CERVICAL CANCER DETECTION USING MACHINE LEARNING MODEL

A PROJECT REPORT

Submitted by

Vadnagara Abhishek(19BECE30004)

Shah Ketul(19BECE30094)

Solanki Shubham(19BECE30185)

Prajapati Yash(19BECE30212)

Submitted to

Prof.Barkha Bhavsar

In fulfillment for the award of the degree

of

BACHELOR OF ENGINEERING

in

Computer Engineering

LDRP Institute of Technology and Research, Gandhinagar

Kadi Sarva Vishwavidyalaya

2022-23

LDRP INSTITUTE OF TECHNOLOGY AND RESEARCH, GANDHINAGAR

CE Department

CERTIFICATE

This is to certify that the Project Work entitled **"Cervical Cancer Detection Using Machine Learning Model"** has been carried out by **VADNAGARA ABHISHEK (19BECE30004)** under my guidance in fulfilment of the degree of Bachelor of Engineering in Computer Engineering Semester-7 of Kadi Sarva Vishwavidyalaya University during the academic year 2022-23.

Prof.Barkha Bhavsar	Dr.Sandip Modha
Internal Guide	**Head of the Department**
LDRP ITR	**LDRP ITR**

CE Department

CERTIFICATE

This is to certify that the Project Work entitled **"Cervical Cancer Detection Using Machine Learning Model"** has been carried out by **SHAH KETUL (19BECE30094)** under my guidance in fulfilment of the degree of Bachelor of Engineering in Computer Engineering Semester-7 of Kadi Sarva Vishwavidyalaya University during the academic year 2022-23.

Prof.Barkha Bhavsar	Dr.Sandip Modha
Internal Guide	**Head of the Department**
LDRP ITR	**LDRP ITR**

CE Department

CERTIFICATE

This is to certify that the Project Work entitled **"Cervical Cancer Detection Using Machine Learning Model"** has been carried out by **SOLANKI SHUBHAM (19BECE30185)** under my guidance in fulfilment of the degree of Bachelor of Engineering in Computer Engineering Semester-7 of Kadi Sarva Vishwavidyalaya University during the academic year 2022-23.

Prof.Barkha Bhavsar	Dr.Sandip Modha
Internal Guide	**Head of the Department**
LDRP ITR	**LDRP ITR**

CE Department

CERTIFICATE

This is to certify that the Project Work entitled **"Cervical Cancer Detection Using Machine Learning Model"** has been carried out by **PRAJAPATI YASH (19BECE30212)** under my guidance in fulfilment of the degree of Bachelor of Engineering in Computer Engineering Semester-7 of Kadi Sarva Vishwavidyalaya University during the academic year 2022-23.

Prof.Barkha Bhavsar	Dr.Sandip Modha
Internal Guide	**Head of the Department**
LDRP ITR	**LDRP ITR**

ACKNOWLEDGEMENT

We take this opportunity to express our profound gratitude and deep regards to our guide for her exemplary guidance, monitoring and constant encouragement throughout the course of this thesis. The blessing, help and guidance given by her time to time shall carry us a long way in the journey of life on which we are about to embark. We also take this opportunity to express a deep sense of gratitude to University for cordial support, valuable information and guidance, which helped us in completing this task through various stages.

Abhishek Vadnagara(19BECE30004)
Ketul Shah(19BECE30094)
Shubham Solanki(19BECE30185)
Yash Prajapati(19BECE30212)

19BECE30004 | 19BECE30094
19BECE30185 | 19BECE30212

Abstract

Cancer is a leading cause of death and disability word-wide and there is a significant growth of cancer in developing countries. Nearly 70% of Indian population lives in rural India. However, nearly 95% of cancer care facilities are in urban India. Thus, the mortality rates are double in rural areas. In India, cervical cancer contributes to approximately 45-67% of all cancers in women. Basis of this stats we can consider the seriousness of this cancer. There are very few facilities of disseminating cancer awareness, special .trainer are required for manual judging of cervical symptoms they are also very few in number and allocated in most of known urban hospitals, Even the biopsies or the blood samples are sent to cities and the reports take weeks to come back it leads to advancement in diseases , because of all these reasons manual screening may not be feasible for wide-scale implementation. we are using convolutional neural network and concepts of deep learning as it learns meaningful kernels that stimulates the extraction of visual features such as edges ,size ,shape and colors in image classification A deep prediction model is built using CNN network to classify the various grades of cancer normal, mild ,moderate, severe and carcinoma We have taken dataset from pap smear cervical dataset as it is most effective image based cancer detection tool categorizing cervical cells as normal and abnormal and achieved acccracy 86%.

Table of Contents

List of Figueres

List of Table

1). Introduction

1.1 Introduction

1.2 Scope

1.3 Objective

1.4 Overview of Project

1.1 Introduction:-

Cancer of the uterine cervix is one of the most common gynecological cancers and is the leading cause of mortality and morbidity among women worldwide. A malignant tumor occurs when the cervix cells grow and replicate abnormally with uncontrolled cell division and cell death. According to the recent report received from Information Centre on HPV and Cancer ICO/ IARC, cervical cancer is the fourth most common and frequent cancer among women. Recent estimates of ICO/IARC, indicate that every year 527,624 women are diagnosed with cervical cancer, and 265,672 die from cancer. This is due to poor access to screening and treatment services especially for women living in low and middle-income countries (Bruni et al., 2015). Cervical cancer is a slow growing cancer that takes nearly 10 to 20 years to show its symptoms. Therefore, routine usage of Pap smear test in developing and underdeveloped countries helps in reducing mortality and morbidity rate of this cancer. It is completely preventable and curable, if detected and treated the pre-cancerous symptoms at the early stages (Takiar et al., 2010). Thus, automating the process helps the pathologists to analyze a large number of samples in a short period.Traditional machine learning techniques have helped massively in automating the diagnosis process, but they failed in dealing with large volume of data.

Deep learning algorithms (LeCun et al., 2015) and architectures mitigate the issues on big volume of data. The objective of this learning process is to learn a complex and abstract representation of data in a hierarchical manner by passing the data through multiple transformation layers. The convolution neural network (CNN) (Krizhevsky et al., 2012) is one such framework that performs exceptionally well for high dimensional data, as it learns the underlying complex function empirically, and shows better performance than the traditional machine learning algorithms. Hence, the CNN is used for classifying the cervical cells in cytological images by giving whole single cell image as input instead of using manually extracted features of nucleus and cytoplasm of the cervical cell images.

The CNN is used for classifying the cervical cells of cytological images. This deep neural network convolves the learned features from the input images and uses 2D convolution kernels to extract abstract and complex features. The CNN eliminates manual feature extraction by making the network to learn meaningful features from the input data repeatedly, when they train on a collection of labeled images. This automated feature extraction makes the deep learning model highly accurate for image analysis task, but also requires a large dataset to work. We used a new Pap smear database developed by Herlev University Hospital for this work. The single cervical cell images of Herlev Dataset were collected using a digital camera and microscope. The Pap smear database consists of 917 images distributed unequally on seven different classes.

1.2 Scope:-

- Cervical cancer continues to be one of the deadliest cancers among women worldwide, especially with it being the most common cause of death in developing countries. Every year, approximately 500,000 new cases are reported, out of which 85% occur in developing countries, along with approximately 270,000 deaths worldwide.
- For screening, traditional PAP-smear test continues to be prevalent, especially in the developing countries . However, traditional techniques have many drawbacks in terms of being tedious, time-consuming and expensive. That's Why we propose a Automated Pap Smear Cervical Screening based on convolutional neural networks (CNNS)which is the concept of machine learning.
- Hence, the CNN is used for classifying the cervical cells in cytological images by giving whole single cell image as input instead of using manually extracted features of nucleus and cytoplasm of the cervical cell images. after giving input CNN model will generate output from seven class .

1.3 Objective:-

- Successful detection and classification of cervical cell in the cervix using automated pap smear cervical cell screening
- Provides results with acceptable accuracy, sensitivity and specificity
- Most of the researchers emphasize on single cell cervical images
- We shall try to work on both single cell and multi cell images

1.4 Overview of Project:-

- Due to the variable nature of cancer cells, human prediction and classification can be erroneous that tend to improper medication.
- Hence objective of our project is to come up with an image processing and machine learning technique that can detect and classify cervical cancer cell more accurately.
- In this project, for training purpose We have taken large dataset from pap smear cervical dataset as it is most effective image-based cancer detection tool categorizing cervical cells as normal and abnormal.
- In this system we are focusing on Accuracy, Sensitivity, Specificity parameters.

2). Technology and Literature Review

2.1 About Tools and Technology

2.2 Literature Review

2.1 About Tools and Technology:-

2.1.1 PyCharm:-

PyCharm is an integrated development environment (IDE) used in computer programming, specifically for the Python programming language. It is developed by the Czech company JetBrains (formerly known as IntelliJ). It provides code analysis, a graphical debugger, an integrated unit tester, integration with version control systems (VCSes), and supports web development with Django as well as data science with Anaconda.

PyCharm is cross-platform, with Windows, macOS and Linux versions. The Community Edition is released under the Apache License, and there is also Professional Edition with extra features – released under a subscription-funded proprietary license and also an educational version.

Features:-

- Coding assistance and analysis, with code completion, syntax and error highlighting, linter integration, and quick fixes
- Project and code navigation: specialized project views, file structure views and quick jumping between files, classes, methods and usages
- Python refactoring: includes rename, extract method, introduce variable, introduce constant, pull up, push down and others
- Support for web frameworks: Django, web2py and Flask [professional edition only]
- Integrated Python debugger
- Integrated unit testing, with line-by-line code coverage
- Google App Engine Python development [professional edition only]
- Version control integration: unified user interface for Mercurial, Git, Subversion, Perforce and CVS with change lists and merge
- Support for scientific tools like matplotlib, numpy and scipy [professional edition only]

2.1.2 Python 3.8.10 or above versions(64-bit):-

According to the release calendar specified in PEP 569, Python 3.8.10 is the final regular maintenance release. Starting now, the 3.8 branch will only accept security fixes and releases of those will be made in source-only form until October 2024.

Compared to the 3.7 series, this last regular bugfix release is relatively dormant at 92 commits since 3.8.9. Version 3.7.8, the final regular bugfix release of Python 3.7, included 187 commits. But there's a bunch of important updates here regardless, the biggest being Big Sur and Apple Silicon build support. This work would not have been possible without the effort of Ronald Oussoren, Ned Deily, Maxime Bélanger, and Lawrence D'Anna from Apple.
Major new features of the 3.8 series, compared to 3.7:-

- PEP 572, Assignment expressions
- PEP 570, Positional-only arguments

- PEP 587, Python Initialization Configuration (improved embedding)
- PEP 590, Vectorcall: a fast calling protocol for CPython
- PEP 578, Runtime audit hooks
- PEP 574, Pickle protocol 5 with out-of-band data
- Typing-related: PEP 591 (Final qualifier), PEP 586 (Literal types), and PEP 589 (TypedDict).
- Parallel filesystem cache for compiled bytecode
- Debug builds share ABI as release builds
- f-strings support a handy = specifier for debugging
- continue is now legal in finally: blocks
- on Windows, the default asyncio event loop is now ProactorEventLoop
- on macOS, the spawn start method is now used by default in multiprocessing
- multiprocessing can now use shared memory segments to avoid pickling costs between processes
- typed_ast is merged back to CPython
- LOAD_GLOBAL is now 40% faster
- pickle now uses Protocol 4 by default, improving performance.

2.1.3 OpenCV Library:-

OpenCV is the huge open-source library for the computer vision, machine learning, and image processing and now it plays a major role in real-time operation which is very important in today's systems. By using it, one can process images and videos to identify objects, faces, or even handwriting of a human. When it integrated with various libraries, such as NumPy, python is capable of processing the OpenCV array structure for analysis. To Identify image pattern and its various features we use vector space and perform mathematical operations on these features.

Applications of OpenCV: There are lots of applications which are solved using OpenCV,:-

- face recognition
- Automated inspection and surveillance
- number of people – count (foot traffic in a mall, etc)
- Vehicle counting on highways along with their speeds
- Interactive art installations
- Anamoly (defect) detection in the manufacturing process (the odd defective products)
- Street view image stitching
- Video/image search and retrieval
- Robot and driver-less car navigation and control
- object recognition
- Medical image analysis
- Movies – 3D structure from motion
- TV Channels advertisement recognition

OpenCV-Functionality:-

- Image/video I/O, processing, display (core, imgproc, highgui)
- Object/feature detection (objdetect, features2d, nonfree)
- Geometry-based monocular or stereo computer vision (calib3d, stitching, videostab)
- Computational photography (photo, video, superres)
- Machine learning & clustering (ml, flann)
- CUDA acceleration (gpu)

2.1.4 Keras:-

- Simple - but not simplistic. Keras reduces developer cognitive load to free you to focus on the parts of the problem that really matter.
- Flexible - Keras adopts the principle of progressive disclosure of complexity: simple workflows should be quick and easy, while arbitrarily advanced workflows should be possible via a clear path that builds upon what you've already learned.
- Powerful - Keras provides industry-strength performance and scalability: it is used by organizations and companies including NASA, YouTube, or Waymo.

2.2 Literature Review:-

2.2.1 Deep Convolution Neural Network for Malignancy Detection and Classification in Microscopic Uterine Cervix Cell Images:-

Objective:- Automated Pap smear cervical screening is one of the most effective imaging based cancer detection tools used for categorizing cervical cell images as normal and abnormal. Traditional classification methods depend on hand-engineered features and show limitations in large, diverse datasets. Effective feature extraction requires an efficient image preprocessing and segmentation, which remains prominent challenge in the field of Pathology. In this paper, a deep learning concept is used for cell image classification in large datasets.[1]

Methods:- This relatively proposed novel method, combines abstract and complicated representations of data acquired in a hierarchical architecture. Convolution Neural Network (CNN) learns meaningful kernels that simulate the extraction of visual features such as edges, size, shape and colors in image classification. A deep prediction model is built using such a CNN network to classify the various grades of cancer: normal, mild, moderate, severe and carcinoma. It is an effective computational model which uses multiple processing layers to learn complex features. A large dataset is prepared for this study by systematically augmenting the images in Herlev dataset.[1]

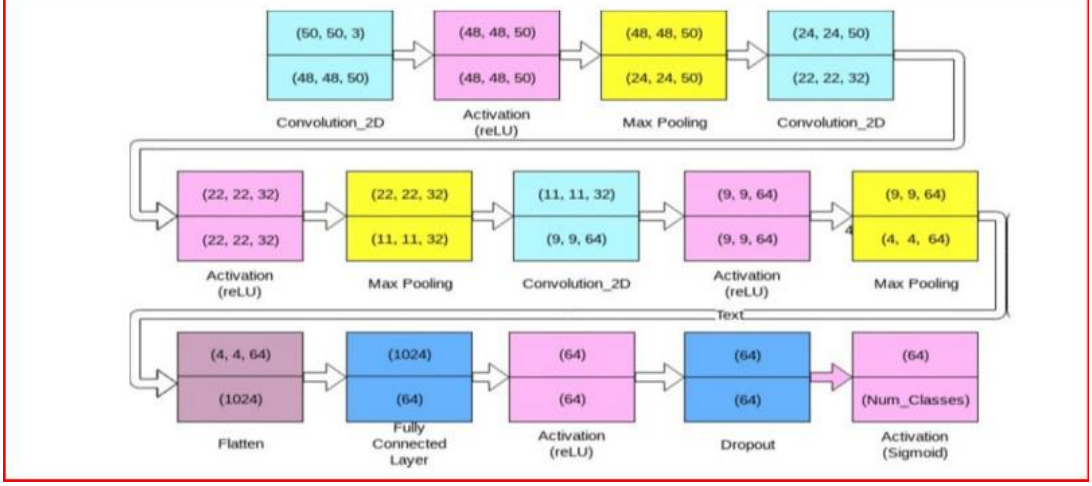


Fig.1 Architecture of Convolutional Neural Network Model Used for the Classification of Cervical Cell Images[1]

Result:- Among the three sets considered for the study, the first set of single cell enhanced original images achieved an accuracy of 94.1% for 5 class, 96.2% for 4 class, 94.8% for 3 class and 95.7% for 2 class problems. The second set includes contour extracted images showed an accuracy of 92.14%, 92.9%, 94.7% and 89.9% for 5, 4, 3 and 2 class problems. The third set of binary images showed 85.07% for 5 class, 84% for 4 class, 92.07% for 3 class and highest accuracy of 99.97% for 2 class problems.[1]

Conclusion:- The experimental results of the proposed model showed an effective classification of different grades of cancer in cervical cell images, exhibiting the extensive potential of deep learning in Pap smear cell image classification.[1]

2.2.2 Considerations for a PAP Smear Image Analysis System with CNN Features:-

It has been shown that for automated PAP-smear image classification, nucleus features can be very informative. Therefore, the primary step for automated screening can be cell-nuclei detection followed by segmentation of nuclei in the resulting single cell PAP-smear images. We propose a patch based approach using CNN for segmentation of nuclei in single cell images. We then pose the question of ion of segmentation for classification using representation learning with CNN, and whether low-level CNN features may be useful for classification. We suggest a CNN-based feature level analysis and a transfer learning based approach for classification using both segmented as well full single cell images. We also propose a decision-tree based approach for classification. Experimental results demonstrate the effectiveness of the proposed algorithms individually (with low-level CNN features), and simultaneously proving the sufficiency of cell-nuclei detection (rather than accurate segmentation) for classification. Thus, we propose a system for analysis of multi-cell PAP-smear images consisting of a simple nuclei detection algorithm followed by classification using transfer learning.[2]

Thus, noting the above aspects, for an overall system development, we propose algorithms for 1) Detection of nuclei in multi-cell images in single cell images, 2) Segmentation of nuclei in single cell images, 3) Classification strategies considering both accurate nuclei segmentation, and nuclei detection (involving some cellular background pixels), and also considering CNN features from different layers.[2]

2.2.3 Automation of Detection of Cervical Cancer Using Convolutional Neural Networks:-

Classification of digital cervical images acquired during visual inspection with acetic acid (VIA) is an important step in automated image-based cervical cancer detection. Many algorithms have been developed for classification of cervical images based on extracting mathematical features and classifying these images. Deciding the suitability of a feature and learning the algorithm is a complex task. On the other hand, convolutional neural networks (CNNs) self-learn most suitable hierarchical features from the raw input image. In this paper, we demonstrate the feasibility of using a shallow layer CNN for classification of image patches of cervical images as cancerous or not cancerous. We used cervix images acquired after the

application of 3%–5% acetic acid using an Android device in 102 women. Of these, 42 cervix images belonged in the VIA-positive category (pathologic) and 60 in the VIA-negative category (healthy controls). A total of 275 image patches of 15 × 15 pixels were manually extracted from VIA-positive areas, and we considered these patches as positive examples. Similarly, 409 image patches were extracted from VIA-negative areas and were labeled as VIA negative. These image patches were classified using a shallow layer CNN composed of a layer each of convolutional, rectified linear unit, pooling, and two fully connected layers. A classification accuracy of 100% is achieved using shallow CNN.[3]

2.2.4 A Novel Automation-assisted Cervical Cancer reading method based on Convolutional Neural Network:-

Cervical cytology screening using Pap smear or liquid-based cytology is one of the most Q2 widely followed and accepted method. Automation-assisted screening based on cervical cytology has become a necessity due to the manual screening method operated by a visual analysis for cervical cell specimen under the microscope of the glass slide is usually labor-intensive and time-consuming. While automation-assisted reading system can improve efficiency, their performance often relies on the success of accurate cell segmentation and hand-craft feature extraction. This paper presents an efficient and totally segmentation-free method for automated cervical cell screening that utilizes modern object detector to directly detect cervical cells or clumps, without the design of specific hand-crafted feature. Specifically, we use the state-of-the-art CNN-based object detection methods, YOLOv3, as our baseline model. In order to improve the classification performance of hard examples which are four highly similar categories, we cascade an additional task-specific classifier. We also investigate the presence of unreliable annotations and coped with them by smoothing the distribution of noisy labels. We comprehensively evaluate our methods on our test set which is consisted of 1014 annotated cervical cell images with size of 4000*3000 and complex cellular situation corresponding to 10 categories. Our model achieves 97.5% sensitivity (Sens) and 67.8% specificity (Spec) on cervical cell image-level screening. Moreover, we obtain a best mean average precision (mAP) of 63.4% on cervical cell-level diagnosis, and improve the average precision (AP) of hard examples which are the most valuable but most difficult to distinguish. Our automation-assisted cervical cell reading system not only achieves cervical cell image-level classification but also provides more detailed location and category reference information of abnormal cells. The results indicate feasible performance of our method, together with the efficiency and robustness, providing a new idea for future development of computer-assisted reading systems in clinical cervical screening.[4]

2.2.5 Conclusion:-

AUTHOR	PAPER	IMAGE ACQUISITION	PRE-PROCESSING	FEATURE SELECTION	CLASSIFICATION	LIMITATION
Shanthi P at el [1]	Deep Convolution Neural Network for Malignancy Detection and Classification on in Microscopic Uterine Cervix Cell Images	Pap smear single cell dataset	Bi-Histogram Equalization with adaptive sigmoid function method combined with sobel horizontal, vertical maskContour extractionBinarization	Convolution Neural Networ	Convolution Neural Network	The binarized shows poor performance for 5, 4 and 3 class problems
Srishti Gautam at el[2]	Considerations for a PAP Smear Image Analysis System with CNN Features	Pap smear single cell dataset(Herlev dataset) (Multi cell dataset)Aindra dataset	Median filter, contrasat-limited adptive histogram equalization(CLAHE),Global threshold	Convolution Neural Network	Decision-Tree classifier	-

AUTHOR	PAPER	IMAGE ACQUISITION	PRE-PROCESSING	FEATURE SELECTION	CLASSIFICATION	LIMITATION
Vidya Kudva at el[3]	Automation of Detection of Cervical Cancer Using Convolutional Neural Networks	Cervix images captured by Android device. VIA-positive category:42 images VIA-negative category:60 images Total of 275 image patches were from VIA-positive areas and labeled as positive, Similarly, 409 image patches from VIA-negative areas and were labeled as VIA negative	-	Random subset feature selection (RSFS) method was used to select relevant features.	Convolution Neural Network	-
Yao Xiang[4]	A novel automation-assisted cervical cancer reading method based on convolutional neural network	our own dataset of multicell images captured by digital camera, The dataset used in this paper is consisted of 12,909 cervical images with 58,995 ground truth boxes and contains 10 categories.	-	Darknet-53, which has 53 layers network trained on ImageNet to extract features	Yolo detector, InceptionV3	-

3). System Requirements Study

3.1 User Characteristics

3.2 Hardware and Software Requirements

3.3 Constraints

3.4 Assumptions and Dependencies

3.1 User Characteristics:-

Analysis user characteristics is an important aspect of any project. It allows us to clearly define and focus on who the end users are for the project. Also, it allows checking the progress of the project to ensure that we are still developing the system for the end users. The user must have following characteristics:

- User should understand the use of all modules.
- User must have basic knowledge of Computers.
- User can easily interact with the proposed system.
- User must know the technical terms used in the company for performing different tasks specially related to call logs, payment details, transportation details and report retrieval.
- User should be also being aware about the running process of the system.

3.2 Hardware and Software Requirements:-

3.2.1 Hardware Requirement:-

The following describes the hardware needed in order to execute and develop the Virtual Mouse application:

- Computer Desktop or Laptop:-
  The computer desktop or a laptop will be utilized to run the visual software in order to display what webcam had captured. A notebook which is a small, lightweight and inexpensive laptop computer is proposed to increase mobility.
- Webcam:-
  Webcam is utilized for image processing, the webcam will continuously taking image in order for the program to process the image and find pixel position
  System will be using:-

Name Of Components	Specification
Processor	Pentium III 630MHz, CORE i3, CORE i5 etc.
RAM	128 MB , 8 GB, 4GB, 6GB etc
Hard Disk	20 GB, 1TB.
Monitor	15” color monitor
Keyboard	122 keys

3.2.2 Software Requirement:-

Software and Hardware Requirements are used to describe the minimum hardware and software requirements to run the Software. These requirements are described below.

- Open CV Library:-
  OpenCV are also included in the making of this program. OpenCV (Open Source Computer Vision) is a library of programming functions for real time computer vision. OpenCV have the utility that can read image pixels value, it also have the ability to create real time eye tracking and blink detection.

- Python:-
  Python is an interpreted high-level general-purpose programming language. Its design philosophy emphasizes code readability with its use of significant indentation. Its language constructs as well as its object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects.
  Python is dynamically-typed and garbage-collected. It supports multiple programming paradigms, including structured (particularly, procedural), object-oriented and functional programming. It is often described as a "batteries included" language due to its comprehensive standard library.

- Keras:-
  Simple but not simplistic. Keras reduces developer cognitive load to free you to focus on the parts of the problem that really matter.
  Flexible-Keras adopts the principle of progressive disclosure of complexity: simple workflows should be quick and easy, while arbitrarily advanced workflows should be possible via a clear path that builds upon what you've already learned.
  Powerful Keras provides industry-strength performance and scalability: it is used by organizations and companies including NASA, YouTube, or Waymo.

- Tkinter Programming:-
  Tkinter is the standard GUI library for Python. Python when combined with Tkinter provides a fast and easy way to create GUI applications. Tkinter provides a powerful object-oriented interface to the Tk GUI toolkit.
  Creating a GUI application using Tkinter is an easy task. All you need to do is perform the following steps −
    - Import the Tkinter module.
    - Create the GUI application main window.
    - Add one or more of the above-mentioned widgets to the GUI application.
    - Enter the main event loop to take action against each event triggered by the user.

Software will be using:

Name Of Components	Specification
Operating System	Windows or Linux, Window 7 or above versions Ultimate 64-bit
Language	Python
Libraries	OpenCV,Tensorflow,Tkinter,keras

3.3 Constraints:-

3.3.1 Parallel Operations

The project is on basis of multi-user. This is used for carrying out updating as well as entry by preventing the redundancy of the data.

3.3.2 Reliability Requirements

Reliability requirements of the system are one of the prime ones in the list. The system is needed to be highly reliable in terms of performance and capable of delivering robust performance. If the reports are generated within 5 seconds then the system is said to be reliable.

3.3.3 Criticality of the Application

The system can stop working on computers with very low internet connection. Other than that there won't be any issues. Apart from these the system should be able to make updates at regular time intervals.

3.3.4 Safety and Security Consideration

Safety and security too are other major concerns of any system. It is necessary to provide safety and security as the system is some kind of application and might be intrude by security threats from the internet. Thus, the code needs to be encrypted and any transaction needs to be done securely.

3.3.5 Hardware Limitations

Hardware Limitations are other constraint of the system. Hardware Limitations should be overcome for better performance of the system. This can be achieved by using minimum and only necessary Hardware.

3.4 Assumptions and Dependencies:-

3.4.1 Assumptions:

- User is the person having enough knowledge for the traversing operation.
- The Libraries used for data operating is always secured.
- Libraries transactions are assumed to be secure and reliable.
- We will provide a user friendly interface so that any user can easily navigate through the system.

3.4.2 Dependencies:

- All the users of the system will be assigned a specific role. According to these roles each and every user will be allowed to access predefined set of feature.
- The system is dependent upon the user's microscopic photo. If microscopic photo is not properly visible then system will not perform properly it may be given wrong prediction.

4). System Analysis

4.1 Study of Current System

4.2 Problem and Weaknesses of Current System

4.3 Analysis of New System

4.4 Activity Diagram

4.1 Study of Current System:-

Automated Pap smear cervical screening is one of the most effective imaging based cancer detection tools used for categorizing cervical cell images as normal and abnormal. Traditional classification methods depend on hand-engineered features and show limitations in large, diverse datasets. Effective feature extraction requires an efficient image preprocessing and segmentation, which remains prominent challenge in the field of Pathology. In this paper, a deep learning concept is used for cell image classification in large datasets.

This relatively proposed novel method, combines abstract and complicated representations of data acquired in a hierarchical architecture. Convolution Neural Network (CNN) learns meaningful kernels that simulate the extraction of visual features such as edges, size, shape and colors in image classification. A deep prediction model is built using such a CNN network to classify the various grades of cancer: normal, mild, moderate, severe and carcinoma. It is an effective computational model which uses multiple processing layers to learn complex features. A large dataset is prepared for this study by systematically augmenting the images in Herlev dataset.

The experimental results of the proposed model showed an effective classification of different grades of cancer in cervical cell images, exhibiting the extensive potential of deep learning in Pap smear cell image classification.

4.2 Problem and Weaknesses of Current System:-

- For lack of dataset the system is not able to detect non cervical cancerous cell.
- The system only able to classify cervical cancer cell into 7 classes.
- The system only able to classify single cervical cancer cell into 7 classes.

4.3 Analysis of New System:-

4.3.1 Image Aquisition:-

For training purpose we have taken large dataset from herlev pap smear cervical dataset. The total images are 916 and are classified into 7 classes i.e. normal columnar, normal intermediate, normal superficial, mild dysplasia, moderate dysplasia, severe dysplasia, carcinoma in situ.

4.3.2 Image Pre-processing:-

The captured image is first converted into gray image. After that gray image has been equalized using histogram equalization. This equalized image is again converted into colour image as CNN focuses on colour.

4.3.3 Feature Extraction and Selection:-

Feature extraction and selection has been done by CNN layers as it focuses on shape, size and colour.

4.3.4 Classification:-

Finally, CNN classifies cervical cancer cell into 7 classes.

4.4 Activity Diagram:-

Activity diagram is another important diagram in UML to describe the dynamic aspects of the system.

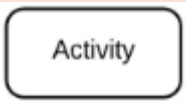
Activity diagram is basically a flowchart to represent the flow from one activity to anotheractivity. The activity can be described as an operation of the system.

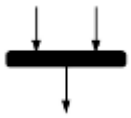
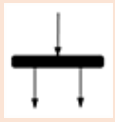
The control flow is drawn from one operation to another. This flow can be sequential, branched, or concurrent. Activity diagrams deal with all type of flow control by using different elements such as fork, join, etc.

The purpose of an activity diagram can be described as −

- Draw the activity flow of a system.
- Describe the sequence from one activity to another.
- Describe the parallel, branched and concurrent flow of the system.

Activity Diagram Symbols:-

Symbol	Name	Description
	Start symbol	Represents the beginning of a process or workflow in an activity diagram. It can be used by itself or with a note symbol that explains the starting point.
	Activity symbol	Indicates the activities that make up a modelled process. These symbols, which include short descriptions within the shape, are the main building blocks of an activity diagram.
	Connector symbol	Shows the directional flow, or control flow, of the activity.An incoming arrow starts a step of an activity; once the step is completed, the flow continues with the outgoing arrow.

	Joint symbol/ Synchronization bar	Combines two concurrent activities and re-introduces them to a flow where only one activity occurs at a time. Represented with a thick vertical or horizontal line.
	Fork symbol	Splits a single activity flow into two concurrent activities. Symbolized with multiple arrowedlines from a join.
	Decision symbol	Represents a decision and always has at least two paths branching out with condition text to allow users to view options. This symbol represents the branching or merging of various flows with the symbol acting as a frame or container.
	Note symbol	Allows the diagram creators or collaborators to communicate additional messages that don't fit within the diagram itself. Leave notes for added clarity and specification.
	Send signal symbol	Indicates that a signal is being sent to a receiving activity.
	Receive signal symbol	Demonstrates the acceptance of an event. After the event is received, the flow that comes from this action is completed.
	Shallowhistory pseudostate symbol	Represents a transition that invokes the last active state.
	Option loop symbol	Allows the creator to model a repetitive sequence within the option loop symbol.
	Flow final symbol	Represents the end of a specific process flow. This symbol shouldn’t represent the end of all flows in an activity; in that instance, you would use the end symbol.The flow final symbol should be placed at the end of a process in a single activity flow.
[Condition]	Condition text	Placed next to a decision marker to let you know under what condition an activity flow should split off in that direction.

	End symbol	Marks the end state of an activity and represents the completion of all flows of a process.

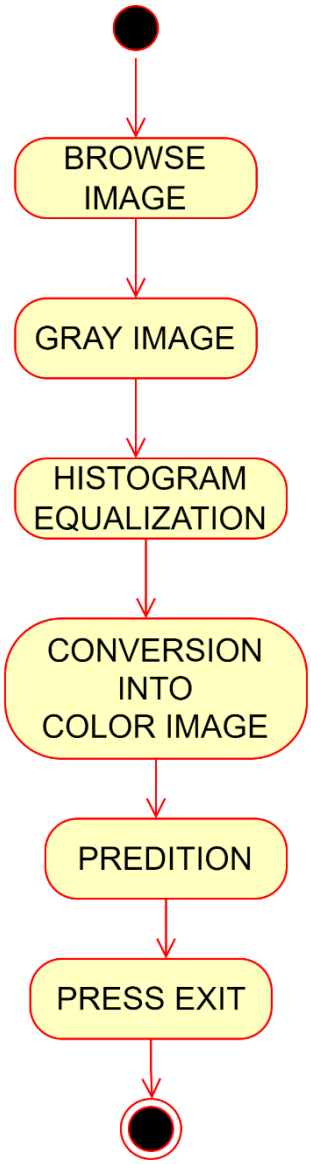


Fig.2 Activity Diagram of System

5). System Design

5.1 System Model

5.2 Flow Chart

5.1 System Model:-

For this project we'll be using the Spiral Model methodology approach in developing the application **Spiral model** is one of the most important Software Development Life Cycle models, which provides support for **Risk Handling**. In its diagrammatic representation, it looks like a spiral with many loops. The exact number of loops of the spiral is unknown and can vary from project to project. Each loop of the spiral is called a **Phase of the software development process.** The exact number of phases needed to develop the product can be varied by the project manager depending upon the project risks. As the project manager dynamically determines the number of phases, so the project manager has an important role to develop a product using the spiral model.

The Radius of the spiral at any point represents the expenses(cost) of the project so far, and the angular dimension represents the progress made so far in the current phase.

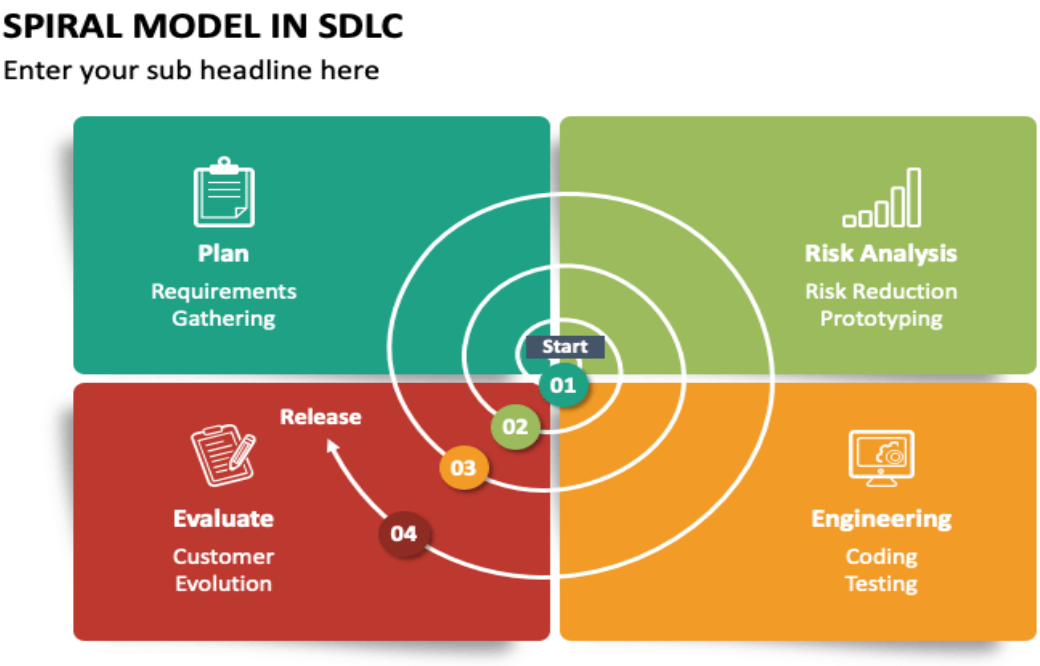


Fig.3 Spiral Model

Each phase of the Spiral Model is divided into four quadrants as shown in the above figure.

The Spiral model is called a Meta-Model because it subsumes all the other SDLC models. For example, a single loop spiral actually represents the Iterative Waterfall Model. The spiral model incorporates the stepwise approach of the Classical Waterfall Model. The spiral model uses the approach of the Prototyping Model by building a prototype at the start of each phase as a risk-handling technique. Also, the spiral model can be considered as supporting the Evolutionary model – the iterations along the spiral can be considered as evolutionary levels through which the complete system is built.

Advantages of Spiral Model:

Below are some advantages of the Spiral Model.

1. Risk Handling: The projects with many unknown risks that occur as the development proceeds, in that case, Spiral Model is the best development model to follow due to the risk analysis and risk handling at every phase.
2. Good for large projects: It is recommended to use the Spiral Model in large and complex projects.
3. Flexibility in Requirements: Change requests in the Requirements at later phase can be incorporated accurately by using this model.
4. Customer Satisfaction: Customer can see the development of the product at the early phase of the software development and thus, they habituated with the system by using it before completion of the total product.

Disadvantages of Spiral Model:

Below are some main disadvantages of the spiral model.

1. Complex: The Spiral Model is much more complex than other SDLC models.
2. Expensive: Spiral Model is not suitable for small projects as it is expensive.
3. Too much dependability on Risk Analysis: The successful completion of the project is very much dependent on Risk Analysis. Without very highly experienced experts, it is going to be a failure to develop a project using this model.
4. Difficulty in time management: As the number of phases is unknown at the start of the project, so time estimation is very difficult.

5.2 Flow chart:-

Flowchart – These flowcharts show physical control or resource level control. These flowcharts show data flow, but unlike data flowcharts, it also shows how decisions can control events. It shows data flow through major components like data entry, programs, storage media, and processes.

Symbols:-

A system flowchart is mainly used for physical modeling of the system showing inputs, outputs, and processes. Analysts frequently use them to illustrate the whole system graphically.

There are multiple symbols used in a system flowchart. All symbols are unique and represent a different process, input, output, or media item.

A system flowchart begins and ends with an oval symbol. This is also called the 'terminator' and indicates the start and end of the processes mentioned.

A rectangle then represents a process or an activity.

If any data is input into the computer, then a parallelogram is used. The data input can be in many forms, including keyboard, microphone, mouse, scanner, and multiple other kinds of sensors.

Then there are multiple types of output like printers, display devices, or speakers.

One of the major symbols used in system flowcharts is the decision boxes represented by a diamond shape. They represent the decisions made, and all arrows coming out of these decision boxes show alternate pathways for data.

Lastly, all these shapes are connected through arrows showing the direction of the flow and how data moves along the entire system.

However, just knowing the flowchart symbols is simply not enough for understanding a system flowchart. To understand it, you need to trace the pathway from one symbol to another. If you land on a decision symbol showing two alternate paths, you need to trace both of them to see where they end.

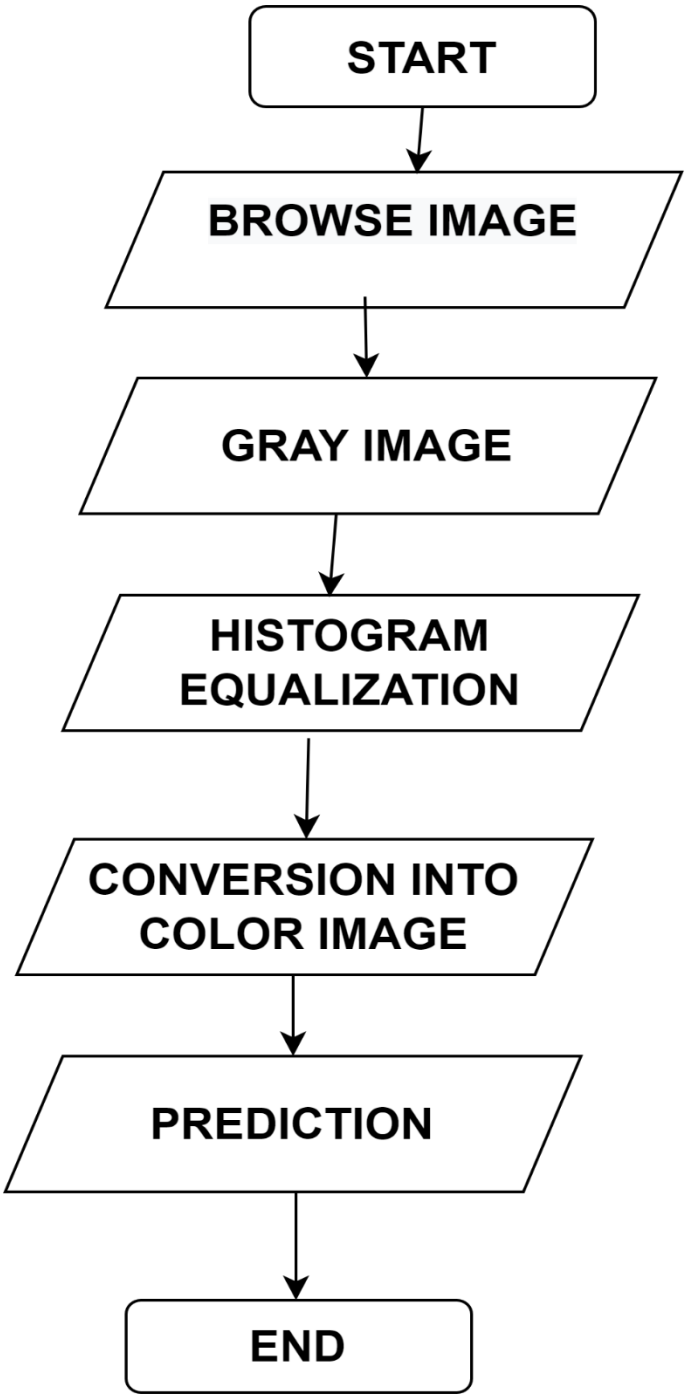


Fig.4 Flow Chart of System

6). System Testing

6.1 Test Report

6.2 Testing Planning Step

6.3 Testing Strategies

6.1 Test Report:-

Test Report is a document which contains a summary of all test activities and final test results of a testing project. Test report is an assessment of how well the Testing is performed. Based on the test report, stakeholders can evaluate the quality of the tested product and make a decision on the software release.

For example, if the test report informs that there are many defects remaining in the product, stakeholders can delay the release until all the defects are fixed.

- What does a test report contain?

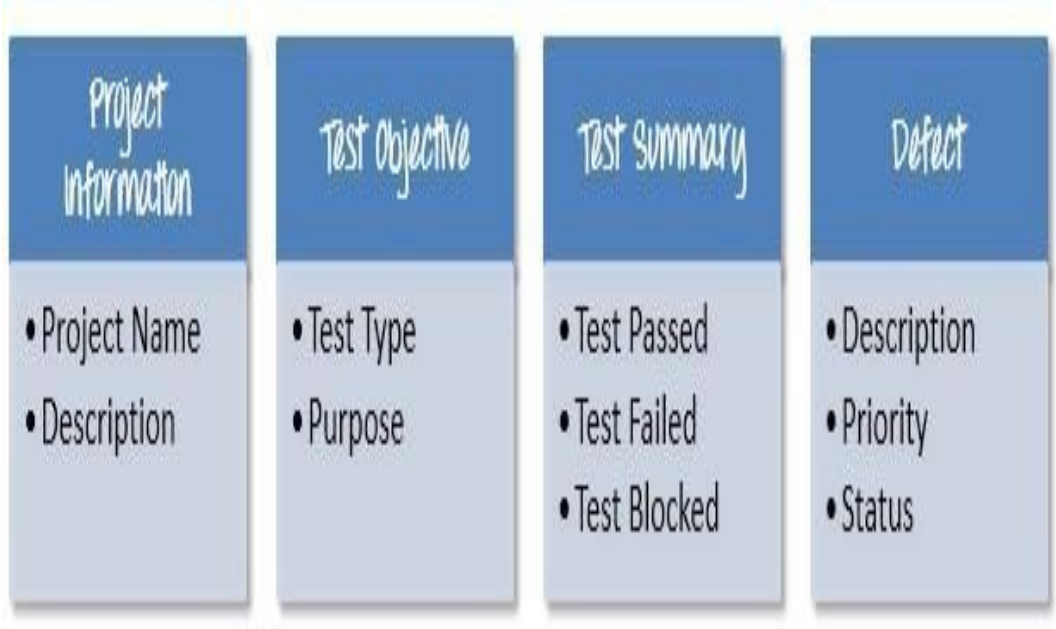


Fig. 5 test report diagram

Project Information:

All information of the project such as the project name, product name, and version should be described in the test report.

Test Objective:
As mentioned in Test Planning tutorial, Test Report should include the objective of each round of testing, such as Unit Test, Performance Test, System Test …etc.

Test Summary:
This section includes the summary of testing activity in general. Information detailed here includes.

6.2 Testing Planning Step:-

1) Functionality Testing

2) Usability testing
3) Interface testing
4) Performance testing
5) Security testing

6.2.1 Functionality Testing:

Test for all the links in web pages, database connection, forms used in the web pages for submitting or getting information from user, Cookie testing. Check all the links:

- Test the outgoing links from all the pages from specific domain under test.
- Test all internal links.
- Test links jumping on the same pages.
- Test links used to send the email to admin or other users from web pages.
- Test to check if there are any orphan pages.
- Lastly in link checking, check for broken links in all above-mentioned links.
- Test forms in all pages: Forms are the integral part of any web site. Forms are used to get information from users and to keep interaction with them. So, what should be checked on these forms?
- First check all the validations on each field.
- Check for the default values of fields.
- Wrong inputs to the fields in the forms.
- Options to create forms if any, form delete, view or modify the forms.

6.2.2 Usability Testing:

- **Test for navigation**: Navigation means how the user surfs the web pages, different controls like buttons, boxes or how user using the links on the pages to surf different pages. Usability testing includes: Web site should be easy to use. Instructions should be provided clearly. Check if the provided instructions are correct means whether they satisfy purpose. Main menu should be provided on each page. It should be consistent.
- **Content:** Content should be logical and easy to understand. Check for spelling errors. Use of dark colours annoys users and should not be used in site theme. You can follow some standards that are used for web page and content building. These are common accepted standards like as I mentioned above about annoying colours, fonts, frames etc.

  Content should be meaningful. All the anchor text links should be working properly. Images should be placed properly with proper sizes. These are some basic standards that should be followed in web development. Your task is to validate all for UI testing.

6.2.3 Interface Testing:-

The main interfaces are:

- Web **server** and application server interface
- Application server and Database server interface.

Check if all the interactions between these servers are executed properly. Errors are handled properly. If database or web server returns any error message for any query by application server then application server should catch and display these error messages appropriately to users.

Check what happens if user interrupts any transaction in-between? Check what happens if connection to web server is reset in between

6.2.4 Performance testing:-

Web application should sustain to heavy load. Web performance testing should include:

- Web Load Testing
- Web Stress Testing

Test application performance on different internet connection speed. In the web load testing test if many users are accessing or requesting the same page. Can system sustain in peak load times? Site should handle many simultaneous user requests, large input data from users, Simultaneous connection to DB, heavy load on specific pages etc.

Stress testing: Generally, stress means stretching the system beyond its specification limits. Web stress testing is performed to break the site by giving stress and checked how system reacts to stress and how system recovers from crashes.

Stress is generally given on input fields, login and sign up areas. In web performance testing web site functionality on different operating systems, different hardware platforms are checked for software, hardware memory leakage errors.

6.2.5 Security Testing:

Following are some test cases for web security testing:

- Test by pasting internal url directly into browser address bar without login. Internal pages should not open.
- If you are logged in using username and password and browsing
- internal pages then try changing url options directly. I.e. If you are checking some publisher site statistics with publisher site ID= 123.
- Try some invalid inputs in input fields like login username, password, input text boxes. Check the system reaction on all invalid inputs.
- Web directories or files should not be accessible directly unless given download option.

- Test if SSL is used for security measures. If used proper message should get displayed when user switch from non-secure http:// pages to secure https:// pages and vice versa.
- All transactions, error messages, security breach attempts should get logged in log files somewhere on web server.

6.3 Testing Strategies:-

6.3.1 White Box Testing:

If we go by the definition, "White box testing" (also known as clear, glass box or structural testing) is a testing technique which evaluates the code and the internal structure of a program.

White box testing involves looking at the structure of the code. When you know the internal structure of a product, tests can be conducted to ensure that the internal operations performed according to the specification. And all internal components have been adequately exercised.

White box testing coverage specifications:

1. **Code coverage**
2. **Segment coverage:** Ensure that each code statement is executed once.
3. **Branch Coverage or Node Testing:** Coverage of each code branch in from allpossible was.
4. **Compound Condition Coverage**: For multiple conditions test each condition with multiple paths and combination of the different path to reach that condition.
5. **Basis Path Testing:** Each independent path in the code is taken for testing.
6. **Data Flow Testing (DFT):** In this approach you track the specific variables through each possible calculation, thus defining the set of intermediate paths through the code. DFT tends to reflect dependencies but it is mainly through sequences of data manipulation. In short, each data variable is tracked and its use is verified. This approach tends to uncover bugs like variables used but not initialize, or declared but not used, and so on.
7. **Path Testing:** Path testing is where all possible paths through the code are defined and covered. It's a time-consuming task.
8. **Loop Testing:** These strategies relate to testing single loops, concatenated loops, and nested loops. Independent and dependent code loops and values are tested by this approach.

Limitations:

Not possible for testing each and every path of the loops in the program. This means exhaustive testing is impossible for large systems.

This does not mean that WBT is not effective. By selecting important logical paths and data structure for testing is practically possible and effective.

6.3.2 Black Box Testing:

Black Box Testing is also known as behavioral, opaque-box, closed-box, specification-based or eye-to-eye testing.

It is a Software Testing method that analyses the functionality of a software/application without knowing much about the internal structure/design of the item that is being tested and compares the input value with the output value.

The main focus in Black Box Testing is on the functionality of the system as a whole. The term 'Behavioral Testing' is also used for Black Box Testing. Behavioral test design is slightly different from the black-box test design because the use of internal knowledge isn't strictly forbidden, but it's still discouraged.

Each testing method has its own advantages and disadvantages. There are some bugs that cannot be found using the only black box or only white box technique.

Majority of the applications are tested by Black Box method. We need to cover the majority of test cases so that most of the bugs will get discovered by a Black- Box method.This testing occurs throughout the software development and Testing Life Cycle i.e., in Unit, Integration, System, Acceptance, and Regression Testing stages.

Advantages:

- The tester need not have a technical background. It is important to test by being in the user's shoes and think from the user's point of view.
- Testing can be started once the development of the project/application is done. Both the testers and developers work independently without interfering in each other's space.
- It is more effective for large and complex applications.
- Defects and inconsistencies can be identified at the early stage of testing.

Disadvantages:

- Without any technical or programming knowledge, there are chances ofignoring possible conditions of the scenario to be tested.
- In a stipulated time, there are possibilities of testing less and skipping allpossible inputs and their output testing.
- A Complete Test Coverage is not possible for large and complex projects.

7). Conclusion and Bibliography

7.1 Conclusion

7.2 Bibliography

7.1 Conclusion and Future Work:-

Due to the variable nature of cancer cells, human prediction and classification can be erroneous that tend to improper medication. Hence, we have developed our system using an image processing and machine learning technique that can detect and classify cervical cancer cell more accurately. Currently this system is working on single cell images it can be developed for multicell images however it is difficult. High accuracy can be achieved if more number of images are available

.

7.2 Bibliography:-

1. B, Shanthi & Faruqi, Faraz & S., Hareesha & Kudva, Ranjini. (2019). Deep Convolution Neural Network for Malignancy Detection and Classification in Microscopic Uterine Cervix Cell Images. Asian Pacific Journal of Cancer Prevention. 20. 3447-3456. 10.31557/APJCP.2019.20.11.3447.

2. Gautam, Srishti & K., Harinarayan & Jith O U, Nirmal & Sao, Anil & Bhavsar, Arnav & Natarajan, Adarsh. (2018). Considerations for a PAP Smear Image Analysis System with CNN Features.

3. Kudva, Vidya & Prasad, Keerthana & Guruvare, Shyamala. (2018). Automation of Detection of Cervical Cancer Using Convolutional Neural Networks. Critical Reviews in Biomedical Engineering. 46. 135-145. 10.1615/CritRevBiomedEng.2018026019.

4. Xiang, Yao & Sun, Wanxin & Pan, Changli & Yan, Meng & Yin, Zhihua & Liang, Yixiong. (2020). A novel automation-assisted cervical cancer reading method based on convolutional neural network. Biocybernetics and Biomedical Engineering. 40. 10.1016/j.bbe.2020.01.016.