A

Major Project Report

On

DETECTION OF BLOOD CANCER

(Submitted in partial fulfillment of the requirements for the award of degree)

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING

By

NALLA VIVEK	**(187R1A0538)**
N. RITESH REDDY	**(187R1A0527)**
SANGARU RACHANA	**(187R1A0548)**
KETHAVATH VAMSHI NAIK	**(187R1A0529)**

Under the Guidance of

Dr. B.LAXMAIAH

(Associate Professor)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CMR TECHNICAL CAMPUS

UGC AUTONOMOUS

(Accredited by NAAC, NBA, Permanently Affiliated to JNTUH, Approved by AICTE, New Delhi)
Recognized Under Section 2(f) & 12(B) of the UGC Act.1956,
Kandlakoya (V), Medchal Road, Hyderabad-501 401.

2018-2022

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

This is to certify that the project entitled "**DETECTION OF BLOOD CANCER**" being submitted by **NALLA VIVEK (187R1A0538), N. RITESH REDDY (187R1A0527), SANGARU RACHANA (187R1A0548), KETHAVATH VAMSHI NAIK (187R1A0529)** in partial fulfillment of the requirements for the award of the degree of B.Tech in Computer Science and Engineering to the Jawaharlal Nehru Technological University Hyderabad, is a record of bonafide work carried out by him/her under the guidance and supervision during the year 2021-2022.

The results embodied in this thesis have not been submitted to any other University or Institution for the award of any degree or diploma.

Dr. B.Laxmaiah
Associate Professor
INTERNAL GUIDE

Dr. A. Raji Reddy
DIRECTOR

Dr. K. Srujan Raju
HOD

EXTERNAL EXAMINER

Submitted for viva voice Examination held on _____

ACKNOWLEDGEMENT

Apart from the efforts of us, the success of any project depends largely on the encouragement and guidelines of many others. We take this opportunity to express our gratitude to the people who have been instrumental in the successful completion of this project. We take this opportunity to express my profound gratitude and deep regard to my guide

Dr. B.Laxmaiah, Assistant Professor for her exemplary guidance, monitoring and constant encouragement throughout the project work. The blessing, help and guidance given by her 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 Project Review Committee (PRC) Coordinators: **Mr. J. Narasimha Rao, Dr. T. S. Mastan Rao, Mr. A. Uday Kiran, Mr. A. Kiran Kumar, Mrs. G. Latha** for their cordial support, valuable information and guidance, which helped us in completing this task through various stages.

We are also thankful to the **Dr. K. Srujan Raju,** Head, Department of Computer Science and Engineering for providing encouragement and support for completing this project successfully.

We are obliged to **Dr. A. Raji Reddy,** Director for being cooperative throughout the course of this project. We also express our sincere gratitude to our Sri. **Ch. Gopal Reddy,** Chairman for providing excellent infrastructure and a nice atmosphere throughout the course of this project.

The guidance and support received from all the members of **CMR Technical Campus** who contributed to the completion of the project. We are grateful for their constant support and help.

Finally, we would like to take this opportunity thank our family for their constant encouragement without which this assignment would not be completed. We sincerely acknowledge and thank all those who gave support directly and indirectly in the completion of this project.

NALLA VIVEK (187R1A0538)
N. RITESH REDDY (187R1A0527)
SANGARU RACHANA (187R1A0548)
KETHAVATH VAMSHI NAIK (187R1A0529)

ABSTRACT

Leukemia is a fatal disease that threatens the lives of many patients. Early detection can effectively improve its rate of remission. This project proposes two automated classification models based on blood microscopic images to detect leukemia by employing transfer learning, rather than traditional approaches that have several disadvantages. In the first model, blood microscopic images are pre-processed; then, features are extracted by a pre-trained deep convolutional neural network MODELS, which makes classifications according to numerous well-known classifiers. In the second model, after pre-processing the images, neural network models are fine-tuned for both feature extraction and classification. Experiments were conducted on a dataset consisting of different images confirming that the second model performs better than the first because of 100% classification accuracy.

LIST OF FIGURES

LIST OF SCREENSHOTS

TABLE OF CONTENTS

1. INTRODUCTION

1. INTRODUCTION

1.1 PROJECT SCOPE

This project is titled as "Detection of Blood Cancer". This project develops a system which detects cancer from blood cell images at a level exceeding practical medical personnel. In this project we have implemented convolutional neural network which extracts features from an image. The input is given to CNN which recognizes blood cells in the image and detects the type of cancer accurately.

1.2 PROJECT PURPOSE

The purpose of our project is to develop a system that can automatically detect cancer from the blood cell images. This system uses a convolutional neural network that inputs a blood cell images and outputs whether the cell is infected with cancer or not.

1.3 PROJECT FEATURES

The main features of this project are that the designer now functions as a problem solver and tries to sort out the difficulties that the enterprise faces. The solutions are given as proposals. The proposal is then weighed with the existing system analytically and the best one is selected. The proposal is presented to the user for an endorsement by the user. The proposal is reviewed on user request and suitable changes are made. This is loop that ends as soon as the user is satisfied with proposal.

2. LITERATURE SURVEY

2. LITERATURE SURVEY

2.1 MACHINE LEARNING

Machine learning is a type of artificial intelligence that allows software applications to become more accurate at predicting outcomes without being explicitly programmed. Machine Learning algorithms use historical data as input to predict new output values. Optimization techniques is required to get accurate results. Optimization techniques involve tuning the hyperparameters to reach an optimum result. As it trains over the examples, again and again, it is able to identify patterns in order to make decisions more accurately. Whenever any new input is introduced to the ML model, it applies its learned patterns over the new data to make future predictions. Based on the final accuracy, one can optimize their models using various standardized approaches. In this way, Machine Learning model learns to adapt to new examples and produce better results.

Types of Machine Learning Algorithms:

- Supervised Learning.
- Unsupervised Learning.
- Reinforcement Learning.

2.1.1 SUPERVISED LEARNING

Supervised learning is the most popular paradigm for machine learning. It is the easiest to understand and the simplest to implement. It is the task of learning a function that maps an input to an output based on example input-output pairs. It infers a function from labelled training data consisting of a set of training examples. In supervised learning, each example is a pair consisting of an input object (typically a vector) and a desired output value (also called the supervisory signal). A supervised learning algorithm analyses the training data and produces an inferred function, which can be used for mapping new examples. Supervised Learning is very similar to teaching a child with the given data and that data is in the form of examples with labels, we can feed a learning algorithm with these example-label pairs one by one, allowing the algorithm to predict the right answer or not. Over time, the algorithm will learn to approximate the exact nature of the relationship between examples and their labels. When fully trained, the supervised learning algorithm will be able to observe a new, never-before-seen example and predict a good label for it.

Two types of Supervised Learning are:

- **Regression**: Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting. Regression can be used to estimate/ predict continuous values (Real valued output). For example, given a picture of a person then we have to predict the age on the basis of the given picture.
- **Classification**: Classification means to group the output into a class. If the data is discrete or categorical then it is a classification problem. For example, given data about the sizes of houses in the real estate market, making our output about whether the house "sells for more or less than the asking price" i.e., Classifying houses into two discrete categories.

2.1.2 UNSUPERVISED LEARNING

Unsupervised Learning is a machine learning technique, where you do not need to supervise the model. Instead, you need to allow the model to work on its own to discover information. It mainly deals with the unlabeled data and looks for previously undetected patterns in a data set with no pre-existing labels and with a minimum of human supervision. In contrast to supervised learning that usually makes use of human-labelled data, unsupervised learning, also known as self-organization, allows for modelling of probability densities over inputs. Unsupervised machine learning algorithms infer patterns from a dataset without reference to known, or labelled outcomes. It is the training of machine using information that is neither classified nor labelled and allowing the algorithm to act on that information without guidance. Here the task of machine is to group unsorted information according to similarities, patterns, and differences without any prior training of data. Unlike supervised learning, no teacher is provided that means no training will be given to the machine. Therefore, machine is restricted to find the hidden structure in unlabeled data by our-self. For example, if we provide some pictures of dogs and cats to the machine to categorized, then initially the machine has no idea about the features of dogs and cats so it categorizes them according to their similarities, patterns and differences. The Unsupervised Learning algorithms allows you to perform more complex processing tasks compared to supervised learning. Although, unsupervised learning can be more unpredictable compared with other natural learning methods.

Unsupervised learning problems are classified into two categories of algorithms:

- **Clustering:** A clustering problem is where you want to discover the inherent groupings in the data, such as grouping customers by purchasing behavior.
- **Association:** An association rule learning problem is where you want to discover rules that describe large portions of your data, such as people that buy X also tend to buy Y.

2.1.3 REINFORCEMENT LEARNING:

Reinforcement Learning (RL) is a type of machine learning technique that enables an agent to learn in an interactive environment by trial and error using feedback from its own actions and experiences. Machine mainly learns from past experiences and tries to perform best possible solution to a certain problem. It is the training of machine learning models to make a sequence of decisions. Though both supervised and reinforcement learning use mapping between input and output, unlike supervised learning where the feedback provided to the agent is correct set of actions for performing a task, reinforcement learning uses rewards and punishments as signals for positive and negative behavior. Reinforcement learning is currently the most effective way to hint machine's creativity.

2.2 NEURAL NETWORKS

Neural Network (or Artificial Neural Network) has the ability to learn by examples. ANN is an information processing model inspired by the biological neuron system. ANN biologically inspired simulations that are performed on the computer to do a certain specific set of tasks like clustering, classification, pattern recognition etc. It is composed of a large number of highly interconnected processing elements known as the neuron to solve problems. It follows the non-linear path and process information in parallel throughout the nodes. A neural network is a complex adaptive system. Adaptive means it has the ability to change its internal structure by adjusting weights of inputs. Artificial Neural Networks can be best viewed as weighted directed graphs, where the nodes are formed by the artificial neurons and the connection between the neuron outputs and neuron inputs can be represented by the directed edges with weights. The ANN receives the input signal from the external world in the form of a pattern and image in the form of a vector. These inputs are then mathematically designated by the notations x(n) for every n number of inputs. Each of the input is then multiplied by its corresponding weights (these weights are the details used by the artificial neural networks to solve a certain problem). These weights typically represent the strength of the interconnection amongst neurons inside the artificial neural network. All the weighted inputs are summed up inside the computing unit.

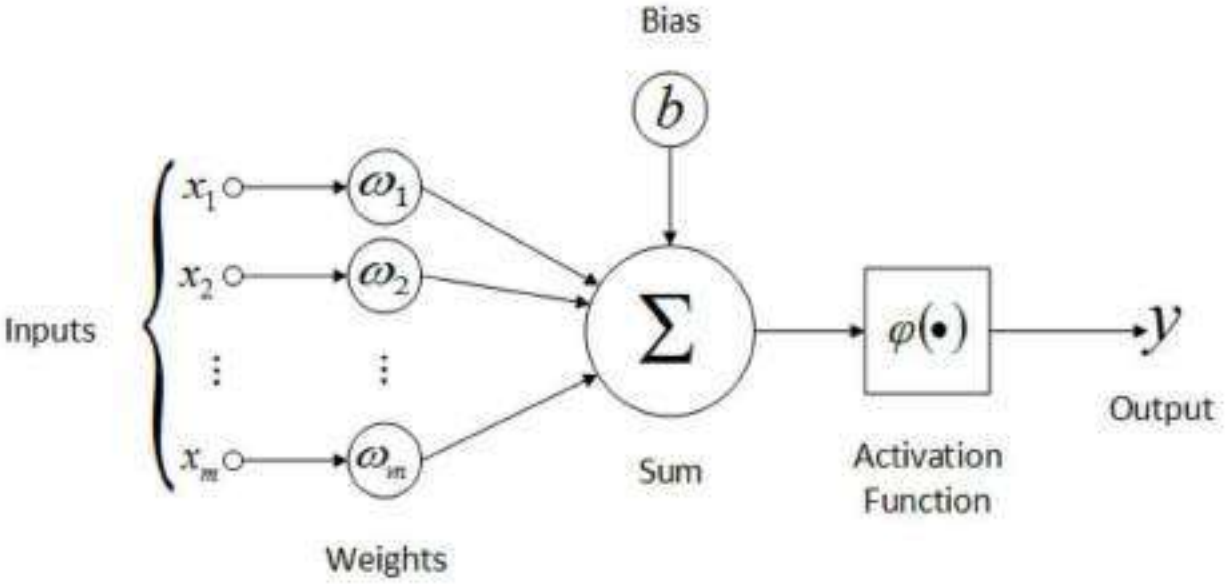


Figure 2.1: Artificial Neuron Model

The Artificial Neural Network contains three layers. They are:

- **Input Layer:** The input layers contain those artificial neurons (termed as units) which are to receive input from the outside world. This is where the actual learning on the network happens or corresponding happens else it will process.
- **Hidden Layer:** The hidden layers are mentioned hidden in between input and the output layers. The only job of a hidden layer is to transform the input into something meaningful that the output layer/unit can use in some way. Most of the artificial neural networks are all interconnected, which means that each of the hidden layers is individually connected to the neurons in its input layer and also to its output layer leaving nothing to hang in the air. This makes it possible for a complete learning process and also learning occurs to the maximum when the weights inside the artificial neural network get updated after each iteration.
- **Output Layer:** The output layers contain units that respond to the information that is fed into the system and also whether it learned any task or not.

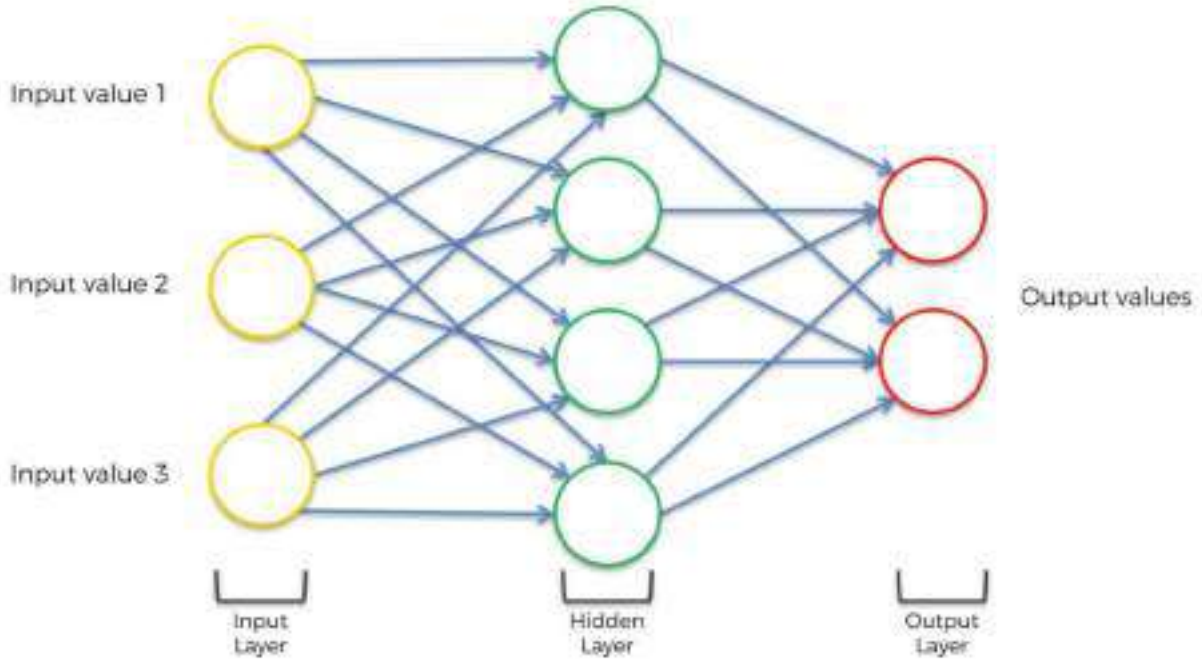


Figure 2.2: Basic Neural Network

2.3 DEEP LEARNING

Deep learning is a branch of machine learning which is completely based on artificial neural networks. 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. It has a greater number of hidden layers and known as deep neural learning or deep neural network. Deep learning has evolved hand-in-hand with the digital era, which has brought about an explosion of data in all forms and from every region of the world. This data, known simply as big data, is drawn from sources like social media, internet search engines, e-commerce platforms, and online cinemas, among others. This enormous amount of data is readily accessible and can be shared through fintech applications like cloud computing. However, the data, which normally is unstructured, is so vast that it could take decades for humans to comprehend it and extract relevant information. Companies realize the incredible potential that can result from unravelling this wealth of information and are increasingly adapting to AI systems for automated support. Deep learning learns from vast amounts of unstructured data that would normally take humans decades to understand and process. Deep learning and utilizes a hierarchical level of

artificial neural networks to carry out the process of machine learning. The artificial neural networks are built like the human brain, with neuron nodes connected like a web. While traditional programs build analysis with data in a linear way, the hierarchical function of deep learning systems enables machines to process data with a nonlinear approach.

2.4 DEEP NEURAL NETWORK

It is a neural network with a certain level of complexity (having multiple hidden layers in between input and output layers). They are capable of modelling and processing non-linear relationships.

2.5 CONVOLUTIONAL NEURAL NETWORK

Convolutional neural networks are specialized deep neural networks which can process the data that has input shape like a 2d matrix. Images are easily represented as a 2d matrix and CNN is very useful in working with images. CNN is basically used for image classifications and identifying if an image is a bird, a plane or superman, etc. CNNs are regularized versions of multilayer perceptron's. Multilayer perceptron's usually mean fully connected networks, that is, each neuron in one layer is connected to all neurons in the next layer. The "full connectivity" of these networks makes them prone to overfitting data. Typical ways of regularization, or preventing overfitting, include: penalizing parameters during training (such as weight decay) or trimming connectivity (skipped connections, dropout, etc.) CNNs take a different approach towards regularization: they take advantage of the hierarchical pattern in data and assemble patterns of increasing complexity using smaller and simpler patterns embossed in their filters. Therefore, on a scale of connectivity and complexity, CNNs are on the lower extreme. Convolutional networks were inspired by biological processes in that the connectivity pattern between neurons resembles the organization of the animal visual cortex. Individual cortical neurons respond to stimuli only in a restricted region of the visual field known as the receptive field. The receptive fields of different neurons partially overlap such that they cover the entire visual field. CNNs use relatively little pre-processing compared to other image classification algorithms. This means that the network learns to optimize the filters (or kernels) through automated learning, whereas in traditional algorithms these filters are hand engineered. This independence from prior knowledge and human intervention in feature extraction is a major advantage.

2.6 DENSENET

A Densenet is a type of convolutional neural network that utilises dense connections between layers, through Dense Blocks, where we connect all layers (with matching feature-map sizes) directly with each other. Densenet is a new CNN architecture that reached State-Of-The-Art (SOTA) results on classification datasets (CIFAR, SVHN, ImageNet) using less parameters. Thanks to its new use of residual it can be deeper than the usual networks and still be easy to optimize.

2.7 INCEPTION RESNET V2

Inception-resnet-v2 is a convolutional neural architecture that builds on the Inception family of architectures but incorporates residual connections. Inception-resnet-v2 is a convolutional neural network that is trained on more than a million images from the ImageNet database [1]. The network is 164 layers deep and can classify images into 1000 object categories, such as keyboard, mouse, pencil, and many animals. ResNet50 is a variant of Resnet model which has 48 Convolution layers along with 1-maxpool and 1-average pool layer. ResNet-50 is a convolutional neural network that is 50 layers deep. You can load a pretrained version of the network trained on more than a million images from the ImageNet database. The pretrained network can classify images into 1000 object categories, such as keyboard, mouse, pencil, and many animals.

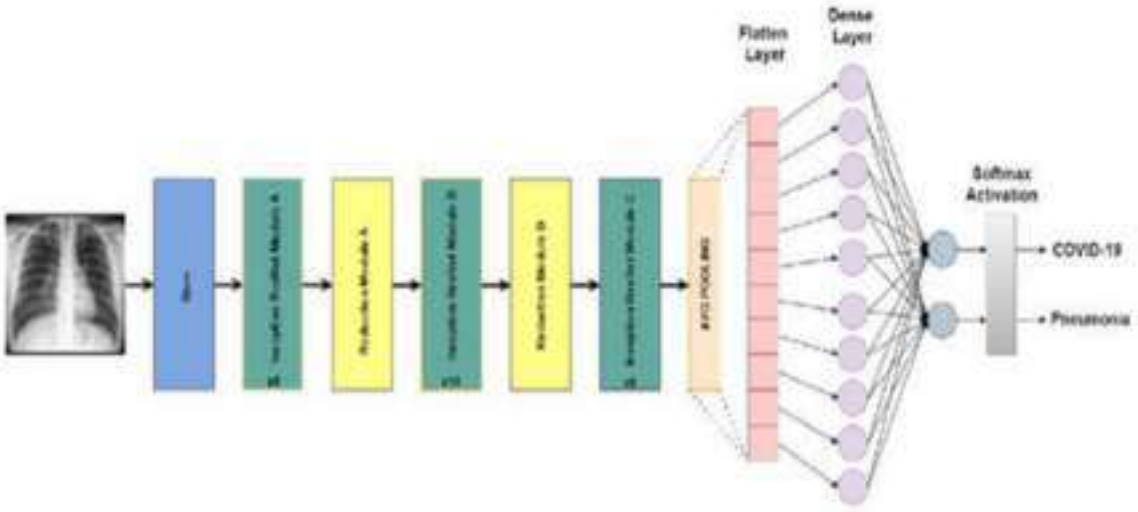


Figure 2.3: Inception Resnet V2 Model

2.8 INCEPTION V3 MODEL

Inception v3 is an image recognition model that has been shown to attain greater than 78.1% accuracy on the ImageNet dataset. The model is the culmination of many ideas developed by multiple researchers over the years. It is based on the original paper: "Rethinking the Inception Architecture for Computer Vision" by Szegedy, et. al. The model itself is made up of symmetric and asymmetric building blocks, including convolutions, average pooling, max pooling, concatenations, dropouts, and fully connected layers. Batch normalization is used extensively throughout the model and applied to activation inputs. Loss is computed using SoftMax.

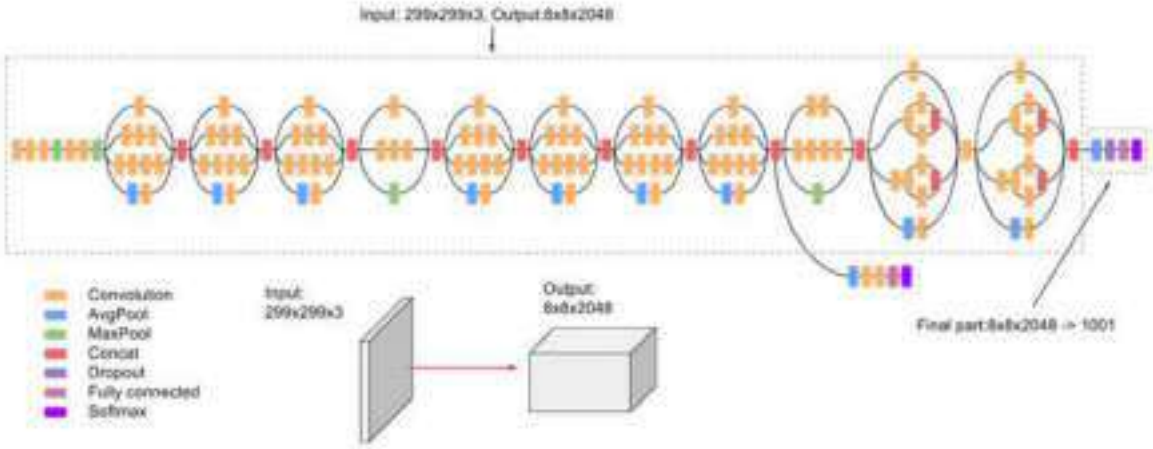


Figure 2.4: Inception V3 Model

2.9 DIAGNOSING LEUKEMIA IN BLOOD SMEAR IMAGES USING AN ENSEMBLE OF CLASSIFIERS AND PRE-TRAINED CONVOLUTIONAL NEURAL NETWORKS

Leukemia is a worldwide disease. In this paper we demonstrate that it is possible to build an automated, efficient and rapid leukemia diagnosis system. We demonstrate that it is possible to improve the precision of current techniques from the literature using the description power of well-known Convolutional Neural Networks (CNNs). We extract features from a blood smear image using pre-trained CNNs in order to obtain an unique image description. Many feature selection techniques were evaluated and we chose PCA to select the features that are in the final descriptor.

To classify the images on healthy and pathological we created an ensemble of classifiers with three individual classification algorithms (Support Vector Machine, Multilayer Perceptron and Random Forest). In the tests we obtained an accuracy rate of 100%. Besides the high accuracy rate, the tests showed that our approach requires less processing time than the methods analyzed in this paper, considering the fact that our approach does not use segmentation to obtain specific cell regions from the blood smear image.

2.10 GLAUCOMA DETECTION BASED ON DEEP CONVOLUTIONAL NEURAL NETWORK

Glaucoma is a chronic and irreversible eye disease, which leads to deterioration in vision and quality of life. In this paper, we develop a deep learning (DL) architecture with convolutional neural network for automated glaucoma diagnosis. Deep learning systems, such as convolutional neural networks (CNNs), can infer a hierarchical representation of images to discriminate between glaucoma and non-glaucoma patterns for diagnostic decisions. The proposed DL architecture contains six learned layers: four convolutional layers and two fully-connected layers. Dropout and data augmentation strategies are adopted to further boost the performance of glaucoma diagnosis. Extensive experiments are performed on the ORIGA and SCES datasets. The results show area under curve (AUC) of the receiver operating characteristic curve in glaucoma detection at 0.831 and 0.887 in the two databases, much better than state-of-the-art algorithms. The method could be used for glaucoma detection.

2.11 MULTI-RESOLUTION-TRACT CNN WITH HYBRID PRE-TRAINED AND SKIN-LESION TRAINED LAYERS

Correctly classifying a skin lesion is one of the first steps towards treatment. We propose a novel convolutional neural network (CNN) architecture for skin lesion classification designed to learn based on information from multiple image resolutions while leveraging pretrained CNNs. While traditional CNNs are generally trained on a single resolution image, our CNN is composed of multiple tracts, where each tract analyzes the image at a different resolution simultaneously and learns interactions across multiple image resolutions using the same field-of-view. We convert a CNN, pretrained on a single resolution, to work for multi-resolution input. The entire network is fine-tuned in a fully learned end-to-end optimization with auxiliary loss functions. We show how

our proposed novel multi-tract network yields higher classification accuracy, outperforming state-of-the-art multi-scale approaches when compared over a public skin lesion dataset.

3. SYSTEM ANALYSIS

3. SYSTEM ANALYSIS

3.1 PROBLEM DEFINTION

According to the Leukemia & Lymphoma Society, one person in the U.S. is diagnosed with blood cancer approximately every 3 minutes and an estimated total of 174,250 people in the U.S. are expected to be diagnosed with leukemia, lymphoma or myeloma in 2018. The estimated new cases in 2019 are around 61,780 and according to the National Cancer Institute, the percentage of all new cancer cases is 3.5 percent. As in acute leukemia, if the treatment is not done in a precise time, the person died within a few months. And it is very necessary to detect cancer in the early stages to treat this type of cancer or any type of cancer. It takes more time and effort to do the detection process by technicians manually and it costs more with the help of the instrument.

3.2 EXISTING SYSTEM

Many traditional computer-aided systems use image processing and machine-learning techniques that usually involve several steps, including pre-processing, segmentation, feature extraction, and classification. However, the success of each step depends on the success of the preceding step. For example, the success of classification depends on the success of the preceding feature extraction, which itself depends on the success of the preceding segmentation. Hence, high classification accuracy requires the success of all steps, each of which is non-trivial and problem-dependent.

3.2.1 DISADVANTAGES OF EXISTING SYSTEM

- Human expert is needed for counting the blood cells.
- High classification accuracy requires the success of all steps.

3.3 PROPOSED SYSTEM

This paper proposes two classification models that are based on transfer learning and can distinguish between healthy and unhealthy blood smear images with high accuracy. These models employ RESNET50, which is a deep CNN that achieved huge success in the image classification challenge, ImageNet.

3.3.1 ADVANTAGES OF PROPOSED SYSTEM

- No Manual Effort is needed for testing blood cancer and counting the number of cells in blood.
- This technology can improve healthcare delivery and increase access to medical imaging expertise in parts of the world where access to skilled medical personnel is limited

3.4 FEASIBILITY STUDY

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. Three key considerations involved in the feasibility analysis are

- Economic Feasibility
- Technical Feasibility
- Behavioral Feasibility

3.4.1 ECONOMIC FEASIBILITY

The developing system must be justified by cost and benefit. Criteria to ensure that effort is concentrated on project, which will give best, return at the earliest. One of the factors, which affect the development of a new system, is the cost it would require. The following are some of the important financial questions asked during preliminary investigation:

- The costs conduct a full system investigation.
- The cost of the hardware and software.
- The benefits in the form of reduced costs or fewer costly errors.

Since the system is developed as part of project work, there is no manual cost to spend for the proposed system. Also, all the resources are already available, it gives an indication of the system is economically possible for development.

3.4.2 TECHNICAL FEASIBILITY

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

3.4.3 BEHAVIORAL FEASIBILITY

This includes the following questions:

- Is there sufficient support for the users?
- Will the proposed system cause harm?

The project would be beneficial because it satisfies the objectives when developed and installed. All behavioral aspects are considered carefully and conclude that the project is behaviorally feasible.

3.5 HARDWARE AND SOFTWARE REQUIREMENTS

3.5.1 HARDWARE REQUIREMENTS

Hardware interfaces specifies the logical characteristics of each interface between the software product and the hardware components of the system. The following are some hardware requirements.

- System : Intel Core i3
- Hard Disk : 10 GB or more
- Input Devices : Keyboard, Mouse
- Ram : 4GB

3.5.2 SOFTWARE REQUIREMENTS

Software Requirements specifies the logical characteristics of each interface and software components of the system. The following are some software requirements.

- Operating system : Windows 7 or higher
- Coding Language : Python
- Tool : Jupyter Notebook or Anaconda

4. ARCHITECTURE

4. ARCHITECTURE

4.1 PROJECT ARCHITECTURE

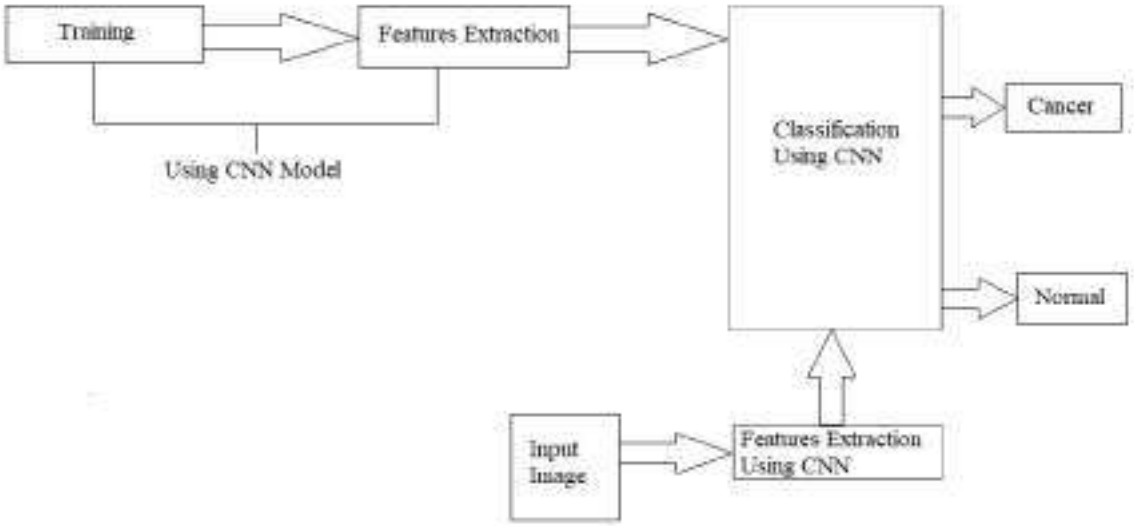


Figure 4.1: System Architecture

4.2 DATASET

Images used in this project were obtained from Kaggle dataset which is a public dataset available online. This dataset was divided into 2 classes. There was total 4961 training images where 2483 images were from healthy patients and 2478 images were from patients affected with blood cancer. We tested the model with total 1240 images 620 from each class. These images had resolution of 320*240.

4.3 FEATURE EXTRACTION

Feature extraction is a type of dimensionality reduction where a large number of pixels of the image are efficiently represented in such a way that interesting parts of the image are captured effectively. The process of feature extraction is useful when you need to reduce the number of resources needed for processing without losing important or relevant information. Feature extraction can also reduce the amount of redundant data for a given analysis. Also, the reduction of the data and the machine's efforts in building variable combinations (features) facilitate the speed of learning and generalization steps in the machine learning process.

4.4 CLASSIFICATION USING CNN

Neural networks are used in the automatic detection of cancer in blood samples. Neural network is chosen as a classification tool due to its well-known technique as a successful classifier for many real applications. The training and validation processes are among the important steps in developing an accurate process model using CNN's. The dataset for training and validation processes consists of two parts; the training features set which are used to train the CNN model; whilst a testing features sets are used to verify the accuracy of the trained using the feed- forward back propagation network. In the training part, connection weights were always updated until they reached the defined iteration Number or suitable error. Neural networks are used in the automatic detection of cancer in blood samples. Neural network is chosen as a classification tool due to its well-known technique as a successful classifier for many real applications. The training and validation processes are among the important steps in developing an accurate process model using CNN's.

4.5 USE CASE DIAGRAM

A use case diagram is a graphical depiction of a user's possible interactions with a system. A use case diagram shows various use cases and different types of users the system has and will often be accompanied by other types of diagrams as well. The use cases are represented by either circles or ellipses. The actors are often shown as stick figures.

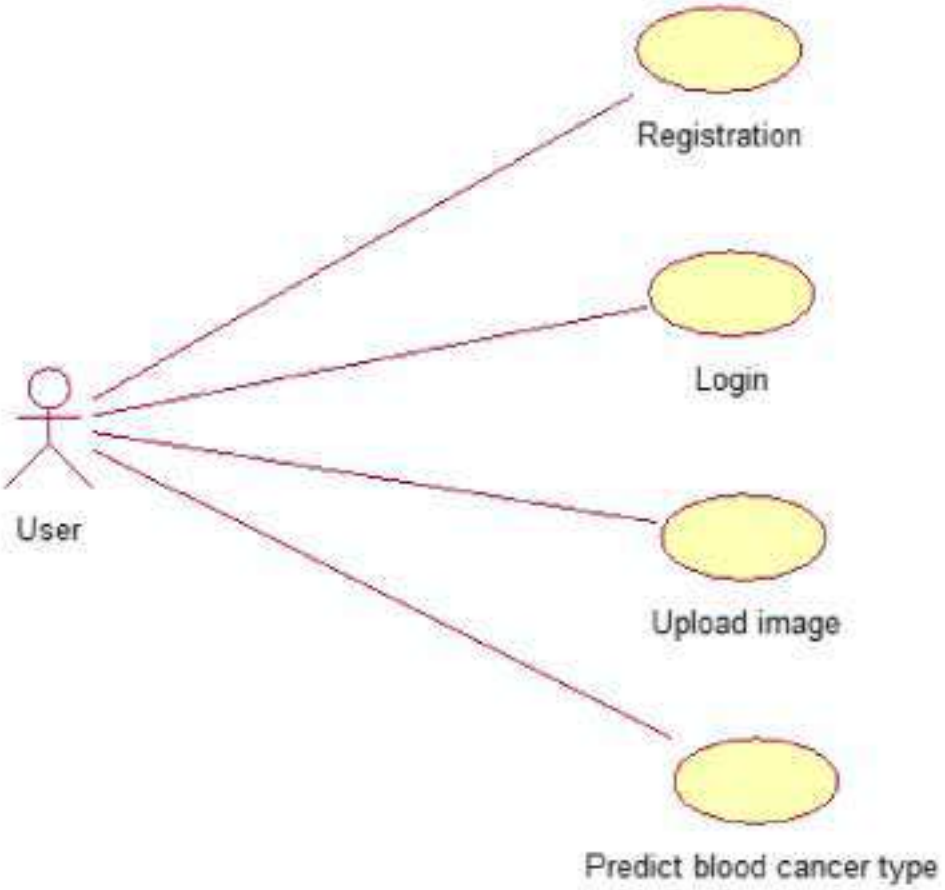


Figure 4.2: Use Case Diagram

4.6 CLASS DIAGRAM

A class diagram in the Unified Modeling Language is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations, and the relationships among objects.

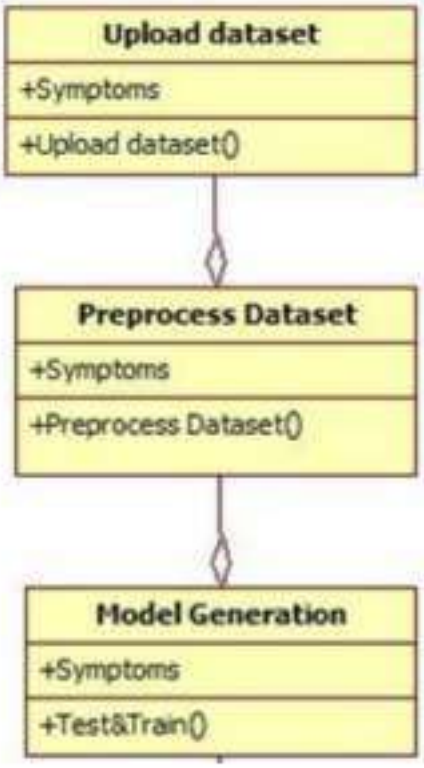


Figure 4.3: Class Diagram

4.7 SEQUENCE DIAGRAM

A sequence diagram or system sequence diagram shows object interactions arranged in time sequence in the field of software engineering. It depicts the objects involved in the scenario and the sequence of messages exchanged between the objects needed to carry out the functionality of scenario.

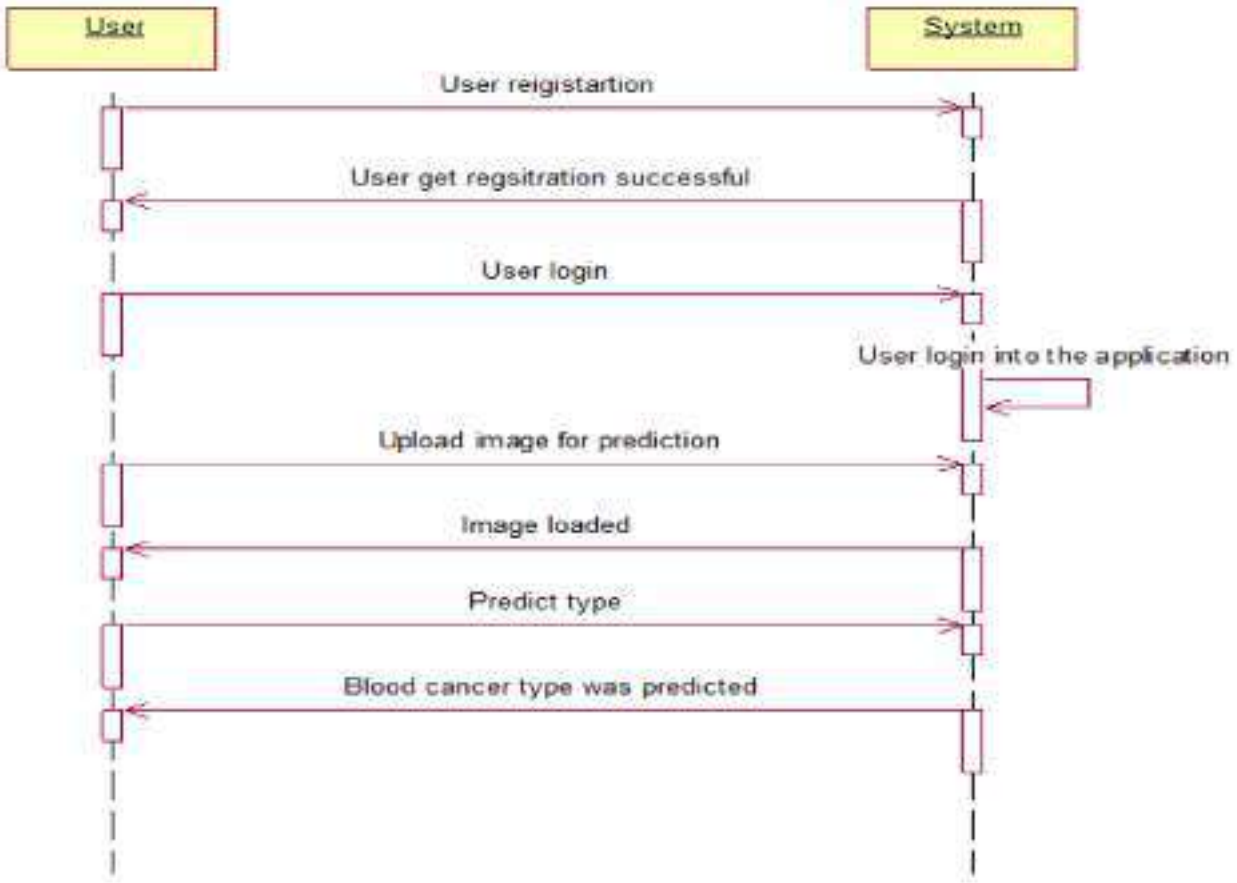


Figure 4.4: Sequence Diagram

4.8 ACTIVITY DIAGRAM

Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams are intended to model both computational and organizational processes (i.e., workflows), as well as the data flows intersecting with the related activities. Although activity diagrams primarily show the overall flow of control, they can also include elements showing the flow of data between activities through one or more data stores.

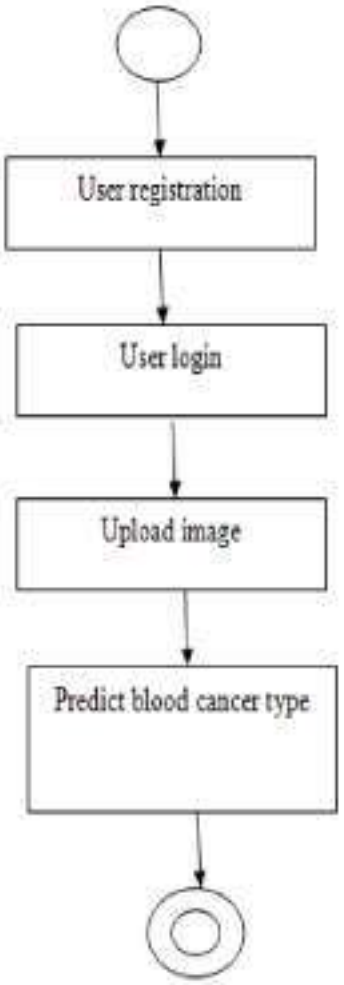


Figure 4.5: Activity Diagram

4.9 DATA FLOW DIAGRAM

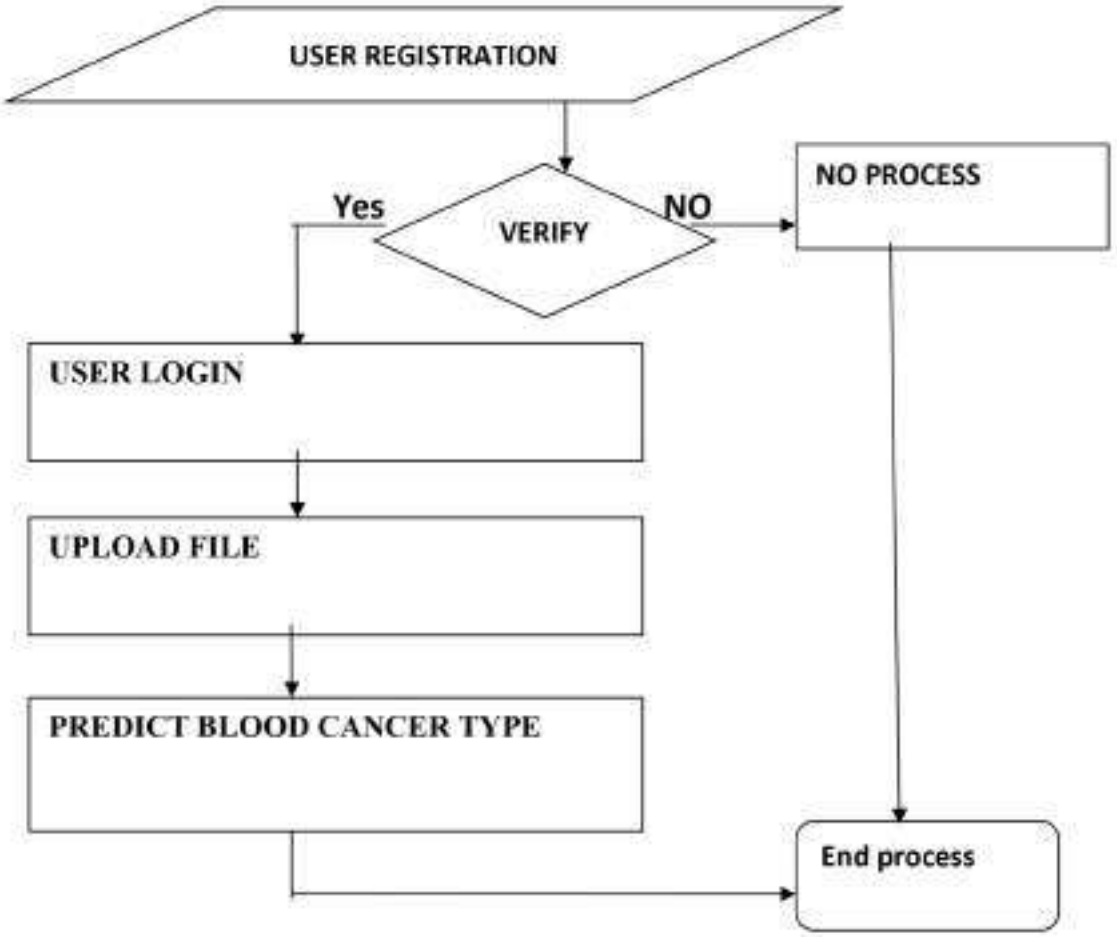


Figure 4.6: Data Flow Diagram

The Data flow diagram is also called as bubble chart. It is a simple graphical formalism that can be used to represent a system in terms of input data to the system, various processing carried out on this data, and the output data is generated by this system. The data flow diagram is one of the most important modeling tools. It is used to model the system components. These components are the system process, the data used by the process, an external entity that interacts with the system and the information flows in the system. Data flow diagram shows how the information moves through the system and how it is modified by a series of transformations. It is a graphical technique that depicts information flow and the transformations that are applied as data moves from input to output.

5. IMPLEMENTATION

5. IMPLEMENTATION

Sample Code:

```
import tensorflow
tensorflow.__version__
from tensorflow.keras.layers import Dense, Flatten, Input, Lambda
from tensorflow.keras.models import Model
from tensorflow.keras.applications.vgg16 import VGG16
from tensorflow.keras.applications.resnet50 import ResNet50
from tensorflow.keras.applications.resnet50 import preprocess_input
from tensorflow.keras.applications.vgg16 import preprocess_input
from tensorflow.keras.preprocessing import image
from tensorflow.keras.applications.vgg19 import preprocess_input
from tensorflow.keras.applications.inception_v3 import InceptionV3
from tensorflow.keras.preprocessing import image
from tensorflow.keras.models import Model
from tensorflow.keras.layers import Dense, GlobalAveragePooling2D
from tensorflow.keras.preprocessing import image
from tensorflow.keras.preprocessing.image import ImageDataGenerator, load_img
from tensorflow.keras.models import Sequential
from glob import glob
import matplotlib.pyplot as plt
import numpy as np
import tensorflow as tf
#SEARCHING FOR P100
import os
import time
x=!nvidia-smi
count=0
```

```
for i in x:
    if "==============" in i:
        count+=1
        break
    count+=1
if 'p100' in x[count].lower():
    print("found")
else:
    print(x[count])
    time.sleep(1)
    #os._exit(00)
import tensorflow
from tensorflow.python.client import device_lib
print(device_lib.list_local_devices())
# Resizinig all the images to (224,224)
IMAGE_SIZE = [224,224]
train_path = 'train'
test_path = 'test'
# Scaling all the images between 0 to 1
train_datagen=ImageDataGenerator(rescale=1./255,shear_range=0.2,zoom_range=0.2,horizontal_f
lip=False)
test_datagen = ImageDataGenerator(rescale=1./255)
train_set = train_datagen.flow_from_directory(train_path,
                                              target_size=(224,224),
                                              batch_size=32,
                                              class_mode = 'categorical')


test_set = test_datagen.flow_from_directory(test_path,
                                            target_size=(224,224),
```

```
                                batch_size=32,
                                class_mode='categorical')
resnet = ResNet50(input_shape = IMAGE_SIZE + [3], weights='imagenet', include_top=False)
for layer in resnet.layers:
  layer.trainable = False
x = Flatten()(resnet.output)
prediction = Dense(4, activation='softmax')(x)
model = Model(inputs = resnet.inputs, outputs = prediction)
model.summary()
model.compile(loss = 'categorical_crossentropy', optimizer='adam', metrics=['accuracy'])
callback = tf.keras.callbacks.EarlyStopping(monitor='loss', patience=3)
hist=model.fit(train_set,validation_data=test_set,epochs=20,steps_per_epoch=1,validation_steps=1
,callbacks=[callback])
model.save('model1.h5')
import matplotlib.pyplot as plt
x=hist
plt.figure(figsize=(20,10))
plt.subplot(1, 2, 1)
plt.suptitle('Optimizer : adam', fontsize=10)
plt.ylabel('Loss', fontsize=16)
plt.plot(x.history['loss'], label='Training Loss')
plt.plot(x.history['val_loss'], label='Validation Loss')
plt.legend(loc='upper right')
plt.subplot(1, 2, 2)
plt.ylabel('Accuracy', fontsize=16)
plt.plot(x.history['accuracy'], label='Training Accuracy')
plt.plot(x.history['val_accuracy'], label='Validation Accuracy')
plt.legend(loc='lower right')
plt.show()
```

6. SCREENSHOTS

6. SCREENSHOTS

Figure 6.1: Home Page

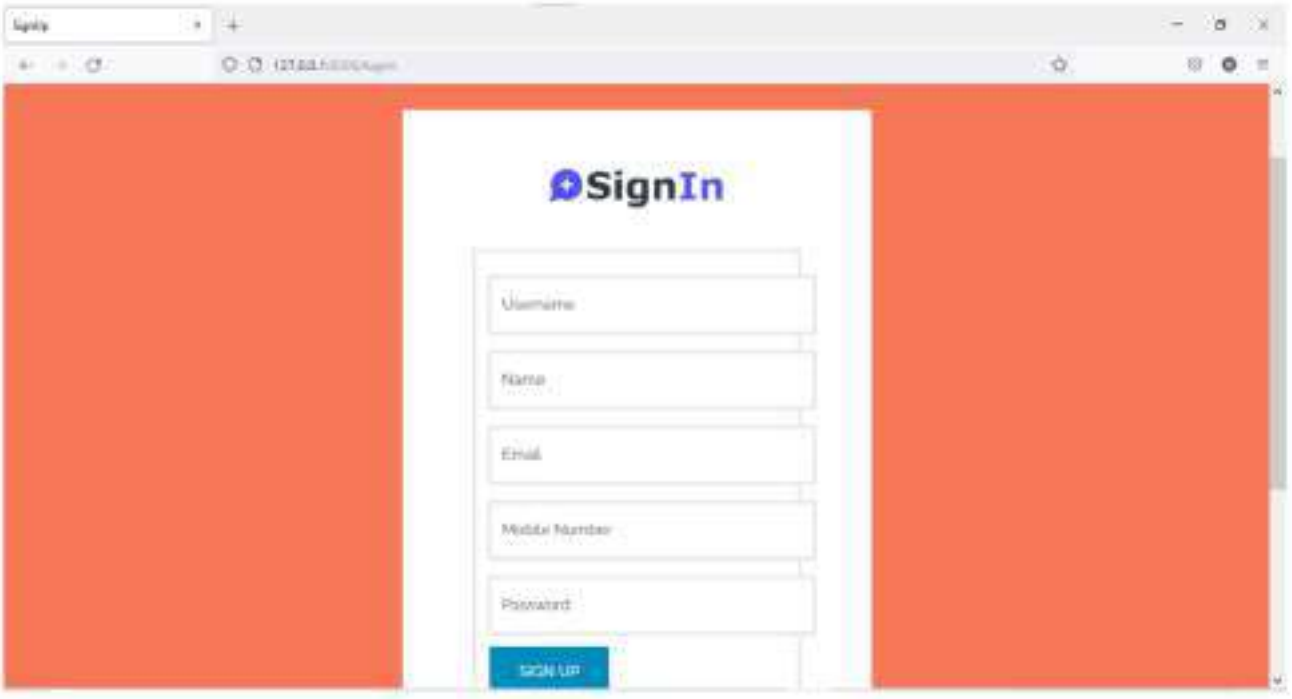


Figure 6.2: Sign Up Page

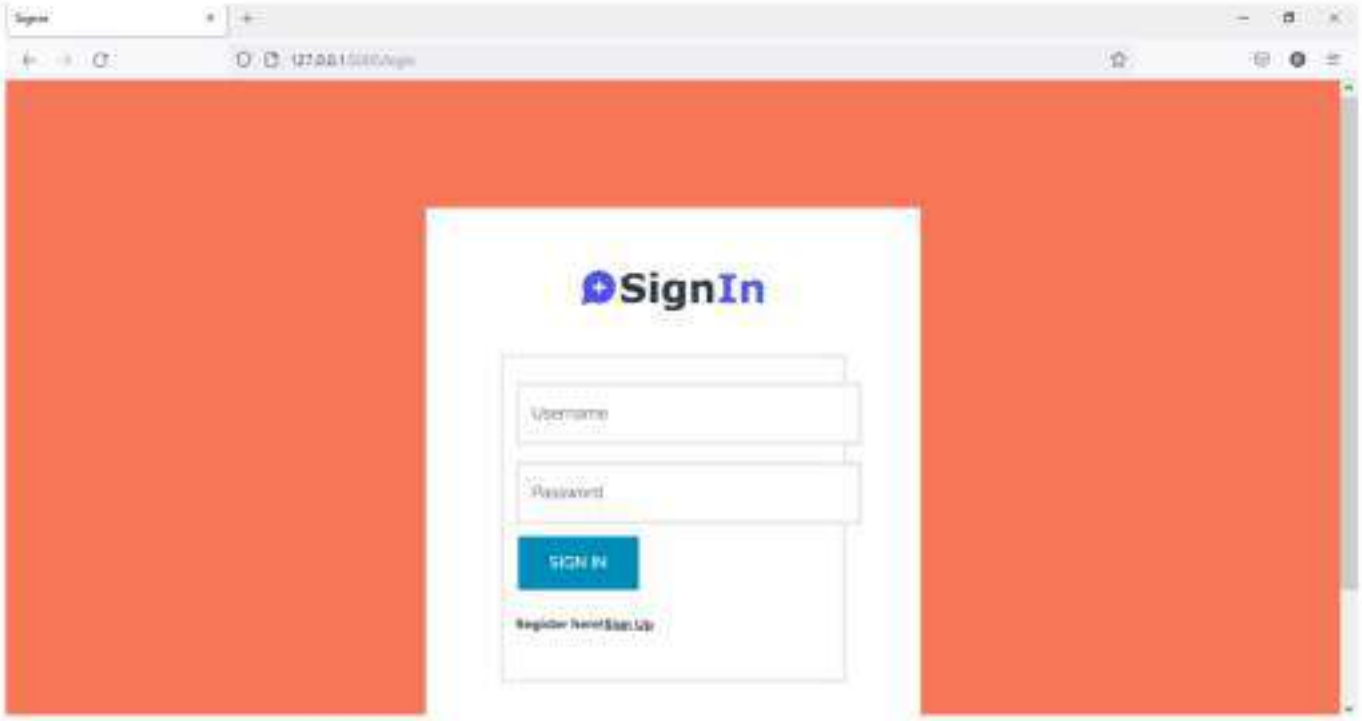


Figure 6.3: Sign In Page

Figure 6.4: Upload Image Page

Figure 6.5: Result

7. TESTING

7. TESTING

7.1 INTRODUCTION TO TESTING

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, subassemblies, assemblies and/or a finished product. It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

7.2 TYPES OF TESTING

7.2.1 UNIT TESTING

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .it is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

7.2.2 INTEGRATION TESTING

Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfaction, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

7.2.3 FUNCTIONAL TESTING

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals. Functional testing is centered on the following items:

Valid Input	: identified classes of valid input must be accepted.
Invalid Input	: identified classes of invalid input must be rejected.
Functions	: identified functions must be exercised.
Output	: identified classes of application outputs must be exercised.
Systems/Procedures	: interfacing systems or procedures must be invoked.

Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic coverage pertaining to identify Business process flows; data fields, predefined processes.

7.2.4 USER INTERFACE TESTING

User interface testing, a testing technique used to identify the presence of defects is a product/software under test by Graphical User interface.

7.3 TEST CASES

S.NO	INPUT	If available	If not available
1	User registration	User get register successfully	There is no process
2	User login	User get login into the application	There is no process
3	Upload image	Image loaded for getting prediction	There is no process
4	Predict	Blood cancer type predicted	There is no process

8. CONCLUSION AND FUTURE SCOPE

8. CONCLUSION AND FUTURE SCOPE

8.1 PROJECT CONCLUSION

The early detection of blood cancer can help effectively in its treatment. This study proposed two classification models distinguishing between leukemia-free and leukemia-affected blood microscopic images. Experiments demonstrated the superiority of the SVM classifier. The second model employs transfer learning models for both feature extraction and classification. Experiments for this model demonstrated its superiority to the first model with respect to various performance metrics.

8.2 FUTURE SCOPE

In future, the scope of the work can be extended so that the system can be more efficiently used by all the researchers. Everything has the potential to be scaled up in terms of power and complexity. With technological advancements, we can make CPUs and GPUs cheaper and/or faster, enabling the production of bigger, more efficient algorithms. We can also design neural nets capable of processing more data, or processing data faster, so it may learn to recognize patterns with just 1,000 examples, instead of 10,000. Unfortunately, there may be an upper limit to how advanced we can get in these areas but we haven't reached that limit yet, so we will likely strive for it in the near future.

9. BIBLOGRAPHY

9. BIBLOGRAPHY

9.1 REFERENCES

1. Wang, D.; Khosla, A.; Gargeya, R.; Irshad, H.; Beck, A.H. Deep learning for identifying metastatic breast cancer. arXiv 2016, arXiv:1606.05718.
2. Agaian, S.; Madhukar, M.; Chronopoulos, A.T. Automated screening system for acute myelogenous leukemia detection in blood microscopic images. IEEE Syst. J. 2014, 8, 995–1004.
3. Vogado, L.H.; Veras, R.M.; Araujo, F.H.; Silva, R.R.; Aires, K.R. Leukemia diagnosis in blood slides using transfer learning in CNNs and SVM for classification. Eng. Appl. Artif. Intell. 2018, 72, 415–422.

9.2 WEBSITES

1. https://docs.python.org/3.8/
2. https://www.tensorflow.org/api_docs/python/tf/all_symbols/
3. https://keras.io/guides/
4. https://numpy.org/doc/
5. https://jupyterlab.readthedocs.io/en/stable/

10. PAPER PUBLICATION

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 06 | June - 2022 Impact Factor: 7.185 ISSN: 2582-3930

DETECTION OF BLOOD CANCER

Nalla Vivek, N. Ritesh Reddy, Sangaru Rachana, Kethavath Vamshi Naik

NALLA VIVEK (CSE, CMR TECHNICAL CAMPUS)

N. RITESH REDDY (CSE, CMR TECHNICAL CAMPUS)

SANGARU RACHANA (CSE, CMR TECHNICAL CAMPUS)

KETHAVATH VAMSHI NAIK (CSE, CMR TECHNICAL CAMPUS)

ABSTRACT:

Leukemia is a fatal disease that threatens the lives of many patients. Early detection can effectively improve its rate of remission. This paper proposes two automated classification models based on blood microscopic images to detect leukemia by employing transfer learning, rather than traditional approaches that have several disadvantages. In the first model, blood microscopic images are pre-processed; then, features are extracted by a pre-trained deep convolutional neural network MODELS, which makes classifications according to numerous well-known classifiers. In the second model, after pre-processing the images, neural network models are fine-tuned for both feature extraction and classification. Experiments were conducted on a dataset consisting of different images confirming that the second model performs better than the first because of 100% classification accuracy.

Keywords: *Convolution neural network*

1. INTRODUCTION

Diagnosis is performed by a physician to detect the presence or absence of a certain disease in a patient according to a particular dataset, which may include signs, symptoms, medical images, and exams. An incorrect diagnosis can have adverse consequences, for example, prescription of drugs with side effects, on a patient's health. As well as increasing the costs of treatment, incorrect diagnoses may complicate treatment procedures. To help physicians achieve high diagnostic accuracy, many assistant systems were proposed. Many diseases, including glaucoma, skin cancer, breast cancer, and leukemia are already addressed by such systems. Early and accurate diagnoses could effectively reduce treatment costs, increase the probability of remission, or even prolong the lives of patients. Leukemia is a common fatal disease that threatens the lives of many teenagers and children. Infants younger than five years of age are at increased risk. A 2012 study showed that about 352,000 adults and children all over the world develop leukemia, which starts in the bone marrow and is distinguished by the number of white cells increasing in an abnormal manner.

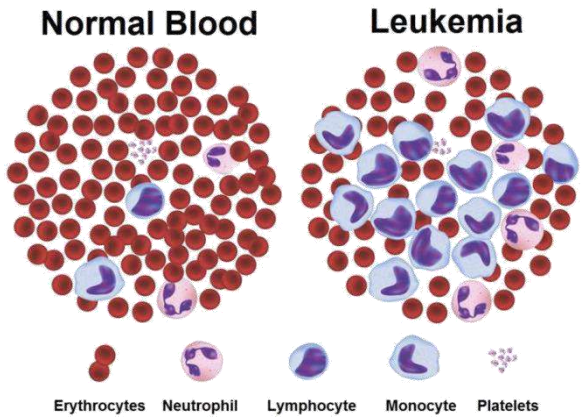


Fig.1: Example for detection

This disease has several causes, such as exposure to radiation and certain chemicals, as well as family history. Diagnoses can be performed via a variety of tests, such as physical examination, blood test, blood count, and bone marrow biopsy. Microscopic analysis is considered the most cost-effective procedure for initial diagnoses, but it is usually performed manually by an operator who is vulnerable to fatigue that could result from having to perform many tests in a single day. Moreover, such manual diagnoses are unreliable in themselves, as they are tedious, time-consuming, and subject to inter-observer variations. Hence, there is a need to build automated, low-cost systems that can differentiate between healthy and unhealthy blood smear images with high accuracy but without manual intervention.

2. LITERATURE REVIEW

2.1 Diagnosing leukemia in blood smear images using an ensemble of classifiers and pre-trained convolutional neural networks

Leukemia is a worldwide disease. In this paper we demonstrate that it is possible to build an automated, efficient and rapid leukemia diagnosis system. We demonstrate that it is possible to improve the precision of current techniques from the literature using the description power of well-known Convolutional Neural Networks (CNNs). We extract features from a blood smear image using pre-trained CNNs in order to obtain an unique image description. Many feature selection techniques were evaluated and we chose PCA to select the features that are in the final descriptor. To classify the images on healthy and pathological we created an ensemble of classifiers with three individual classification algorithms (Support Vector Machine, Multilayer Perceptron and Random Forest). In the tests we obtained an accuracy rate of 100%. Besides the high accuracy rate, the tests showed that our approach requires less processing time than the methods analyzed in this paper, considering the fact that our approach does not use segmentation to obtain specific cell regions from the blood smear image.

2.2 Glaucoma detection based on deep convolutional neural network

Glaucoma is a chronic and irreversible eye disease, which leads to deterioration in vision and quality of life. In this paper, we develop a deep learning (DL) architecture with convolutional neural network for automated glaucoma diagnosis. Deep learning systems, such as convolutional neural networks (CNNs), can infer a hierarchical representation of images to discriminate between glaucoma and non-glaucoma patterns for diagnostic decisions. The proposed DL architecture contains six learned layers: four convolutional layers and two fully-connected layers. Dropout and data augmentation strategies are adopted to further boost the performance of glaucoma diagnosis. Extensive experiments are performed on the ORIGA and SCES datasets. The results show area under curve (AUC) of the receiver operating characteristic curve in glaucoma detection at 0.831 and 0.887 in the two databases, much better than state-of-the-art algorithms. The method could be used for glaucoma detection.

2.3 Multi-Resolution-Tract CNN with Hybrid Pretrained and Skin-Lesion Trained Layers

Correctly classifying a skin lesion is one of the first steps towards treatment. We propose a novel convolutional neural network (CNN) architecture for skin lesion classification designed to learn based on information from multiple image resolutions while leveraging pretrained CNNs. While traditional CNNs are generally trained on a single resolution image, our CNN is composed of multiple tracts, where each tract analyzes the image at a different resolution simultaneously and learns interactions across multiple image

resolutions using the same field-of-view. We convert a CNN, pretrained on a single resolution, to work for multi-resolution input. The entire network is fine-tuned in a fully learned end-to-end optimization with auxiliary loss functions. We show how our proposed novel multi-tract network yields higher classification accuracy, outperforming state-of-the-art multi-scale approaches when compared over a public skin lesion dataset.

2.4 Deep learning for identifying metastatic breast cancer.

The International Symposium on Biomedical Imaging (ISBI) held a grand challenge to evaluate computational systems for the automated detection of metastatic breast cancer in whole slide images of sentinel lymph node biopsies. Our team won both competitions in the grand challenge, obtaining an area under the receiver operating curve (AUC) of 0.925 for the task of whole slide image classification and a score of 0.7051 for the tumor localization task. A pathologist independently reviewed the same images, obtaining a whole slide image classification AUC of 0.966 and a tumor localization score of 0.733. Combining our deep learning system's predictions with the human pathologist's diagnoses increased the pathologist's AUC to 0.995, representing an approximately 85 percent reduction in human error rate. These results demonstrate the power of using deep learning to produce significant improvements in the accuracy of pathological diagnoses.

3. IMPLEMENTATION

Many traditional computer-aided systems use image processing and machine-learning techniques that usually involve several steps, including pre-processing, segmentation, feature extraction, and classification. However, the success of each step depends on the success of the preceding step. For example, the success of classification depends on the success of the preceding feature extraction, which itself depends on the success of the preceding segmentation. Hence, high classification accuracy requires the success of all steps, each of which is non-trivial and problem-dependent.

Disadvantages:

1. High classification accuracy requires the success of all steps.

This paper proposes two classification models that are based on transfer learning and can distinguish between healthy and unhealthy blood smear images with high accuracy. These models employ RESNET50, which is a deep CNN that achieved huge success in the image classification challenge, ImageNet.

Advantages:

1. High accuracy.

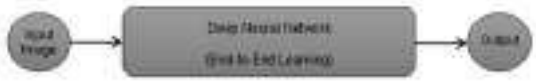


Fig.2: System architecture

Classification models are proposed here to distinguish between microscopic images depicting healthy tissue and leukemia. Transfer learning was adopted for both models, which employed pre-trained deep neural networks. Transfer learning eliminates the time and effort needed to design and train such networks from scratch. According to Castelluccio et al., there are two methods to apply transfer learning. The first method includes obtaining features extracted from the input images by obtaining the values of the last fully connected layer (FC) of the net, before using another classifier for classification. The second method involves modifying the structure of the network by eliminating the high-level layers. This process is known as network fine-

tuning. In this study, methods were adopted and implemented by our proposed models.

4. ALGORITHMS

DENSENET:

A DenseNet is a type of convolutional neural network that utilises dense connections between layers, through Dense Blocks, where we connect all layers (with matching feature-map sizes) directly with each other.

DenseNet is a new CNN architecture that reached State-Of-The-Art (SOTA) results on classification datasets (CIFAR, SVHN, ImageNet) using less parameters. Thanks to its new use of residual it can be deeper than the usual networks and still be easy to optimize.

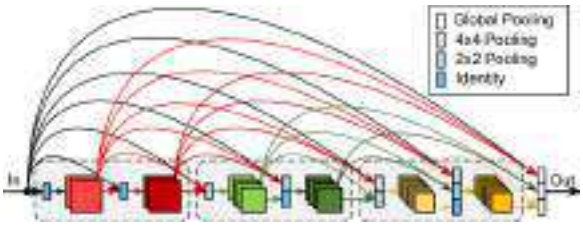


Fig.3: DenseNet model

INCEPTION RESNET V2:

Inception-ResNet-v2 is a convolutional neural architecture that builds on the Inception family of architectures but incorporates residual connections Inception-ResNet-v2 is a convolutional neural network that is trained on more than a million images from the ImageNet database [1]. The network is 164 layers deep and can classify images into 1000 object categories, such as keyboard, mouse, pencil, and many animals.

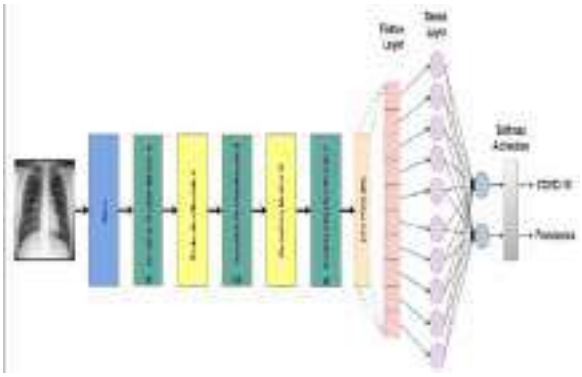


Fig.4: Inception Resnet V2 model

ResNet50 is a variant of ResNet model which has 48 Convolution layers along with 1 MaxPool and 1 Average Pool layer. It has 3.8 x 10^9 Floating points operations. It is a widely used ResNet model and we have explored ResNet50 architecture in depth. ResNet-50 is a convolutional neural network that is 50 layers deep. You can load a pretrained version of the network trained on more than a million images from the ImageNet database. The pretrained network can classify images into 1000 object categories, such as keyboard, mouse, pencil, and many animals.

Fig.5: Resnet50 model

5. EXPERIMENTAL RESULTS

Fig.6: Home page

Fig.7: Signin

Fig.8: Admin login

Fig.9: Upload image

Fig.10: Predict result

6. CONCLUSION

The early detection of leukemia can help effectively in its treatment. This study proposed two classification models distinguishing between leukemia-free and leukemia-affected blood microscopic images. Experiments demonstrated the superiority of the SVM classifier. The second model employs transfer learning models for both feature extraction and classification. Experiments for this model demonstrated its superiority to the first model with respect to various performance metrics.

7. FUTURE SCOPE

A future study could be extended to differentiate among the different types of leukemia rather than simply marking images as leukemia-free or leukemia-affected.

ACKNOWLEDGEMENT

We thank CMR Technical Campus for supporting this paper titled with "DETECTION OF BLOOD CANCER", which provided good facilities and support to accomplish our work. Sincerely thank to our Chairman, Director, Deans, Head of the Department, Department of Computer Science and Engineering, Guide and Teaching and Non-Teaching faculty members for giving valuable suggestions and guidance in every aspect of our work.

REFERENCES

1. Vogado, L.H.S.; Veras, R.D.M.S.; Andrade, A.R.; De Araujo, F.H.D.; e Silva, R.R.V.; Aires, K.R.T. Diagnosing leukemia in blood smear images using an ensemble of classifiers and pre-trained convolutional neural networks. In Proceedings of the 2017 IEEE 30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), Niteroi, Brazil, 17–20 October 2017; pp. 367–373.

2. Chen, X.; Xu, Y.; Wong, D.W.K.; Wong, T.Y.; Liu, J. Glaucoma detection based on deep convolutional neural network. In Proceedings of the 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, Italy, 25–29 August 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 715–718.

3. Kawahara, J.; Hamarneh, G. Multi-resolution-tract CNN with hybrid pretrained and skin-lesion trained layers. In Proceedings of the International Workshop on Machine Learning in Medical Imaging, Athens, Greece, 17 October 2016; Springer: Cham/Canton of Zug, Switzerland, 2016; pp. 164–171.

4. Wang, D.; Khosla, A.; Gargeya, R.; Irshad, H.; Beck, A.H. Deep learning for identifying metastatic breast cancer. arXiv 2016, arXiv:1606.05718.

5. Agaian, S.; Madhukar, M.; Chronopoulos, A.T. Automated screening system for acute myelogenous leukemia detection in blood microscopic images. IEEE Syst. J. 2014, 8, 995–1004. [CrossRef]

6. Thanh, T.T.P.; Vununu, C.; Atoev, S.; Lee, S.-H.; Kwon, K.-R. Leukemia blood cell image classification using convolutional neural network. Int. J. Comput. Theory Eng. 2018, 10, 54–58. [CrossRef]

7. Imran Razzak, M.; Naz, S. Microscopic blood smear segmentation and classification using deep contour aware CNN and extreme machine learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 49–55.

8. Sajjad, M.; Khan, S.; Jan, Z.; Muhammad, K.; Moon, H.; Kwak, J.T.; Rho, S.; Baik, S.W.; Mehmood, I. Leukocytes classification and segmentation in microscopic blood smear: A resource-aware healthcare service in smart cities. IEEE Access 2016, 5, 3475–3489. [CrossRef]

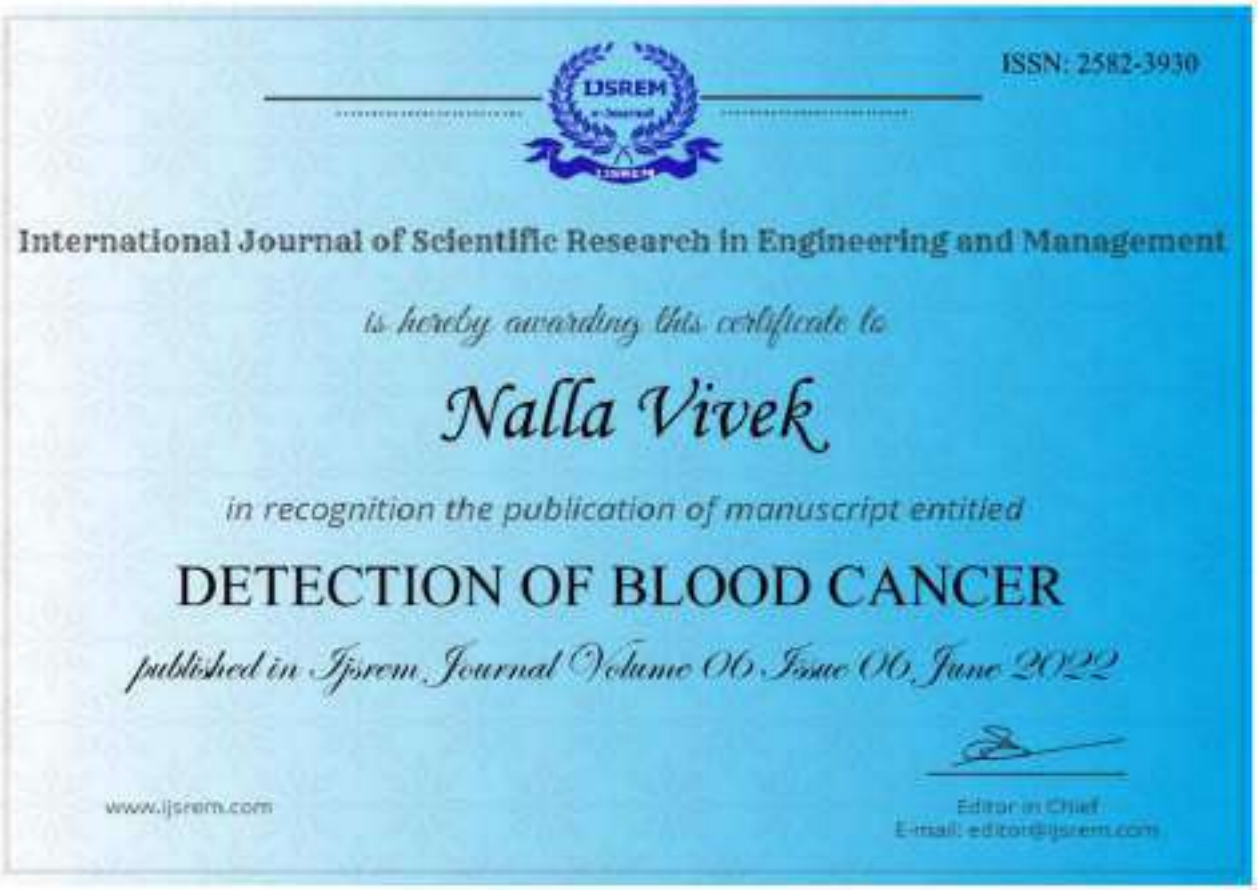
9. Abdeldaim, A.M.; Sahlol, A.T.; Elhoseny, M.; Hassanien, A.E. Computer-aided acute lymphoblastic leukemia diagnosis system based on image analysis. In Advances in Soft Computing and Machine Learning in Image Processing; Springer: Berlin/Heidelberg, Germany, 2018; pp. 131–147.

10. Vogado, L.H.; Veras, R.M.; Araujo, F.H.; Silva, R.R.; Aires, K.R. Leukemia diagnosis in blood slides using transfer

learning in CNNs and SVM for classification. Eng. Appl. Artif. Intell. 2018, 72, 415–422.

11. CERTIFICATES

11. CERTIFICATES



ISSN: 2582-3930

International Journal of Scientific Research in Engineering and Management

is hereby awarding this certificate to

Nalla Vivek

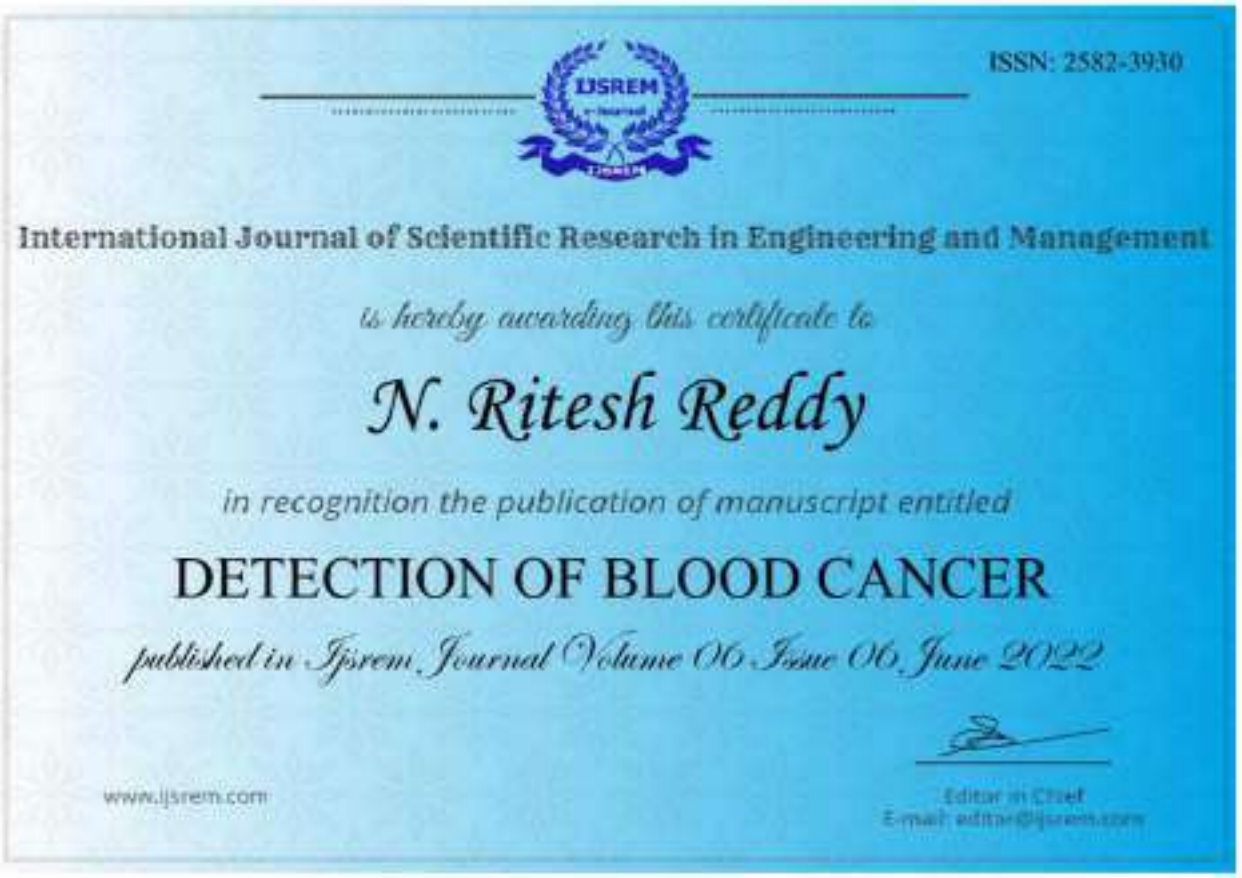
in recognition the publication of manuscript entitled

DETECTION OF BLOOD CANCER

published in Ijsrem Journal Volume 06 Issue 06, June 2022

www.ijsrem.com

Editor in Chief
E-mail: editor@ijsrem.com



ISSN: 2582-3930

International Journal of Scientific Research in Engineering and Management

is hereby awarding this certificate to

N. Ritesh Reddy

in recognition the publication of manuscript entitled

DETECTION OF BLOOD CANCER

published in Ijsrem Journal Volume 06 Issue 06 June 2022

www.ijsrem.com

Editor in Chief

ISSN: 2582-3930

International Journal of Scientific Research in Engineering and Management

is hereby awarding this certificate to

Sangaru Rachana

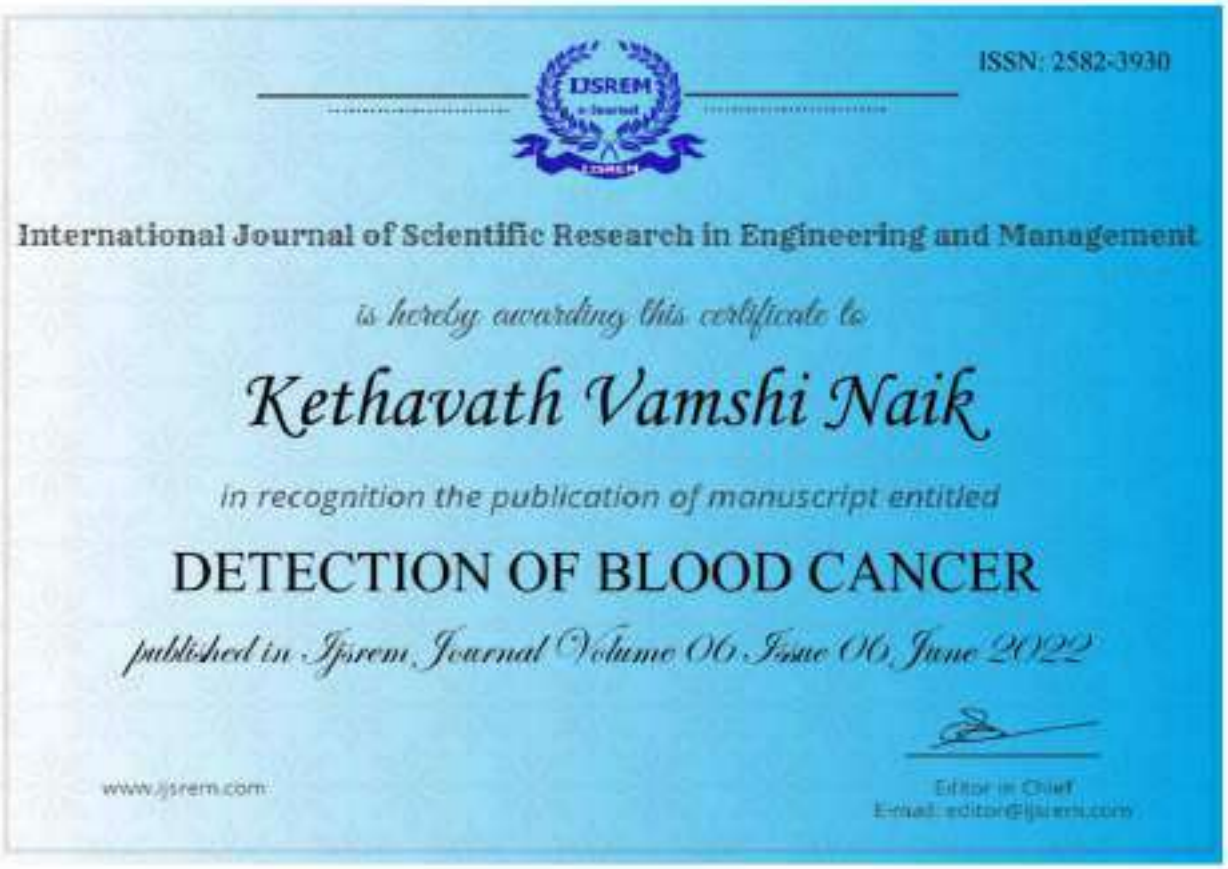
in recognition the publication of manuscript entitled

DETECTION OF BLOOD CANCER

published in Ijsrem Journal Volume 06 Issue 06 June 2022

www.ijsrem.com

Editor in Chief
E-mail: editor@ijsrem.com



ISSN: 2582-3930

International Journal of Scientific Research in Engineering and Management

is hereby awarding this certificate to

Kethavath Vamshi Naik

in recognition the publication of manuscript entitled

DETECTION OF BLOOD CANCER

published in Ijsrem Journal Volume 06 Issue 06 June 2022

www.ijsrem.com

Editor in Chief
E-mail: editor@ijsrem.com