International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 24 Issue 09, SEP, 2024

A SYSTEMATIC REVIEW ON RECENT ADVANCEMENTS IN DEEP AND MACHINE LEARNING BASED DETECTION AND CLASSIFICATION OF ACUTE LYMPHOBLASTIC LEUKEMIA

Aroosa Aymen[1], Dr. K. Palani[2]
[1]PG Scholar, Department of CSE, Shadan Women's College of Engineering and Technology, Hyderabad, aroosaymen005@gmail.com
[2]Professor, Department of CSE, Shadan Women's College of Engineering and Technology, Hyderabad, Principalswcet2020@gmail.com

ABSTRACT

Automatic leukemia or blood cancer detection is a difficult job that is crucially needed in medical facilities. It is crucial for early diagnosis and the planning of treatments. A hematological condition called leukemia affects white blood cells (WBCs) and originates in the bone marrow. The preferred method for early identification of leukemia is microscopic analysis of WBCs because it is less invasive and expensive. There haven't been many studies in the literature that offer a thorough examination of deep and machine learning-based acute lymphoblastic leukemia (ALL) diagnosis. An organized review of the most recent developments in this field of knowledge is provided in this article. Here, multiple AI-based ALL detection methods are systematically evaluated with respect to their benefits and drawbacks. This scheme's reviews are carried out in an organized manner. For this reason, segmentation strategies are broadly divided into three groups: deep learning-based strategies, traditional machine learning-based strategies, and Traditional machine learning-based approaches to categorization are given as supervised and unsupervised machine learning. Additionally, convolution neural networks (CNN), recurrent neural networks (RNN), and auto encoders are subcategories of deep learning-based classification techniques. Then, traditional CNN, transfer learning, and other developments in CNN are further divided into CNN-based classification schemes. These classification schemes are also offered, along with a brief explanation of their significance. Additionally, a critical analysis is carried out to give a clear picture of the most recent studies in this area. Finally, The discussion of a number of difficult challenges and potential future directions may help readers create fresh research questions in this field.

INTRODUCTION

Leukemia is a blood malignancy that interferes with the bone marrow's ability to create white blood cells (WBCs). It increases the quantity of aberrant WBCs, which causes immunity to decline. WBC is a vital component of blood, just like erythrocytes (red blood cells) and platelets, which are also two other essential components. Leukemia is usually divided into two types: acute and chronic. WBCs have a nucleus and cytoplasm. While chronic leukemia takes considerably longer to worsen, acute leukemia develops swiftly and progresses to the worst stage. The French-American-British (FAB) classification model divides acute leukemia into two subgroups: acute lymphoblastic leukemia (ALL) and acute myeloid leukemia (AML). Chronic lymphocytic leukemia (CLL) and chronic myeloid leukemia (CML) are the two subtypes of chronic leukemia, respectively. Leukemia thus has four different subtypes: ALL, AML, CLL, and CML. The bone marrow, blood, and extra medullar sites are all significantly affected by ALL, a blood malignancy with a fatty growth pattern. B lymphatic cells are more prevalent than T cells in ALL. B-cells guard against germ infection, but T-cells eliminate contaminated cells.

OBJECTIVE

The type of camera, microscope, light source, camera angle, illumination change, and noise all affect how well microscopic images are captured. Stain normalization, an essential preprocessing procedure that deals with changes in capture environments, notably variations in lighting conditions, normalizes all stain slides as a result. It improves segmentation and classification performance by reducing lighting and color fluctuations brought on by various capture environments for microscopic pictures collected from various laboratories. Histogram equalisation, or modified forms of it such as adaptive histogram equalisation and contrast-limited adaptive histogram equalisation, is a straightforward solution to this problem. Two U-Net-style modules are used in the effective coupled, self-supervised architecture Gimlet et al. have presented.

PROBLEM STATEMENT

This article provides a succinct review of recent developments in ALL detection and categorization using deep and machine learning. In order to effectively detect ALL, various segmentation, feature extraction, and classification techniques currently in use have been examined. This review also led us to the conclusion that unsupervised machine learning schemes are preferable for segmentation tasks, while supervised machine learning schemes are recommended for classification challenges. However, as it produces good performance even in tiny datasets, deep learning, particularly transfer learning, and has emerged as a preferred strategy for automatic and more reliable detection and classification of ALL.

EXISTING SYSTEM

Leukocytes, also referred to as white blood cells or WBCs, are an essential part of the immune system that defends against bacteria, viruses, and other foreign substances. These blood-circulating, nucleated cells originate in the bone marrow and are distributed throughout the lymphatic system. Leukocytes are different from other blood cells like red blood cells and platelets; thus, they are primarily divided into two categories based on lymphoid and myeloid cell inheritance and granulocyte and granulocyte cell structure.

Disadvantage of Existing System

When the target classes overlap and the data set has more sound, it does not operate very well. The support vector machine will perform poorly when the number of attributes for each data point exceeds the number of training data specimens.

PROPOSED SYSTEM

This article provides a succinct review of recent developments in ALL detection and categorization using deep and machine learning. We have examined numerous segmentation, feature extraction, and classification techniques currently in use that is used to effectively detect ALL. This review also led us to the conclusion that unsupervised machine learning schemes are preferable for segmentation tasks, while supervised machine learning schemes are recommended for classification challenges. However, as it produces good performance even in tiny datasets, deep learning, particularly transfer learning, and has emerged as a preferred strategy for automatic and more reliable detection and classification of ALL.

Advantages of Proposed System

The MobileNetV2 models are significantly faster than the MobileNetV1 models. They require 30% fewer parameters, use two times fewer processes, and are around 30–40% faster on a Google Pixel phone.

RELATED WORKS

Along with transfer learning models, several other effective, quicker CNN models have recently been proposed. The improved, faster CNN models You only look once (YOLO), YOLOv2, YOLOv3, and YOLOv4 are some of the most well-known. YOLO is a successful CNN technique that improves computing efficiency by using CNN architecture to carry out localization and classification tasks at the same time. As a result, faster object detection results. In order to forecast the location of a bounding box, a convolution layer is applied. However, it has a problem with localization errors. In order to reduce localization mistakes, YOLOv2 is a modified version of YOLO that uses anchor boxes rather than the convolution layer. Better item detection and classification follow as a result. YOLOv3 is an enhanced version of YOLO that precisely detects bounding boxes by using a logistic regression-based prediction approach. YOLOv2 has been used by Ai-Quad and Seen to efficiently and healthily classify ALL. A modified version of YOLO called YOLOv2 increases both speed and accuracy. They observe that Yolov2 with random resizing performs better than Yolov3 and Yolov2 without random resizing.

METHODLGY OF PROJECT

The machine learning techniques for recognizing delirium risk and its occurrence during hospitalization are briefly introduced in this section. Because of their superior performance in numerous clinical application areas, machine learning techniques, particularly artificial neural networks (ANNs), support vector machines (SVMs), and random forests (RFs), have been widely employed to evaluate delirium risk. The seven most pertinent studies that sought to create a model for recognizing delirium risk in clinical processes are summarized in Table I. In order to predict postoperative delirium, Devoid et al. employed EHR data that contained 70 preoperative predictors, retrospectively gathered from 51457 patients admitted for more than 24 hours after any type of inpatient operational treatment.

MODULES DESCSRIPTION:

1 Dataset:

For the first module, we created a method to obtain the input dataset for training and testing. The dataset was obtained from the blood cancer detection programme.6516 images in total make up the dataset for blood cancer detection.

2 Importing the necessary libraries:

Python will be used for this. The first step is to import the relevant libraries, including keras (for creating the primary model), sklearn (for separating training and test data), PIL (for turning images into arrays of numbers), numpy (for splitting the training and test data), matplotlib (for plotting the data), and tensor flow.

3 Retrieving the images:

We'll get the images back together with their labels. The photographs should then be resized to 224x224 since all images must be the same size for identification. Then transform the images into a numpy array.

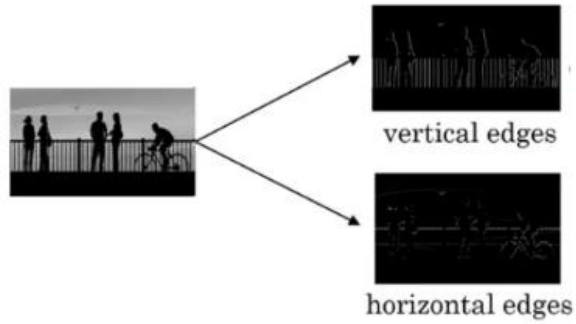
4 Splitting the dataset:
Dividing the dataset into train and test halves 80 percent train data, 20 percent test data

A. MobileNetV2 Model
A convolutional neural network 53 layers deep is called MobileNet-v2.You can load a retrained version of the network from the Image Net database, which has been trained on more than a million photos and can categories images into 1,000 object categories. An excellent feature extractor for object recognition and segmentation is MobileNetV2. For instance, when used in conjunction with the recently released SSDLite, the new model for detection performs around 35% faster than MobileNetV1 while maintaining the same level of accuracy.

Computer Vision
The computer vision issues that we shall address in this essay include some of the following: Image classification, first

1. Detection of objects
2. Transfer in the neural style the fact that the input data might grow significantly is one of the main issues with computer vision. Consider a picture that is 68 x 68 x 3 pixels in size. The input feature dimension changes to 12,288 as a result. If we use bigger photos, like those that are 256 x 256 x 3, this will get even larger. Now, depending on the number of hidden layers and hidden units, if we give a neural network such a large input, the number of parameters will skyrocket. More processing and memory requirements will arise from this, which most people cannot handle, can deal with.

Edge Detection Illustration
In the prior article, we observed that a neural network's initial layers can identify edges in a picture. Deeper levels may be able to identify the origin of the objects, and further deeper layers may identify the origin of full things.
We will concentrate on identifying edges in an image in this part. Let's say we are shown the picture below:



5 Building the model:
We'll use the sequential model from the Keras library for construction. Then, to create a convolutional neural network, we will add the additional 53 deep layers. 32 filters were employed in the first two Conv2D layers, and the kernel size was (5, 5). We left the MaxPool2D layer's pool size at (2, 2), which means it will choose the highest value from each section of the image that is 2 x 2 pixels. By doing this, the image's dimensions will shrink by a factor of 2. We kept the dropout rate in the dropout layer at 0.25, which indicates that 25% of neurons are eliminated arbitrarily. We reapply these three layers with a few tweaks to the parameters. Next, we apply a flatten layer to turn 2-D data into a vector in 1-D space. Following this layer are a dense layer, a dropout layer, and another dense layer. Two nodes, either yes or not, are produced by the final dense layer. This layer makes use of the soft max activation function to estimate which of the two possibilities has the highest likelihood and to provide a probability value.

6. Apply the model and plot the graphs for accuracy and loss:
The model will be built, and the fit function will be used to apply it. There will be two batches. The graphs for accuracy and loss will then be plotted. The average training accuracy was 100%, while the average validation accuracy was 99.99%.

7. Accuracy on test set:
We achieved 100% test set accuracy.

8. Saving the Trained Model:
The first thing to do is save your trained and tested model in an environment where it is ready for production. h5
Make sure your environment has it installed.
The model will now be dumped into an a.h5 file after we load the module.

ALGORITHM USED IN PROJECT
A convolutional neural network 53 layers deep is called MobileNet-v2.

- The Image Net database contains a pre-trained version of the network that has been trained on more than a million photos.
- The trained network is capable of classifying photos into 1,000 different item categories.
- The object identification and segmentation feature extractor, MobileNetV2, is particularly efficient. For instance, the new model is approximately 35% faster for detection when used with the recently released SSDLite while maintaining the same accuracy as MobileNetV1.
- The model is publicly available through the Tensor Flow Object Detection API

Benefits of NLP

- In compared to MobileNetV1, the MobileNetV2 models are substantially faster.
- They use 2 times less operations, are more accurate, use 30% fewer parameters, and run roughly 30% to 40% faster on a Google Pixel phone.

DATA FLOW DIAGRAM

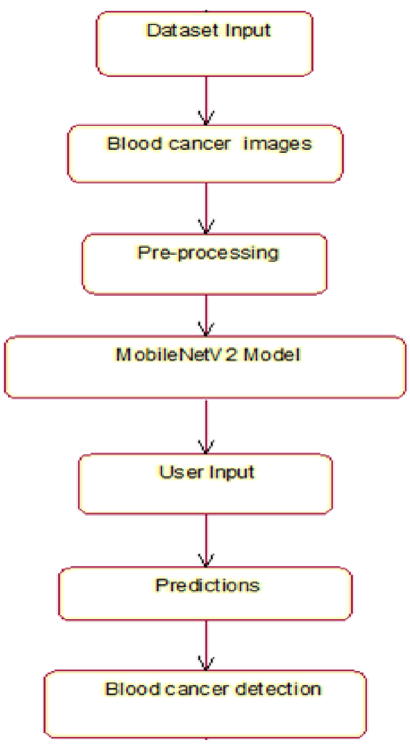


Fig: 7 Flow Diagrams of Modules

Impediments of DL

It is a modified version of the Mobile-Net model that includes the introduction of an inverted residual bottleneck structure. To make the model computationally quick, it additionally makes use of width and resolution multipliers as well as depth-wise separable convolutions. Two blocks, MobileNetV2 block 1 (MVB1) and MobileNetV2 block 2 (MVB2), are proposed in this architecture to increase speed and performance. As shown, A skip connection is offered between two bottleneck layers in the inverted residual bottleneck structure with stride 1 that is represented by MVB1 as illustrated.

SYSTEM ARCHITECTURE

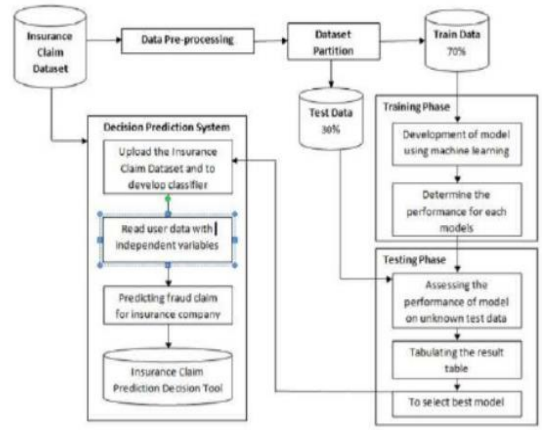
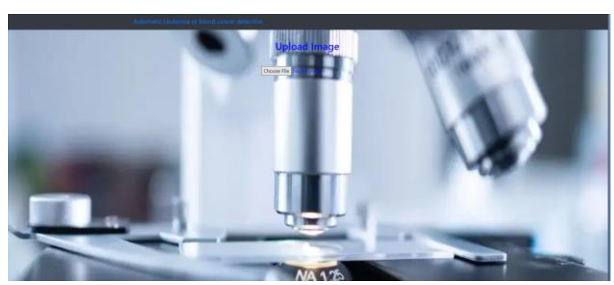
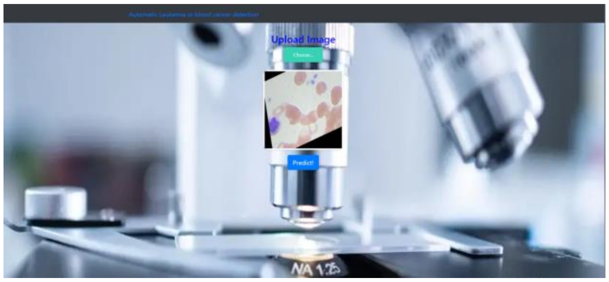


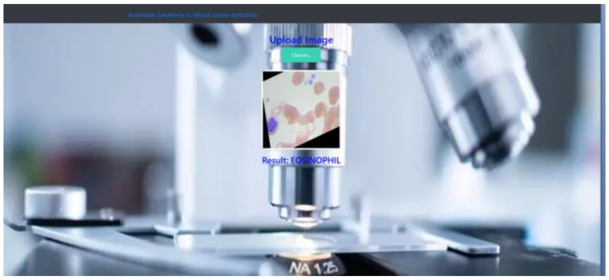
Fig: 8 SYSTEM ARCHITECTURE OF PROJECT

RESULTS AND DISCUSSION

The existing effective models can be modified to create more complex deep learning or transfer learning-based systems that are faster and more accurate. To simplify optimisation and reduce co-adaptation among classification layer parameters, for instance, more effective classification layers can be used in place of the soft-max layer or the fully connected classification layer. By choosing more important and desired connections, it is possible to reduce the number of connections while enhancing the interclass angles. Thus, the system can operate more quickly and efficiently.







FUTURE ENHANCEMENT

By combining a powerful active contour/level set approach with the marker-based watershed algorithm or the level set method with deep learning techniques, segmentation performance can be improved. A method based on efficient deep learning (especially transfer learning) may produce segmentation that is more accurate. It is possible to conduct research to create an ALL-segmentation system based on deep learning (especially transfer learning). By altering the current effective models, more sophisticated deep learning or transfer learning-based systems that are faster and more accurate can be created. To simplify optimization and reduce co-adaptation among classification layer parameters, for instance, more effective classification layers can be used in place of the softmax layer or fully connected classification layer. By choosing more important and desired connections, it is possible to reduce the number of connections while enhancing the interclass angles. Thus, the system can operate more quickly and efficiently. By recommending more effective and reliable machine learning, deep learning, or transfer learning techniques, classification performance can be enhanced.

CONCLUSION

This article provides a succinct review of recent developments in ALL detection and categorization using deep and machine learning. In order to effectively detect ALL, various segmentation, feature extraction, and classification techniques currently in use have been examined. This review also led us to the conclusion that unsupervised machine learning schemes are preferable for segmentation tasks, while supervised machine learning schemes are recommended for classification challenges. However, as it produces good performance even in tiny datasets, deep learning, particularly transfer learning, has emerged as a preferred strategy for automatic and more reliable detection and classification of ALL. We also learned from this study that the MobileNetV2-ResNet18 architecture, which combines the advantages of both techniques, produces the best ALL detection performance in the ALLIDB1 dataset. By combining the advantages of both strategies, MobileNetV2-SVM exhibits admirable classification performance in the ALLIDB2 dataset. We have also spoken about the difficult problems and the potential for this research topic in the future. This article is intended to aid in the analysis of current developments in ALL detection and to stimulate additional investigation.

REFERENCES:

[1] P. K. Das, S. Meher, R. Panda, and A. Abraham, ‘‘An efficient blood-cell segmentation for the detection of hematological disorders,’’ IEEE Trans. Cybern., early access, Mar. 18, 2021, doi: 10.1109/TCYB.2021.3062152.

[2] P. K. Das and S. Meher, ‘‘An efficient deep convolutional neural network based detection and classification of acute lymphoblastic leukemia,’’ Expert Syst. Appl., vol. 183, Nov. 2021, Art. no. 115311.

[3] P. P. Banik, R. Saha, and K.-D. Kim, ‘‘An automatic nucleus segmentation and CNN model based classification method of white blood cell,’’ Expert Syst. Appl., vol. 149, Jul. 2020, Art. no. 113211.

[4] L. H. Vogado, R. M. S. Veras, F. H. D. Araujo, R. R. V. Silva, and K. R. T. Aires, ‘‘Leukemia diagnosis in blood slides using transfer learning in CNNs and SVM for classification,’’ Eng. Appl. Artif. Intell., vol. 72, pp. 415–422, Jun. 2018.

[5] P. K. Das, B. Nayak, and S. Meher, ‘‘A lightweight deep learning system for automatic detection of blood cancer,’’ Measurement, vol. 191, Mar. 2022, Art. no. 110762.

[6] P. K. Das, A. Pradhan, and S. Meher, ‘‘Detection of acute lymphoblastic leukemia using machine learning techniques,’’ in Machine Learning, Deep Learning and Computational Intelligence for Wireless Communication. Singapore: Springer, 2021, pp. 425–437.

[7] P. K. Das and S. Meher, ‘‘Transfer learning-based automatic detection of acute lymphocytic leukemia,’’ in Proc. Nat. Conf. Commun. (NCC), Jul. 2021, pp. 1–6.

[8] S. Mishra, B. Majhi, and P. K. Sa, ‘‘Texture feature based classification on microscopic blood smear for acute lymphoblastic leukemia detection,’’ Biomed. Signal Process. Control, vol. 47, pp. 303–311, Jan. 2019.

[9] P. K. Das, S. Meher, R. Panda, and A. Abraham, ‘‘A review of automated methods for the detection of sickle cell disease,’’ IEEE Rev. Biomed. Eng., vol. 13,

pp. 309–324, 2020. VOLUME 10, 2022 81759 P. K. Das et al.: Systematic Review on Recent Advancements in Deep and Machine Learning.

[10] P. K. Das, P. Jadoun, and S. Meher, ''Detection and classification of acute lymphocytic leukemia,'' in Proc. IEEE-HYDCON, Sep. 2020, pp. 1–5.

[11] N. Patel and A. Mishra, ''Automated leukaemia detection using microscopic images,'' Proc. Comput. Sci., vol. 58, pp. 635–642, Jan. 2015.

[12] J. M. Bennett, D. Catovsky, M.-T. Daniel, G. Flandrin, D. A. G. Galton, H. R. Gralnick, and C. Sultan, ''Proposals for the classification of the acute leukaemias French-American-British (FAB) co-operative group,'' Brit. J. Haematol., vol. 33, no. 4, pp. 451–458, Aug. 1976. T. Terwilliger and M. Abdul-Hay, ''Acute lymphoblastic leukemia: A comprehensive review and 2017 update,'' Blood Cancer J., vol. 7, no. 6, p. e577, Jun. 2017.

[14] K. AL-Dulaimi, J. Banks, K. Nugyen, A. Al-Sabaawi, I. Tomeo-Reyes, and V. Chandran, ''Segmentation of white blood cell, nucleus and cytoplasm in digital haematology microscope images: A review–challenges, current and future potential techniques,'' IEEE Rev. Biomed. Eng., vol. 14, pp. 290–306, 2021.

[15] S. Shafique and S. Tehsin, ''Acute lymphoblastic leukemia detection and classification of its subtypes using pretrained deep convolutional neural networks,'' Technol. Cancer Res. Treatment, vol. 17, Sep. 2018, Art. no. 1533033818802789.

[16] S. Mishra, B. Majhi, and P. K. Sa, ''GLRLM-based feature extraction for acute lymphoblastic leukemia (ALL) detection,'' in Recent Findings in Intelligent Computing Techniques. Singapore: Springer, 2018, pp. 399–407.

[17] S. Mishra, B. Majhi, P. K. Sa, and L. Sharma, ''Gray level co-occurrence matrix and random forest based acute lymphoblastic leukemia detection,'' Biomed. Signal Process. Control, vol. 33, pp. 272–280, Mar. 2017.

[18] J. Laosai and K. Chamnongthai, ''Classification of acute leukemia using medical-knowledge-based morphology and CD marker,'' Biomed. Signal Process. Control, vol. 44, pp. 127–137, Jul. 2018.

[19] Z. Moshavash, H. Danyali, and M. S. Helfroush, ''An automatic and robust decision support system for accurate acute leukemia diagnosis from blood microscopic images,'' J. Digit. Imag., vol. 31, no. 5, pp. 702–717, Oct. 2018.

[20] S. Kumar, S. Mishra, and P. Asthana, ''Automated detection of acute leukemia using K-mean clustering algorithm,'' in Advances in Computer and Computational Sciences. Singapore: Springer, 2018, pp. 655–670.