A Project Report on

## BLOOD CANCER DETECTION USING MICROSCOPIC IMAGES

Submitted in partial fulfillment of the requirement for the award the degree of

Bachelor of Technology in Computer Science & Engineering by

|  |  |
| --- | --- |
| B. HARISH | 17VV1A0507 |
| M. HEMALATHA | 17VV1A0526 |
| P. PRANAYA DURGA | 17VV1A0534 |
| B. RHEMA SRAVYA | 17VV1A0508 |

### Under the Esteemed Guidance of

**Dr. P. ARUNA KUMARI**

**Assistant Professor**

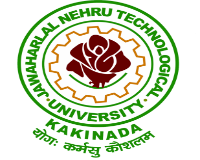


Department of Computer Science & Engineering

**JNTUK - University College of Engineering Vizianagaram**

VIZIANAGARAM – 535 003, A.P, INDIA

(2017-2021)



**JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA**

**UNIVERSITY COLLEGE OF ENGINEERING VIZIANAGARAM**

Dwarapudi (post), Vizianagaram Dist-Phone No. 08922-277911.

Website: [www.jntukucev.ac.in](http://www.jntukucev.ac.in)

**INSTITUTE**

**VISION**

To emerge as a premier technical Institution in the field of engineering and research with a focus to produce professionally competent and socially sensitive engineers capable of working in a multidisciplinary global environment.

**MISSION**

* To provide high quality technical education through a creative balance of academia and Industry by adopting highly effective teaching learning processes
* To promote multidisciplinary research with a global perspective to attain professional excellence
* To establish standards that inculcate ethical and moral values that contribute to growth in the Career and development of society.

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**VISION**

To achieve academic excellence in Computer Science and Engineering by imparting comprehensive knowledge to the students, promoting research activities and professional ethics to outfit the ever changing industrial demands and societal needs.

**MISSION**

* To produce the best quality Computer Science professionals by imparting quality
* teaching and learning process, training and value education. To strengthen links with industry through partnerships and collaborative developmental works.
* To attain self-sustainability and overall development through Research and Consultancy activities.
* To inculcate work ethics and commitment in students for their future endeavors to serve the society.

**PROGRAM OUTCOMES(POS)**

Program Outcomes (POS) form a set of individually assessable outcomes that are the components indicative of the graduate potential to acquire competence to practice at the appropriate level. The POs is exemplars of the attributes expected of a graduate from an institution.

**PO Description**

POI: **Engineering Knowledge** Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.

PO2**: Problem Analysis** Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

PO3: **Design/development of Solutions** Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.

PO4: **Conduct Investigations of Complex Problems** Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

PO5: **Modern Tool Usage** Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.

PO6: **The Engineer and Society** Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.

PO7: **Environment and Sustainability** Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

PO8: **Ethics** Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

PO9: **Individual and Team Work** Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.

PO10: Communication: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.

PO11: **Project Management and Finance** Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.

PO12: **Life-Long Learning** Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

**PROGRAM EDUCATIONAL OBJECTIVES (PEOS)**

PEO 1: The graduates of the Program will be prepared for their careers

in the software industry, public sector or pursue higher studies and continue to develop their professional knowledge.

PEO 2: The graduates will acquire capability to apply their knowledge of Computer science and engineering to solve real world problems using latest technologies

PEO 3: The graduates will inculcate in professional attitude, inter-disciplinary approach, ethics and ability to relate computer engineering issues with social awareness.

PEO 4: The graduates of Computer Science Engineering will have soft skills to adapt to the diverse global environment through lifelong learning.

**PROGRAM SPECIFIC OUTCOMES (PSOs)**

PSO 1: Able to apply the knowledge of programming languages, data structures and algorithms, data science, networks and software engineering principles for software product development

PSO 2: Able to analyze and formulate solutions to real world and socially relevant problems over multidisciplinary domains by using latest technologies.

PSO3: Able to be a technically competent employee, researcher, entrepreneur, excel in competitive exams and zest for higher studies.

**ACKNOWLEDGEMENT**

This acknowledgement transcends the reality of formality when we express deep gratitude and respect to all those people behind the screen who inspired and helped us in the completion of this project work.

We also take the privilege to express my heartfelt gratitude to my guide **Dr. P. Aruna Kumari,** Asst. Prof., CSE, of JNTUK UCEV for her valuable suggestions and constant motivation that greatly helped me in successful completion of the project. We express our sincere thanks to **Prof.A.S.N.Chakravarthy,** HOD**-**Dept. of Computer Science and Engineering for his continuous support. We express our sincere thanks to our respected Principal **Prof. G. Swami Naidu** with a great sense of pleasure and immense sense of gratitude that we acknowledge the help of these individuals. We owe many thanks to a many people who helped and supported us during the writing of this report.

We are thankful to all faculty members for extending their kind cooperation and assistance. Finally, we are extremely thankful to our parents and friends for their constant help and moral support.

**ABSTRACT**

The main motto of this work is to detect the Leukemia at earlier stage with the help of image processing techniques. Leukemia means blood cancer which is featured by uncontrolled and abnormal production of white blood cells(leukocytes) by the bone marrow in the blood. Acute Lymphoblastic Leukemia (ALL) is a type of Leukemia which is more common in children due to its nonspecific nature of the symptoms and signs of ALL lead’s wrong diagnosis. Even hematologist finds it difficult to classify the Leukemia cells, there manual classification of blood cells is not only time consuming but also inaccurate therefore early identification of Leukemia yields in providing the appropriate treatment to the patient. Detection through the images is fast and cheap method. Identification of Blood disorders is practiced by visualization of blood samples through a microscope by the naked eye of a human. In this work, a computerized technique has been developed to help the doctor in identifying different types of Leukemia.

**This project is mapped with following PO’S and PO’S:**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **PO1** | **PO2** | **PO3** | **PO4** | **PO5** | **PO6** | **PO7** | **PO8** | **PO9** | **PO10** | **PO11** | **PO12** |
|  |  |  |  |  |  |  |  |  |  |  |  |

|  |  |  |
| --- | --- | --- |
| **PSO1** | **PSO2** | **PSO3** |
|  |  |  |

**DECLARATION**

We, B.HARISH, M.HEMALATHA, P.PRANAYA DURGA, B.RHEMA SRAVYA hereby declare that the project report titled "Blood Cancer Detection Using Microscopic Images" submitted to JNTUK University College of Engineering Vizianagaram, in partial fulfillment of the requirements for the award of the degree of B.Tech in COMPUTER SCIENCE AND ENGINEERING is award of original and independent research work done by us during the academic year 2020 - 2021 under the supervision of Assistant Prof. Dr. P. ARUNA KUMARI and it has not formed the basis of any Degree/Diploma/Associate ship/ Fellowship or other similar title to any candidate in University.

# Signature of the Candidates

1. B. HARISH (17VV1A0507)
2. M. HEMALATHA (17VV1A0526)

1. P. PRANAYA DURGA (17VV1A0534)

1. B. RHEMA SRAVYA (17VV1A0508)

Place: Vizianagaram

Date:

**TABLE OF CONTENTS**

**S.NO. CHAPTER PAGE**

**1. INTRODUCTION 01-18**

1.1 BLOOD CANCER FACTS 04

1.2 IMAGE PROCESSING 05

1.2.1 PURPOSE OF IMAGE PROCESSING 05

1.2.2 TYPES 05

1.3 FEATURE EXTRACTION 07

1.4 MORPHOLOGICAL FEATURES 08

1.4.1 SHAPE FEATURES 08

1.4.2 BENDING ENERGY 08

1.4.3 ROUNDNESS RATIO 09

1.4.4 CHAIN CODE 09

1.5 TEXTURE AND INTENSITY FEATURES 09

1.5.1 GLCM 09

1.5.2 HISTOGRAM 09

1.5.3 GABOR TEXTURE EXTRACTION 10

1.5.4 COLOR INTENSITY FEATURES 10

1.5.5 FRACTAL DIMENSION FEATURES 10

1.5.6 ENTROPY 10

1.5.7 HAUSDROFF DIMENSION 10

1.5.8 LOCAL BINARY PATTERN 10

1.6 DISADVANTAGES OF FEATURE EXTRACTION 11

1.7 DEEP LEARNING 11

1.8 NEURAL NETWORKS 12

1.9 CONVOLUTION NEURAL NETWORKS 12

1.9.1 LAYERS IN CNN 13

1.9.1.1 CONVOLUTION LAYER 13

1.9.1.2 POOLING LAYER 14

1.9.1.3 FULLY CONNECTED LAYER 16

1.9.2 ADVANTAGES OF CNN 18

1.10 WHY WE NEED TO DO THIS PROBLEM 18

**2. LITERATURE SURVEY** **20-25**

**3. SYSTEM ANALYSIS** **27-34**

3.1 EXISTING MODEL 27

3.1.1 DRAWBACKS 28

3.2 PROPOSED MODEL 28

3.3 PROBLEM STATEMENT 29

3.4 SYSTEM ARCHITECTURE 29

3.4.1 IMAGE LOADING 30

3.4.2 PRE-PROCESSING 30

3.4.3 PREDICTION OF CANCER BASED ON CNN 32

3.5 DATASET DESCRIPTION 33

3.6 SYSTEM REQUIREMENTS 33

3.7 SOFTWARE REQUIREMENTS 34

**4. METHODOLOGY** **36-43**

4.1 PROBLEM STATEMENT 36

4.2 MODULES 36

4.3 PREDICTION OF CANCER BASED ON CNN MODEL 39

4.4 INITIAL MODEL 41

4.5 FINAL MODEL 42

**5. IMLEMENTATION** **45-48**

5.1 USER INTERFACE DESIGN 45

5.2 USER MANUAL 45

5.3 SAMPLE CODE 46

5.3.1 IMPORTING NECESSARY LIBRARIES 46

5.3.2 MODEL BUILDING 47

5.3.3 MODEL COMPILATION 48

**6. RESULTS** **49-56**

6.1 RESULT ANALYSIS 50

6.2 RESULT OF OPTIMIZED MODEL 51

**7. CONCLUSION** **57-58**

7.1 CONCLUSION 58

7.3 FUTURE WORKS 58

**8. REFERENCES** **59-76**

**CHAPTER 1**

**INTRODUCTION**

## 1.INTRODUCTION

Cancer has been one of the leading reasons for deaths worldwide and blood cancer is a widely prevalent type of cancer. Most of the functions and productions of blood cells are affected by blood cancer. It is seen that most of the cancers begin from the place where blood is produced i.e., bone marrow. In this case, the process of development of normal blood cells is disrupted by the growth of an abnormal type of cell. These cancerous blood cells stop your blood from performing their primary functions like preventing blood loss, fighting against infections, etc.

Blood consists of plasma, and three different types of cells and they are: White Blood Cells, Red Blood Cells and Platelets and each of these performs particular task. Red blood cells transport oxygen from the lungs to the tissues of the body and vice versa. White blood cells help the body to fight against diseases and infections. Platelets help to clot and control bleeding. Leukemia is cancer of blood cells in which number of white cells is increases numerously and those are immature cells that interfere with other blood cells, usually red blood cells and platelets. Our body’s white blood cell ratio is 1000:1. It means that between 1000 red blood cells there is 1 white blood cell.

There are two types of white blood cells that get turn into leukemia and they are:

* Lymphoid cells
* Myeloid cells

Leukemia that caused due to lymphoid cells is called lymphocytic or lymphoblastic leukemia and if it is caused due to myeloid cells then it is known as myelogenous or myeloid leukemia. Leukemia is grouped in two ways: acute or chronic, grouped according to how fast the cells are growing. The abnormal blood cells in acute leukemia are usually immature blasts (young cells) that do not work properly. These cells are growing fast. Acute leukemia gets worse quickly unless it is immediately treated. Young blood cells are present in chronic leukemia, but also mature functional cells are produced. Blasts are growing slowly in chronic leukemia. It takes the disease longer to get worse.

The four major forms of leukaemia are:

1. Acute lymphoblastic leukaemia (ALL)
2. Acute myelogenous leukaemia (AML)
3. Chronic lymphocytic leukaemia (CLL) and
4. Chronic myelogenous leukaemia (CML)

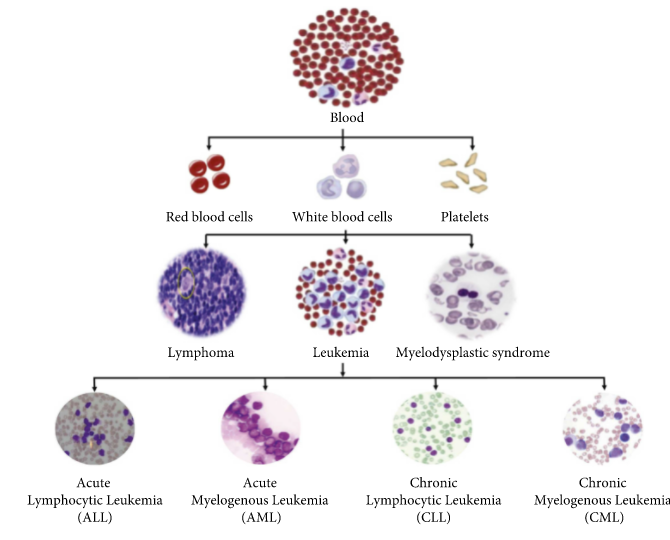
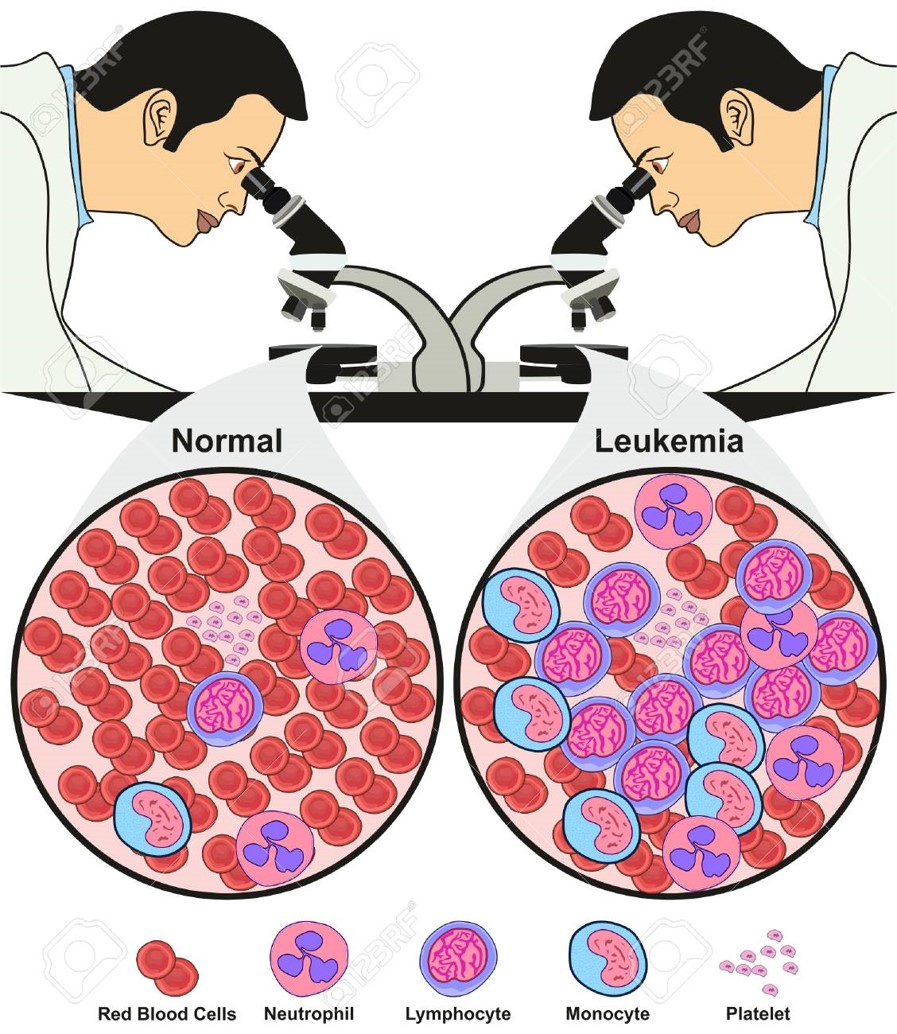


FIG.1.1 BLOOD CELL CLASSIFICATION

* 1. **Blood Cancer Facts**
* Leukemia is most commonly found in small children.
* Lymphoma is commonly found in the people who fall in the age group of 16 to 24 years.
* About 31% of males suffer from leukemia as compared to females.

Note that blood cancer prognosis is dependent on different factors like age, how severe it is, and at which stage it has been diagnosed.



### Fig.1.2 Images of Normal Blood Cells and Leukemia

### 1.2 Image Processing

### Image processing is a method to convert an image into digital form and perform some operations on it, in order to get an enhanced image or to extract some useful information from it. It is a type of signal dispensation in which input is image, like video frame or photograph and output may be image or characteristics associated with that image. Usually, Image Processing system includes treating images as two-dimensional signals while applying already set signal processing methods to them.

### It is among rapidly growing technologies today, with its applications in various aspects of a business. Image Processing forms core research area within engineering and computer science disciplines too.

### Image processing basically includes the following three steps.

### Importing the image with optical scanner or by digital photography.

### Analyzing and manipulating the image which includes data compression and image enhancement and spotting patterns that are not to human eyes like satellite photographs.

### Output is the last stage in which result can be altered image or report that is based on image analysis.

### 1.2.1 Purpose of Image processing

### The purpose of image processing is divided into 5 groups. They are:

### 1. Visualization – Observe the objects that are not visible.

### 2. Image sharpening and restoration – To create a better image.

### 3. Image retrieval – Seek for the image of interest.

### 4. Measurement of pattern – Measures various objects in an image.

### 5. Image Recognition – Distinguish the objects in an image.

### 1.2.2 Types

### The two types of methods used for Image Processing are Analog and Digital Image Processing. Analog or visual techniques of image processing can be used for the hard copies like printouts and photographs. Image analysts use various fundamentals of interpretation while using these visual techniques. The image processing is not just confined to area that has to be studied but on knowledge of analyst. Association is another important tool in image processing through visual techniques. So, analysts apply a combination of personal knowledge and collateral data to image processing.

### Digital Processing techniques help in manipulation of the digital images by using computers. As raw data from imaging sensors from satellite platform contains deficiencies. To get over such flaws and to get originality of information, it has to undergo various phases of processing. The three general phases that all types of data have to undergo while using digital technique are Pre- processing, enhancement and display, information extraction.

### FIG.1.3 FLOW DIAGRAM OF IMAGE PREPROCESSING

### 1.3 Feature Extraction

### In digital image processing and machine learning, features are the information we retrieve from the computational problem, which helps us to efficiently solve the task. Features can be some specific structure of the image like its shape, points, texture, edges, and so forth. Feature’s extraction is the interpretation of this information to reduce the dimension of image in such a way that is more informative and less redundant. This technique is very effective when the algorithm has a large set of data and that data can be also redundant; then the data is minimized to reduce set of features that carry the most of information of the image and is easy to compute by the algorithm. The selection of subset of these relevant features is called features selection. These features will contain the most relevant information and this subset is used instead of complete feature set. By minimizing the information into reduced set of features, a classifier will give better result by interpreting fewer amounts of data with more relevant information.

### For acute lymphoblastic leukemia detection, features extraction plays vital role because blast cells may have lot of information including different characteristics of their nucleus and cytoplasm. Different features have been extracted in the recent study. These features can be divided into two broad categories, morphological features and texture features.

### 1.4. Morphological Features

### In medical image processing, morphological features are very effective to analyse the information of the blood cell. In acute lymphoblastic leukaemia, blast cell has unique shape based features because every type of cell has unique area, perimeter, and circular rounding. So, by extracting morphological features from the blood cells we can efficiently perform classification of these cells.

### 1.4.1 Shape Features

### Shape based features play very important role for the acute leukaemia cell detection. Different shape based features like area, perimeter, circulatory, solidity, eccentricity, and so forth have been extracted to classify the blast cells of leukaemia.

### 1.4.2 Bending Energy

### Bending energy is also an essential feature that helps in efficient detection of acute lymphoblastic leukaemia. This parameter is used to detect the curvature of blast’s cell boundary which can help in ALL classification.

### 1.4.3 Roundness Ratio

### Roundness ratio is also an important feature that is widely used in leukemia detection. Because of increased variance in the circular shape of blast cells, roundness ratio is an efficient feature for the better classification of leukemia cell and its subtypes. Also, it helps in counting of WBCs which is also an important factor for the leukemia occurrence.

### 1.4.4 Chain Code

### Chain code features are widely used in acute lymphoblastic leukemia detection. These features separate the boundaries of nucleus and cytoplasm of blast cell which will help us to trace out the nucleus and cytoplasm.

### 1.5. Texture and Intensity Features

### In medical image processing, texture is an important characteristic for the identification of blast cells. By analyzing texture of an image, we can easily pick our region of interest and it also describes the spatial intensity distribution and specific colors in that ROI. For leukemia detection, important information including texture and intensity has been extracted from blood smear images, which can help in better classification of blast cells for leukemia detection and blast cell identification.

### 1.5.1 GLCM (Grey Level Co-occurrence Matrix)

### GLCM is a statistical method used for the examining of texture in which spatial relationship between the pixels is considered. For leukemia detection, GLCM is very useful to utilize the texture of the input blood smear image and extract features based on the texture and intensity of blast cells.

### 1.5.2 Histogram

### Histogram features including entropy, energy, mean, standard deviation, skewness, and kurtosis are extracted from the blood smear image to get enough relevant information. These types of features are also called 1st-order statistical features that are calculated by utilizing original pixels and excluding neighbor pixels.

### 1.5.3 Gabor Texture Extraction

### Gabor texture extraction method proposed by Dennis Gabor is very useful to extract relevant information of a blood smear image by analyzing its texture. Gabor features can be extracted after applying Gabor filter.

### 1.5.4 Color Intensity Features

### Color features are very useful for fetching relevant information from blood cell nucleus. So, mean color values from different color models like RGB, HSV, HIS, and so forth are extracted as a feature and input to the classifier for better classification of blast cells.

### 1.5.5 Fractal Dimension Features

### Fractal dimension is widely used in medical image processing to measure different quantitative information. To identify whether a leukemia cell is blast or normal, the roughness of its nucleus is being measured over spatial distribution by using fractal geometry.

### 1.5.6 Entropy

### By performing nucleus texture measurement, we can extract the entropy as a feature vector that is used to measure the randomness of the nucleus from the blood smear image. Different entropy measurements are used for the acute lymphoblastic leukemia cell detection.

### 1.5.7 Hausdorff Dimension

### It is also an essential feature for microscopic blood image analysis, which is used with fractal dimension to extract relevant information by measuring roughness of nucleus.

### Local Binary Pattern

### Local Binary Pattern is a texture classification technique that is used to extract texture features of an image. Because of fast computational speed of LBP, it is highly recommended for leukemia detection, where speed is an important factor. Also, it provides significant information about the illumination changes, which also helps in detection of blasted leukemia cells. In Discriminative Robust Local Binary Pattern is proposed for features extraction, which provides very encouraging results.

### 1.6 Disadvantages of Feature Extraction

### There are mainly two disadvantages of Feature Extraction they are:

### Loss data Interpretability

### The transformation may be expensive

### DEEP LEARNING

### Deep learning is an artificial intelligence function that imitates the workings of the human brain in processing data and creating patterns for use in decision making. Deep learning is a subset of machine learning in artificial intelligence (AI) that has networks capable of learning unsupervised from data that is unstructured or unlabeled. Also known as deep neural learning or deep neural network. Deep learning learns from vast amounts of unstructured data that would normally take humans decades to understand and process.

There are some popular deep leaning algorithms. They are:

* Convolution Neural Networks (CNNs)
* Long Short Term Memory Networks (LSTMs)
* Recurrent Neural Network (RNNs)
* Generative Adversarial Networks
* Radial Basis Function Networks
* Multilayer Perceptrons
* Self Organizing Maps
* Deep Belief Networks
* Restricted Boltzman Machines
* Autoencoders

## 

## 1.8 NEURAL NETWORKS

A neural network is a series of algorithms that endeavors to recognize underlying relationships in a set of data through a process that mimics the way the human brain operates. In this sense, neural networks refer to systems of neurons, either organic or artificial in nature. Neural networks can adapt to changing input; so, the network generates the best possible result without needing to redesign the output criteria.

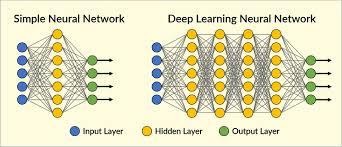


Fig. 1.4 SIMPLE AND DEEP NEURAL NETWORK

**1.9 Convolutional Neural Network**

A Convolutional Neural Network (Convent/CNN) is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other. The pre-processing required in a Convent is much lower as compared to other classification algorithms. While in primitive methods filters are hand-engineered, with enough training, ConvNets have the ability to learn these filters/characteristics. The architecture of a ConvNet is analogous to that of the connectivity pattern of Neurons in the Human Brain and was inspired by the organization of the Visual Cortex. Individual neurons respond to stimuli only in a restricted region of the visual field known as the Receptive Field. A collection of such fields overlaps to cover the entire visual area. The advancements in Computer Vision with Deep Learning have been constructed and perfected with time, primarily over one particular algorithm a Convolutional Neural Network.

**1.9.1 Layers in CNN**

There are three types of layers that make up the CNN which are the convolutional layers, pooling layers, and fully-connected (FC) layers. When these layers are stacked, a CNN architecture will be formed. In addition to these three layers, there are two more important parameters which are the dropout layer and the activation function which are defined below.

**1.9.1.1 Convolutional Layer**

This layer is the first layer that is used to extract the various features from the input images. In this layer, the mathematical operation of convolution is performed between the input image and a filter of a particular size MxM. By sliding the filter over the input image, the dot product is taken between the filter and the parts of the input image with respect to the size of the filter (MxM).

The output is termed as the Feature map which gives us information about the image such as the corners and edges. Later, this feature map is fed to other layers to learn several other features of the input image.

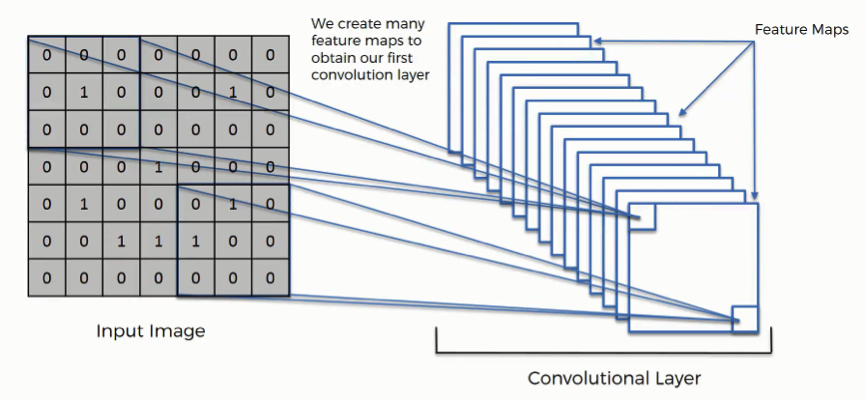


FIG.1.5 CONVOLUTION LAYER

**1.9.1.2 Pooling Layer**

In most cases, a Convolutional Layer is followed by a Pooling Layer. The primary aim of this layer is to decrease the size of the convolved feature map to reduce the computational costs. This is performed by decreasing the connections between layers and independently operates on each feature map. Depending upon method used, there are several types of Pooling operations.

In Max Pooling, the largest element is taken from feature map. Average Pooling calculates the average of the elements in a predefined sized Image section. The total sum of the elements in the predefined section is computed in Sum Pooling. The Pooling Layer usually serves as a bridge between the Convolutional Layer and the FC Layer.

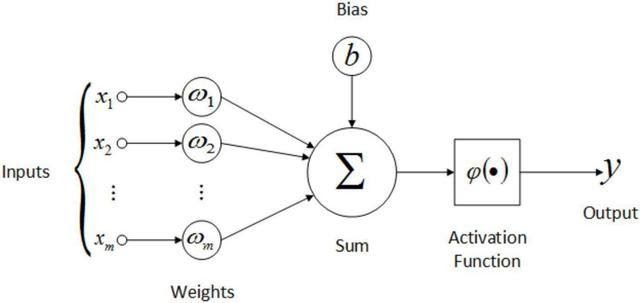


**FIG.1.6 POOLING**

***Activation Function***

Activation functions are mathematical equations that determine the output of a neural network. The function is attached to each neuron in the network, and determines whether it should be activated (“fired”) or not, based on whether each neuron’s input is relevant for the model’s prediction.

Activation functions also help normalize the output of each neuron to a range between 1 and 0 or between -1 and 1.



**Fig.1.7 ACTIVATION FUNCTION**

***Non-Linear Layers***

Neural networks in general and CNNs in particular rely on a non- linear “trigger” function to signal distinct identification of likely features on each hidden layer. CNNs may use a variety of specific functions —such as rectified linear units (ReLUs) and continuous trigger (non-linear) functions—to efficiently implement this non- linear triggering.

***ReLU***

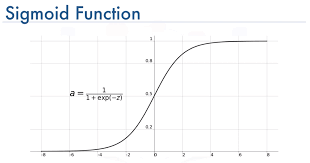
A ReLU implements the function y = max(x,0), so the input and output sizes of this layer are the same. It increases the nonlinear properties of the decision function and of the overall network without affecting the receptive fields of the convolution layer. In comparison to the other non-linear functions used in CNNs (e.g., hyperbolic tangent, absolute of hyperbolic tangent, and sigmoid), the advantage of a ReLU is that the network trains many times faster. ReLU functionality is illustrated in Figure below, with its transfer function plotted above the arrow.

***Continuous Trigger (Non-Linear) Function***

The non-linear layer operates element by element in each feature. A continuous trigger function can be hyperbolic tangent, absolute of hyperbolic tangent or sigmoid and demonstrates how non-linearity gets applied element by element.

***Sigmoid Function***

The sigmoid activation function, also called the logistic function, is traditionally a very popular activation function for neural networks. The input to the function is transformed into a value between 0.0 and 1.0. Inputs that are much larger than 1.0 are transformed to the value 1.0, similarly, values much smaller than 0.0 are snapped to 0.0. The shape of the function for all possible inputs is an S-shape from zero up through 0.5 to 1.0. For a long time, through the early 1990s, it was the default activation used on neural networks.

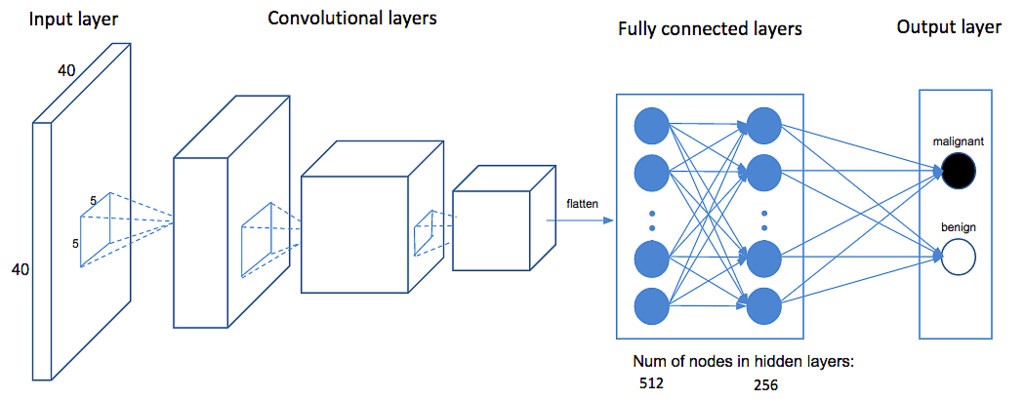


**Fig. 1.8 SIGMOID FUNCTION**

**1.9.1.3 Fully Connected Layer**

The Fully Connected (FC) layer consists of the weights and biases along with the neurons and is used to connect the neurons between two different layers. These layers are usually placed before the output layer and form the last few layers of a CNN Architecture.

In this, the input image from the previous layers is flattened and fed to the FC layer. The flattened vector then undergoes few more FC layers where the mathematical functions operations usually take place. In this stage, the classification process begins to take place.



**Fig 1.9 FULLY CONNECTED LAYER**

***Dropout***

Usually, when all the features are connected to the FC layer, it can cause overfitting in the training dataset. Overfitting occurs when a particular model works so well on the training data causing a negative impact in the model’s performance when used on a new data.

To overcome this problem, a dropout layer is utilized wherein a few neurons are dropped from the neural network during training process resulting in reduced size of the model. On passing a dropout of 0.3, 30% of the nodes are dropped out randomly from the neural network.

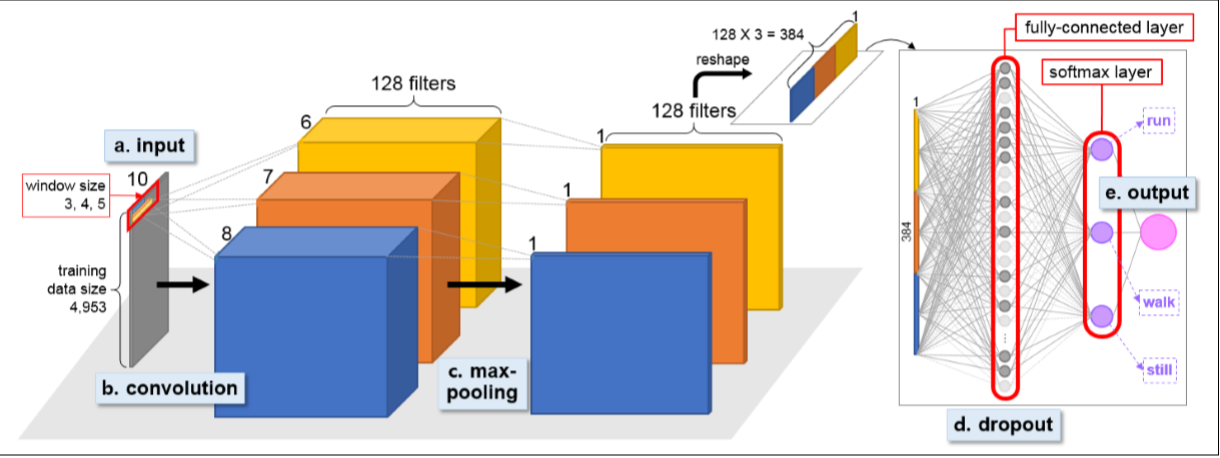


Fig. 1.10 MODEL OF CNN

**1.9.2** **Advantages of CNN**

There are a lot of algorithms that people used for image classification before CNN became popular. People used to create features from images and then feed those features into some classification algorithm like SVM. Some algorithm also used the pixel level values of images as a feature vector too.

CNNs can be thought of automatic feature extractors from the image. While if we use an algorithm with pixel vector, we lose a lot of spatial interaction between pixels, a CNN effectively uses adjacent pixel information to effectively down sample the image first by convolution and then uses a prediction layer at the end.

This concept was first presented by Yann le cun in 1998 for digit classification where he used a single convolution layer. It was later popularized by Alex net in 2012 which used multiple convolution layers to achieve state of the art on image net. Thus, making them an algorithm of choice for image classification challenges henceforth.

**1.10 Why We Need to Do This Problem**

Leukaemia means blood cancer which is featured by uncontrolled and abnormal production of white blood cells(leukocytes) by the bone marrow in the blood. Acute Lymphoblastic Leukaemia (ALL) is a type of Leukaemia which is more common in children due to its non-specific nature of the symptoms and signs of ALL lead’s wrong diagnosis. Even haematologist finds it difficult to classify the Leukaemia cells, there manual classification of blood cells is not only time consuming but also inaccurate therefore early identification of Leukaemia yields in providing the appropriate treatment to the patient. And it is very necessary to detect cancer in the early stages to treat this type of cancer or any type of cancer.

**CHAPTER 2**

**LITERATURE SURVEY**

**2.LITERATURE SURVEY**

**Literature Survey**

In this Literature Survey we are going to briefly describe about the concepts such as Leukemia classification Using deep Learning models which describe about the evaluation of white blood cell differential counts using hematology rules which plays an important role in the immune system, Microscopic smear segmentation and classification using deep contour aware CNN and extreme machine learning which describes presentation of efficient contour aware segmentation approach based on fully conventional network . There are some more topics to be described in this module will be of, of detection of sub type blood cells using deep learning which deals about that A CNN-based framework is built to automatically classify the blood cell images into sub types of the cells, and A deep convolutional neural network for classification of red blood cells in sickle cell anemia which deals about the blood vessel occlusion accompanied by painful episodes and even death.

**2.1 White Blood Cells’ Segmentation for the Detection of Acute Lymphoblastic Leukemia**

**Authors:** Seminal D. Joshi, Prof. A. H. Karode

**Description**

Acute lymphoblastic leukemia (ALL) is the most common hematological neoplasia of childhood. The characterization of ALL is figured out by rapid uncontrolled growth of immature leukemic cells (named as blast cells) in bone marrow, lymphoid organs etc. The most important preliminary step in the diagnosis of ALL is the morphological analysis of the blood and bone marrow smear under the microscope. Nowadays, this is performed manually by skilled operators. The process has numerous drawbacks, such as slowness of the analysis, non-standard accuracy, dependency on the operator skills etc. Also, the nonspecific nature of the signs and symptoms of ALL often leads to wrong diagnosis. The low cost and efficient process is to use a computerized system for image analysis of stained blood microscopic images for quantitative examination and ALL detection. This paper presents a description of a robust segmentation technique combining a few segmentation subsystems, which makes the further process of feature extraction and classification of blood microscopic images easier for diagnosing the presence of ALL and will help out to make a perfect automated system in future.

**Keywords:** Acute lymphoblastic leukemia (ALL), Image segmentation, Automatic cell identification, Cell features extraction, Lymphocyte classification, Image segmentation, classification, feature extraction, pattern recognition, KNN classifier.

**Limitations:** White blood cells should be localized from large number of data sets.

**2.2 Unsupervised Blood Microscopic Image Segmentation and Leukemia Detection using Color based Clustering.**

**Authors:** Subrajeet Mohapatra1, Dipti Patra1 and Sangha Mitra Satpathy

**Description:**

In real clinical trial there is no substitute for final assessment of any disease independent of diagnostic test. Diagnostic hematology is a specialty that deals with the understanding of the essential pathological processes of the blood through appropriate morphological or biochemical analysis. Microscopic analysis of peripheral blood and bone marrow by a hematologist are subjected to various shortcomings like inter observer variations, slowness, operator experience and tiredness. Whereas biochemical tests, immunophenotyping, molecular probing etc. are expensive for routine examination. Thus, microscopic image analysis serves as an impressive automated diagnostic tool for hematological disorders i.e., leukemia, malaria, psoriasis, AIDS etc. Acute Lymphoblastic Leukemia (ALL) is a serious hematological disorder of blood which needs to be diagnosed early for faster cure. This paper introduces a comparative approach to Acute Lymphoblastic Leukemia (ALL) detection based on WBC nucleus image segmentation and morphological analysis. Color based clustering is employed for segregating various blood components and obtaining the nucleus of the white blood cell. Further fractal geometry, contour signature and texture-based techniques are employed for nucleus feature extraction which leads to automatic leukemia detection using a Support Vector Machine (SVM) classifier. The proposed approach is validated with the collected blood microscopic images and satisfactory results have been obtained.

**Keywords**: Data Clustering, Acute Leukemia Detection, Hausdorff Distance, Quantitative Microscopy, Blood Image Analysis

**Limitations:** K- Medoid Feature Space Clustering Results accuracy failed for many times.

**2.3 Automatic Segmentation of Leukocytes for the Detection of Leukemia Using a New Computing Algorithm**

**Authors**: Aldrin Karunaharan K\* and Om Prakash

**Description**

Hematological disorders refer to the diseases caused with the changes in blood cells or blood system such as Leukemia, Anemia, Malaria and Azotemia. Leukocytes are the cells of immune system derived in the bone marrow as hematopoietic stem cell. The presence of immature cells changes the granularity and geometry of leukocytes. Detection of white blood cells plays an important role in the diagnosis of diseases like leukemia. Features such as nucleus and cytoplasm area, average color co-ordinates and number of pixels in the nuclear perimeter are used. Accurate classification of human blood cells plays a decisive role in the diagnosis and treatment of diseases. Hematological disorders refer to the diseases caused with the changes in blood cells or blood system such as Leukemia, Anemia, Malaria and Azotemia. This paper explores the techniques used in the automatic segmentation of leukocytes using a new computing technique

**Keywords:** Acquisition, Segmentation, Thresholding, Morphology, Soft computing, Threshold segmentation mathematical morphology, Fuzzy and cellular neural networks

**Limitations:** Binary image segmentation subjecting to errors for some trails and Algorithm accuracy is low.

**2.4 Hematological Image Analysis for Acute Lymphoblastic Leukemia Detection and Classification**

**Authors:** P. Sob Revilla, E. Montseny & J. Keller

**Description**

Microscopic analysis of peripheral blood smear is a critical step in detection of leukemia. However, this type of light microscopic assessment is time consuming, inherently subjective, and is governed by hematopathology’s clinical acumen and experience. To circumvent such problems, an efficient computer aided methodology for quantitative analysis of peripheral blood samples is required to be developed. In this thesis, efforts are therefore made to devise methodologies for automated detection and subclassification of Acute Lymphoblastic Leukemia (ALL) using image processing and machine learning methods. Choice of appropriate segmentation scheme plays a vital role in the automated disease recognition process. Accordingly, to segment the normal mature lymphocyte and malignant lymphoblast images into constituent morphological regions novel schemes have been proposed. In order to make the proposed schemes viable from a practical and real–time stand point, the segmentation problem is addressed in both supervised and unsupervised framework. These proposed methods are based on neural network, feature space clustering, and Markov random field modeling, where the segmentation problem is formulated as pixel classification, pixel clustering, and pixel labeling problem respectively. A comprehensive validation analysis is presented to evaluate the performance of four proposed lymphocyte image segmentation schemes against manual segmentation results provided by a panel of hematopathology’s.

It is observed that morphological components of normal and malignant lymphocytes differ significantly. To automatically recognize lymphoblasts and detect ALL in peripheral blood samples, an efficient methodology is proposed. Morphological, textural and color features are extracted from the segmented nucleus and cytoplasm regions of the lymphocyte images. An ensemble of classifiers represented as EOC3 comprising of three classifiers shows highest classification accuracy of 94.73% in comparison to individual members.

**Keywords:** Automated leukemia detection, Acute lymphoblastic leukemia, Quantitative microscopy, Lymphocyte image segmentation, Hematological image analysis, Machine learning.

**Limitations:** The adjacent leukocytes segmentation phase may be affected. Threshold segmentation also effected.

**2.5 A Robust Segmentation Method for Acute Lymphoblastic Leukemia Detection**

**Authors:** Patil Tejashri G.& V. B. Raskar

**Description**

Acute Lymphoblastic Leukemia is a malignant disorder of lymphoid cell, which affect both children & adults of different ages. The 80% cure is possible in children by effective treatment. Pathologists use leukocytes for the identification of various diseases. The methods like Fluorescent in Situ Hybridization (FISH), flow cytometry, immunophenotyping are used for leukemia detection. These methods have specific roles in the diagnosis & management of various hematological neoplasms. Careful examination of blood & bone marrow is fundamental in all hematological diagnosis. The major role of cytochemistry is in the diagnosis of acute myeloid leukemia & myelodysplastic syndromes and the major role of immunophenotyping is in the diagnosis of chronic lymphoproliferative disorder & acute leukemia cytogenetic analysis has role in confirming the diagnosis of chronic granulocyte leukemia and gives important information in the acute leukemias. But these methods give slow analysis and less accuracy. This paper presents fully automatic method for identification and classification of WBCs from microscopic images. In the proposed segmentation method Otsu’s thresholding method is used to segment normal and ALL lymphocytes from the blood cell component. The whole work has been done by using MATLAB software. Keywords—Acute Lymphoblastic Leukemia, White Blood Cells, Segmentation.

**Keywords:** Acute Lymphoblastic Leukemia, Segmentation, MATLAB

**Limitations:** These methods give slow analysis and less accuracy.

**CHAPTER 3**

**SYSTEM ANALYSIS**

**3.SYSTEM ANALYSIS**

System analysis is a process of collecting and interpreting facts, identifying the problems, and decomposition of a system into its components. System analysis is conducted for the purpose of studying a system or its parts in order to identify its objectives. It is a problem-solving technique that improves the system and ensures that all the components of the system work efficiently to accomplish their purpose. Analysis specifies what the system should do.

To know if the cell has cancer, we need to give the microscopic image of blood cell to the system. The system can extract features from this image and recognize the patterns in the image. The recognized patterns compared with the already existing features of the images that are present in our training dataset in order to make predictions.

Let us discuss the methodology of the system.

**3.1 Existing Model**

Previously related leukemia detection techniques mainly include KNN algorithm, segmentation, Thresholding, Fuzzy and cellular neural networks, clustering etc. We briefly review and discuss in this section.

* The core idea of the KNN algorithm is that if most of the k most adjacent samples in a feature space belong to a certain category. This method determines the class in which the sample is to be classified based on the category of the nearest samples in determining the classification decision. The KNN method is only relevant to a very small number of neighboring samples in the category decision. The recognition rate was 80.6%.
* Aldrin Karunaharan K\* and Om Prakash used both segmentation entropy based and iterative thresholding methods to divide cells and detected the cancer cells with mathematical morphology techniques, with a recognition rate of 85.14%.
* Subrajeet Mohapatra1 in 2017 first time used color-based clustering techniques to divide blood cancer cells. He used Hausdorff Distance to detect distance between each cell.
* Patil Tejashri used a robust segmentation method Acute Lymphoblastic leukemia detection. A Robust segmentation algorithm using morphological operators for detection of in order cells in blood. This algorithm uses Fuzzy Clustering Based on Kernel-Induced Distance Measure to detect the distance between cells. This particular method applied on nearly 105 images and got accuracy up to 81%.

**3.1.1 Drawbacks**

Although these methods can be used to generate good classification and segmentation engines, they still have some drawbacks. Some of the segmentation and thresholding techniques working slowly which leads to slow analysis and less accuracy. For Automated techniques the adjacent leukocytes segmentation phase and threshold segmentation may be affected. K-medoid featured space clustering results accuration failed for many times. Binary image segmentation generally subjecting to many errors. In All these methods the output can be improved by the cost effective and robust automated system for the screening of leukemia.

**3.2 Proposed Model**

The deep learning algorithm effectively solves this problem. It can automatically learn the effective features of the image. We propose a CNN based model which gives 90.57% accuracy for the 2 labels i.e., Cancer and Normal. Our model can automatically classify the blood cell type and detects unorder cell types in order to save time and enhance clinical efficiency. CNN is a category of Artificial Neural Network (ANN), which has been proven to be useful in image classification and object recognition. It takes the raw pixels as input and produces an outcome indicating the probabilities that the input belongs to different classes. CNN gains its high reputation by advocating an innovative architecture that may address a big hurdle existing in the conventional neural network: a massive number of parameters to tune in the training stage. Instead of implementing the fully-connected structure in every layer, CNN imposes two additional layers, convolution and pooling, that may significantly reduce the magnitude of parameters. CNN can be thought of automatic feature extractors from the image. While if we use algorithm with pixel vector, we lose a lot of spatial interaction between pixels, a CNN effectively uses adjacent pixel information to effectively down sample the images first by convolution and then uses a prediction layer at the end. Convolution layer is entitled by the convolutional operation, with the purpose of extracting the features from the Blood Cell Image Classification Using Convolutional Neural Network input images.

**3.3 Problem statement:**

Blood cell segmentation and identification is a vital in the study of blood as a health indicator. A complete blood count is used to determine the state of a person’s health based on the contents of the blood in particular the white blood cells and the red blood cells. The main problem arises when massive amounts of blood samples are required to be processed by the hematologist or Medical Laboratory Technicians. Leukemia is a term used for diseases in which abnormal cells divide without control and are able to invade other tissues. Leukemic cells can spread to other parts of the body through the blood and lymph systems

**3.4 System Architecture Modules of Proposed System**

The system architecture of the project is as follows it contains the below modules and they are:

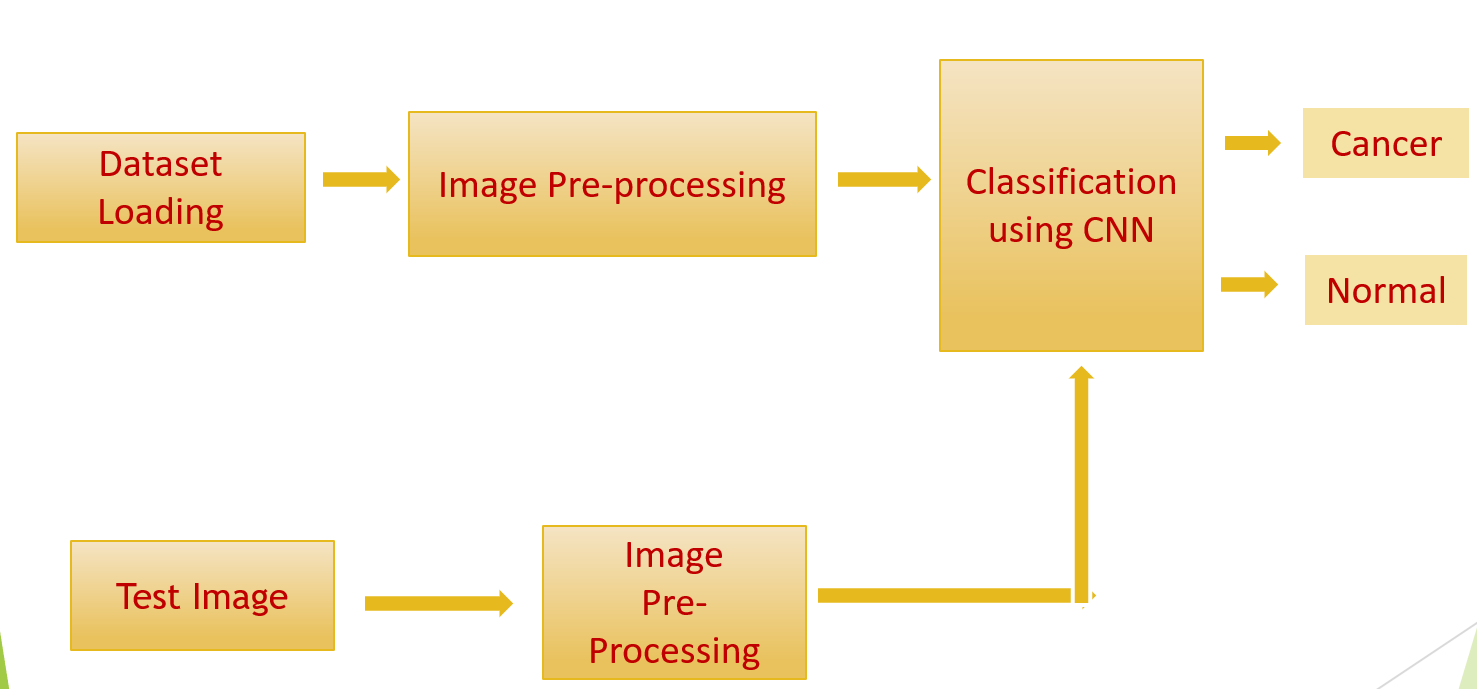


FIG.3.1 SYSTEM ARCHITECTURE

**3.4.1 Image Loading**

Loading dataset is the first step, next the loaded dataset is given to the preprocessing, Here the dataset used is leukemia classification dataset.

**3.4.2 Pre-processing**

Pre-processing has the following steps to make data set cleaner and get better result. They are:

**3.4.2.1 Noise Removal**

Noise reduction is a very essential step in digital image processing for getting better quality images. Medical imaging is a valuable tool in the field of medicine.Noise is caused due to various sources which include many external causes in transmission system and environmental factors which includes noise like Gaussian, Poisson, Blurred, Speckle and salt-and-pepper noise. Noise removing method has become an important factor in medical imaging applications and the most commonly used filters are Median filter, Gaussian filter, Weiner filter which gives the best result for the respective noises.

The need for the smoothening of images has becomes essential which is required to remove the noise and for that best filters or standard filters are used in most of the image processing applications. The important asset of a good image de-noising model is to remove the noise from the image and also preserve the edges.

There are two types of models which are used for de-noising i.e linear model and non-liner model and generally, linear models are used because of its speed and limitation is that it is not able to preserve the edges in an efficient manner.

**3.4.2.2 Segmentation:**

In digital image processing and computer vision, **image segmentation** is the process of partitioning a digital image into multiple segments (sets of pixels, also known as image objects). The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze.

***Algorithm***

Adaptive binarization processing is different from fixed threshold processing. The threshold of each pixel depends on the grayscale of its neighboring pixels. In order to obtain the threshold T(x,y) of the (x,y) point, we need to perform the following processing .

1.Select a bxb area around this pixel, which is designated by the user.

2.Calculate the weighted average of this bxb area. OpenCV provides two methods to calculate this weighted average, one is the arithmetic average method, and the other is the Gaussian weighted average method. The closer to the center of the region when the latter calculates the average, the greater the weight. We will calculate the weighted mean as WA(x,y)

3.The threshold value T(x,y) is obtained by subtracting the above-mentioned weighted average value and a fixed parameter. The fixed parameter is set to param1, and the threshold value of (x,y) point can be calculated by the following formula:

T(x,y)=WA(x,y)-param1

**3.4.2.3 Dataset Augmentation:**

The image count disproportion was be balanced through image augmentation which artificially creates training images through different ways of processing or combination of multiple processing, such as random rotation, shifts, shear. It can help overcome the increasingly large requirements of Deep Learning models. It acts as a regularize and helps reduce overfitting when training a deep learning model.

**3.4.3 Prediction of Cancer based on CNN model**

In model building we try learn what happens at each step in the layers so it can be explained as here we are using the convolutional neural network model the proposed model has the layers given below

***Convolutional Layer***

A convolutional neural network (CNN, or ConvNet) is a class of deep, artificial neural networks that found application mostly in image and sound analysis. CNN usually analyze part or region of the sample which called receptive field.

The convolution layer has different filters such as 16, 32, 64 and kernel size of (3,3) and an activation function used is ReLU.

***Pooling Layer***

This layer collects the output of a cluster of neurons and combines it into single neuron of the next layer. Pooling can use different aggregation functions (max, average, etc.) to choose the value that represents the cluster. Pooling layer reduces the size of the feature maps which has applied in the convolution layer.

***Fully connected Layer (or) Dense Layer***

This layer has properties of traditional perceptron neural network in which every neuron of one layer is connected to every neuron in the next layer. Dense layers are the last layers that perform the final classification.

***Drop Out Layer***

Overfitting occurs when a particular model works so well on the training data causing a negative impact in the model’s performance when used on a new data.

To overcome this problem, a dropout layer is utilized wherein a few neurons are dropped from the neural network during training process resulting in reduced size of the model.

**3.5 Dataset description**

The dataset is publicly available and taken from the Kaggle. Acute lymphoblastic leukemia (ALL) is the most common type of childhood cancer and accounts for approximately 25% of the pediatric cancers.

These cells have been segmented from microscopic images and are representative of images in the real-world because they contain some staining noise and illumination errors, although these errors have largely been fixed in the course of acquisition.

The task of identifying immature leukemic blasts from normal cells under the microscope is challenging due to morphological similarity and thus the ground truth labels were annotated by an expert oncologist.

In total there are 15,135 images from 118 patients with two labelled classes:

• Normal cell;

• Leukemia blast.

**3.6 System Requirements**

System requirement specification is a structured collection of information that embodies the requirements of a system. it is the most important document of reference in developing a design. It is to produce the specification analysis of the task and also to establish the complete information about the requirement, behavior and other constraints such as functional performance and so on. the goal of system requirement specification is to completely specify the technical requirements for the product in a concise and unambiguous manner.

**3.7** **Software Requirements**

Software requirements are the system services and constraints the requirements that are generated during engineering process. The requirements mentioned are the software used in the project.

**Operating System:** Windows 10

**Tools:** Google Collab, Visual Studio 2019,

**Hardware Requirements:** Hardware Requirement is the system capacity levels in the process of engineering.

**System:** Intel Core i5 CPU Hard

**Work:** 500GB

**RAM:** 4GB

**Speed:** 2.4GH

**CHAPTER 4**

**METHODOLOGY**

## 4.METHODOLOGY

### 4.1 Problem Statement

### Blood cell segmentation and identification is a vital in the study of blood as a health indicator. A complete blood count is used to determine the state of a person’s health based on the contents of the blood in particular the white blood cells and the red blood cells. The main problem arises when massive amounts of blood samples are required to be processed by the hematologist or Medical Laboratory Technicians.

Leukemia is a term used for diseases in which abnormal cells divide without control and are able to invade other tissues. Leukemic cells can spread to other parts of the body through the blood and lymph systems.

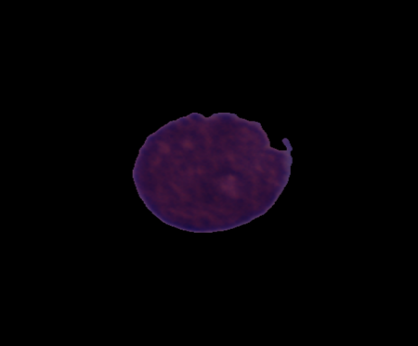
* 1. **Modules**

The system is composed of many modules they are:

* Image Loading
* Preprocessing
* Segmentation
* Model Building
* Layers in Model
* Convolutional Layer
* Pooling Layer
* Fully Connected Layer

### 4.2.1 Image Loading

### Image Loading is the ﬁrst step and here the dataset which was used is Luekemia Classification and that is publicly available from Kaggle This dataset comprises of 15,135 original images from 118 patients with two labelled classes that contains images of normal cells and leukemia blast cells. The cell pictures are labelled in the CSV ﬁle attached to the image data.



**CANCER IMAGE NORMAL IMAGE**

### Fig.4.1 Microscopic Blood Cell Images

### 4.2.2 Preprocessing

### Preprocessing is considered as most important step when dealing with the data, data cleaning is important to make model more accurate and give results with less precession this step is not done properly we cannot expect better results the steps involved in the preprocessing are down below.

### *Noise Removal*

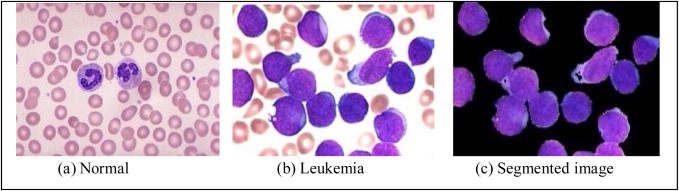
### Noise reduction is a very essential step in digital image processing for getting better quality images. Medical imaging is a valuable tool in the field of medicine. Noise is caused due to various sources which include many external causes in transmission system and environmental factors which includes noise like Gaussian, Poisson, Blurred, Speckle and salt-and-pepper noise. Noise removing method has become an important factor in medical imaging applications and the most commonly used filters are Median filter, Gaussian filter, Weiner filter which gives the best result for the respective noises.

There are two types of models which are used for de-noising i.e linear model and non-liner model and generally, linear models are used because of its speed and limitation is that it is not able to preserve the edges in an efficient manner.

### *Segmentation*

### Image segmentation is the process of partitioning a digital image into multiple segments (sets of pixels, also known as image objects). The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain characteristics.

### The result of image segmentation is a set of segments that collectively cover the entire image, or a set of contours extracted from the image (see edge detection). Each of the pixels in a region are similar with respect to some characteristic or computed property, such as color, intensity, or texture. Adjacent regions are significantly different color respect to the same characteristic(s). When applied to a stack of images, typical in medical imaging, the resulting contours after image segmentation can be used to create 3D reconstructions with the help of interpolation algorithms like marching cubes.



### FIG .4.1 SEGMENTATION

***Algorithm***

Adaptive binarization processing is different from fixed threshold processing. The threshold of each pixel depends on the grayscale of its neighboring pixels. In order to obtain the threshold T(x,y) of the (x,y) point, we need to perform the following processing .

1.Select a bxb area around this pixel, which is designated by the user.

2.Calculate the weighted average of this bxb area. OpenCV provides two methods to calculate this weighted average, one is the arithmetic average method, and the other is the Gaussian weighted average method. The closer to the center of the region when the latter calculates the average, the greater the weight. We will calculate the weighted mean as WA(x,y)

3.The threshold value T(x,y) is obtained by subtracting the above-mentioned weighted average value and a fixed parameter. The fixed parameter is set to param1, and the threshold value of (x,y) point can be calculated by the following formula:

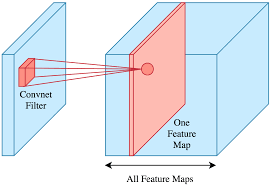
T(x,y)=WA(x,y)-param1.

**4.3 Prediction of Cancer based on CNN model**

In model building we try learn what happens at each step in the layers so it can be explained as here we are using the convolutional neural network model the proposed model has the layers given below

**4.3.1 Convolutional Layer**

This layer applies convolution operation to the input of the receptive ﬁeld and passes the result to the next layer. Usually comprises of several ﬁlters that detect diﬀerent features. First convolutional layer usually detects simple features.



### Fig.4.2 CONVOLUTIONAL NET

### 4.3.2 Pooling Layer

### CNN uses **max pooling** to replace output with a max summary to reduce data size and processing time. This allows you to determine features that produce the highest impact and reduces the risk of overfitting.

Max pooling takes two **hyperparameters**: stride and size. The stride will determine the skip of value pools while the size will determine how big the value pools in every skip.

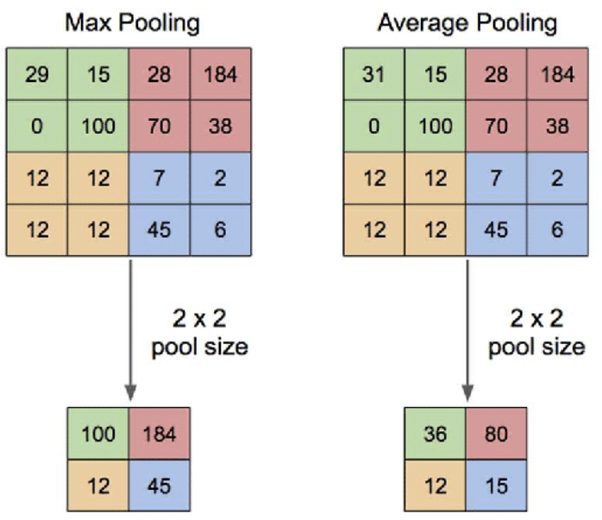


FIG.4.3 POOLING LAYER

### 4.3.3 Fully Connected Layer

### A fully connected CNN (Full ConvNet) is built by stacking Convolutional layers (CONV), Pooling layers (POOL) and at the output a Fully-Connected layer (FC).

The CONV layer consists of a set of learnable filters or kernels. Each filter slides across the input image and compute the discrete convolution between the filter and the receptive field of the input. If the input has multiple channels or depth (For example an RGB image), then the convolution outputs are summed together, followed by the non-linearity.

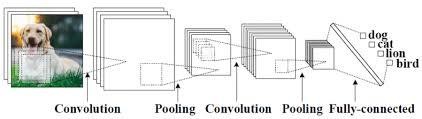


FIG. 4.4 FULL CONVNET ARCHITECTURE

***Activation Function (ReLU and Sigmoid)***

After each convolutional and max pooling operation, we can apply Rectified Linear Unit (ReLU). The ReLU function mimics our neuron activations on a “big enough stimulus” to introduce nonlinearity for values x>0 and returns 0 if it does not meet the condition. This method has been effective to solve diminishing gradients. Weights that are very small will remain as 0 after the ReLU activation function

### *Dropout*

This layer helps to overcome overﬁtting through dropping out a random set of activations and setting them to zero. The network trained with dropout layer must develop more redundancy.

i.e., classify even if some neurons are deactivated. Dropout layer is only used in the training phase.

**4.4 Initial Model**

There is set of recommendations to start the initial CNN architecture. The most common practice suggests to stack multiple times a pattern of convolution layer that is followed by ReLU - activation layer which is followed by a pooling layer. After C-R-P layers a ﬂatten layer is introduced that is followed by at least one fully connected layer that performs the ﬁnal classiﬁcation with ReLU activation function. The initial model used in this study was C-R-PF-FC.

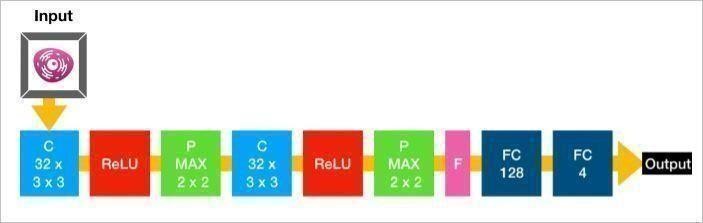


Fig. 4.4 INITIAL MODEL

### 

### 4.5 Final Model

Finding the best architecture for image classiﬁcation is often a trial-and-error process. In this study several architectures and parameters were tested starting.

|  |  |  |
| --- | --- | --- |
| Parameter | Description | Used values |
|  |  |  |
| Optimizer | Algorithm to minimize a loss function and build parametrized models based on data | Adam |
| Loss function | Algorithm that measures error of the model | Binary-cross entropy |
| Stride | The step taken by a filter when convolves around input | 1 |
| Padding | Number of additional pixels of zero that help to preserve the original size | Valid (no padding) |
| Dropout rate | Fraction of the input units that is dropped | 0.3 |
| Batch size | Number images trained per step | 128 |

Table 4.1 OPTIMIZED MODEL PARAMETER

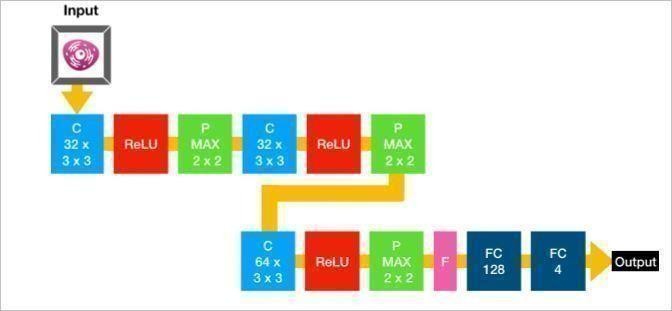
All the parameters were summarized During the training phase 20% of training data was hold for validation. Validation loss was the parameter that was monitored for model optimization.

FIG.4.6 OPTIMIZED MODEL

# 

# CHAPTER 5

# IMPLEMENTATION

**5.IMPLEMENTATION**

**5.1 USER‌ ‌INTERFACE‌ ‌DESIGN‌ ‌**

A‌ ‌user‌ ‌interface,‌ ‌also‌ ‌called‌ ‌a‌ ‌"UI"‌ ‌or‌ ‌simply‌ ‌an‌ ‌"interface,"‌ ‌is‌ ‌the‌ ‌means‌ ‌in‌ ‌which‌ ‌a‌ ‌person‌ ‌controls‌ ‌a‌ ‌software‌ ‌application‌ ‌or‌ ‌hardware‌ ‌device.‌ ‌A‌ ‌good‌ ‌user‌ ‌interface‌ ‌provides‌ ‌a‌ ‌"user-friendly"‌ ‌experience,‌ ‌allowing‌ ‌the‌ ‌user‌ to‌ ‌interact‌ ‌with‌ ‌the‌ ‌software‌ ‌or‌ ‌hardware‌ ‌in‌ ‌a‌ ‌natural‌ ‌and‌ ‌intuitive‌ ‌way.‌

When‌ ‌the ‌user‌ ‌opens‌ ‌the‌ ‌website‌, ‌first‌ of all, ‌the‌ ‌“Home”‌ ‌page‌ ‌is‌ ‌displayed,‌ ‌which‌ ‌has‌ ‌detailed‌ ‌vision‌ ‌on‌ ‌leukaemia‌ ‌for‌ ‌the‌ ‌user.‌ ‌It‌ ‌can‌ ‌provide‌ ‌user‌ ‌about‌ ‌how‌ ‌we‌ ‌are‌ ‌working‌ ‌with‌ ‌microscopic‌ ‌blood‌ ‌cell‌ ‌images‌ ‌to‌ ‌detect‌ ‌a‌ ‌person‌ which ‌is‌ ‌suffering‌ ‌from‌ ‌leukaemia.‌ ‌If‌ ‌the‌ ‌user‌ ‌wants‌ ‌to‌ ‌explore‌ ‌more‌ ‌about‌ ‌the‌ ‌project‌ ‌model‌ ‌starting‌ ‌from‌ ‌zero,‌ ‌he/she‌ ‌can‌ ‌visit‌ ‌“About”‌ ‌page.‌ ‌It‌ ‌layouts‌ ‌the‌ ‌problem‌ ‌statement‌ ‌to‌ ‌begin‌ ‌with‌ ‌and‌ what‌ ‌has‌ ‌stimulated‌ ‌to‌ ‌take‌ ‌over‌ ‌this‌ ‌project‌ .

‌For‌ ‌diagnosis,‌ ‌the‌ ‌user‌ ‌has‌ ‌to‌ ‌go‌ ‌for ‌ “Diagnosis”‌ ‌page‌ ‌and‌ ‌should‌ ‌“Click”‌ ‌the‌ ‌Upload‌ ‌button,‌ when ‌‌he/she‌ ‌can‌ ‌upload‌ ‌microscopic‌ ‌image‌ ‌of‌ ‌blood‌ ‌cells and after‌ ‌submitting‌ ‌the‌ ‌image‌‌,‌ ‌it‌ ‌navigates‌ ‌to‌ ‌page‌ ‌that‌ ‌shows‌ ‌final‌ ‌result‌ ‌whether‌ ‌the‌ ‌image‌ ‌has‌ ‌cancer‌ ‌or‌ ‌not,‌ ‌if‌ ‌detected‌ ‌it‌ ‌displays‌ ‌ “Image‌ ‌has‌ ‌detected‌ ‌cancer”‌ ‌if‌ ‌not‌ ‌it‌ ‌displays‌ ‌ “Image‌ ‌is‌ ‌normal”.

Finally, there‌ ‌is‌ ‌“More‌ ‌info”‌ ‌page‌ ‌contains‌ ‌sources‌ ‌for‌ ‌datasets‌ ‌we‌ ‌used,‌ ‌Presentations‌ ‌we‌ ‌are‌ ‌prepared,‌ ‌GitHub‌ ‌profiles‌ ‌of‌ ‌project‌ ‌members.‌ ‌User‌ ‌can‌ ‌visit‌ ‌this‌ ‌page‌ ‌and‌ ‌get‌ ‌to‌ ‌know‌ ‌more‌ ‌about‌ ‌project‌ ‌and‌ ‌methodology‌ ‌we‌ ‌used‌ ‌in‌ ‌our‌ ‌system‌ ‌and‌ ‌information‌ ‌about‌ ‌contributors.‌ ‌ ‌

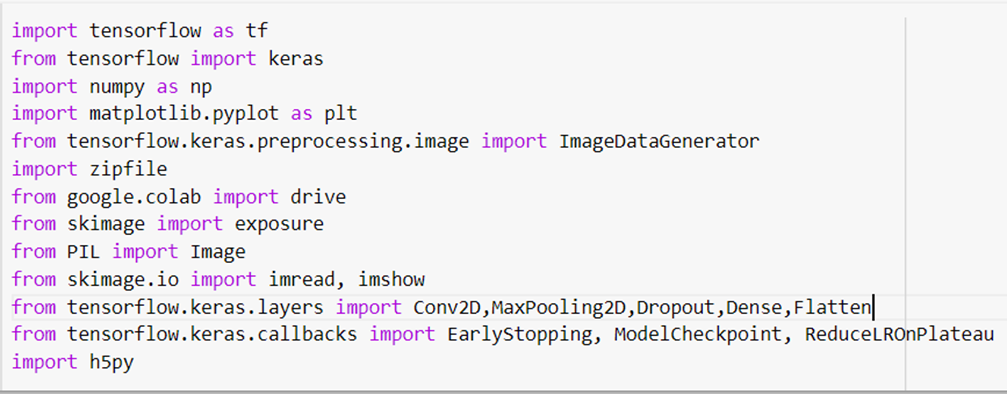
**5.2 User manual**

This project model code was run on the Google Collab note book and this note book provides access to many in built libraries like pandas, shutil, CV2, NumPy etc. Later on, we have to download the Keras, which is a high-level neural networks API to build and mainly used to train the deep learning models and it is capable of running on the top of TensorFlow, CNTK or thearo. we also need tensor flow as it provides a graph protocol buffer and a runtime that executes the distributed graphs.

**5.3 Sample Code:**

The whole project is implemented in python. To understand briefly we use functions that specifies what code does. The functions are:

**5.3.1 IMPORTING NECESSARY LIBRARIES:**



**5.3.1.1 MATPLOTLIB:**

Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python.

**5.3.1.2 NUMPY:**

NumPy is a general-purpose array-processing package. It provides a high-performance multidimensional array object, and tools for working with these arrays. It is the fundamental package for scientific computing with Python.

**5.3.1.3 KERAS:**

Keras is a minimalist Python library for deep learning that can run on top of Theano or TensorFlow. It was developed to make implementing deep learning models as fast and easy as possible for research and development. Keras is a high-level neural networks API, written in Python and capable of running on top of TensorFlow, CNTK, or Theano. ... Use Keras if you need a deep learning library that: Allows for easy and fast prototyping (through user friendliness, modularity, and extensibility).

**5.3.1.4 TENSORFLOW:**

Created by the Google Brain team, TensorFlow is an open-source library for numerical computation and large-scale machine learning. TensorFlow bundles together a slew of machine learning and deep learning (aka neural networking) models and algorithms and makes them useful by way of a common metaphor.

**5.3.1.5 EARLYSTOPPING:**

Keras supports the early stopping of training via a callback called Early Stopping. This callback allows you to specify the performance measure to monitor, the trigger, and once triggered, it will stop the training process. The Early Stopping callback is configured when instantiated via arguments.

**5.3.1.6 MODELCHECKPOINT:**

The Model Checkpoint callback class allows you to define where to checkpoint the model weights, how the file should name and under what circumstances to make a checkpoint of the model. The API allows you to specify which metric to monitor, such as loss or accuracy on the training or validation dataset.

**5.3.2 MODEL BUILDING:**

model = tf.keras.Sequential([

tf.keras.layers.Conv2D(16,(3,3), activation = 'relu',

input\_shape=(img\_width, img\_height, dimensions)),

tf.keras.layers.MaxPooling2D(2,2),

tf.keras.layers.Conv2D(32,(3,3), activation = 'relu'),

tf.keras.layers.MaxPooling2D(2,2),

tf.keras.layers.Conv2D(64,(3,3), activation = 'relu'),

tf.keras.layers.MaxPooling2D(2,2),

tf.keras.layers.Dropout(0.3),

tf.keras.layers.Flatten(),

tf.keras.layers.Dense(512,activation='relu'),

tf.keras.layers.Dense(1,activation = 'sigmoid') ])

**5.3.3 MODEL COMPILATION:**

Keras model provides a method, compile() to compile the model.

**model.compile(loss = 'binary\_crossentropy', optimizer ='adam' , metrics = ['accuracy'])**

**Arguments:**

•**Loss:** Binary cross entropy compares each of the predicted probabilities to actual class output which can be either 0 or 1. It then calculates the score that penalizes the probabilities based on the distance from the expected value. That means how close or far from the actual value.

•**Metrics:** The accuracy metric computes the accuracy rate across all predictions. y\_true represents the true labels while y\_pred represents the predicted ones.

•**Optimizer:** Adam(Adaptive Moment Estimation) works with momentums of first and second order. The intuition behind the Adam is that we don’t want to roll so fast just because we can jump over the minimum, we want to decrease the velocity a little bit for a careful search.

**5.3.4 Image Datagenerator**

train\_datagen = ImageDataGenerator(rescale = 1.0/255.,

                                   zoom\_range = 0.2,

                                   rotation\_range = 40,

                                   horizontal\_flip = True)

                                   fill\_mode ='nearest')

validation\_datagen = ImageDataGenerator(rescale = 1.0/255.)

**CHAPTER 6**

**RESULTS**

### 6.RESULTS

### 6.1 RESULT ANALYSIS

### There are many optimizers available out of which four optimizers are considered and compared as follows:

### 

|  |  |  |  |
| --- | --- | --- | --- |
| S.No | Optimizer | Epochs | Accuracy |
| 1 | Adadelta | 10 | 62.34 |
| 2 | Adadelta | 20 | 68.83 |
| 3 | Adamax | 10 | 70.41 |
| 4 | Adamax | 20 | 79.13 |
| 5 | Nadam | 10 | 82.14 |
| 6 | Nadam | 20 | 87.37 |
| 7 | Adam | 10 | 88.90 |
| 8 | Adam | 20 | 90.09 |

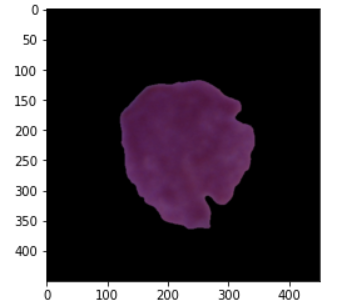
### Table 6.1 Comparison of Results Using different Optimizers

### From the above table, optimizer Adam has the best accuracy when compared to other optimizers. Based on this, an additional epochs are considered and after executing that, it has observed an accuracy of 92.03 at 50th epoch. So, the model with optimizer Adam, is considered as the Optimized Model.

**6.2 Results of Optimized model**

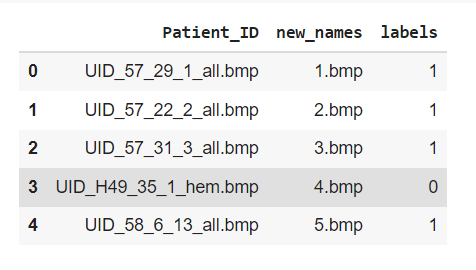
**Output-1:**

Below Image describes the preprocessed Image after Segmentation:



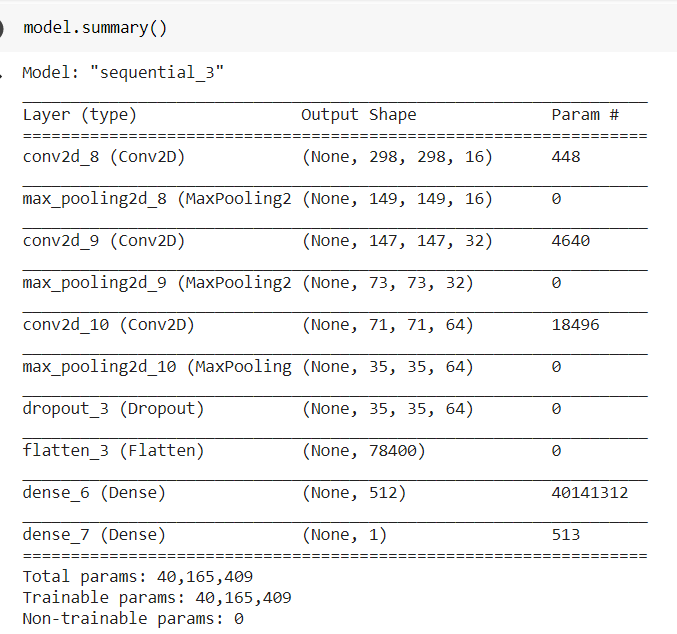
**Output-2:**

The sample validation data in CSV file is displayed as below:



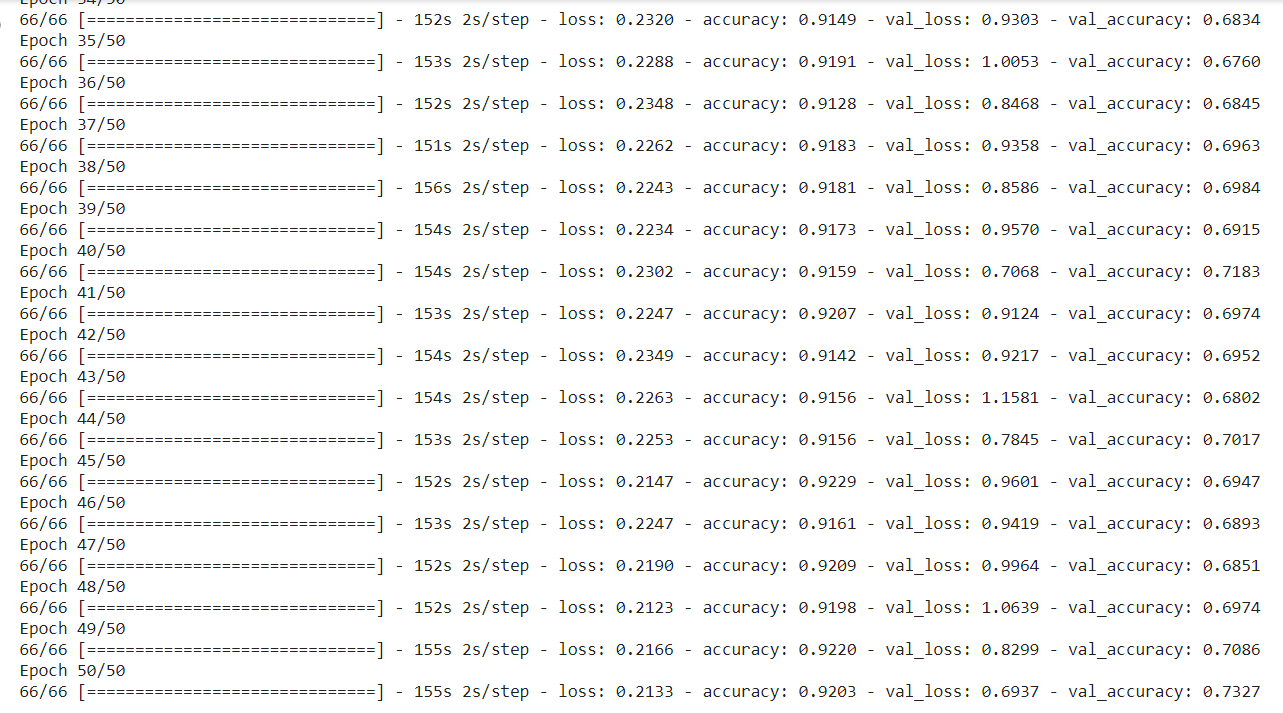
**Output-3:**

The Summary of the CNN model is described as below:



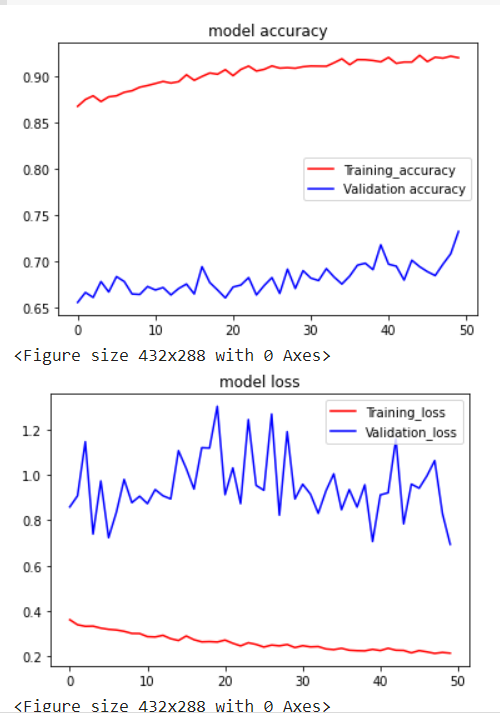
**Output-4:**

Below image describes the epochs of training data accuracy:



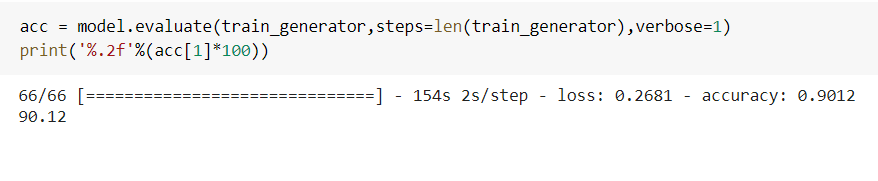
**Output-5:**

Below image describes the model accuracy and validation accuracy graphs:



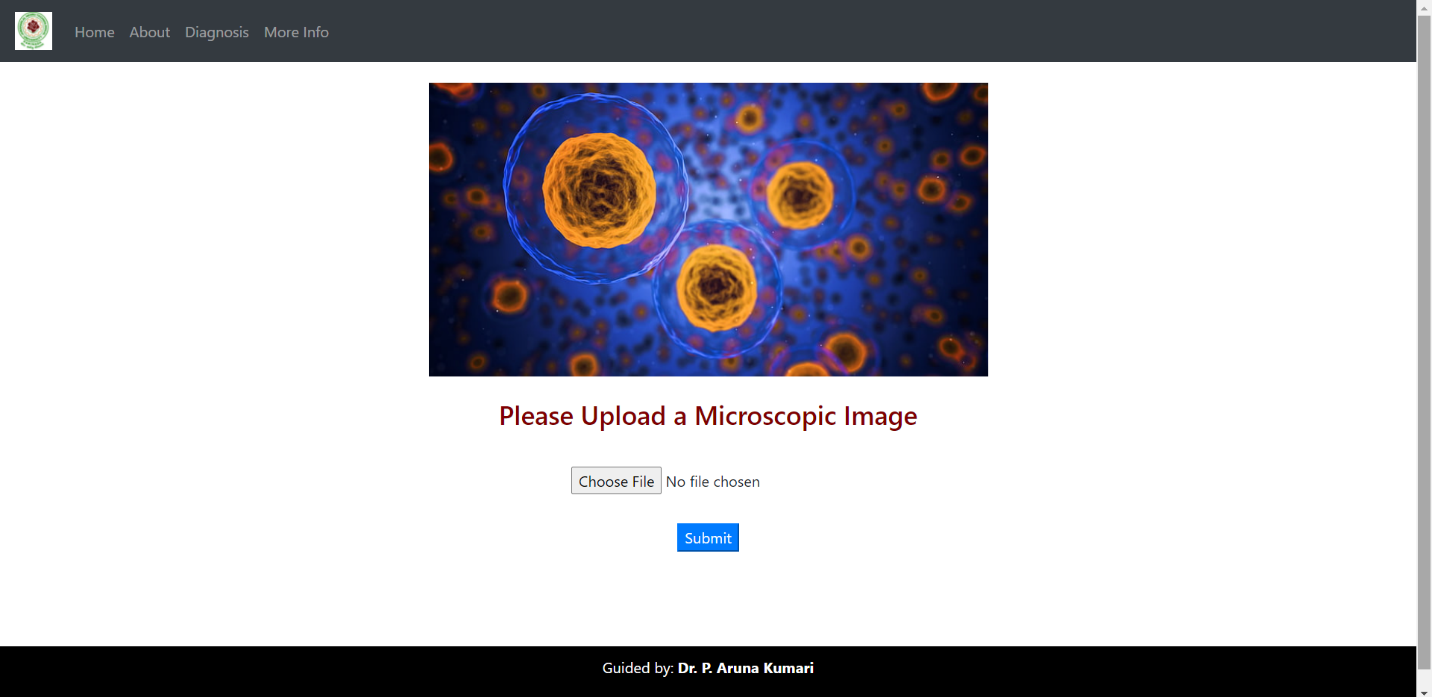
**Output-6:**

The accuracy for the test data of the model is described below:



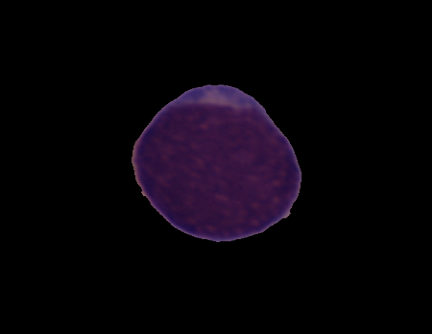
**Output-7:**

User interface for diagnosis:

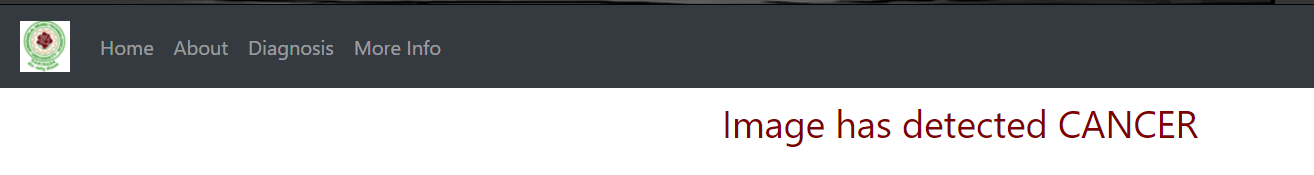


**OUTPUT-8:**

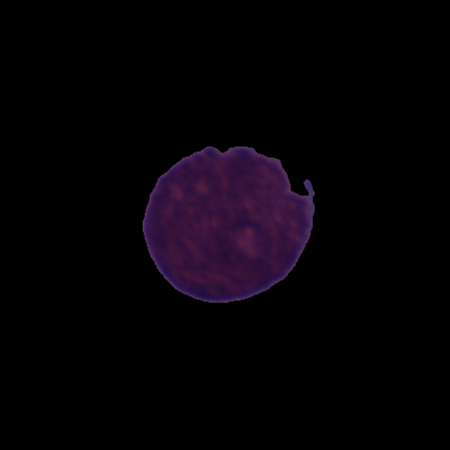
**Input Image:**



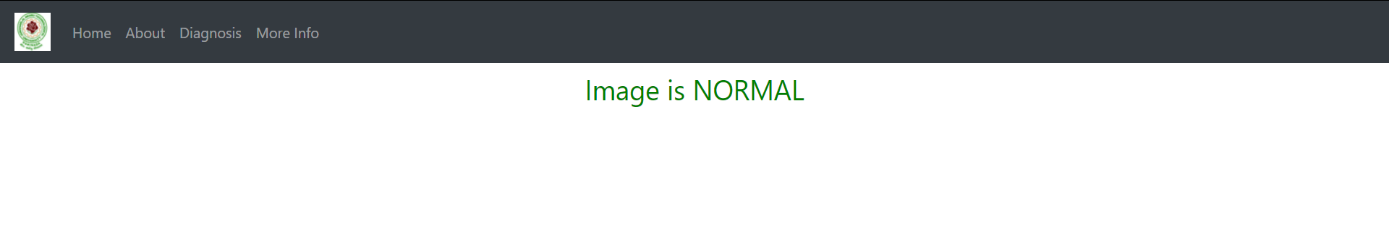
**Output:**



**OUTPUT-9:**



**OUTPUT:**



**CHAPTER-7**

**Conclusion and Future Work**

## 7.CONCLUSION

## 7.1 CONCLUSION

The final model comprised of 3 CRP stacks, flatten layer and 2 dense layers and was trained during 20 epochs on batch size = 128 with Adam as an optimizer. It shows loss and accuracy curves for training and validation sets. The final model scored on average 0.90 in precision and in recall, which means that the final model classifies correctly 90 out of random 100 images and correctly, and 90 images out of 100 images of one class. The final model’s architecture seems working well on this data set. It trained relatively fast, and quickly achieved accuracy of around 90%. This data set is available online on Kaggle.com where one can find that it was trained on more complex architectures with similar final results. This study showed that one can reduce model capacity and therefore computational complexity without compromising accuracy.

**This project is mapped with following PO’S and PO’S:**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **PO1** | **PO2** | **PO3** | **PO4** | **PO5** | **PO6** | **PO7** | **PO8** | **PO9** | **PO10** | **PO11** | **PO12** |
|  |  |  |  |  |  |  |  |  |  |  |  |

|  |  |  |
| --- | --- | --- |
| **PSO1** | **PSO2** | **PSO3** |
|  |  |  |

Signature of Supervisor

(Dr. P. Aruna Kumari)

**7.2 Future Work**

The project can be extended by classifying the types of Cancer. The type of cancers helps a lot in the medical fields with this we can find the what is the problem with the person. With a vast variety of data sets we can also upgrade the program to identify the classification of cancer that show deficiency in their structure and composition. As a result, it can be used to caution the person to recover and take required measures.

**CHAPTER-8**

**REFERENCES**

**8.REFERENCES**

[1]Mini puri Jyoti Mana, used image processing technique using spatial and frequency domain techniques for their paper “A cancer detection technique using image processing - A review”. 2015

[2] Subrajeet Mohapatra1, Dipti Patra1 and Sanghamitra Satpathy, attempted unsupervised technique for their paper and recognized blood cancer cells “Unsupervised Blood Microscopic Image Segmentation and Leukemia Detection using Color based Clustering”. 2012

[3] Patil Tejashri G.& V. B. Raskar, used a unique and robust segmentation

method for blood cancer in lymphoid cells and got successful results for their

paper “Robust Segmentation Method for Acute Lymphoblastic Leukemia Detection”.2015

[4] P. Sobrevilla, E. Montseny & J. Keller, successfully recognized blood cells in bone marrow images “White Blood Cell Detection in Bone Marrow Images”.2016

[5] M. U. Sabino, L. F. Costa∗, E. G. Rizzatti, M. A. Zago, used clor combined

with texure and proved better results for their paper “Toward leukocyte

recognition using morphometry, texture and color”. 2011

[6] Gupta, A, & Gupta, R. (2019). ALL Challenge dataset of ISBI 2019[Dataset]. The Cancer Imaging Archive <https://doi.org/10.7937/tcia.2019.dc4i45r>