Cervical Cytology Classification

A PROJECT REPORT

Submitted by

Hardik Gupta (22BCE11278)
Khush Kedawat (22BCE11095)
Md.Hashir Hussain (22BCE10195)
Abhinav Dixit (22BCE11085)
Mayank Singh Bisht (22BCE11328)

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING

VIT BHOPAL UNIVERSITY

KOTHRIKALAN, SEHORE
MADHYA PRADESH - 466114

April 2024

VIT BHOPAL UNIVERSITY,KOTHRIKALAN, SEHORE

MADHYA PRADESH – 466114

BONAFIDE CERTIFICATE

Certified that this project report titled **"Cervical Cytology Classification"** is the bonafide work of **" Hardik Gupta (22BCE11278), Khush Kedawat (22BCE11095), Md.Hashir Hussain (22BCE10195), Abhinav Dixit (22BCE11085), Mayank Singh Bisht (22BCE11328) "** who carried out the project work under my supervision. Certified further that to the best of my knowledge the work reported here does not form part of any other project / research work on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

PROJECT SUPERVISOR
Dr. Sanat Jain,
Assistant Professor
School of Computer Science and Engineering
VIT BHOPAL UNIVERSITY

The Project Exhibition II Examination is held on _____

ACKNOWLEDGEMENT

First and foremost I would like to thank the Lord Almighty for His presence and immense blessings throughout the project work.

I wish to express my heartfelt gratitude to Dr. Sandip Mal, Head of the Department, School of Computer Science and Engineering for much of his valuable support encouragement in carrying out this work.

I would like to thank my internal guide Dr. Sanat Jain, for continually guiding and actively participating in my project, giving valuable suggestions to complete the project work.

I would like to thank all the technical and teaching staff of the School of Computer Science and Engineering, who extended directly or indirectly all support.

Last, but not the least, I am deeply indebted to my parents who have been the greatest support while I worked day and night for the project to make it a success.

ABSTRACT

Cervical cytology classification is a critical component of cervical cancer screening programs, aimed at identifying cellular abnormalities indicative of pre-cancerous or cancerous lesions. This report provides an in-depth analysis of cervical cytology classification methodologies, including traditional systems such as the Bethesda System and emerging technologies utilizing artificial intelligence and machine learning algorithms.

Through a comprehensive review of key terminology, diagnostic criteria, and quality assurance measures, this report elucidates the intricacies of cervical cytology interpretation and its clinical implications.

Furthermore, the report explores the motivation behind advancements in cervical cytology classification, highlighting the significance of improving diagnostic accuracy, enhancing screening efficiency, and addressing disparities in access to healthcare resources.

By synthesizing current research and best practices, this report serves as a valuable resource for healthcare professionals, researchers, and policymakers involved in cervical cancer screening and prevention efforts.

TABLE OF CONTENTS

CHAPTER: 1 INTRODUCTION

INTRODUCTION

Cervical cancer is a significant global health challenge characterized by the abnormal growth of cells in the cervix, the lower part of the uterus. It is primarily caused by persistent infection with high-risk strains of the human papillomavirus (HPV), a sexually transmitted infection. Cervical cancer typically progresses slowly, often without symptoms in its early stages, making regular screening through methods such as Pap smears or HPV testing critical for early detection and intervention.

When diagnosed early, cervical cancer is highly treatable, with various options including surgery, radiation therapy, and chemotherapy. However, in regions with limited access to healthcare and screening services, cervical cancer remains a leading cause of cancer-related mortality among women. Efforts to increase awareness, improve access to screening and vaccination programs, and advance research into new prevention and treatment modalities are essential in the ongoing fight against cervical cancer.

Cervical cytology, commonly known as Pap smear, is a crucial screening tool in the detection of cervical cancer and pre-cancerous lesions. The classification system for cervical cytology plays a pivotal role in accurately interpreting cellular abnormalities, guiding clinical management, and ultimately improving patient outcomes.

MOTIVATION FOR THE WORK

Our work's motivation stems from its profound impact on women's health worldwide. Accurate classification is instrumental in identifying cellular abnormalities indicative of cervical cancer or pre-cancerous lesions, enabling early intervention and improved patient outcomes. Moreover, advancements in classification methodologies hold the promise of enhancing the sensitivity and specificity of screening tests, thereby reducing false positives and unnecessary interventions while ensuring that no cases of cervical abnormalities are overlooked. By contributing to the refinement of classification systems, researchers and healthcare professionals can play a pivotal role in the ongoing efforts to optimize cervical cancer screening programs, particularly in underserved populations where access to healthcare resources may be limited.

PROBLEM STATEMENT

There are two main problems with cervical cancer in the current situation:

1. **Lack of access to screening and prevention:** Cervical cancer is highly preventable through vaccination against the HPV virus and regular screening with Pap smears and HPV tests. However, many women, particularly in low- and middle-income countries, don't have access to these preventive measures.
2. **Late diagnosis:** Early detection is crucial for successful treatment of cervical cancer. When diagnosed at an early stage, cervical cancer is one of the most treatable cancers. However, many women are diagnosed with cervical cancer at a late stage, when it is more difficult to treat. These problems contribute to a higher number of cervical cancer cases and deaths than could be prevented with better access to screening and prevention programs.

OBJECTIVE OF THE WORK

The project's objective is to create a sophisticated classification system specifically tailored for analyzing cervical cytology images. This system will utilize advanced image processing and machine learning techniques to accurately distinguish between normal and abnormal cytology samples. By leveraging large datasets of annotated cytology images, the classification algorithm will be trained to recognize subtle cellular abnormalities indicative of pre-cancerous or cancerous lesions. The primary aim of developing this robust classification system is to augment the capabilities of healthcare professionals in interpreting cervical cytology results.

SUMMARY

In the exploration of cervical cytology classification, it becomes evident that it plays a crucial role in identifying cervical abnormalities, aiding in the early detection of cancerous or pre-cancerous conditions. The motivation to work on improving this classification system stems from the profound impact it has on women's health worldwide. By developing a robust classification system for cervical cytology images, healthcare professionals can more accurately identify abnormal cases, thus enhancing the efficiency and effectiveness of cervical cancer screening programs. This advancement not only improves diagnostic accuracy but also addresses disparities in access to healthcare resources, particularly in underserved populations. Ultimately, the goal is to contribute to the early detection and prevention of cervical cancer, ensuring better health outcomes for women globally.

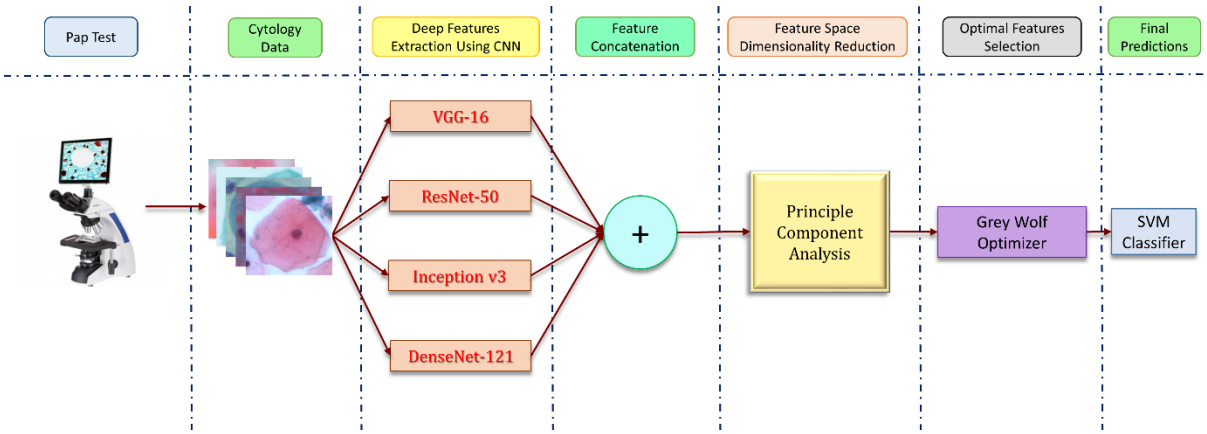
CHAPTER:2 LITERATURE SURVEY

LITERATURE REVIEW

INTRODUCTION

Machine learning models play a crucial role in the field of cervical cytology classification, offering significant benefits in terms of accuracy, efficiency, and scalability. By leveraging large datasets of annotated cytology images, machine learning algorithms can learn complex patterns and relationships that may not be easily discernible through traditional methods. This allows for the development of highly accurate classification systems capable of detecting subtle cellular abnormalities indicative of pre-cancerous or cancerous lesions.

FLOWCHART:



DIFFERENT TECHNOLOGIES USED:

Convolutional Neural Network(CNN):- Convolutional Neural Networks (CNNs) are deep learning algorithms primarily used for image classification and recognition tasks. They consist of convolutional layers that apply learnable filters to input images to detect features like edges and textures. Pooling layers reduce spatial dimensions while retaining important information. Activation functions introduce non-linearity, aiding in learning complex

relationships. Fully connected layers make final predictions based on learned features. CNNs are trained using backpropagation to adjust weights and minimize prediction errors. Transfer learning, utilizing pre-trained models, is common due to computational costs and dataset size requirements. CNNs have revolutionized computer vision, showing exceptional performance in various tasks like medical image analysis and facial recognition, and are widely used in research and industry applications.

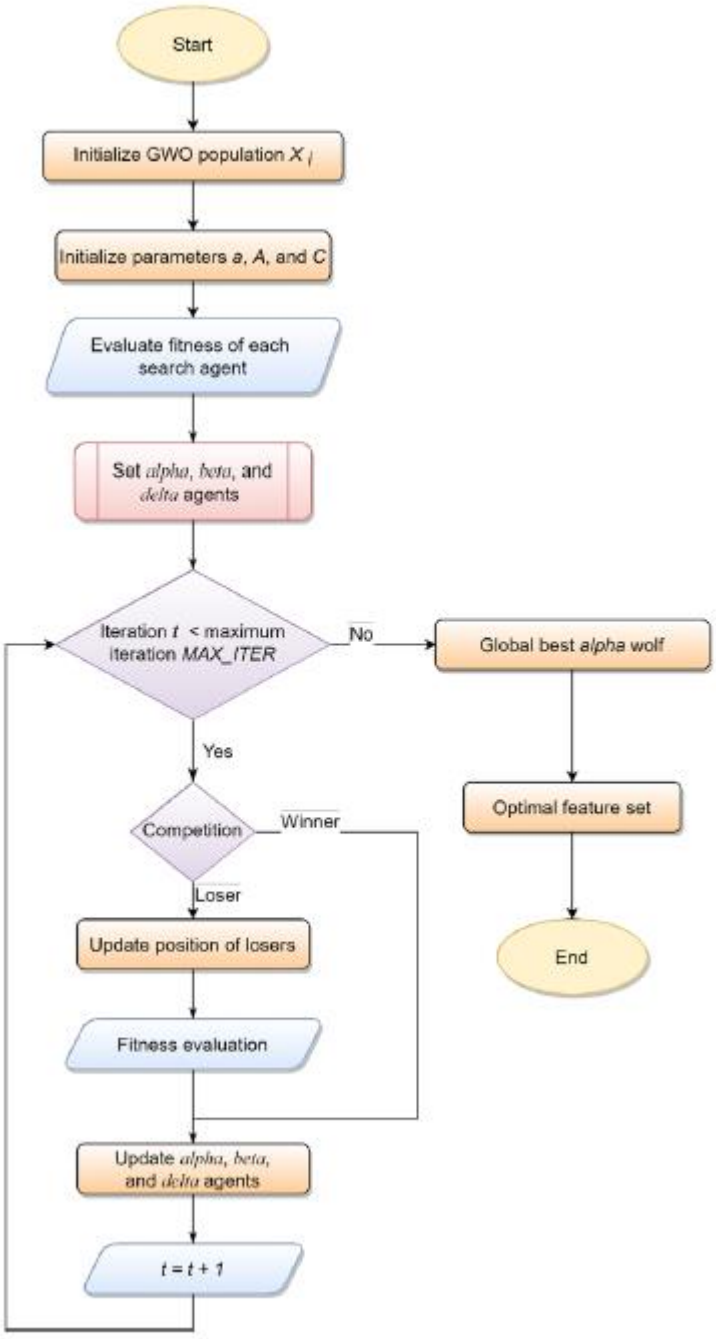
ResNet50:- ResNet-50 is a deep convolutional neural network architecture that belongs to the ResNet (Residual Network) family. Introduced by Microsoft Research in 2015, it is renowned for its exceptional performance in image classification tasks. ResNet-50 consists of 50 layers, including convolutional layers, pooling layers, and fully connected layers. Its key innovation lies in the use of residual connections, which allow the network to learn residual mappings instead of directly fitting desired mappings. This approach mitigates the vanishing gradient problem and enables the training of very deep networks effectively. ResNet-50 is pre-trained on the ImageNet dataset, making it suitable for transfer learning in various computer vision applications. With its deep architecture and residual connections, ResNet-50 achieves state-of-the-art accuracy on a wide range of image classification benchmarks while maintaining relatively low computational complexity compared to other deep networks.

Principal Component Analysis(PCA):- Principal Component Analysis (PCA) is a widely used technique in data analysis and dimensionality reduction. Its primary goal is to transform a dataset of possibly correlated variables into a new set of linearly uncorrelated variables called principal components. PCA achieves this by identifying the directions (or principal components) along which the data varies the most.
The first principal component captures the maximum variance in the data, with each subsequent component capturing as much of the remaining variance as possible, while also being orthogonal to the previous components. This ensures that the components are uncorrelated with each other. PCA is particularly useful for reducing the dimensionality of high-dimensional datasets while preserving as much information as possible. By selecting only a subset of the principal components that capture the majority of the variance in the data, PCA can simplify

complex datasets and facilitate easier visualization, interpretation, and analysis. Additionally, PCA can be employed for data preprocessing, noise reduction, and feature extraction. It finds applications in various fields such as image processing, genetics, finance, and signal processing. However, it's important to note that PCA assumes linear relationships among variables and may not perform optimally if this assumption is violated. Moreover, interpretation of the principal components may not always be straightforward, especially in high-dimensional spaces. Despite these limitations, PCA remains a powerful and widely used tool in exploratory data analysis and dimensionality reduction.

Grey Wolf Optimizer(GWO):- Grey Wolf Optimizer (GWO) is a metaheuristic optimization algorithm inspired by the social behavior and hunting strategies of grey wolves. Developed by Mirjalili et al. in 2014, GWO mimics the hunting behavior of grey wolves, where the pack collaborates to capture prey efficiently. In the GWO algorithm, potential solutions to optimization problems are represented as wolf positions in a search space. The positions of alpha, beta, and delta wolves correspond to the best, second-best, and third-best solutions found so far. Other wolves in the pack adjust their positions based on the alpha, beta, and delta wolves to explore and exploit the search space effectively. GWO utilizes four main operations: encircling prey, attacking prey, searching for prey, and updating wolf positions. These operations allow the wolves to iteratively refine their positions and converge towards optimal solutions.

One of the key advantages of GWO is its simplicity and ease of implementation. It requires minimal tuning of parameters and exhibits good convergence properties across a wide range of optimization problems. GWO has been successfully applied to various optimization tasks, including engineering design, parameter tuning, and machine learning. Despite its effectiveness, GWO may struggle with multimodal optimization problems and may require additional strategies to prevent premature convergence. Overall, GWO is a powerful and versatile optimization algorithm with potential applications in diverse domains.



Flowchart showing the workflow of the grey wolf optimization algorithm used

Support Vector Machine(SVM):- Support Vector Machine (SVM) is a supervised machine learning algorithm used for classification and regression tasks. It works by finding the optimal hyperplane that best separates data points into different classes in a high-dimensional space. SVM aims to maximize the margin, which is the distance between the hyperplane and the nearest data points from each class. This margin ensures better generalization and robustness to noise. SVM is particularly effective in dealing with high-dimensional data and is capable of handling both linearly separable and non-linearly separable datasets through the use of kernel functions. Commonly used kernel functions include linear, polynomial, radial basis function (RBF), and sigmoid kernels. One of the key advantages of SVM is its ability to handle complex decision boundaries while minimizing the risk of overfitting. Additionally, SVM has been widely used in various fields, including text classification, image recognition, and bioinformatics, due to its versatility and high performance.

Challenges and Trends:

Principal Component Analysis (PCA) faces challenges in handling non-linear data distributions, as it assumes linear relationships among variables. Additionally, interpreting the principal components may be challenging, especially in high-dimensional spaces. A trend in PCA is the development of non-linear variants such as Kernel PCA, which can capture non-linear relationships in the data. Support Vector Machines (SVM) encounter challenges with large-scale datasets and high computational complexity, particularly when using non-linear kernels. A trend in SVM is the exploration of kernel approximation techniques and distributed computing frameworks to improve scalability and efficiency. Convolutional Neural Networks (CNNs) struggle with overfitting, especially when dealing with small datasets, and require substantial computational resources for training deep architectures. A trend in CNNs is the development of regularization techniques, transfer learning strategies, and lightweight architectures for deployment on resource-constrained devices. ResNet-50 may face challenges with training instability, particularly in deeper architectures, and requires careful hyperparameter tuning. A trend in ResNet-50 is the exploration of residual connections in other network architectures and the development of deeper and more efficient models. Grey Wolf Optimizer (GWO) may encounter challenges with premature convergence and poor exploration-exploitation balance, especially in multimodal optimization problems. A trend in GWO is the investigation of hybrid and improved variants, such as enhanced exploration

strategies and adaptive parameter settings, to overcome these challenges and improve convergence performance.

CONCLUSION:

In Conclusion, Principal Component Analysis (PCA) is a dimensionality reduction technique used to identify significant features in a dataset and represent them in a lower-dimensional space. While it struggles with non-linear data distributions, trends like Kernel PCA address this limitation. Support Vector Machine (SVM) is a supervised learning algorithm for classification and regression tasks, optimizing hyperplanes to separate data points into classes. Challenges include scalability and complexity, with kernel approximation and distributed computing being trends to improve efficiency. Convolutional Neural Networks (CNNs) are deep learning architectures for image tasks, detecting features through convolutional layers. Overfitting and resource consumption are challenges, addressed by regularization and transfer learning trends. ResNet-50, a specific CNN, excels in depth and performance but faces training instability. Grey Wolf Optimizer (GWO), inspired by wolf behavior, optimizes problems with challenges like premature convergence, improved by hybrid variants.

CHAPTER: 3 SYSTEM ANALYSIS

INTRODUCTION

A pap smear, also called a Pap test, is a routine screening for cervical cancer. It checks for abnormal cells on your cervix, the opening to your womb. These abnormal cells could be precancerous, meaning they could turn into cancer if left untreated.

During the test, a healthcare professional will gently scrape a small sample of cells from your cervix. This might feel slightly uncomfortable, but it shouldn't cause any lasting pain. The cells are then examined in a lab for abnormalities. Pap smears are important because cervical cancer is highly preventable when caught early. Typically, women between the ages of 25 and 64 need Pap smears regularly, but the recommended schedule can vary depending on your age and health history. Talk to your doctor about when and how often you should have a Pap smear.

DISADVANTAGES & LIMITATIONS OF EXISTING SYSTEMS

Pap smears, while a crucial tool for early detection of cervical cancer, have certain limitations.

Here are some key points to consider:

1. **Incomplete Detection:** Pap smears aren't perfect and can miss precancerous cells, especially those caused by certain strains of the human papillomavirus (HPV). This means some women with developing cervical cancer might receive a negative Pap smear result.

2. **Inconclusive Results:** Sometimes, the collected cells during a Pap smear test might be unclear due to factors like inflammation or blood. This can lead to inconclusive results, causing anxiety and requiring a repeat test.
3. **Accuracy in Younger Women:** Pap smears might be less accurate for young women under 25. This is because their cervixes naturally undergo cell changes as part of development, which can mimic precancerous abnormalities.
4. **Psychological Impact:** A positive Pap smear result, although not always indicative of cancer, can understandably cause significant worry and stress. This highlights the importance of discussing the test with your doctor beforehand to understand the follow-up process.
5. **HPV Prevention**: The Pap smear itself doesn't prevent HPV infection, the underlying cause of most cervical cancers. While it can detect precancerous changes caused by HPV, it doesn't eliminate the virus itself. Vaccination against certain HPV strains is recommended to reduce the risk of cervical cancer.

It's important to remember that despite these limitations, Pap smears remain an essential screening method for cervical cancer. Discussing these limitations with your doctor can empower you to make informed decisions about your cervical health and explore additional screening options, if recommended, for a more comprehensive approach.

PROPOSED SYSTEMS

Pap smears are a valuable tool, but limitations exist. Our machine learning model that incorporates Principal Component Analysis (PCA), Grey Wolf Optimizer (GWO), Support Vector Machine (SVM), ResNet50, and Convolutional Neural Networks (CNN) offers exciting advancements.

Here's how this new approach might improve upon the traditional Pap smear test:

1. **Enhanced Accuracy**: By integrating deep learning techniques like ResNet50 and CNN, the model could potentially analyze cell images with greater precision, leading to fewer missed precancerous cases compared to traditional Pap smears.
2. **Reduced Inconclusive Results:** PCA, a dimensionality reduction technique, might help extract the most critical features from cell images, potentially leading to clearer classifications and fewer inconclusive results requiring repeat tests.
3. **Improved Efficiency:** GWO, an optimization algorithm, could streamline the model's training process, potentially leading to faster analysis times and quicker results.
4. **Potential for Automation**: The model's reliance on machine learning techniques opens doors for further automation. This could streamline the analysis process, potentially reducing human error and workload in cytology labs.
5. **Standardized Interpretation**: Machine learning models can offer consistent analysis, potentially reducing discrepancies in interpretation that can sometimes occur with traditional Pap smears evaluated by different cytologists.

These potential improvements suggest our machine learning model holds promise for advancing Pap smear analysis. Further research and clinical trials will be crucial to validate its effectiveness and pave the way for its use in real-world settings.

SUMMARY

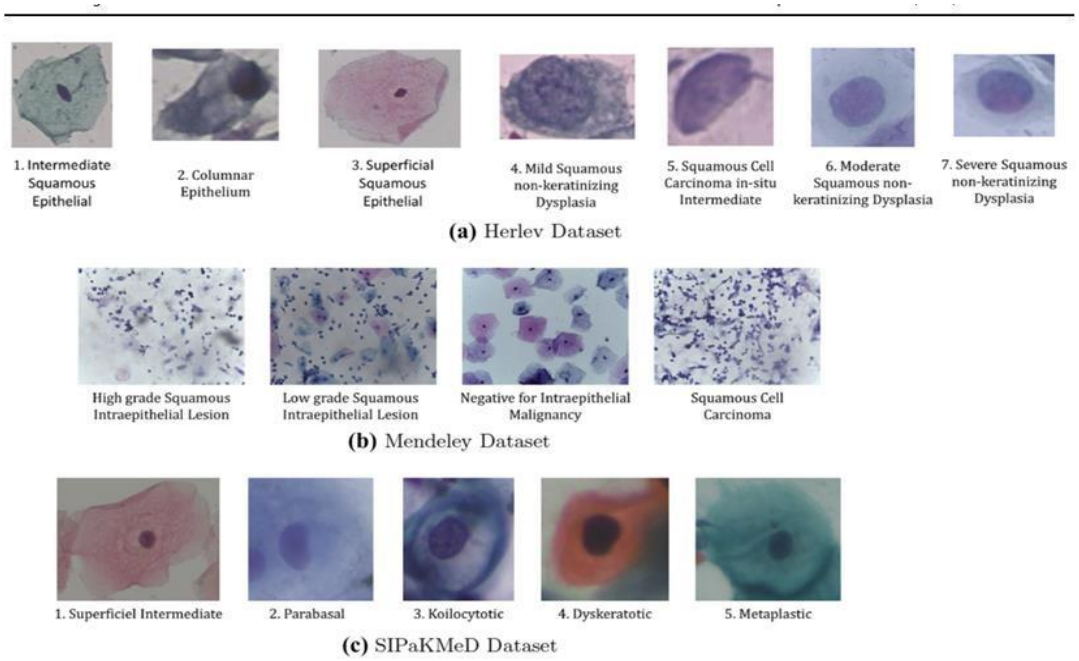
Pap smears are a cornerstone of cervical cancer prevention, but they aren't without limitations. While they effectively detect precancerous cells in many cases, there's a chance they might miss some, particularly those linked to specific HPV strains. Inconclusive results due to unclear cell samples can also lead to repeat tests and anxiety. Additionally, Pap smears might be less accurate for younger women due to normal cervical cell changes. Psychologically, a positive Pap smear can cause significant worry, even though most abnormalities are not cancerous. Finally, the Pap smear itself doesn't prevent HPV infection. However, our development of a machine learning model for Pap smear analysis offers exciting possibilities. This model incorporates techniques like PCA, GWO, SVM, ResNet50, and CNN. These techniques have the potential to significantly improve Pap smear analysis. Deep learning approaches within the model could analyze cell images with greater precision, potentially leading to fewer missed precancerous cases. PCA might help extract the most critical information from cell images, leading to clearer classifications and fewer inconclusive results. Additionally, the model's training process could be streamlined through GWO, potentially resulting in faster analysis times. The model's reliance on machine learning also opens doors for automation, potentially reducing human error and workload in cytology labs. Finally, machine learning models can offer consistent analysis, potentially reducing discrepancies in interpretation that can sometimes occur with traditional Pap smears evaluated by different cytologists
In conclusion, our machine learning model holds significant promise for advancing Pap smear analysis. While further research and clinical trials are necessary to validate its effectiveness, this approach has the potential to improve accuracy, efficiency, and potentially reduce the psychological impact associated with Pap smears.

CHAPTER: 4 SYSTEM DESIGN AND IMPLEMENTATION

DATASETS:-

We use three publicly available cervical cytology data sets in this study for evaluating the proposed classification framework. These datasets are described in brief in the following subsections.

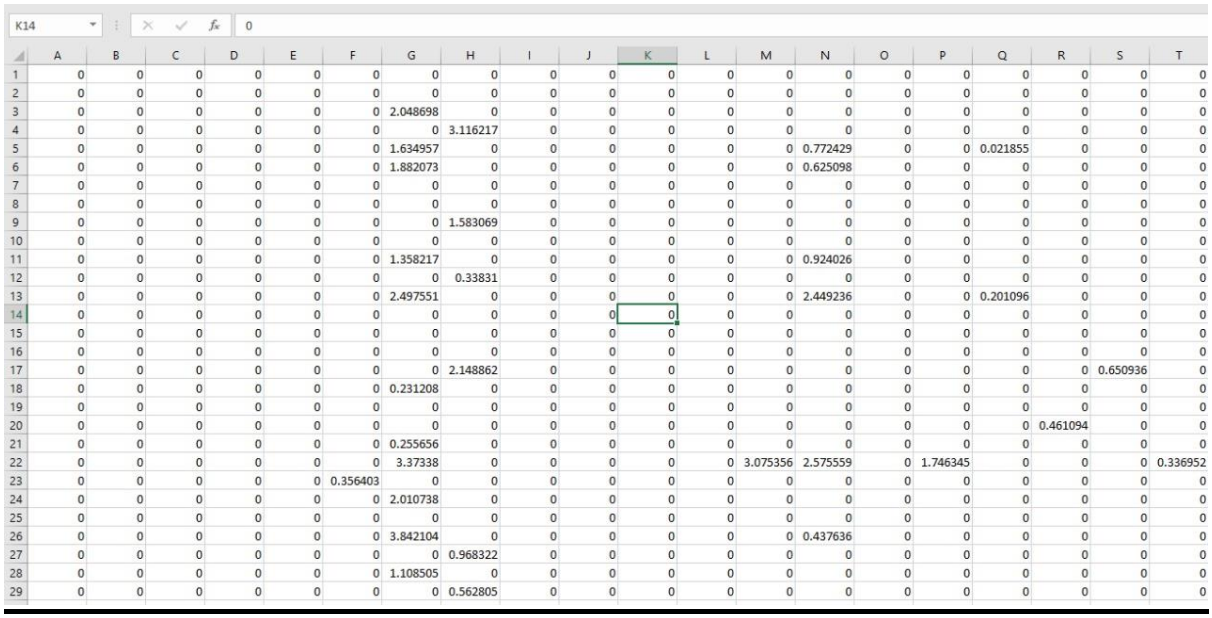
1. **Herlev Pap Smear Dataset:-** The Herlev Pap Smear dataset is a publicly available benchmark dataset consisting of 917 single cell images distributed unevenly among 7 different classes. The distribution of images in each class are tabulated in Table
2. **Mendeley Liquid Based Cytology Dataset:-** The Mendeley LBC dataset developed at Obstetrics and Gynaecology department of Guwahati Medical College and Hospital, consists of 963 whole slide images of cervical cytology distributed unevenly in four different classes.
3. **SIPaKMeD Pap Smear Dataset:-** The SIPaKMeD Pap Smear dataset by Plissiti et al. consists of 4049 images of isolated cells (extracted from 966 whole slide images) categorized into five different classes based on their cytomorphological features.



Dataset	Class	Category	Cell type	Num-ber of images
Herlev Pap Smear (total: 917)	1	Normal	Intermediate squamous epithelial	70
	2	Normal	Columnar epithelial	98
	3	Normal	Superficial squamous epithelial	74
	4	Abnormal	Mild squamous non-keratinizing dysplasia	182
	5	Abnormal	Squamous cell carcinoma in-situ intermediate	150
	6	Abnormal	Moderate squamous non-keratinizing dysplasia	146
	7	Abnormal	Severe squamous non-keratinizing dysplasia	197
Mendeley LBC (total: 963)	1	Normal	Negative for intraepithelial malignancy	613
	2	Abnormal	Low grade squamous intraepithelial lesion (LSIL)	163
	3	Abnormal	High grade squamous intraepithelial lesion (HSIL)	113
	4	Abnormal	Squamous cell carcinoma (SCC)	74
SIPaKMeD Pap Smear (total: 4049)	1	Normal	Superficial-intermediate	831
	2	Normal	Parabasal	787
	3	Abnormal	Koilocytotic	825
	4	Abnormal	Dyskeratotic	813
	5	Benign	Metaplastic	793

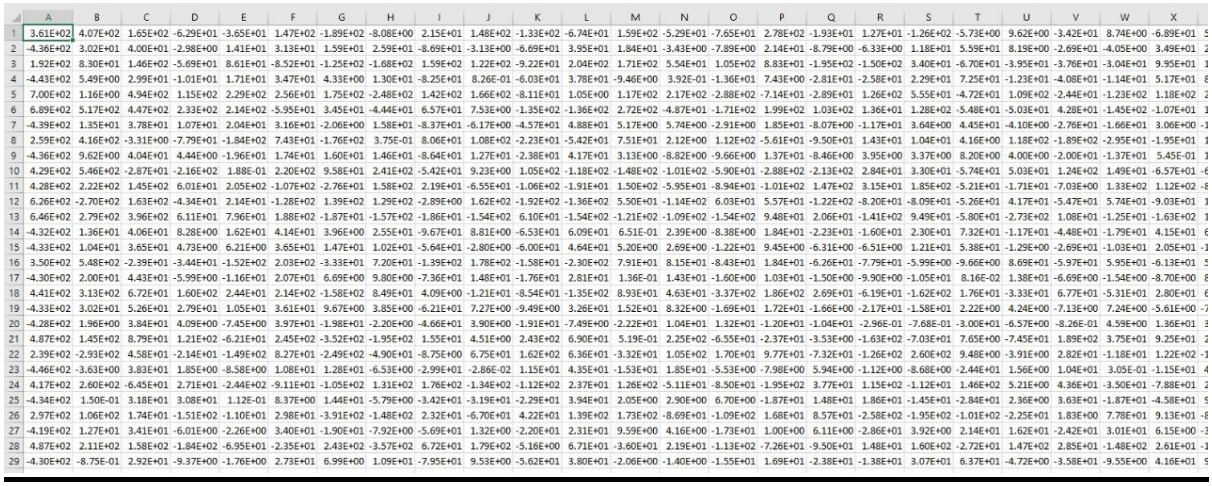
IMPORTANT CODE OUTPUT

First we will be using ResNet50 (a type of CNN) to extract features from the dataset, it has 50 layers in the neural network. We are getting this csv file as the output after running it.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	2.048698	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	3.116217	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	1.634957	0	0	0	0	0	0	0.772429	0	0	0.021855	0	0	0
6	0	0	0	0	0	0	1.882073	0	0	0	0	0	0	0.625098	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	1.583069	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	1.358217	0	0	0	0	0	0	0.924026	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0.33831	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	2.497551	0	0	0	0	0	0	2.449236	0	0	0.201096	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	2.148862	0	0	0	0	0	0	0	0	0	0	0.650936	0
18	0	0	0	0	0	0	0.231208	0	0	0	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.461094	0	0
21	0	0	0	0	0	0	0.255656	0	0	0	0	0	0	0	0	0	0	0	0	0
22	0	0	0	0	0	0	3.37338	0	0	0	0	0	3.075356	2.575559	0	1.746345	0	0	0	0.336952
23	0	0	0	0	0	0.356403	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24	0	0	0	0	0	0	2.010738	0	0	0	0	0	0	0	0	0	0	0	0	0
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
26	0	0	0	0	0	0	3.842104	0	0	0	0	0	0	0.437636	0	0	0	0	0	0
27	0	0	0	0	0	0	0	0.968322	0	0	0	0	0	0	0	0	0	0	0	0
28	0	0	0	0	0	0	1.108505	0	0	0	0	0	0	0	0	0	0	0	0	0
29	0	0	0	0	0	0	0	0.562805	0	0	0	0	0	0	0	0	0	0	0	0

After we are getting the csv file from extracting features, we are running main.py which is used to concatenate the features. Then we will apply PCA (Principal Component Analysis) onto the concatenated features. After PCA we get output like this

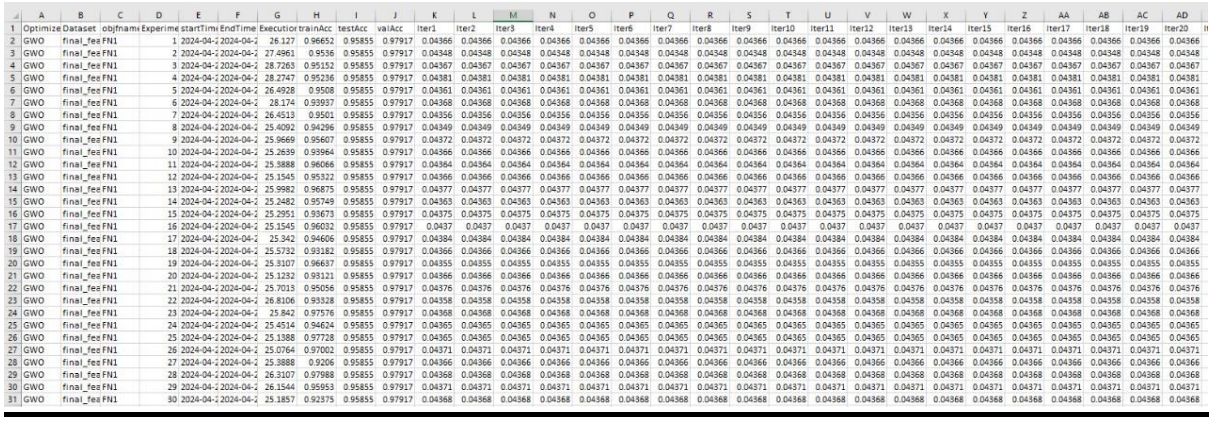


	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
1	3.61E+02	4.07E+02	1.65E+02	-6.29E+01	-3.65E+01	1.47E+02	-1.89E+02	-8.08E+00	2.15E+01	1.48E+02	-1.33E+02	-6.74E+01	1.59E+02	-5.29E+01	-7.65E+01	2.78E+02	-1.93E+01	1.27E+01	-1.26E+02	-5.73E+00	9.62E+00	-3.42E+01	8.74E+00	-6.89E+01
2	-4.36E+02	3.02E+01	4.00E+01	-2.98E+00	1.41E+01	3.13E+01	1.59E+01	2.59E+01	-8.69E+01	-3.13E+00	-6.69E+01	3.95E+01	1.84E+01	-3.43E+00	-7.89E+00	2.14E+01	-8.79E+00	-6.33E+00	1.18E+01	5.59E+01	8.19E+00	-2.69E+01	-4.05E+00	3.49E+01
3	1.92E+02	8.30E+01	1.46E+02	-5.69E+01	8.61E+01	-8.52E+01	-1.25E+02	-1.68E+02	1.59E+02	1.22E+02	-9.22E+01	2.04E+02	1.71E+02	5.54E+01	1.05E+02	8.83E+01	-1.95E+02	-1.50E+02	3.40E+01	-6.70E+01	-3.95E+01	-3.76E+01	-3.04E+01	9.95E+01
4	-4.43E+02	5.49E+00	2.99E+01	-1.01E+01	1.71E+01	3.47E+01	4.33E+00	1.30E+01	-8.25E+01	8.26E-01	-6.03E+01	3.78E+01	-9.46E+00	3.92E-01	-1.36E+01	7.43E+00	-2.81E+01	-2.58E+01	2.29E+01	7.25E+01	-1.23E+01	-4.08E+01	-1.14E+01	5.17E+01
5	7.00E+02	1.16E+00	4.94E+02	1.15E+02	2.29E+02	2.56E+01	1.75E+02	-2.48E+02	1.42E+02	1.66E+02	-8.11E+01	1.05E+00	1.17E+02	2.17E+02	-2.88E+02	-7.14E+01	-2.89E+01	1.26E+02	5.55E+01	-4.72E+01	1.09E+02	-2.44E+01	-1.23E+02	1.18E+02
6	6.89E+02	5.17E+02	4.47E+02	2.33E+02	2.14E+02	-5.95E+01	3.45E+01	-4.44E+01	6.57E+01	7.53E+00	-1.35E+02	-1.36E+02	2.72E+02	-4.87E+01	-1.71E+02	1.99E+02	1.03E+02	1.36E+01	1.28E+02	-5.48E+01	-5.03E+01	4.28E+01	-1.45E+02	-1.07E+01
7	-4.39E+02	1.35E+01	3.78E+01	1.07E+01	2.04E+01	3.16E+01	-2.06E+00	1.58E+01	-8.37E+01	-6.17E+00	-4.57E+01	4.88E+01	5.17E+00	5.74E+00	-2.91E+00	1.85E+01	-8.07E+00	-1.17E+01	3.64E+00	4.45E+01	-4.10E+00	-2.76E+01	-1.66E+01	3.06E+00
8	2.59E+02	4.16E+02	-3.31E+00	-7.79E+01	-1.84E+02	7.43E+01	-1.76E+02	3.75E-01	8.06E+01	1.08E+02	-2.23E+01	-5.42E+01	7.51E+01	2.12E+00	1.12E+02	-5.61E+01	-9.50E+01	1.43E+01	1.04E+01	4.16E+00	1.18E+02	-1.89E+02	-2.95E+01	-1.95E+01
9	-4.36E+02	9.62E+00	4.04E+01	4.44E+00	-1.96E+01	1.74E+01	1.60E+01	1.46E+01	-8.64E+01	1.27E+01	-2.38E+01	4.17E+01	3.13E+00	-8.82E+00	-9.66E+00	1.37E+01	-8.46E+00	3.95E+00	3.37E+00	8.20E+00	4.00E+00	-2.00E+01	-1.37E+01	5.45E-01
10	4.29E+02	5.46E+02	-2.87E+01	-2.16E+02	1.88E-01	2.20E+02	9.58E+01	2.41E+02	-5.42E+01	9.23E+00	1.05E+02	-1.18E+02	-1.48E+02	-1.01E+02	-5.90E+01	-2.88E+02	-2.13E+02	2.84E+01	3.30E+01	-5.74E+01	5.03E+01	1.24E+02	1.49E+01	-6.57E+01
11	4.28E+02	2.22E+02	1.45E+02	6.01E+01	2.05E+02	-1.07E+02	-2.76E+01	1.58E+02	2.19E+01	-6.55E+01	-1.06E+02	-1.91E+01	1.50E+02	-5.95E+01	-8.94E+01	-1.01E+02	1.47E+02	3.15E+01	1.85E+02	-5.21E+01	-1.71E+01	-7.03E+00	1.33E+02	1.12E+02
12	6.26E+02	-2.70E+02	1.63E+02	-4.34E+01	2.14E+01	-1.28E+02	1.39E+02	1.29E+02	-2.89E+00	1.62E+02	-1.92E+02	-1.36E+02	5.50E+01	-1.14E+02	6.03E+01	5.57E+01	-1.22E+02	-8.20E+01	-8.09E+01	-5.26E+01	4.17E+01	-5.47E+01	5.74E+01	-9.03E+01
13	6.46E+02	2.79E+02	3.96E+02	6.11E+01	7.96E+01	1.88E+02	-1.87E+01	-1.57E+02	-1.86E+01	-1.54E+02	6.10E+01	-1.54E+02	-1.21E+02	-1.09E+02	-1.54E+02	9.48E+01	2.06E+01	-1.41E+02	9.49E+01	-5.80E+01	-2.73E+02	1.08E+01	-1.25E+01	-1.63E+02
14	-4.32E+02	1.36E+01	4.06E+01	8.28E+00	1.62E+01	4.14E+01	3.96E+00	2.55E+01	-9.67E+01	8.81E+00	-6.53E+01	6.09E+01	6.51E-01	2.39E+00	-8.38E+00	1.84E+01	-2.23E+01	-1.60E+01	2.30E+01	7.32E+01	-1.17E+01	-4.48E+01	-1.79E+01	4.15E+01
15	-4.33E+02	1.04E+01	3.65E+01	4.73E+00	6.21E+00	3.65E+01	1.47E+01	1.02E+01	-5.64E+01	-2.80E+00	-6.00E+01	4.64E+01	5.20E+00	2.69E+00	-1.22E+01	9.45E+00	-6.31E+00	-6.51E+00	1.21E+01	5.38E+01	-1.29E+00	-2.69E+01	-1.03E+01	2.05E+01
16	3.50E+02	5.48E+02	-2.39E+01	-3.44E+01	-1.52E+02	2.03E+02	-3.33E+01	7.20E+01	-1.39E+02	1.78E+02	-1.58E+01	-2.30E+02	7.91E+01	8.15E+01	-8.43E+01	1.84E+01	-6.26E+01	-7.79E+01	-5.99E+00	-9.66E+00	8.69E+01	-5.97E+01	5.95E+01	-6.13E+01
17	-4.30E+02	2.00E+01	4.43E+01	-5.99E+00	-1.16E+01	2.07E+01	6.69E+00	9.80E+00	-7.36E+01	1.48E+01	-1.76E+01	2.81E+01	1.36E-01	1.43E+01	-1.60E+00	1.03E+01	-1.50E+00	-9.90E+00	-1.05E+01	8.16E-02	1.38E+01	-6.69E+00	-1.54E+00	-8.70E+00
18	4.41E+02	3.13E+02	6.72E+01	1.60E+02	2.44E+01	2.14E+02	-1.58E+02	8.49E+01	4.09E+00	-1.21E+01	-8.54E+01	-1.35E+02	8.93E+01	4.63E+01	-3.37E+02	1.86E+02	2.69E+01	-6.19E+01	-1.62E+02	1.76E+01	-3.33E+01	6.77E+01	-5.31E+01	2.80E+01
19	-4.33E+02	3.02E+01	5.26E+01	2.79E+01	1.05E+01	3.61E+01	9.67E+00	3.85E+00	-6.21E+01	7.27E+00	-9.49E+00	3.26E+01	1.52E+01	8.32E+00	-1.69E+01	1.72E+01	-1.66E+00	-2.17E+01	-1.58E+01	2.22E+00	4.24E+00	-7.13E+00	7.24E+00	-5.61E+00
20	-4.28E+02	1.96E+00	3.84E+01	4.09E+00	-7.45E+00	3.97E+01	-1.98E+01	-2.20E+00	-4.66E+01	3.90E+00	-1.91E+01	-7.49E+00	-2.22E+01	1.04E+01	1.32E+01	-1.20E+01	-1.04E+01	-2.96E-01	-7.68E-01	-3.00E+01	-6.57E+00	-8.26E-01	4.59E+00	1.36E+01
21	4.87E+02	1.45E+02	8.79E+01	1.21E+02	-6.21E+01	2.45E+02	-3.52E+02	-1.95E+02	1.55E+01	4.51E+00	2.43E+02	6.90E+01	5.19E-01	2.25E+02	-6.55E+01	-2.37E+01	-3.53E+00	-1.63E+02	-7.03E+01	7.65E+00	-7.45E+01	1.89E+02	3.75E+01	9.25E+01
22	2.39E+02	-2.93E+02	4.58E+01	-2.14E+01	-1.49E+02	8.27E+01	-2.49E+02	-4.90E+01	-8.75E+00	6.75E+01	1.62E+02	6.36E+01	-3.32E+01	1.05E+02	1.70E+01	9.77E+01	-7.32E+01	-1.26E+02	2.60E+02	9.48E+00	-3.91E+00	2.82E+01	-1.18E+01	1.22E+02
23	-4.46E+02	-3.63E+00	3.83E+01	1.85E+00	-8.58E+00	1.08E+01	1.28E+01	-6.53E+00	-2.99E+01	-2.86E-02	1.15E+01	4.35E+01	-1.53E+01	1.85E+01	-5.53E+00	-7.98E+00	5.94E+00	-1.12E+00	-8.68E+00	-2.44E+01	1.56E+00	1.04E+01	3.05E-01	-1.15E+01
24	4.17E+02	2.60E+02	-6.45E+01	2.71E+01	-2.44E+02	-9.11E+01	-1.05E+02	1.31E+02	1.76E+02	-1.34E+02	-1.12E+02	2.37E+01	1.26E+02	-5.11E+01	-8.50E+01	-1.95E+02	3.77E+01	1.15E+02	-1.12E+01	1.46E+02	5.21E+00	4.36E+01	-3.50E+01	-7.88E+01
25	-4.34E+02	1.50E-01	3.18E+01	3.08E+01	1.12E-01	8.37E+00	1.44E+01	-5.79E+00	-3.42E+01	-3.19E+01	-2.29E+01	3.94E+01	2.05E+00	2.90E+00	6.70E+00	-1.87E+01	1.48E+01	1.86E+01	-1.45E+01	-2.84E+01	2.36E+00	3.63E+01	-1.87E+01	-4.58E+01
26	2.97E+02	1.06E+02	1.74E+01	-1.51E+02	-1.10E+01	2.98E+01	-3.91E+02	-1.48E+02	2.32E+01	-6.70E+01	4.22E+01	1.39E+02	1.73E+02	-8.69E+01	-1.09E+02	1.68E+01	8.57E+01	-2.58E+02	-1.95E+02	-1.01E+02	-2.25E+01	1.83E+00	7.78E+01	9.13E+01
27	-4.19E+02	1.27E+01	3.41E+01	-6.01E+00	-2.26E+00	3.40E+01	-1.90E+01	-7.92E+00	-5.69E+01	1.32E+00	-2.20E+01	2.31E+01	9.59E+00	4.16E+00	-1.73E+01	1.00E+00	6.11E+00	-2.86E+01	3.92E+00	2.14E+01	1.62E+01	-2.42E+01	3.01E+01	6.15E+00
28	4.87E+02	2.11E+02	1.58E+02	-1.84E+02	-6.95E+01	-2.35E+01	2.43E+02	-3.57E+02	6.72E+01	1.79E+02	-5.16E+00	6.71E+01	-3.60E+01	2.19E+01	-1.13E+02	-7.26E+01	-9.50E+01	1.48E+01	1.60E+02	-2.72E+01	1.47E+02	2.85E+01	-1.48E+02	2.61E+01
29	-4.30E+02	-8.75E-01	2.92E+01	-9.37E+00	-1.76E+00	2.73E+01	6.99E+00	1.09E+01	-7.95E+01	9.53E+00	-5.62E+01	3.80E+01	-2.06E+00	-1.40E+00	-1.55E+01	1.69E+01	-2.38E+01	-1.38E+01	3.07E+01	6.37E+01	-4.72E+00	-3.58E+01	-9.55E+00	4.16E+01

Using PCA allowed us to reduce the number of features from 76288 to 1007. This helps in faster and efficient running of our model.

```
D:\MAIN>python main.py --num_csv 2
Enter name of csv number 1: t2
Enter name of csv number 2: t4
(1281, 76288)
(1281, 1007)
```

Final result



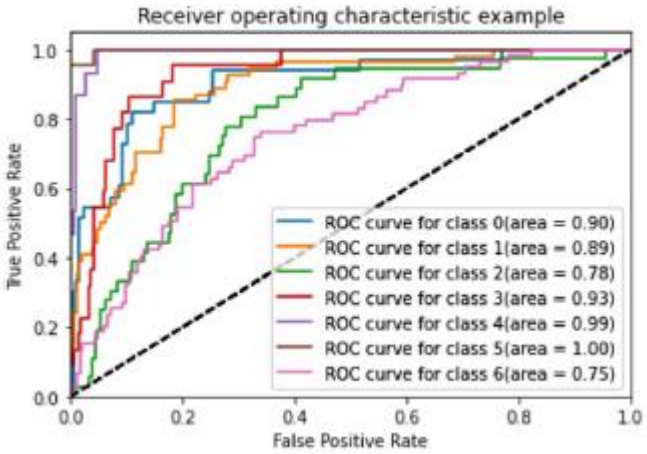
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD
1	Optimize	Dataset	objfnam	Experime	startTime	EndTime	Executior	trainAcc	testAcc	valAcc	Iter1	Iter2	Iter3	Iter4	Iter5	Iter6	Iter7	Iter8	Iter9	Iter10	Iter11	Iter12	Iter13	Iter14	Iter15	Iter16	Iter17	Iter18	Iter19	Iter20	I
2	GWO	final_fea	FN1	1	2024-04-2	2024-04-2	26.127	0.96652	0.95855	0.97917	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	
3	GWO	final_fea	FN1	2	2024-04-2	2024-04-2	27.4961	0.9536	0.95855	0.97917	0.04348	0.04348	0.04348	0.04348	0.04348	0.04348	0.04348	0.04348	0.04348	0.04348	0.04348	0.04348	0.04348	0.04348	0.04348	0.04348	0.04348	0.04348	0.04348	0.04348	
4	GWO	final_fea	FN1	3	2024-04-2	2024-04-2	28.7263	0.95152	0.95855	0.97917	0.04367	0.04367	0.04367	0.04367	0.04367	0.04367	0.04367	0.04367	0.04367	0.04367	0.04367	0.04367	0.04367	0.04367	0.04367	0.04367	0.04367	0.04367	0.04367	0.04367	
5	GWO	final_fea	FN1	4	2024-04-2	2024-04-2	28.2747	0.95236	0.95855	0.97917	0.04381	0.04381	0.04381	0.04381	0.04381	0.04381	0.04381	0.04381	0.04381	0.04381	0.04381	0.04381	0.04381	0.04381	0.04381	0.04381	0.04381	0.04381	0.04381	0.04381	
6	GWO	final_fea	FN1	5	2024-04-2	2024-04-2	26.4928	0.9508	0.95855	0.97917	0.04361	0.04361	0.04361	0.04361	0.04361	0.04361	0.04361	0.04361	0.04361	0.04361	0.04361	0.04361	0.04361	0.04361	0.04361	0.04361	0.04361	0.04361	0.04361	0.04361	
7	GWO	final_fea	FN1	6	2024-04-2	2024-04-2	28.174	0.93937	0.95855	0.97917	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	
8	GWO	final_fea	FN1	7	2024-04-2	2024-04-2	26.4513	0.9501	0.95855	0.97917	0.04356	0.04356	0.04356	0.04356	0.04356	0.04356	0.04356	0.04356	0.04356	0.04356	0.04356	0.04356	0.04356	0.04356	0.04356	0.04356	0.04356	0.04356	0.04356	0.04356	
9	GWO	final_fea	FN1	8	2024-04-2	2024-04-2	25.4092	0.94296	0.95855	0.97917	0.04349	0.04349	0.04349	0.04349	0.04349	0.04349	0.04349	0.04349	0.04349	0.04349	0.04349	0.04349	0.04349	0.04349	0.04349	0.04349	0.04349	0.04349	0.04349	0.04349	
10	GWO	final_fea	FN1	9	2024-04-2	2024-04-2	25.9669	0.95607	0.95855	0.97917	0.04372	0.04372	0.04372	0.04372	0.04372	0.04372	0.04372	0.04372	0.04372	0.04372	0.04372	0.04372	0.04372	0.04372	0.04372	0.04372	0.04372	0.04372	0.04372	0.04372	
11	GWO	final_fea	FN1	10	2024-04-2	2024-04-2	25.2639	0.93984	0.95855	0.97917	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	
12	GWO	final_fea	FN1	11	2024-04-2	2024-04-2	25.5888	0.96066	0.95855	0.97917	0.04364	0.04364	0.04364	0.04364	0.04364	0.04364	0.04364	0.04364	0.04364	0.04364	0.04364	0.04364	0.04364	0.04364	0.04364	0.04364	0.04364	0.04364	0.04364	0.04364	
13	GWO	final_fea	FN1	12	2024-04-2	2024-04-2	25.1545	0.95322	0.95855	0.97917	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	
14	GWO	final_fea	FN1	13	2024-04-2	2024-04-2	25.9982	0.96875	0.95855	0.97917	0.04377	0.04377	0.04377	0.04377	0.04377	0.04377	0.04377	0.04377	0.04377	0.04377	0.04377	0.04377	0.04377	0.04377	0.04377	0.04377	0.04377	0.04377	0.04377	0.04377	
15	GWO	final_fea	FN1	14	2024-04-2	2024-04-2	25.2482	0.95749	0.95855	0.97917	0.04363	0.04363	0.04363	0.04363	0.04363	0.04363	0.04363	0.04363	0.04363	0.04363	0.04363	0.04363	0.04363	0.04363	0.04363	0.04363	0.04363	0.04363	0.04363	0.04363	
16	GWO	final_fea	FN1	15	2024-04-2	2024-04-2	25.2951	0.93673	0.95855	0.97917	0.04375	0.04375	0.04375	0.04375	0.04375	0.04375	0.04375	0.04375	0.04375	0.04375	0.04375	0.04375	0.04375	0.04375	0.04375	0.04375	0.04375	0.04375	0.04375	0.04375	
17	GWO	final_fea	FN1	16	2024-04-2	2024-04-2	25.1545	0.96032	0.95855	0.97917	0.0437	0.0437	0.0437	0.0437	0.0437	0.0437	0.0437	0.0437	0.0437	0.0437	0.0437	0.0437	0.0437	0.0437	0.0437	0.0437	0.0437	0.0437	0.0437	0.0437	
18	GWO	final_fea	FN1	17	2024-04-2	2024-04-2	25.342	0.94606	0.95855	0.97917	0.04384	0.04384	0.04384	0.04384	0.04384	0.04384	0.04384	0.04384	0.04384	0.04384	0.04384	0.04384	0.04384	0.04384	0.04384	0.04384	0.04384	0.04384	0.04384	0.04384	
19	GWO	final_fea	FN1	18	2024-04-2	2024-04-2	25.5732	0.93182	0.95855	0.97917	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	
20	GWO	final_fea	FN1	19	2024-04-2	2024-04-2	25.3107	0.96637	0.95855	0.97917	0.04355	0.04355	0.04355	0.04355	0.04355	0.04355	0.04355	0.04355	0.04355	0.04355	0.04355	0.04355	0.04355	0.04355	0.04355	0.04355	0.04355	0.04355	0.04355	0.04355	
21	GWO	final_fea	FN1	20	2024-04-2	2024-04-2	25.1232	0.95121	0.95855	0.97917	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	
22	GWO	final_fea	FN1	21	2024-04-2	2024-04-2	25.7013	0.95056	0.95855	0.97917	0.04376	0.04376	0.04376	0.04376	0.04376	0.04376	0.04376	0.04376	0.04376	0.04376	0.04376	0.04376	0.04376	0.04376	0.04376	0.04376	0.04376	0.04376	0.04376	0.04376	
23	GWO	final_fea	FN1	22	2024-04-2	2024-04-2	26.8106	0.93328	0.95855	0.97917	0.04358	0.04358	0.04358	0.04358	0.04358	0.04358	0.04358	0.04358	0.04358	0.04358	0.04358	0.04358	0.04358	0.04358	0.04358	0.04358	0.04358	0.04358	0.04358	0.04358	
24	GWO	final_fea	FN1	23	2024-04-2	2024-04-2	25.842	0.97576	0.95855	0.97917	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	
25	GWO	final_fea	FN1	24	2024-04-2	2024-04-2	25.4514	0.94624	0.95855	0.97917	0.04365	0.04365	0.04365	0.04365	0.04365	0.04365	0.04365	0.04365	0.04365	0.04365	0.04365	0.04365	0.04365	0.04365	0.04365	0.04365	0.04365	0.04365	0.04365	0.04365	
26	GWO	final_fea	FN1	25	2024-04-2	2024-04-2	25.1388	0.97728	0.95855	0.97917	0.04365	0.04365	0.04365	0.04365	0.04365	0.04365	0.04365	0.04365	0.04365	0.04365	0.04365	0.04365	0.04365	0.04365	0.04365	0.04365	0.04365	0.04365	0.04365	0.04365	
27	GWO	final_fea	FN1	26	2024-04-2	2024-04-2	25.0764	0.97002	0.95855	0.97917	0.04371	0.04371	0.04371	0.04371	0.04371	0.04371	0.04371	0.04371	0.04371	0.04371	0.04371	0.04371	0.04371	0.04371	0.04371	0.04371	0.04371	0.04371	0.04371	0.04371	
28	GWO	final_fea	FN1	27	2024-04-2	2024-04-2	25.3888	0.9206	0.95855	0.97917	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	0.04366	
29	GWO	final_fea	FN1	28	2024-04-2	2024-04-2	26.3107	0.97988	0.95855	0.97917	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	
30	GWO	final_fea	FN1	29	2024-04-2	2024-04-2	26.1544	0.95953	0.95855	0.97917	0.04371	0.04371	0.04371	0.04371	0.04371	0.04371	0.04371	0.04371	0.04371	0.04371	0.04371	0.04371	0.04371	0.04371	0.04371	0.04371	0.04371	0.04371	0.04371	0.04371	
31	GWO	final_fea	FN1	30	2024-04-2	2024-04-2	25.1857	0.92375	0.95855	0.97917	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	0.04368	

CHAPTER- 5 PERFORMANCE ANALYSIS

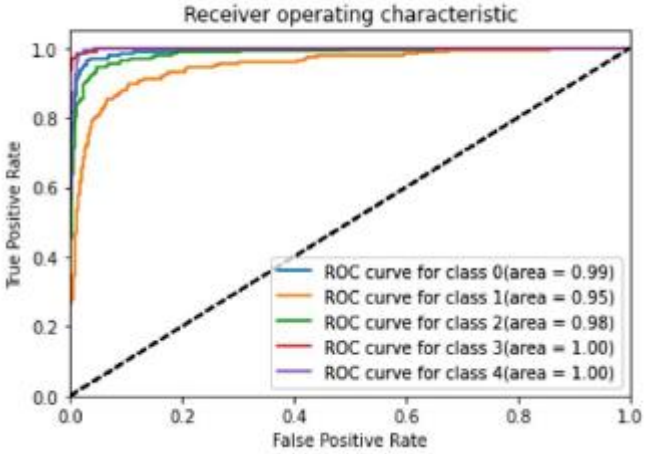
After extracting the features from the dataset using the CNN architectures said in "Materials and Methods", and the features were concatenated. We then use PCA (which retained 99% of the variance of the data) for the reduction in the dimensionality of the feature space and improvements in feature qualities respectively. The Table shows the statistics of reduction in feature dimensionality as well as the improvement of training time after this for the Herlev dataset. Then, we used the GWO algorithm and finally split the dataset and calculated the accuracy score for the training, validation, and testing sets.

To cross-validate the results of the classification task on different datasets and different features, we performed an AUC-ROC test on different datasets. The ROC (Receiver Operating Characteristics) curve is an important analyzing tool for validating the clinical findings of our experiment. The different line segments in the OVA (One Vs. All) ROC represent different classes stating that how good the features and the classifier performance are for classifying the different classes which can be broadly categorized in normal and infected cases. It represents the graphical analysis of the TPR (True Positive Rate) against the FPR (False Positive Rate) as the two operating characteristics criterion of the classifier based on the features selected. A false-positive result is a case when data of a healthy or uninfected class is predicted as an unhealthy or infected case by a classifier and it's a major drawback of the classification task. This is reciprocated by the points lying far above the diagonal line of the ROC curve suggesting that the TPR is significantly high as compared to FPR. Another important feature for analyzing the classification result is the AUC (Area Under Curve) of the ROC curve which was computed considering the 97% of the confidence interval. The analysis using the AUC-ROC curves for different datasets and different features are discussed further.

The ROC Curve obtained by the method for the two of the datasets used:



a)Herlev Pap Smear Dataset

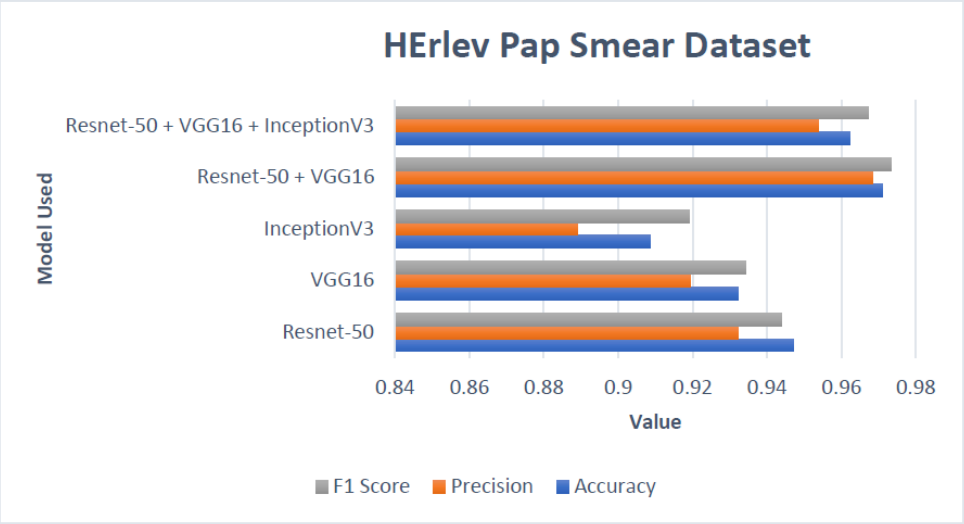


b)SIPaKMed Pap Smear Test

REDUCTION IN FEATURES DIMENSIONS ON THE HERLEV DATASET

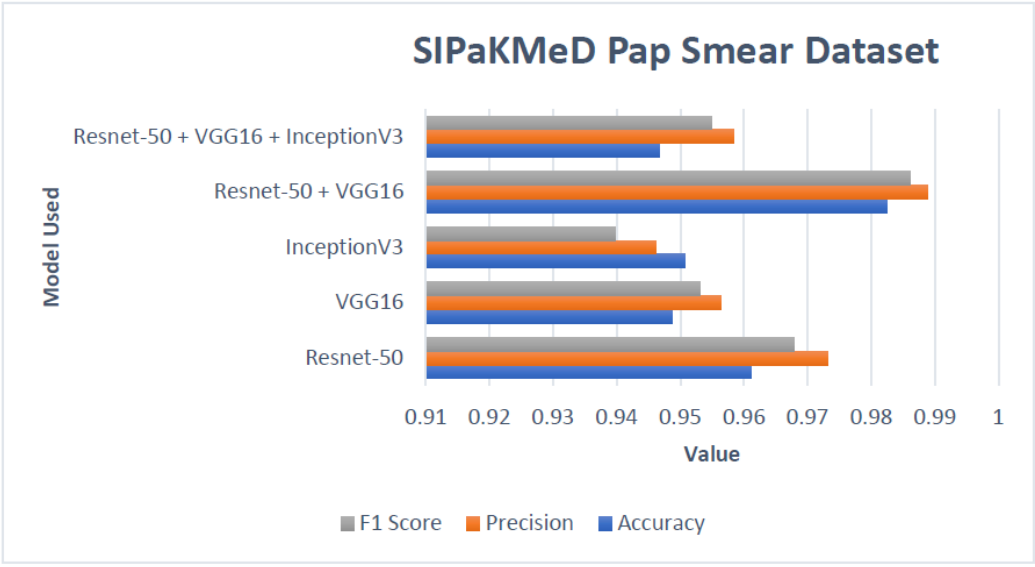
Model used	No.of features (before PCA)	No.of features (after PCA)	Reduction in feature Dimensions(%)
ResNet-50	76288	1007	98.68%
VGG16	18432	140	99.24%
ResNet-50+VGG16	94720	417	99.56%

Results on HeRlev Pap Smear Dataset:



The results obtained on different experiments on the Herlev Pap Smear dataset are shown in Fig. 5. The best classification results observed this dataset was achieved by merging the feature extracted from ResNet-50 and VGG-16 models, which gave the performance metrics as follows: accuracy = 97.11%, precision = 98.85% and F1-score = 97.32%

Results on SIPaKMeD Pap Smear Dataset:



The results obtained on different experiments on the SIPaKMeD Pap Smear datase. The best results on the dataset are obtained by merging features extracted from VGG-16 and ResNet-50: accuracy = 98.07%, precision = 98.86% and F1-score = 98.52%.

Accuracies on Training, Validation and Testing on the proposed method

Dataset	Feature Extraction method	Training Accuracy	Validation Accuracy	Testing Accuracy
HeRlev Pap Smear Dataset	ResNet-50	97.23%	96.41%	96.55%
	VGG16	95.92%	94.38%	95.37%
	IncepionV3	96.13%	95.49%	95.36%
	ResNet-50 + VGG16	**97.62%**	**98.03%**	**97.97%**
SiPaKMeD Pap Smear Dataset	ResNet-50	96.85%	96.77%	95.92%
	VGG16	94.26%	95.03%	94.78%
	IncepionV3	96.02%	95.91%	95.78%
	ResNet-50 + VGG16	**98.48%**	**97.42%**	**97.56%**

COMPARISON WITH EXISTING LITERATURE

Several models have been proposed in the literature for cervical cell classification as discussed in "Related Work". Our proposed work and the results achieved are therefore compared with some of these models that used the same datasets to assess the reliability of the proposed framework and the results are tabulated.

Dataset	Method	Results
HeRlev Pap Smear	Bora et al. [6]	Accuracy: 96.51%
	Win et al. [10]	Accuracy: 90.84%
	Chankong et al. [8]	Accuracy: 93.78%
	Proposed method	**Accuracy: 97.11%**
SiPaKMeD Pap Smear	Win et al. [10]	Accuracy: 94.09%
	Proposed method	**Accuracy: 98.07%**

CHAPTER- 6 FUTURE ENHANCEMENT AND CONCLUSION

INTRODUCTION

The machine learning model your team has developed presents a groundbreaking approach to Pap smear analysis. By incorporating techniques like PCA for data clarity, GWO for optimized training, and deep learning powerhouses like ResNet-50 and CNN for image analysis, the model has the potential to revolutionize cervical cancer screening. Imagine a future where Pap smears boast significantly improved accuracy, leading to fewer missed cases and reduced anxiety associated with inconclusive results. Additionally, faster analysis times through automation could streamline workflow in cytology labs. However, for this model to truly transform Pap smear analysis, further exploration is necessary. Future enhancements could focus on expanding the training data to encompass a wider range of cell presentations and demographics, ensuring generalizability. Additionally, research into improving the model's interpretability would allow healthcare professionals to better understand its reasoning and build trust in its results. By addressing these key areas, your team's machine learning model has the potential to become a powerful tool in the fight against cervical cancer.

LIMITATION/CONSTRAINTS OF THE SYSTEM

The potential of our machine learning model for Pap smear analysis is exciting, but it's important to acknowledge that, like any new technology, it will likely have limitations and constraints. Here are some areas to consider:

1. **Data Dependence**: The model's accuracy relies heavily on the quality and quantity of data used to train it. Biases or limitations in the training data could lead to biased or inaccurate results in real-world use.

2. **Interpretability:** While machine learning models can be powerful, understanding how they arrive at their conclusions can be challenging. This lack of interpretability might make it difficult for healthcare professionals to fully trust or explain the model's results.
3. **Generalizability**: The model might perform well on the specific data it was trained on, but its generalizability to different populations or variations in cell images needs to be evaluated.
4. **Regulatory Hurdles:** Machine learning models in healthcare often face regulatory hurdles before widespread adoption. Ensuring the model meets safety and efficacy standards will be crucial for its real-world implementation.
5. **Human Expertise Integration:** Even with advancements, human expertise in cytopathology will likely remain important. The ideal scenario might involve the model assisting cytologists by highlighting suspicious areas or providing a second opinion, rather than replacing them entirely.

Addressing these limitations will be crucial for ensuring the responsible and effective integration of our machine learning model into Pap smear analysis.

FUTURE ENHANCEMENTS

Some of the future enhancements that we could incorporate in our machine learning model for Pap smear analysis are as follows:

1. **Multimodal Data Integration**: Currently, the model likely focuses on analyzing Pap smear images. Consider incorporating additional data points like a patient's medical history, HPV test results, or demographic information. This multi-modal approach could provide a more holistic view and potentially improve the model's accuracy.
2. **Active Learning and Continuous Improvement**: Develop a feedback loop where the model can learn from real-world data. This "active learning" approach could involve incorporating data from positive Pap smears confirmed through biopsies. The model could then use this data to continuously refine its ability to identify precancerous cells.
3. **Explainable AI (XAI):** While machine learning excels at pattern recognition, improving the model's interpretability is crucial for gaining trust from healthcare professionals. Explore techniques like LIME (Local Interpretable Model-agnostic Explanations) to explain the model's reasoning behind its classifications.
4. **Standardization and Generalizability:** Focus on ensuring the model performs consistently across different laboratories and image acquisition systems. This might involve data normalization techniques or collaborations with cytology labs to ensure generalizability to real-world variations.
5. **Integration with Clinical Workflow**: Consider how the model can seamlessly integrate into existing clinical workflows. This could involve developing user-friendly interfaces for cytologists or designing the model to highlight suspicious regions on Pap smear images for further analysis.

CONCLUSION

The machine learning model we have developed for Pap smear analysis represents a significant leap forward in cervical cancer screening. By harnessing the power of PCA for data clarity, GWO for optimized training, and deep learning techniques like ResNet50 and CNN for image analysis, this model offers a glimpse into a future with enhanced accuracy, reduced human error, and potentially faster turnaround times.

However, for this model to truly transform Pap smear analysis, further exploration and development are necessary. One key area for future enhancements lies in expanding the training data. Currently, the model likely focuses on analyzing Pap smear images. Incorporating additional data points like a patient's medical history, HPV test results, or demographic information could provide a more holistic view and potentially improve the model's generalizability across diverse patient populations.

Another exciting avenue for exploration is "active learning." This approach would allow the model to learn and improve continuously from real-world data. Imagine a system where the model can be updated with data from positive Pap smears confirmed by biopsies. This feedback loop would allow the model to refine its ability to identify precancerous cells with even greater precision over time.

Building trust with healthcare professionals is crucial for successful integration of the model into clinical practice. Here, Explainable AI (XAI) techniques like LIME (Local Interpretable Model-agnostic Explanations) can be invaluable. By explaining the model's reasoning behind its classifications, XAI can bridge the gap between human intuition and machine learning's "black box" nature. Finally, ensuring the model performs consistently across different laboratories and image acquisition systems is paramount. Data normalization techniques or

collaborations with cytology labs can help achieve this goal. Additionally, consider how the model can seamlessly integrate into existing clinical workflows. Developing user-friendly interfaces for cytologists or designing the model to highlight suspicious regions on Pap smear images for further analysis could be valuable steps in this direction.

In conclusion, our machine learning model holds immense promise for revolutionizing Pap smear analysis and ultimately, cervical cancer prevention. By incorporating the future enhancements discussed, this model can evolve from a promising tool to a powerful real-world solution. Further research, coupled with close collaboration between data scientists, healthcare professionals, and regulatory bodies, will be instrumental in paving the way for its successful implementation and ultimately, saving lives.

REFERENCES

1. Akter L, Islam MM, Al-Rakhami MS, Haque MR, et al. Predic tion of cervical cancer from behavior risk using machine learning techniques. SN Comput Sci. 2021;2(3):1–10.
2. AlMubarak HA, Stanley J, Guo P, Long R, Antani S, Thoma G, Zuna R, Frazier S, Stoecker W. A hybrid deep learning and handcrafted feature approach for cervical cancer digital his tology image classification. Int J Healthc Inf Syst Inform. 2019;14(2):66–87.
3. Azaza A, Abdellaoui M, Douik A. Off-the-shelf deep features for saliency detection. SN Comput Sci. 2021;2(2):1–10.
4. Basak H, Kundu R. Comparative study of maturation profiles of neural cells in different species with the help of computer vision and deep learning. In: International symposium on signal pro cessing and intelligent recognition systems. Springer; 2020. p. 352–66.
5. Basak H, Kundu R, Agarwal A, Giri S. Single image super-resolu tion using residual channel attention network. In: 2020 IEEE 15th international conference on industrial and information systems (ICIIS). IEEE; (2020). p. 219–24.
6. Bora K, Chowdhury M, Mahanta LB, Kundu MK, Das AK. Auto mated classification of pap smear images to detect cervical dys plasia. Comput Methods Programs Biomed. 2017;138:31–47.
7. Byriel J. Neuro-fuzzy classification of cells in cervical smears. Master's Thesis, Technical University of Denmark: Oersted-DTU, Automation. 1999.
8. Chankong T, Theera-Umpon N, Auephanwiriyakul S. Automatic cervical cell segmentation and classification in Pap smears. Com put Methods Programs Biomed. 2014;113(2):539–56.
9. Chattopadhyay S, Basak H. Multi-scale attention U-Net (MsAU NeT): a modified U-Net architecture for scene segmentation. 2020. arXiv:200906911. 10. De Jong KA. Analysis of the behaviour of a class of genetic adaptive systems. Technical report. 1975.
10. Win KP, Kitjaidure Y, Hamamoto K, Myo Aung T. Computer assisted screening for cervical cancer using digital image processing of Pap smear images. Appl Sci. 2020;10(5):1800.