using transfer learning.

Table 11 presents the performance metrics of well-known models trained and tested on dataset D1. It includes various metrics. These architectures had a single neuron in the output layer with a sigmoid function and were trained from scratch.

Table 11. Performance Metrics of Well-Known Models Trained and Tested on Dataset D1 with Single Neuron and Sigmoid Activation Function in the Output Layer from Scratch.

				Training		Testing			
Model	Weights	Epoch's	Optimizer	Training Accuracy	Validation Accuracy	Accuracy	Precision	Recall	F1 Score

Inception V3	random	100	Sgd with momentum 0.9	50.00	50.00	43.61	19.02	43.61	26.49
Resnet50	random	100	Sgd with momentum 0.9	50.00	50.00	50.38	51.54	50.38	50.54
vgg16	random	100	Sgd with momentum 0.9	50.00	50.00	46.62	52.57	46.62	40.40

The accuracy of the Resnet50 model is slightly better than the other two models, which indicates that it is better at correctly identifying the positive and negative classes from the dataset. But performance is much low as compared to model M1 with elu activation function.

Table 12 presents the performance metrics of well-known models trained on dataset D1 using transfer learning and tested on portions of dataset D3. These architectures had a single neuron in the output layer with a sigmoid function.

Table 12. Performance Metrics of Well-Known Models Trained with Transfer Learning and Tested on Portion of Dataset D1.

					Training		Testing			
Model	**Weights**	**Method**	**Epoch's**	**Optimizer**	**Training Accuracy**	**Validation Accuracy**	**Accuracy**	**Precision**	**Recall**	**F1 Score**
Inception V3	image net	Transfer Learning	100	Sgd with momentum 0.9	50.00	50.00	46.62	46.06	46.62	46.28
Resnet50	image net		100	Sgd with momentum 0.9	50.00	50.00	47.37	48.68	47.37	47.49
vgg16	image net		100	Sgd with momentum 0.9	50.00	50.00	57.89	57.89	57.89	57.89

The results shown in the table seem to indicate that the models are not performing well. The training accuracy and validation accuracy are both at 50%, which means that the models are essentially guessing randomly. The accuracy, precision, recall, and F1 score are all quite low, indicating that the models are not able to correctly classify the images. In transfer learning vgg16 performed well and obtained accuracy, precision, recall and f1 score of 57.89%,57.89%,57.89%,57.89%. But performance is much low as compared to model M1 with elu activation function.
But performance is much low as compared to model M1 with elu activation function.

From the analysis of the results, it appears that combining two different datasets from the IARC for colposcopy screening has led to an issue of interclass dissimilarity. This could be due to the usage of different imaging methods for capturing images of the cervix, resulting in the combination of various types of cervical images. As a consequence, even after training different neural network architectures on this merged dataset, the training and validation accuracy remained at 50%, while the loss decreased. This indicates that the models were unable to differentiate between the different classes, indicating the presence of an interclass dissimilarity problem. Furthermore, it was found that the performance of the models differed based on their complexity, and the model with fewer layers (m1) performed better than others on the same dataset and with the same settings.

2. The outcome of the training process of Model M2 on Dataset D2 (Training Step 2)

Table 13. Fine-tuning results of Model M2 on Dataset D3.

.

				Training		Testing			
Model	**Weights**	**Epoch's**	**Optimizer**	**Training Accuracy**	**Validation Accuracy**	**Accuracy**	**Precision**	**Recall**	**F1 Score**
LBC (M2)	random	15	Sgd with momentum 0.9	95.39	98.39	**100%**	**100%**	**100%**	**100%**

After training on the Pap smear image dataset, the model's performance was outstanding, high marks on all relevant performance metrics bear this out. During the 15 epochs of training, with an optimizer momentum of 0.9, the authors employed Stochastic Gradient Descent (SGD). The training accuracy achieved was 95.39%, while the validation accuracy was 98.39%. These results demonstrate that the model is capable of effectively identifying and learning the unique features of the Pap smear image dataset. Table 13 shows the results of the fine-tuned model M2 on dataset D3.

The model was also tested on a separate unseen dataset of Pap smear images and achieved 100% accuracy: precision, recall, and F1 score. This indicates that the model is capable of generalizing well on unseen data and could be a promising approach for assisting medical professionals in the accurate identification of cervical lesions.

3. The results obtained after fine-tuning Model M1 and M2 on Dataset D3 (Training Step 3)

The following Table 14 represents the fine-tuning results of Model M1 on Dataset D3 using colposcopy images, along with various performance metrics.

Table 14. Fine-tuning results of Model 1 on Dataset D3 using colposcopy images with various performance metrics.

				Training		Testing			
Model	**Weights**	**Epoch's**	**Optimizer**	**Training Accuracy**	**Validation Accuracy**	**Accuracy**	**Precision**	**Recall**	**F1 Score**
M1 (fine-tuned)	pre-trained on D1 dataset	100	Sgd with momentum 0.9	75.00	72.74	63.64	66.88	63.64	63.03

The model M1 with the ELU activation function performed well on the D1 dataset, and therefore it was selected for fine-tuning on the ASAM Hospital images. Through fine-tuning, the model gained knowledge and was able to classify images more accurately as it was trained on a new dataset. However, compared to the D1 dataset, the ASAM Hospital dataset had fewer images, which limited the model's learning capacity. As a result, the accuracy was not very impressive, as the model could only update its weights based on the limited information it had available. The model was able to attain notable metrics of accuracy, precision, recall, and F1 score, specifically scoring 63.64, 66.88, 63.64, and 63.03, respectively.

The following Table 15 represents the fine-tuning results of Model M2 on Dataset D3 using Pap smear images along with various performance metrics.

Table 15. Fine-tuning results for Model M2 on Dataset D3.

				Training		Testing			
Model	**Weights**	**Epoch's**	**Optimizer**	**Training Accuracy**	**Validation Accuracy**	**Accuracy**	**Precision**	**Recall**	**F1 Score**

M2 (Fine Tuned)	Pretrained on D2 dataset of Pap Smear	15	Sgd with momentum 0.9	97.60	100.00	100.00	100.00	100.00	100.00

Based on the results obtained from fine-tuning model M2 on the D3 dataset comprising pap smear images, it is evident that the model performed well during the fine-tuning process. This is because both the D2 dataset and the Pap smear portion of D3 dataset contain similar images that generated the same kind of features. Moreover, the fact that model M2 had already achieved 100% testing accuracy on the testing set suggests that it was well-generalized. Using the same type of data for fine-tuning the model resulted in the same features that were learned by the model during its training on the D2 dataset. Therefore, the model was able to accept the new dataset with just 15 epochs and achieved 100% testing accuracy,precision,recall and F1 score. On the other hand, the D1 dataset comprising colposcopy images and the D3 dataset containing colposcopy images exhibited some feature differences due to the variations in instruments, biological factors, and adaptations.

5.2. Result of Ensemble Approach

Probabilities of the output classes are extracted and used as features in machine learning methods that incorporate LDA, SVM, random forests, decision trees, and naive Bayes to combine the findings of models M1 and M2. The tabular data below displays the outcomes of the aforementioned experiments. The result of the experiment is given in the following Table 16.

Table 16. Performance of the ensemble approach

.

				Testing			
Feature Extractor Used for Colposcopy Image	**Feature Extractor Used for Pap Image**	**Machine Learning Algorithm**	**Training Accuracy**	**Accuracy**	**Precision**	**Recall**	**F1 Score**
Model M1	Model M2	Random Forest	100.00	97.56	87.50	98.68	92.19
		Decision Tree	100.00	97.56	87.50	98.67	92.21
		KNN	100.00	**100.00**	**100.00**	**100.00**	**100.00**
		LDA	93.68	92.68	46.34	50.00	48.10
		Naive Bayes	86.31	80.48	59.54	74.12	60.95
		SVM	93.68	92.68	46.34	50.00	48.10

On the other hand, LDA and Naive Bayes performed poorly compared to KNN, which obtained perfect accuracy on both the training and testing sets. KNN was the only algorithm to get a perfect score in both accuracy,precision, F1 score and recall. Overall, the ensemble method performed quite well when comparing normal and abnormal Pap smear and colposcopy pictures.

Conclusion

The DeepColpo (M1) model was trained from scratch using a combination of the IARC colposcopy VIA dataset and IARC Colposcopy Image Bank, which were merged into a single dataset D1. The elu activation function was utilized during the training process, and the resulting test results showed an accuracy, precision, recall, and f1 score of 72.18%, 73.07%, 72.18%, and 72.29%, respectively. It was found that the interclass dissimilarity problem was the root cause of the poor performance of the model. This was attributed to including images taken using different

approaches and methodologies. The DeepCyto+ model was trained from scratch on a Liquid-based Cytology Pap smear dataset (D2). The resulting test results showed 100% accuracy, precision, recall, and f1 score, indicating that the model could generalize well. This suggests that the DeepCyto+ model was able to learn and accurately predict new data beyond the training data, demonstrating its potential utility in real-world scenarios.

But to make them robust to handle real-world problems and biological or physical adaptation, real data is collected for fine-tuning the weights of both models. After fine-tuning DeepColpo (M1) achieved the following test result for accuracy, precision, recall, and f1 score of 63.64%, 66.68%, 63.64%, and 63.04%, respectively. Whereas DeepCyto+ retains its perfect score in all matrices because the Liquid-based Cytology Pap smear dataset D2 and D3 (LBC/ pap smear portion) are collected from the same geographical location, most probably using the same approach and methodology. Whereas dataset D1 and the colposcopy portion of D3 were collected from different geographical regions, different methodologies were used to collect samples.

To overcome the limitations of DeepColpo's versatility and DeepCyto+'s lack of robustness, an ensemble approach was adopted that leveraged various machine learning algorithms such as SVM, Decision Tree, Random Forest, KNN, LDA, and Naive Bayes to combine the strengths of both models. After rigorous evaluation, KNN was identified as the optimal algorithm that provided the best results.

The ensemble model was designed by utilizing DeepColpo's multi-dataset and multi-method training, which provided a diverse knowledge base, and DeepCyto+'s high accuracy on a specific image acquisition method, which enabled it to contribute highly reliable predictions. As a result, the ensemble model exhibited improved robustness and accuracy, making it more suitable for handling diverse image datasets.

To assess the ensemble model's real-world applicability, it was tested on colposcopy and cytology images from the same patient. Impressively, the model achieved perfect scores across all evaluation metrics on KNN, demonstrating its exceptional performance in accurately analyzing and diagnosing these images. This highlights the ensemble model's potential as a valuable tool for the clinical diagnosis and management of patients with cervical cancer.

Acknowledgements:

This research was funded by Department of Science and Technology Ministry of Science and Technology, India, grant number TDP/BDTD/29/2021.

References

[1] Anupama Bhan, Divyam Sharma, Sourav Mishra. Computer Based Automatic Segmentation of Pap smear Cells for Cervical Cancer Detection. 2018 5th International Conference on Signal Processing and Integrated Networks (SPIN), Noida, 22-23 February 2018, DOI: 10.1109/SPIN.2018.8474108.

[2] Zhi Lu et. al, Evaluation of Three Algorithms for the Segmentation of Overlapping Cervical Cells. IEEE Journal of Biomedical and Health Informatics, Volume: 21, Issue: 2, March 2017, DOI: 10.1109/JBHI.2016.2519686.

[3] Zaid Alyafeai et. al, A fully automated deep learning pipeline for cervical cancer classification. Expert Systems with Applications, Volume 141, 1 March 2020, 112951, https://doi.org/10.1016/j.eswa.2019.112951.

[4] R. Elakkiya, V. Subramaniyaswamy et al, Cervical Cancer Diagnostics Healthcare System Using Hybrid Object Detection Adversarial Networks. IEEE Journal of Biomedical and Health Informatics, Volume: 26, Issue: 4, April 2022, DOI: 10.1109/JBHI.2021.3094311.

[5] Nina Youneszade et. al, Deep Learning in Cervical Cancer Diagnosis: Architecture, Opportunities, and Open Research Challenges. IEEE Access, Volume: 11, DOI: 10.1109/ACCESS.2023.3235833.

[6] Jennyfer Susan M.B et. al, Design and Development of Webportal for Cervical Cancer Diagnosis using MobileODT Images. 2019 IEEE International Smart Cities Conference (ISC2), doi:10.1109/ISC246665.2019.9071683.

[7] Kaggle. Intel & mobileodt Cervical Cancer Screening, https://www.kaggle.com/c/intel-mobileodt-cervical-cancerscreening, 2017.

[8] Balkin, M.S. Cervical Cancer Prevention and Treatment: Science, Public Health and Policy Overview. In Proceedings of the Challenges and Opportunities for Women's Right to Health, Brussels, Belgium, 27–28 September 2007.

[9] Ferlay J, Bray F, Pisani P et al. Globocan 2002 cancer incidence, mortality, and prevalence worldwide. Version 2.0. 2004: Lyon, France, IARC Press. IARC CancerBase No. 5.

[10] D. M. Eddy, Secondary prevention of cancer: an overview. B World Health Organ, vol. 64, no. 3, p. 421, 1986.

[11] Madhura Kalbhor, Dr. Swati Shinde, et al, DeepCyto: A Hybrid Framework for Cervical Cancer Classification by using Deep Feature Fusion of Cytology Images Mathematical Biosciences and Engineering 2022, Volume 19, Issue 7: 6415-6434. doi: 10.3934/mbe.2022301.

[12] Gay, J D et al. False-negative results in cervical cytologic studies, Acta cytologica vol. 29, 6 (1985): 1043-6, PMID: 3866457.

[13] Bosch, M et al. "Characteristics of false-negative smears tested in the normal screening situation", Acta cytologica vol. 36, 5 (1992): 711-6, PMID: 1523929.

[14] Naryshkin, S. The false-negative fraction for Papanicolaou smears: how often are "abnormal" smears not detected by a "standard" screening cytologist? Archives of pathology & laboratory medicine vol. 121, 3 (1997): 270-2, PMID: 9111116.

[15] Koonmee S, Bychkov A, Shuangshoti S, Bhummichitra K, Himakhun W, Karalak A, Rangdaeng S, False-Negative Rate of Papanicolaou Testing: A National Survey from the Thai Society of Cytology, Acta Cytologica 2017; 61:434-440, doi: 10.1159/000478770.

[16] Anjali Deswal, Sanjeev Dhawan et. al, A Technique for Determining the Early Detection For Cervical Cancer, 2019 5th International Conference on Signal Processing, Computing and Control (ISPCC), DOI: 10.1109/ISPCC48220.2019.8988374.

[17] Romuere Silva et. al, Searching for cell signatures in multidimensional feature spaces, International Journal of Biomedical Engineering and Technology, 2021 Vol.36 No.3, pp.236 - 256, doi:10.1504/IJBET.2021.10040044.

[18] Wenya linda bi et. al, Artificial Intelligence in Cancer Imaging: Clinical Challenges and Applications, A Cancer Journal for Clinicians, 05 February 2019 https://doi.org/10.3322/caac.21552.

[19] Krishna Prasad Battula et.al, Deep Learning based Cervical Cancer Classification and Segmentation from Pap Smears Images using an EfficientNet, International Journal of Advanced Computer Science and Applications (IJACSA), Volume 13 Issue 9, 2022, Doi:10.14569/IJACSA.2022.01309104.

[20] O. Attallah, Cervical Cancer Diagnosis Based on Multi-Domain Features Using Deep Learning Enhanced by Handcrafted Descriptors. Applied Sciences, vol. 13, no. 3, p. 1916, Feb. 2023, doi: 10.3390/app13031916.

[21] Ming Fang; Xiujuan Lei et. al, A Deep Neural Network for Cervical Cell Classification Based on Cytology Images, IEEE Access, Volume: 10, DOI: 10.1109/ACCESS.2022.3230280.

[22] Khaled Mabrouk Amer Adweb et. al, Cervical Cancer Diagnosis Using Very Deep Networks Over Different Activation Functions, IEEE Access, Volume: 9, DOI: 10.1109/ACCESS.2021.3067195.

[23] Venkatesan Chandran et. al, Diagnosis of Cervical Cancer based on Ensemble Deep Learning Network using Colposcopy Images, Hindawi BioMed Research International Volume 2021, Article ID 5584004, 15 pages, https://doi.org/10.1155/2021/5584004.

[24] Liu L, Wang Y, Liu X, Han S, Jia L, et. al, Computer-aided diagnostic system based on deep learning for classifying colposcopy images, Ann Transl Med. 2021 Jul; 9(13): 1045, doi: 10.21037/atm-21-885.

[25] Y. Karasu Benyes, E. C. Welch, A. Singhal, J. Ou, and A. Tripathi, A Comparative Analysis of Deep Learning Models for Automated Cross-Preparation Diagnosis of Multi-Cell Liquid Pap Smear Images, Diagnostics, vol. 12, no. 8, p. 1838, Jul. 2022, doi: 10.3390/diagnostics12081838.

[26] Kyi Pyar Win at. al, Computer-Assisted Screening for Cervical Cancer Using Digital Image Processing of Pap Smear Images, Appl. Sci. 2020, 10, 1800, doi:10.3390/app10051800.

[27] Mohammed Alsalatie et. al, "Analysis of Cytology Pap Smear Images Based on Ensemble Deep Learning Approach", Diagnostics 2022, 12, 2756. https://doi.org/10.3390/diagnostics12112756.

[28] Komala Rayavarapu et. al, Prediction of Cervical Cancer using Voting and DNN Classifiers, 2018 International Conference on Current Trends towards Converging Technologies (ICCTCT), DOI: 10.1109/ICCTCT.2018.8551176.

[29] N. Sompawong et. al, Automated Pap Smear Cervical Cancer Screening Using Deep Learning, 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), doi:10.1109/EMBC.2019.8856369.

[30] D. N. Diniz et al., A Deep Learning Ensemble Method to Assist Cytopathologists in Pap Test Image Classification, Journal of Imaging, vol. 7, no. 7, p. 111, Jul. 2021, doi: 10.3390/jimaging7070111.

[31] Pin Wang et. al, Adaptive Pruning of Transfer Learned Deep Convolutional Neural Network for Classification of Cervical Pap Smear Images, IEEE Access, Volume: 8, doi: 10.1109/ACCESS.2020.2979926.

[32] International Agency for Research on Cancer, IARC Visual Inspection with Acetic Acid (VIA) Image Bank, [Online]. Available: https://screening.iarc.fr/cervicalimagebank.php. [Accessed: Apr. 19, 2023].

[33] International Agency for Research on Cancer, IARC Colposcopy Image Bank, [Online]. Available: https://screening.iarc.fr/cervicalimagebank.php. [Accessed: Apr. 19, 2023].

[34] E. Hussain, L. B. Mahanta, H. Borah, and C. R. Das, Liquid based-cytology Pap smear dataset for automated multi-class diagnosis of pre-cancerous and cervical cancer lesions. Data in Brief, 30, (2020) 105589. doi: 10.1016/j.dib.2020.105589.