International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 12 | Dec - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

Automatic Detection of Blood Cancer Using Deep Learning Neural Networks

[#1]M.V. PHANINDRA, Research Scholar,

[#2]Dr. G. THIPPANNA, Supervisor,

[#1,#2]*Dept. of Computer Science and Engineering*

[#1,#2]NIILM University, Kaithal, Haryana, India

Abstract: The Deep Learning classifier outperforms existing Red Blood Cell cancer CT image classifiers with a good accuracy. The findings of the experiment suggest that the current computer-aided diagnostic method offers considerable promise and security for automated diagnosis of benign and malignant cancers. This classifier dramatically reduces processing complexity. Additionally, the suggested approach might be expanded to identify many types of blood cell abnormalities. The results of the segmentation approach indicate that the online region-based segmentation method outperforms other techniques in terms of segmentation quality. After the initial extraction phase, map segmentation is applied in combination with online region-based segmentation for refined tumour area segmentation. The nodules identified through this segmentation are then used to extract texture features, which include energy, contrast, homogeneity, and correlation. Additionally, histogram-based features such as entropy, kurtosis, mean, skewness, and standard deviation are also extracted. These features are then compared against images of benign and malignant tumours. The DL classifier significantly outperforms other methods, achieving an accuracy of 97.1%. This marks an improvement of 5.4% over BRT, 4.4% over ANN, 2.4% over SVM, and 6.1% over KNN. These findings highlight the advantages of the proposed DL based approach for tumour classification.

1 Introduction

Blood cancer, also known as hematologic cancer, encompasses a group of malignancies that affect the blood, bone marrow, lymphatic system, and spleen. This category includes leukaemia, lymphoma, and myeloma, each of which originates from different types of blood cells and presents distinct clinical challenges. Leukemia primarily impacts white blood cells, leading to their uncontrolled proliferation, while lymphoma involves the malignant growth of lymphocytes, a key component of the immune system. Myeloma, on the other hand, results from the abnormal growth of plasma cells in the bone marrow [1].

All blood cells start their trip in the blood cells, which are the factory where the stem cells make life blood. A mother cell known as a stem cell gives rise to three different kinds of blood cells. These include platelets, white blood cells, red blood cells, and more. White blood cells, namely neutrophils, are classified as myeloid or lymphoid cells (T and B cells). Blood cell growth is usually carefully controlled by genes and the proteins they produce. This power may sometimes go crazy for no apparent reason. This may often lead to the growth of blood malignancy. Your blood cancer type is determined by the stage of development at which the cell is when the controls malfunction. Abnormal cells in leukaemia are often limited to the blood and bone marrow. The lymphatic tissues contain the aberrant cells that cause lymphoma, but in some cases, the bone marrow may also become involved. While there are several related blood-related illnesses, leukaemia and lymphomas are not among them. Most of

them are odd in their youth. Leukemia may develop from several of these disorders, and some have symptoms with blood cancer. Associated ailments include of Aplastic anaemia, Multiple myeloma, Myelodysplastic syndromes, Myeloproliferative disorders [2].

Blood malignancies impact the synthesis and functionality of your blood cells. Most of these malignancies start in the bone marrow, which makes blood. Red blood cells, white blood cells, and platelets are the three kinds of blood cells that are developed from stem cells in your bone marrow. The unchecked proliferation of rare blood cell types disrupts the typical blood cell growth phase in most blood malignancies. These cancerous blood cells, or aberrant blood cells, interfere with the blood's ability to fight infections and stop severe bleeding, among other things. Blood cancer may be of three primary types: Uneven white blood cell formation leads to blood leukaemia, a kind of cancer that develops in the bone marrow and blood. The production of red blood cells and platelets by the bone marrow is restricted when many faulty white blood cells are not able to fight infection [3]. A kind of blood cancer called lymphoma affects the lymphatic system, which clears the body of extra fluid and increases the number of immune cells in it. White blood cells that combat infections include lymphocytes. Your lymph nodes and other organs may have lymphoma cells called abnormal lymphocytes, which proliferate and amass. These cancer cells may eventually weaken the immune system. Myeloma is a malignancy of the plasma cell. White blood cells called plasma cells are what the body uses as antibodies to fight disease and illness. Myeloma cells inhibit the production of natural antibodies, which may impair immunity and increase susceptibility to infection.

Deep learning (DL)[4] offers significant advantages over traditional machine learning (ML) algorithms in predicting blood cancer, especially when dealing with complex, high-dimensional, and unstructured data such as genomic sequences, medical imaging, and clinical records. Unlike traditional algorithms that require manual feature engineering, DL models can automatically extract meaningful patterns from raw data, enabling them to handle complex relationships and large datasets more effectively. DL is particularly useful for integrating multi-modal data (e.g., genomic, imaging, and clinical data) and for analysing time-series data, making it well-suited for tracking disease progression [5]. However, DL models tend to be computationally intensive and may lack interpretability, which is critical in medical applications. On the other hand, traditional ML algorithms are more interpretable and require less computational power, but they may struggle with high-dimensional or unstructured data. In practice, a hybrid approach that combines DL with traditional techniques often yields the best results, balancing predictive power with interpretability and resource constraints [6].

This article is structured as follows: Section II discusses and outlines of various researchers employed work on blood cancer detection. Section III describes the proposed DL algorithms. Section IV discusses dataset description and experiment setup employed in this research. Section V presents the results of the experimental validation for the BCCD dataset. The article concludes with Section VI.

2 Literature review

This chapter reviews the literature on the use of image recognition and computer vision in connection to the medical system, blood cell identification. This paper covers the background and present state of haematological systems before reviewing studies done on both normal and pathological blood samples. Understanding the present state of the work and identifying its flaws are the goals of this examination.

Researchers have conducted some previous studies in this section. [7] Reviewing machine learning algorithms applied to leukaemia detection, this paper provides a comprehensive overview. With its emphasis on standardized datasets and proper evaluation metrics, the review serves as a valuable resource for researchers and 91% in KNN, 91% in Naive Bayes, and 75% in SVM because of the small size of the dataset.

[8] Vogado et al. achieved 100% accuracy in the performance, challenges, and challenges associated with Multilayer Perceptron's, SVM, and RFs. Considering the results of the tests, it can be concluded that segmentation is not necessary to locate specific cells with this method, as it reduces the processing time for creating the blood smear image. Balakumar K *et al.* [9] This paper presents exhibits that combine machine learning algorithms to ensure accurate and early detection of the disease, feature selection techniques and model optimization are employed, found that the accuracy of the classification was 92% in comparison to other standard classifiers.

Khaled A. S. Abu Daqqa *et. al.* [10] proposed a prediction and diagnosis system for leukemia using Classification algorithms. They applied several classification algorithms such as DT, SVM, and KNN. In their approach, decision trees obtained higher accuracy than others. The accuracy that they have found using a DT, SVM, and KNN is 77.30%, 76.82%, and 70.15%, respectively. They show the prediction accuracy of different classification model but did not work on any image processing technique.

3 Methodology

A Convolutional Neural Network (CNN) is a deep learning architecture designed primarily for processing grid-like data, such as images[11]. CNNs have been highly successful in various computer vision tasks, such as image classification, object detection, and segmentation, due to their ability to automatically learn spatial hierarchies of features from raw input data. The key components of a CNN involve the following layers: convolutional layers, activation layers, pooling layers, and fully connected layers. Below is a step-by-step breakdown of how a CNN works:

The input to a CNN is typically an image, which is represented as a multi-dimensional array (tensor). For example, a coloured image might be represented as a 3D array with dimensions (Height, Width, Depth), where Depth corresponds to the number of colour channels. The first key operation in a CNN is the convolution operation. This involves applying a set of filters (also called kernels) to the input image. Each filter is a small matrix that slides over the image in a process called convolution. During convolution, the filter performs element-wise multiplication between its values and the values of the image's pixels, followed by a summation to produce a single value in the output feature map. This operation helps the network detect specific patterns such as edges, textures, or simple shapes. The mathematical operation for convolution is given in Eq. (1). Where I is the input image, K is the convolutional filter and F is the output feature map.

$$F(i,j) = \sum_{m=0}^{k_H-1} \sum_{n=0}^{k_w-1} \sum_{d=0}^{D-1} I(i+m, j+n, d).K(m,n,d) \qquad \text{Eq. (1)}$$

After the convolution operation, the output feature map typically passes through an activation function to introduce non-linearity. The most used activation function in CNNs is ReLU (Rectified Linear Unit), which replaces all negative values in the feature map with zeros, making the network better at learning complex patterns. $A(i,\ j)$ is the activated feature map as given in Eq. (2).

$$A(i,j) = \max(0, F(i,j)) \qquad \text{Eq. (2)}$$

Pooling is a down sampling operation that reduces the spatial dimensions of the feature maps. The most common pooling operation is max pooling, where for each region. Pooling helps to reduce the computational complexity, prevents overfitting, and makes the network invariant to small translations in the image. The max-pooling operation as given by Eq. (3). Where, p_H and p_W are the height and width of the pooling window, s is the stride, or step size of the pooling operation. The fully connected layer combines learned features to make predictions. The SoftMax

converts output scores into probabilities for classification tasks. Loss function and backpropagation measures the prediction error and updates model parameters.

$$P(i,j) = \max_{m\in[0,p_H-1]} \max_{n\in[0,p_W-1]} A(s.i+m, s.j+n) \quad \text{Eq. (3)}$$

4 Datasets Description and Experiment Setup

4.1 Dataset

The BCCD (Blood Cell Count and Detection) dataset consists of 200 benign and 200 malignant blood cell images (JPEG format) from 534 patients, along with corresponding cell type labels. The dataset includes four types of blood cells: neutrophils, eosinophils, lymphocytes, and monocytes. Figure 1 displays the dataset samples of BCCD. Initially, the benign cell images are processed to identify potential cancer cells. The data is split into a training set (80%) and a test set (20%) using random stratified sampling. However, due to the limited number of training samples, deep learning models may face challenges in learning effectively. To address this, we applied image augmentation techniques such as rotating and flipping the white blood cell (WBC) images to artificially expand the training set. This augmentation resulted in 400 images for each type of WBC, which were used to create a more robust training set. For model training, we leveraged pre-trained models, initially trained on the ImageNet dataset, and fine-tuned them by adjusting their weights based on our augmented dataset. To optimize the models, we performed five-fold cross-validation, selecting the best model based on validation accuracy. The performance of the model was then evaluated using the test set. Figure 1 illustrates the BCCD dataset samples.

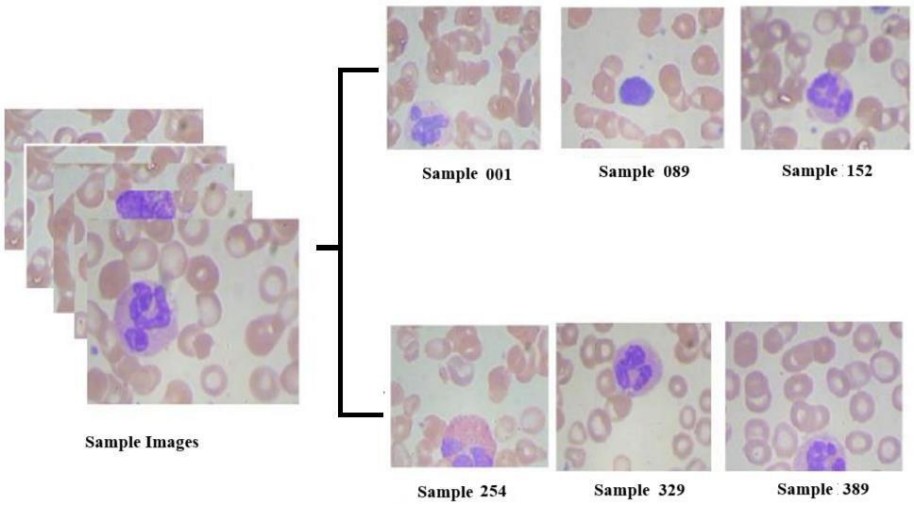
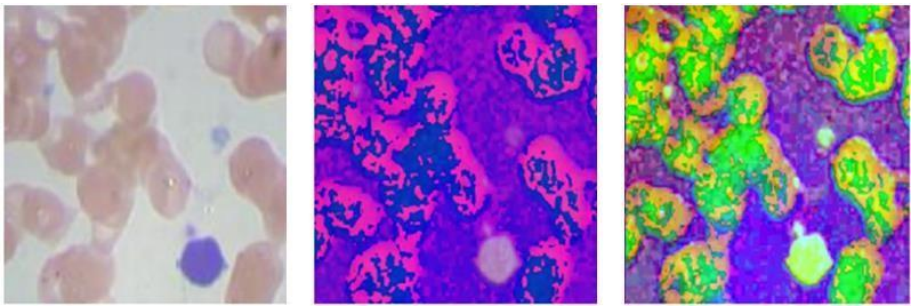


Figure 1: BCCD Dataset Samples

4.2 preprocessing of image

Any categorization requires preprocessing a picture since it sets the image up for further processing. The RGB colour pictures from the CT scan of the red blood sample are first transformed to HSV images for the operation. Colour conversion is the process of combining RBC subtractive coloured layers for a certain application and defined colour combination. Once this is done, the picture is subjected to histogram equalization. Brightness dependent contrast is a phase image-based application, where histogram equalized pictures are used to increase pixel quality for pre-processed images by calculating pixel intensity. The global contrast of the cancerous RBC picture increases with intensity value as shown in Figure 2.



(A) (B) (C)

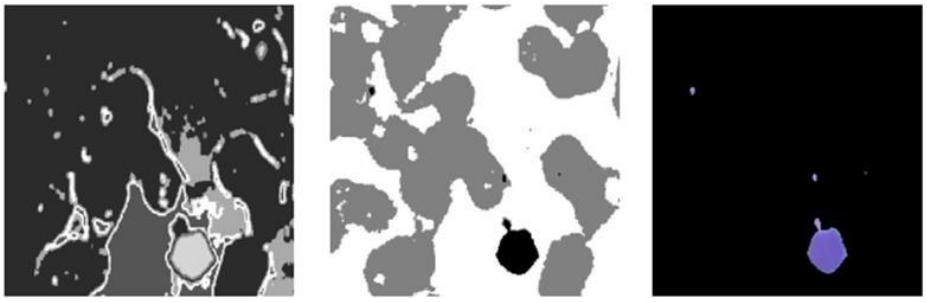
Figure 2: (A) Benign RBC Tumour sample 1 pre-processed picture (B) Histogram-based input colour-converted image and (C) sample image equalization

The colour scale histogram offers the equalization step for detection at the threshold's limited values. For human viewers, the sense of visual borders is growing at a significant pace. The result of histogram-specified equalization has a uniform distribution of intensity and provides intensity values in the input picture. For uniformly based histogram values, it raises the objective of contrast with nominal histogram equalization. Histogram equalization, which modifies image intensities to boost background and foreground contrast, is used to improve performance. When data for constant contrast values at a global level of contrast are available, this approach is used. The resulting improved picture for the segmentation stage is more aesthetically pleasing and features distinct nodules.

4.3 Image segmentation

Following the preprocessing stage, the output from this phase is used for segmentation to remove malignant cells from CT RBC pictures. The disorders that are of important structures with categorization, the segments of different classes of tissue organs, and the imaging at medical condition. The RBC segmentation state is now emerging in the low contrast imaging ambiguity range. Functionally, the suggested map-based segmentation job uses a tagged reference picture and an MRF system to identify feature values.

To effectively recognize cluster pixels, this clustering is identified. K-means centroid begins using the suggested method, effectively assigning from the reference cluster with a decrease in mean aligned between the allocated pixels within the cluster centre. K-means clustering is a technique that effectively analyses online pixel grouping with comparable intensity levels in non-cancerous zones. Radiologists may diagnose patients more accurately by using the suggested approach, which precisely separates malignant cells from non-cancerous cells in fewer affected areas. By determining the minimal area threshold values based on the previous information of the image collected series, the online region-based segmentation is applied. Effective segmentation is provided by the suggested ORBS approach, which processes data live. It makes the foreground pixels' border larger. Segmentation of the picture is applied according to its form and features as shown in Figure 3.



(A) (B) (C)

Figure 3: Benign Red Blood Cell Tumour Sample 1's Segmented Image (A) Segmentation Map (B) K-means Cluster and (C) Suggested ORBS

Even the regular tissue portions are taken into consideration together with all the foreground elements. Segmentation mapping expands the area according to foreground pixel boundaries. A picture may be retrieved based on its form and properties. A preliminary extraction involving minor components exists. The K stands for clustering, which uses processing methods to distinguish between healthy and malignant cells. Using new segmentation, a powerfully simple method splits lighter items at background cells. In smaller blood cells, this divides cancer cells. For use in medical analysis, the suggested online region-based segmentation is carried out automatically. This technique eliminates every flaw in the previous approaches and produces accurate visual results when separating the afflicted RBC cells from the unaffected RBC cells.

Table 1: CT imaging of Benign Red Blood Cell Tumours: Features and Values

Image	Contrast	Clustering Shade values	Clustering prominence values	Correlation density	Dissimilarity coefficient	Probability Values
1	0.453402	-24.35896	324.884	0.928544	0.364405	0.1952
2	0.480787	-23.97541	369.751	0.748975	0.201361	0.1544
3	0.278814	-22.28947	270.458	0.926434	0.160338	0.1613
4	0.425737	-23.97845	296.354	0.954037	0.193531	0.1643
5	0.482967	-24.82447	332.154	0.849975	0.241061	0.1944
6	0.581298	-18.94672	271.394	0.827944	0.173282	0.1552
7	0.491035	-19.35874	264.897	0.754218	0.343656	0.1653
8	0.407573	-17.94246	279.365	0.955328	0.475228	0.1797
9	0.496626	-14.75612	256.684	0.751768	0.203421	0.1861
10	0.595226	-16.38726	222.333	0.959457	0.348617	0.1883
11	0.241714	-21.36726	278.394	0.824956	0.244935	0.1651
12	0.597532	-18.97521	284.654	0.913457	0.396749	0.1766
13	0.523191	-22.97564	345.897	0.745301	0.427157	0.1842
14	0.579228	-24.36994	336.655	0.915996	0.271558	0.1518
15	0.496519	-19.06789	397.254	0.810656	0.378492	0.1371
16	0.395769	-19.87624	378.245	0.708442	0.417742	0.1758
17	0.439558	-17.59251	363.252	0.951269	0.272483	0.1445
18	0.397581	-16.54212	397.845	0.752689	0.274834	0.1650
19	0.394613	-18.42351	387.112	0.870231	0.324856	0.1848
20	0.377559	-23.97511	278.951	0.867869	0.494871	0.1920
21	0.486621	-21.34562	202.646	0.844381	0.374931	0.1990

4.4 Feature extraction

For feature extraction, histogram-based features are often taken into consideration, with feature values determined by the threshold values of the histogram. The classifier uses these properties to classify the illness in the red blood cell image. These database photos are used to generate features based on histograms as given in Table 1.

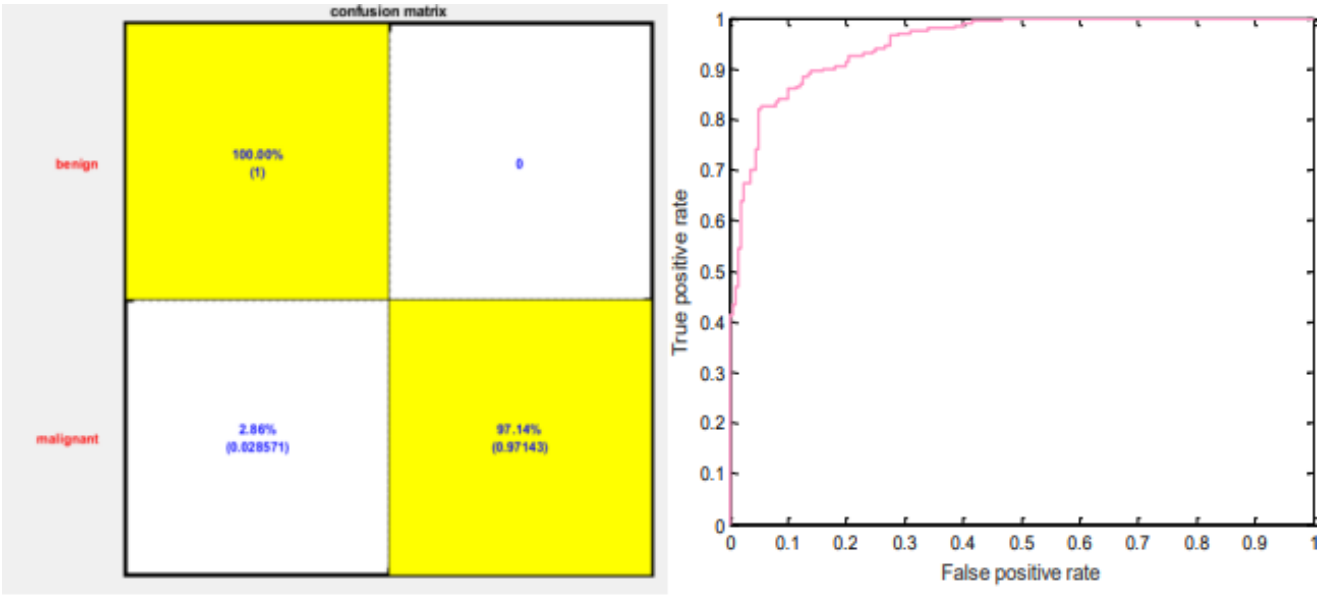
4.5 Experiment Setup

The evaluation was performed by using the statistical tool-rich MATLAB 2018a program for designing framework and to plot the graphs.

5 Results and Discussion

The example pictures in this section are classified using CNN and they are been compared with ANN, KNN, SVM, and BRT classifiers. Efficiency is influenced by sensitivity, specificity, accuracy, precision, and error value, among other things.

For CNN classification, the research study allows ResNet-50 with six fully linked layers in a hidden state. It consists of three layers: hidden, output, and input. The feature extraction section covers ResNet-50, which is made up of building blocks at the output stage that is tightly coupled. Figure 4 shows the connections between each linked block and all other levels. In Figure 4(a), Class 0 denotes a benign red blood cell tumour and Class 1 denotes a malignant red blood cell tumour. With the use of deep learning Histogram-based characteristics, no malignant cases are misclassified, while 50 out of 50 benign cases are accurately classified as benign. Of the fifty instances, 48 have a valid diagnosis of malignancy, whereas the other cases are incorrectly classed as benign (FP). deep learning, with TPs accounting for 48.5 percent, FPs for 51.5 percent, FNs for 64.9 percent, and TNs for 35.1%. The Deep Learning classifier has a 97.1 percent accuracy rate. There are 2.9 percent of persons who are incorrectly categorized.



(a) **(b)**

Figure 4: a) Confusion matrix, b) ROC Curve - CNN classifier

Table 2: Performance values of different classifier

Types	BRT	SVM	KNN	ANN	Proposed
Sensitivity(%)	91.3	94.6	93.1	92.3	96.7
Specificity(%)	93.6	96.7	83.6	93.5	95.6
Precision(%)	95.5	94.3	84.3	83.7	96.4
Accuracy(%)	92.7	96.1	92	93.7	96.1
Time (second)	43	30	27	35	22

In the context of blood cancer detection, the performance of different machine learning models can be assessed through metrics such as sensitivity, specificity, precision, and accuracy. Among the models tested CNN, SVM, K-NN, ANN, and BRT. DL CNN stands out as the most effective model for this task. A comparison of several methodologies' sensitivity, specificity, precision, and accuracy values on CT images of RBC tumours is shown in Figure 5. Here, the X-axis displays the metrics of several classifiers, while the Y-axis shows the results, expressed as a percentage, according to the classification methods.

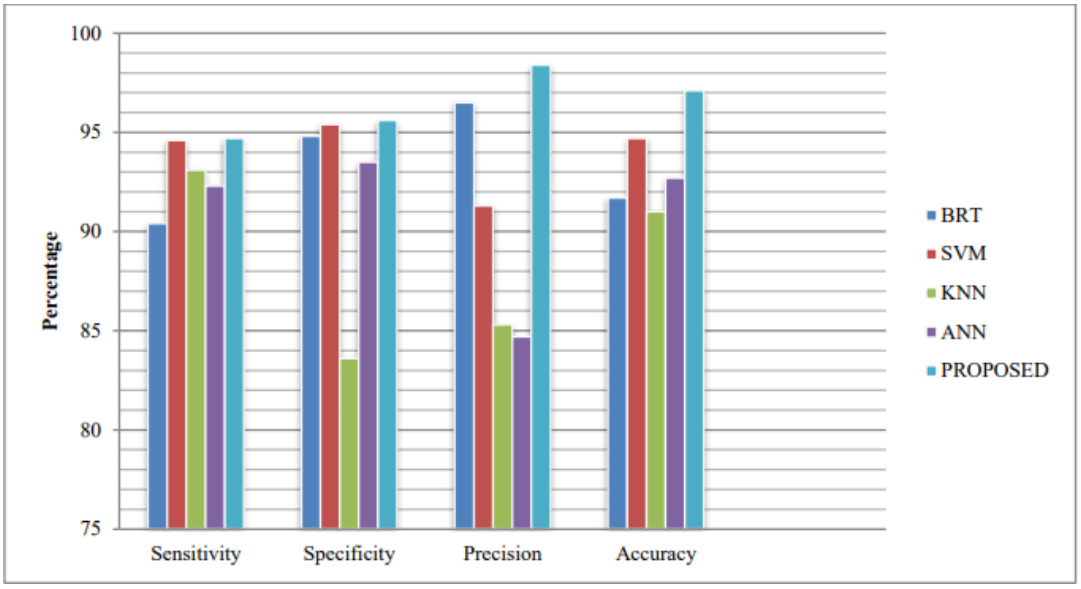


Figure 5: Comparisons of Error values with different classifier

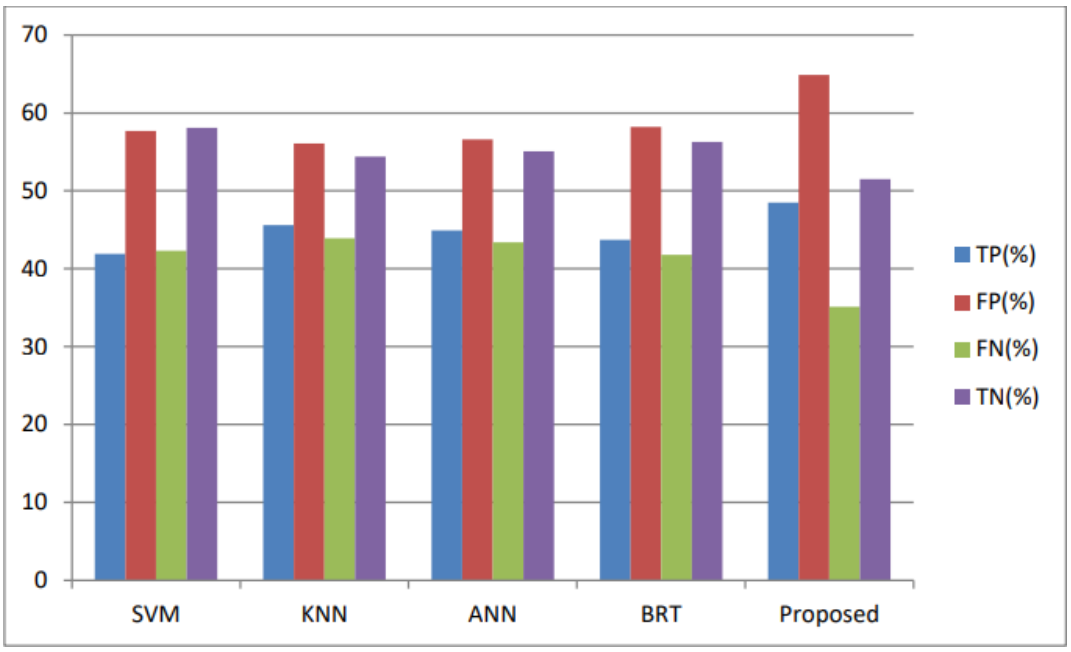


Figure 6: Comparative analyses of confusion matrix values of various classifiers

The comparative study of several classifiers using the TP, FN, FP, and TN values is shown in Figure 6. Comparative study compares the blood tumour database image classification to that of other classifiers. Higher real positive values are shown by the suggested deep learning classifier. Comparative analysis reveals that the classifiers used in the current thesis study provide superior outcomes across a range of classifiers. Accuracy using various classification methods on pictures of RBC cancer is shown in Figure 7. The suggested classifier and several other

classifiers are compared using percentage values. When compared to existing classifiers, the suggested DL-based classifier yields a superior classification accuracy of 97.1%.

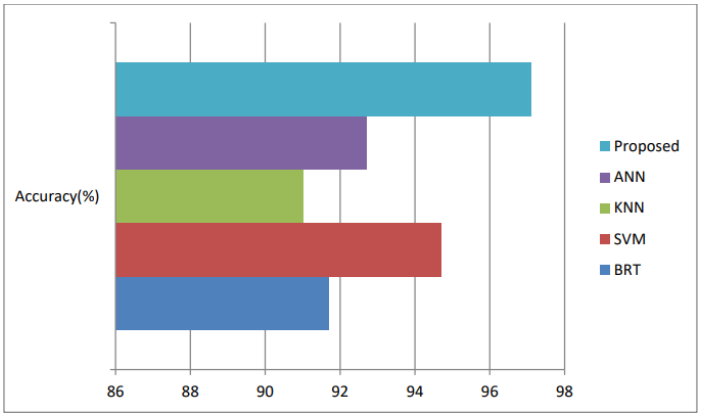


Figure 7: Comparisons of accuracy with different classifier

6 Conclusion

The study presents a comparative analysis of RBC cancer images using various classifiers. The study utilizes images from the BCCD database of RBC cancer and applies preprocessing techniques, including colour conversion and histogram equalization, to enhance image quality. Tumour segmentation is performed using two methods: map segmentation and online region-based segmentation, which are complemented by histogram-based feature extraction. To classify the segmented RBC tumour images as benign or malignant, the ResNet-50 deep learning model is employed. The model's performance is evaluated based on key metrics, including specificity, sensitivity, precision, error rate, and accuracy. The DL approach is compared to other classifiers, such as SVM, KNN, BRT, and ANN. For each classifier, performance is further assessed using ROC curves and confusion matrices. The proposed DL model achieves an impressive accuracy of 97.1%, demonstrating its effectiveness for RBC tumour classification.

References

[1] Bukhari, Maryam, Sadaf Yasmin, Saima Sammad, and Ahmed A. Abd El-Latif. "A deep learning framework for leukemia cancer detection in microscopic blood samples using squeeze and excitation learning." *Mathematical problems in engineering* 2022, no. 1 (2022): 2801227.

[2] Selvakumaran, S., C. Poonguzhali, Elango Baskaran, Ayushi Arya, Deepak Kumar Nayak, and Sandeep Gupta. "Integrated Hybrid Deep Learning Approach for Precise Blood Cancer Detection." In *2024 10th International Conference on Communication and Signal Processing (ICCSP)*, pp. 496-501. IEEE, 2024.

[3] Nasr, Mona, Alaa El Din M. El Ghazali, and Amr I. Shehta. "Deep Learning Models for Early Detection of Blood Cancer Disease." In *International Conference on Advanced Intelligent Systems and Informatics*, pp. 53-65. Springer, Cham, 2024.

[4] Reddy, Dukka KarunKumar, Himansu Sekhar Behera, Janmenjoy Nayak, Pandi Vijayakumar, Bighnaraj Naik, and Pradeep Kumar Singh. "Deep neural network-based anomaly detection in Internet of Things network traffic tracking for the applications of future smart cities." Transactions on Emerging Telecommunications Technologies 32, no. 7 (2021): e4121.

[5] Reddy, D.K.K., Behera, H.S., Nayak, J., Routray, A.R., Kumar, P.S. and Ghosh, U., 2022. A fog-based intelligent secured iomt framework for early diabetes prediction. In *Intelligent Internet of Things for Healthcare and Industry* (pp. 199-218). Cham: Springer International Publishing.

[6] Siddiqui, Shahan Yamin, Iftikhar Naseer, Muhammad Adnan Khan, Muhammad Faheem Mushtaq, Rizwan Ali Naqvi, Dildar Hussain, and Amir Haider. "Intelligent breast cancer prediction empowered with fusion and deep learning." *Computers, Materials and Continua* 67, no. 1 (2021): 1033-1049.

[7] A Comprehensive Study of Machine Learning Algorithms for Predicting Leukemia Based on Biomedical Data, author: Anamika Das Mou and Pratap Kumar Saha, 2019 2nd International Conference on Innovation in Engineering and Technology (ICIET), the year 2019.

[8] L. H. S. Vogado, R. D. M. S. Veras, A. R. Andrade, F. H. D. de Araujo, R. R. V. Silva, and K. R. T. Aires, "Diagnosing Leukemia in Blood Smear Images Using an Ensemble of Classifiers and Pre-Trained Convolutional Neural Networks," 2017 30th SIBGRAPI Conference on Graphics, Patterns, and Images (SIBGRAPI), Niteroi, Brazil, 2017, pp. 367-373.

[9] B. K, G. A. T, N. G, and S. Umamaheswari, "Improving the Performance of Leukemia Detection using Machine Learning Techniques," 2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2022, pp. 867-872.

[10] Daqqa KA, Maghari AY, Al Sarraj WF," Prediction and diagnosis of leukaemia using classification algorithms",8thInternational Conference on Information Technology (ICIT), 2017 May 17 (pp. 638-643).

[11] Chua, L. O. (1997). CNN: A vision of complexity. *International Journal of Bifurcation and Chaos*, *7*(10), 2219-2425.