Journal of Management & Entrepreneurship
ISSN 2229-5348
UGC Care Group I Journal
Vol-10 Issue-02 Dec 2021

An Automatic Method for Detecting White Blood Cell Cancer in Bone Marrow Micrographs using Convolution Neural Networks

Mr. JULURU ESWARANARASIMHA ABHILASH[1], Mrs. BELLAMKONDA PAVANI[2], Mrs. K. RADHA[3], ASSOCIATE PROFESSOR[13] ,ASST PROFESSOR[23] , DEPARTMENT OF ECE,
SWARNANDHRA COLLEGE OF ENGINEERING AND TECHNOLOGY, NARASAPUR

ABSTRACT

The bone marrow is responsible for producing about 1% of all blood cells, which are known as leukocytes. Blood malignancy originates in the unchecked multiplication of these neutrophils. The suggested research offers a reliable method for categorizing three distinct kinds of cancer.Multiple Myeloma (MM) and Acute Lymphoblastic Leukemia (ALL) use the SN-AM dataset. In cases of acute lymphoblastic leukemia (ALL), an abnormally high number of lymphocytes are produced by the bone marrow.However, unlike other cancers, multiple myeloma (MM) causes malignant cells to collect in the bone marrow rather than being disseminated throughout the body. As a result, they inhibit the development of new, healthy blood cells by crowding them out. In the past, this would take a very long time and a lot of effort on the part of a trained expert. Using deep learning methods, specifically convolutional neural networks, the suggested model eliminates the possibility of mistakes in the human process. The algorithm is educated on pictures of cells, so it knows how to best pre-process and retrieve characteristics from those images. Next, the sort of malignancy in the cells is predicted by training the model using an improved version of the Dense Convolutional neural network (DCNN) architecture. While remembering the samples precisely 94 times out of 100, the model accurately reproduced all the data.

I. INTRODUCTION

CELLS The three types of blood cells are white blood cells, red blood cells, and platelets. Each one is made in the bone marrow and continuously released. prompt in the heart and blood vessels. The main cause of blood cancer is the exponential multiplication of abnormal blood cells, which prevents normal blood cells from developing normally. Among blood cancers, leukemia, myeloma, and lymphoma are the three most prevalent types. The white blood cell cancer known as acute lymphocytic leukemia mostly attacks the bone marrow. (EVERYONE). Diseases that progress quickly and have the potential to be fatal in a short period of time if untreated are referred to as "acute" [1, 2]. There are three types of ALL: L1, L2, and L3. Conversely, automatic numbering reduces labor costs but raises the risk of an incorrect total. It implies that both strategies have advantages and disadvantages. An overview of the automated system for categorizing white blood cell cancers is given in this article. Both urban and rural areas can quickly adopt automatic classification, which saves time and money. The proposed approach introduces inconsistencies due to labor-intensive human categorization, the necessity for an expert, and other problems. the errors caused by the uniformity of cell appearance under a microscope, etc. Deep learning-based techniques can help alleviate all of these issues [10] because they can extract valuable features from the original data. When working with a large quantity powerful and

reliable automatic categorization technique for identifying the sort of white blood malignancy, specifically Acute Lymphoblastic Leukemia (ALL) and Multiple Myeloma (MM).Therefore, the paper compares the suggested deep learning model's performance against a set of benchmarks, including accuracy, precision, memory, sensitivity, and specificity.What this piece really adds to the discussion is:To identify cancer types from a limited dataset, the suggested model adopts a generic strategy. Data enhancement has been used for extension. On the recovered and prepped dataset, the suggested DCNN has been shown to outperform some state-of-the-art CNN models through comparison analysis. The suggested model beats state-of-the-art machine learning and pre-trained deep learning models in classifying cancer types despite requiring less processing time and having trainable parameters.

This report followed the following outline: The associated study is discussed in Section II. The suggested method for blood cancer classification, learning metrics, a summary of the dataset, and the evaluation approach are described in Section III. In Section IV, we present the findings of our extensive trials with the suggested model and draw comparisons to more conventional machine learning methods.

II. RELATED WORK

The Three stages are often involved in the analysis of images of infectious blood cells: image preparation, feature extraction, feature selection, and categorization. Numerous cancers have been the subject of extensive research (especially leukemia, lymphoma, and myeloma). Zhang et al. [13] proposed a convolutional neural network model to immediately classify cervical cells into infected and normal cells without segmenting them. Zhao et al. [12] proposed the use of machine learning techniques such CNN, SVM, Random Forests, etc. to classify the various WBCs detected in the body. The proposed Stain Deconvolution Layer (SD Layer) for distinguishing between malignant and healthy cells learned from the images in the Optical Density (OD) space rather than training the images in RGB space.

III.PROPOSED METHODOLOGY

After learning on the training set, the model is applied to the test set to provide predictions. Three layers make up the model: a fully connected layer, a max pool layer, and a convolution layer. group under control. The suggested approach's process flow is shown in Figure 1.

A. DATASET DESCRIPTION

Two separate sections of a wider compilation provide the sample for the proposed study [29]. The first images in the series are of individuals with B-ALL. In particular, 90 images depict B-Lineage Acute Lymphoblastic Leukemia. Figure 2 displays the backdrop mask and nucleus mask of the ALL picture. The second part of the collection consists of the remaining 100 images, which are from people who have been diagnosed with multiple myeloma.

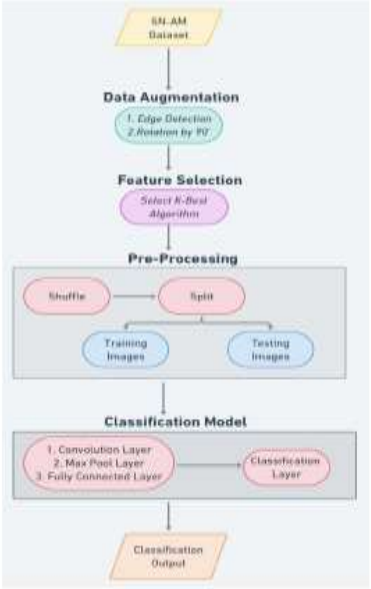


FIGURE 1: Proposed Methodology.

The masks for the backdrop, nuclei, and cytoplasm of plasma cells for the matching MM picture are displayed in Figure 3. 2560 x 1920 resolution images please.The collection is comprised of images in BMP format. The suggested CNN model for classifying cancer cells as either ALL or MM is trained using the mixed version.

B. DATA AUGMENTATION

We rotate and crop the image before adding it to the SN-AM collection. After then, the jumbled images are divided into two groups: the exam group and the instruction set. For the model to perform successfully in testing, there must be a major, a significant amount of data that is available, and the target item present in a range of sizes, locations, and lighting conditions.

From existent data, data can be created using various visual editing techniques. Image 4.

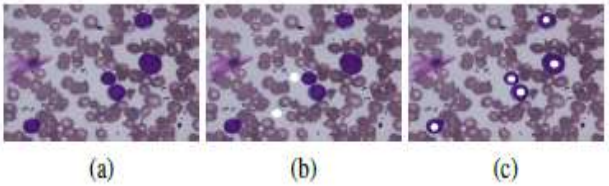
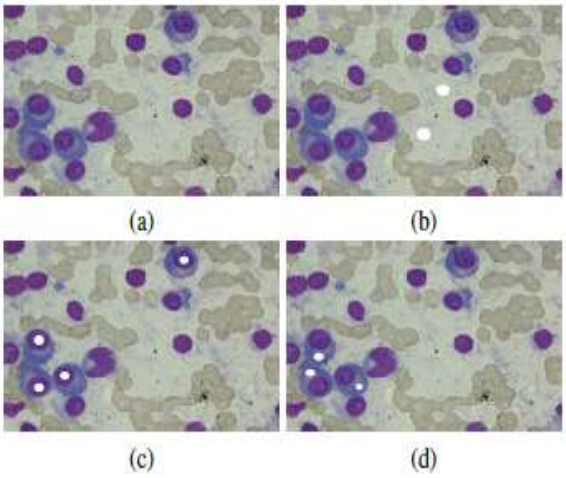


FIGURE 2: Images from the FULL Dataset as examples. Sample picture (a), backdrop mask (b), and nucleus mask (c) in their entirety.Acute lymphoblastic leukemia is abbreviated as "ALL."

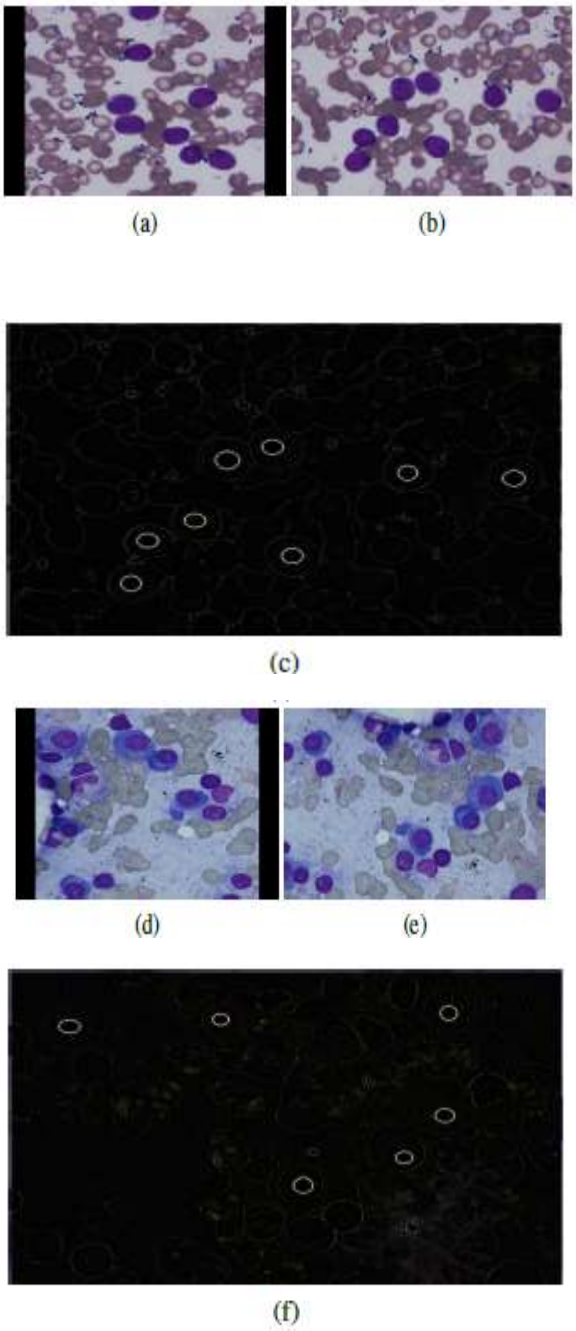


Images from the MM Collection (Figure 3). Masks representing the cytoplasm (d) and nucleus (c) of an MM sample image (a). Multiple myeloma is the diagnosis indicated by the letters MM. shows the final images following the application of data enhancement. The first technique involves simply shifting the images by 90 degrees [30]. We spin all the images by 90 degrees to make sure our model can recognize the object in any direction, as seen in Figures 4(b) and 4(c).(e). The next method, which is shown in Figures 4(c) and 4(d), screens the image and produces a filtered version of the image that only includes the original image's borders or limits. (f). Without data augmentation, overfitting happens, making it difficult for the model to

C. FEATURE SELECTION

Feature selection is a crucial aspect of Deep Learning and greatly influences the effectiveness of the models. A model's accuracy decreases with an excessive number of features. The process of choosing which attributes to employ to affect the result is known as feature selection. One of the key advantages of this strategy is that it reduces over-fitting because fewer duplicate data eliminates the probability of projections based on noise. Less training material implies less effort is required, and precision rises when abnormalities and false information are removed. The recommended research made use of univariate feature selection. In univariate feature selection, the correlation between each feature and the dependent variable is examined separately to ascertain its strength.

The Select Best makes use of a set of



Micrographs with added data (Figure 4).

A UNIQUE ORIGINAL IMAGE (a). (B) The original image has been rotated ninety degrees counterclockwise. (d) the original MM picture; (c) the identified marginal areas in the original ALL picture. (e) The original MM photo was rotated ninety degrees counterclockwise. (f) The edges discovered in the MM picture. Depending on the patient's disease class (all stand for acute lymphoblastic leukemia; mm for multiple myeloma), K-specific features might be selected. The suggested model applies the Chi-square (chi2) test. In the field of statistics, the chi-square test is used to establish the independence of two occurrences. We can determine a "predicted count" of E and a "observed count" of O given information about two components. The discrepancy between the expected (E) and actual (O) values is examined using chi-square analysis. We select traits count, giving us a lower Chi-square number. Therefore, the large Chi-square number disproves the independence theory. Therefore, if the Chi-square value is high, the feature is highly response-dependent and should be chosen for model training. Select Best, in its simplest form, computes scores for each feature and then discards all but the K highest-scoring features.

The solution for the chi-squared test is shown in Equation (1).

$$\chi^2 = \frac{(O-E)^2}{E} \qquad (1)$$

If there were no correlation between the characteristic and the result, then the anticipated number of class observations would be E, where O is the total number of class observations.

D. PRE-PROCESSING OF IMAGES

Data pre-processing entails a wide range of operations, such as dealing with missing values, one-hot encoding, normalization, multi-co linearity, scaling, randomization, and division. The suggested research will involveafter feature selection and transformation, the resulting data is first standardized before being randomly split into training and testing groups. The sample is split 75/25 in favor of model training and testing.

E. PROPOSED CONVOLUTION NEURAL NETWORKAND ARCHITECTURE

To improve this procedure, we suggest a novel CNN technique in this study. It has been determined that the malignancy is either MM or ALL. Our deep learning model is trained using the. Figure 5 gives an illustration of it. Convolutional neural networks, or CNNs, are one kind of neural network. mostly employed in visual media studies. CNNs' significant contribution to image classification algorithms. The team is efficient and precise in their picture classification work. When compared to other picture classification methods, they require comparatively less preparation. An input layer, a hidden layer, an intermediate layer, and an output layer make up a typical CNN model. An unprocessed image is fed into the algorithm this study proposes..

$$(a-k+2p)/s+1 \times (b-k+2p)/s+1.$$

Furthermore, different training pictures will result in different filtering granularities. Two convolution layers with softmax activation are present in the suggested model.Routines and then a sharing layer. Equation gives the usual definition of the softmax function _: K! K. (2).Convolution of the input picture with the kernel function yields the activation map, which is given by Eq. (3).

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \qquad (2)$$

$$A[x,y] = (I * f)[x,y] = \sum_i \sum_j I[i,j] f[x-i, y-i] \qquad (3)$$

where I is the input picture, f is the kernel function, x and y are the row and column counts of the output matrix, i and j are the row and column counts of the input matrix, andthe picture data array in each case.

2) Pooling Layer

Another key component of CNNs is the pooling layer, which serves as a non-linear downsampling layer. This method allocates resources using the non-linear max-pooling function. provides the greatest number for each of the non-overlapping regions that make up the image. By lowering the amount of parameters and processing power needed by the network, the pooling layer aims to prevent overfitting. Max-pooling takes a picture of size a _ b and uses a kernel of size k and a step size s to find the highest value inside a specified area. The quantity of the product that this process produces is (ak)=(s+1) _ (Bk) =(s+1). Figure 6 illustrates how a max-pooling

technique down samples an image.

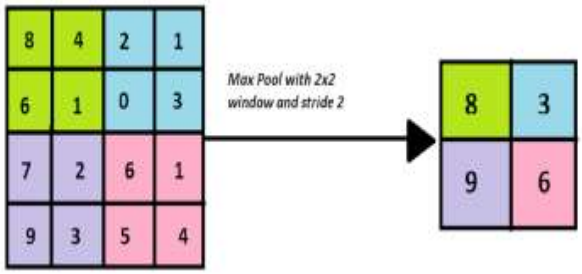


FIGURE 6: Maxpool Function

3) Fully Connected Layer

Multi-Layer Perceptions have completely linked layers. After a series of convolution layers, a thick layer is applied. A completely linked layer, as the term implies,establishes a connection between every cell in two or more layers. They use the characteristics gleaned from the convolutions in the same way that conventional neural networks do: to categorize pictures.The mistake is computed and back-propagated at this stage.The suggested model has five interconnected levels, the fifth of which is the output layer. All four completely linked levels use the softmax triggering algorithm. The sigmoid activation function is present in the output layer, and it generates a chance between 0 and 1 for each of the categorization categories that the model is attempting to forecast. This equation gives the definition of the sigmoid function. (4).

$$Sig(z) = \frac{1}{1+e^{-z}} = \frac{e^{z}}{e^{z}+1} \qquad (4)$$

Here, z is the input vector.

TABLE 1: Network Architecture.

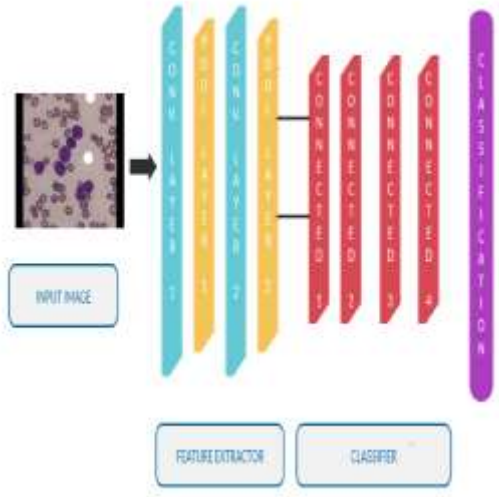


FIGURE 5: A Suggested Framework for a CNN Model.

Pseudo-code for the deep learning-based picture classification method is presented in method 1.

IV. RESULTS AND ANALYSIS

The classification algorithm is created using the open-source Tensor Flow end-to-end framework. 424 images were used to train a binary categorization model over 1000 trials. Optimizes the loss function using the Adam Optimizer to reach a low loss at the end of each cycle. The cancer subtypes shown in the images were then deduced using the algorithm that had been taught. An K80 GPU is utilized to generate the model. Initially, the results of the proposed paradigm will be explained. The outcomes of these comparisons between the proposed model and the state-of-the-art in deep learning and machine learning are discussed.

A. DEEP LEARNING APPROACH

In the CNN part, we utilize Adam Optimizer to train our network with a learning rate of 0.01 and a weight-changing method. (5). It is implemented to use the sigmoid loss function. The cross-entropy loss function's optimal value is determined by the Adam algorithm. Figure 8 illustrates how damage diminishes with increasing repeat count. The chart shows that the training loss stabilizes at a set number after 800 training rounds. We get at the lowest number, 0.5034, after 1000 cycles. Figure 7 shows the confusion matrix for the CNN-based binary classification of aggressive white blood cancer. The accuracy and precision of the CNN are 97.25% and 95.19%, respectively, as indicated in Table 2.

In this case, w stands for the weight matrix, and _

B. ANALYSIS USING MACHINE LEARNING

We report our comparison study's findings in the section that follows, showing that deep learning performs better with larger data sets. We have tested several photo classification techniques from Differencing Machines in order to assess the recommended deep learning approach. The histogram and the Discrete Fourier Transform (DFT) must be extracted from the images before utilizing any of the machine learning techniques. These attributes are then used as input for all models during their learning process. With the use of the 'RBF' kernel and the supervised learning technique SVM, the categorization model was built [31]. The probability predictor Naive Bayes uses the Gaussian distribution for both types of cancer [32]. The decision tree classification model's objective [33]

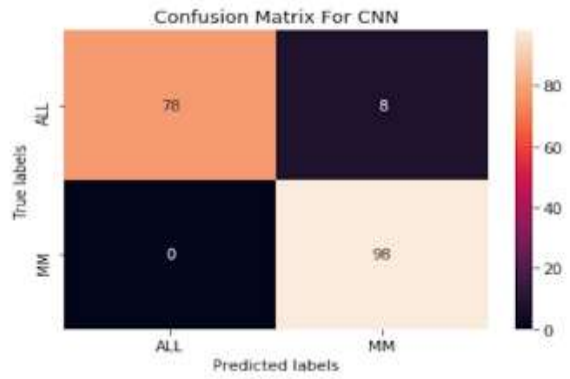


Figure 7: Proposed DCNN Model Confusion Matrix

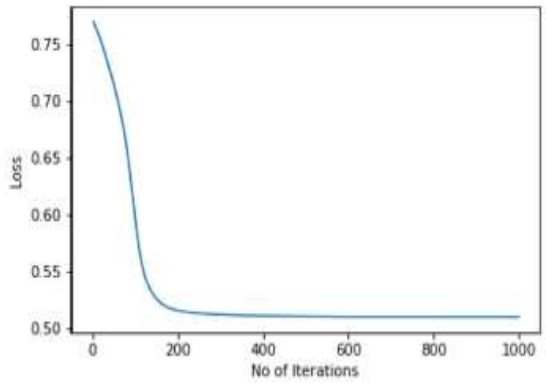


FIGURE 8: Loss vs. Training Iterations Scatter plot

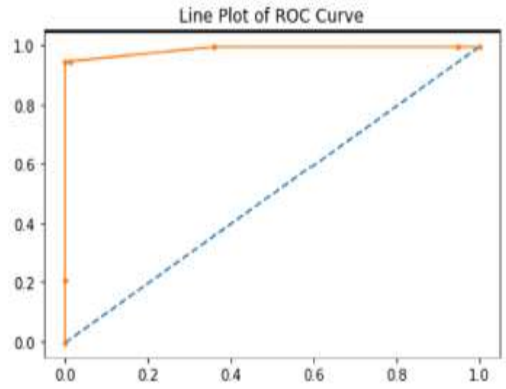


FIGURE 9: ROC Curve for proposed CNN model

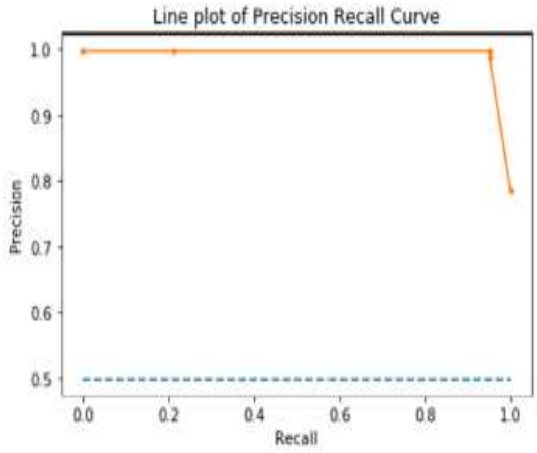


FIGURE 10: Precision-Recall for proposed CNN model

C. ANALYSIS USING TRANSFER LEARNING

In this study, we show that the proposed approach reduces the computing cost of addressing similar problems more effectively than VGG-16 and other transferred learning techniques. With only three layers, the VGG-16 is a convolutional neural network example that has been simplified. Softmax is employed for categorization, and the model consists of two 4096-node totally linked layers [35]. These measures have all been compared: F1 Score, Accuracy, Precision, Recall, and Specificity. The percentage of samples that were correctly assigned to the positive group is known as the TP. But TN stands for the percentage of samples that were correctly classified as belonging to the negative class. False positives and false negatives are samples where the model predicts the positive class (FN) incorrectly and samples where it forecasts the negative class (FP)

incorrectly.

$$AC = (TP + TN)/(TP + TN + FP + FN) \quad (6)$$

$$P = TP/(TP + FP) \quad (7)$$

$$R = TP/(TP + FN) \quad (8)$$

$$S = TN/(TN + FP) \quad (9)$$

$$F = (2 * P * R)/(P + R) \quad (10)$$

The outcomes of using various machine learning algorithms are displayed in Table 2.

The SVM confusion vectors are depicted in Figure 11.Images can be classified into two groups using Naive Bayes, Decision Trees, and Random Forests. True positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) are the four categories represented in a confusion matrix.(FN).In our system, pictures of ANY type of cancer cell are considered positive, while those of MMcancer cells are considered negative. With an impressive precision of 96.83%, Random Forests has successfully categorized both MM and ALL pictures. Very little preprocessing of the data is required, and it works well with binary or category characteristics.

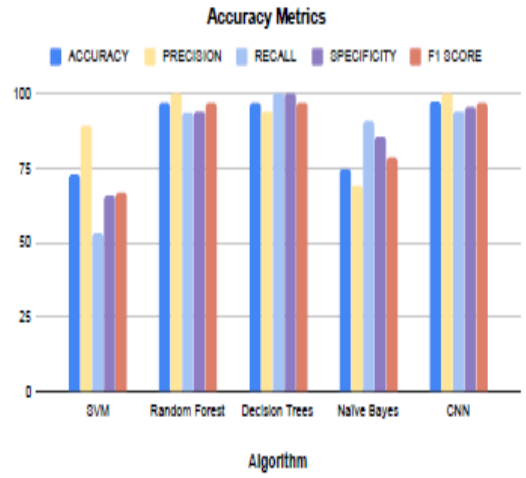


FIGURE 12: Accuracy of the suggested CNN algorithm in terms of recall

Table 3 displays the results of a comparison study of the suggested model applied to various datasets.The first collection of statistics is collected from fertilizing wasps [36]. The information was compiled from footage shot in June 2017 at a bee colony's entryway. The suggested algorithm successfully separated fertilizing pollinators from the rest of the population 82% of the time. Hematoxylin and eosin (H&E) stained slides of blood cancer cells is the next group [37].There are 116 standardized damaged pictures in the collection. The precision of the suggested algorithm in distinguishing between normal and cancerous data sets is 87%. Finally, 84,495 X-Ray pictures from a collection of retinal OCT images are used for comparison [38]. The suggested algorithm successfully distinguished between infectious and uninfected retinal pictures with an accuracy of 87%.

V. CONCLUSION

The proposed technique eliminates the risk of human mistake in the conventional method by using a deep learning approach, specifically convolutional neural networks. This is the prototype. It preforms the photos and extracts the most important elements before building the model using the updated Convolutional Neural Network Architecture. In the end, it can identify the type of cancer that is shown in an image. A accuracy rating of 97.2% was assigned to the program. Additionally, we contrasted our findings with those attained by utilizing cutting-edge methods like Naive Bayes, Decision Trees, Random Forests, Support Vector Machines (SVMs), etc. The proposed model performed better than the previously discussed methods. We show the improved accuracy of the model on three different datasets and compare it with other proposed models. Consequently, the model

REFERENCES

[1] ZenaWerb, Kai Kessenbrock, and Vicki Plaks. Tumor microenvironment regulators: matrix metalloproteinases. 2010; Cell, 141(1):52–67.

[2] ZahidMehmood, HoshangKolivand, Tanzila Saba, Amjad Rehman, Naveed

Abbas, and Syed IjazurRahman. Deep learning-based classification of acute lymphoblastic leukemia. Research and Technique in Microscopy, 81(11):1310–1317, 2018.
[3] SamabiaTehsin and Sarmad Shafique. Pretrained deep convolutional neural networks are used in the identification and classification of subtypes of acute lymphoblastic leukemia. Research and treatment with technology in cancer, 17:1533033818802789, 2018.
[4] Kenneth C. Anderson, P. Leif Bergsagel, and Michael Hallek. There is growing evidence that multiple myeloma involves a multi-step transformation process. Blood, American Society of Hematology Journal, 91(1):3–21, 1998.
[5] J Hirsch, Nina Hageman, Alan R. Giles, Peter B. Neame, Peter Powers, and John G. Kelton. A comparison between two