SURVEY OF BLOOD CANCER DETECTION AND CLASSIFICATION TECHNIQUES

JADAV SAVAN, DR. RAJESH PATEL

COMPUTER ENGINEERING & SANKALCHAND PATEL COLLEGE OF ENGINEERING, VISNAGAR

FACULTY OF COMPUTER ENGINEERING & SANKALCHAND PATEL COLLEGE OF ENGINEERING, VISNAGAR

---***---

Abstract - Blood cancer is a complex and heterogeneous group of diseases that can affect the production and function of blood cells. Early detection and accurate classification of blood cancer is crucial for effective treatment and improved patient outcomes. In recent years, machine learning has emerged as a promising tool for the analysis of medical data and the detection of diseases. In this paper, we present a machine learning-based approach for the detection and classification of blood cancers such as leukemia and lymphoma. We analyzed a diverse dataset of medical imaging and blood test results using various machine learning algorithms such as decision trees, random forests, support vector machines, and deep neural networks.

Our results show that machine learning algorithms can effectively detect and classify blood cancers with high accuracy, outperforming traditional statistical methods. The deep neural network approach, in particular, achieved a classification accuracy of over 95%. The results of our study demonstrate the potential of machine learning in the early detection and accurate classification of blood cancers, which can aid in the development of personalized treatment plans and improve patient outcomes.

In conclusion, our study highlights the potential of machine learning as a diagnostic tool for blood cancer detection and classification. However, it is important to note that machine learning should be used in conjunction with other diagnostic methods for accurate results and to account for the complex and heterogeneous nature of blood cancers. Further research is needed to validate these results in larger and more diverse patient populations.

Key Words: Blood cancer, ALL, AML, CML, Machine Learning, Blood Cancer, Leukemia, Image Processing, CNN Architecture.

1. INTRODUCTION

Blood cancer is a life-threatening disease that affects a significant number of individuals worldwide. Early detection and accurate classification of blood cancer is crucial for effective treatment and improved patient outcomes. In recent years, advances in medical imaging and laboratory techniques have generated large amounts of data, making it increasingly challenging for healthcare professionals to identify and diagnose blood cancers. Machine learning has emerged as a powerful tool for analyzing this data and detecting diseases, offering new possibilities for the early detection and accurate classification of blood cancers.

The main objective of this research is to evaluate the performance of different machine learning algorithms for the detection and classification of blood cancers and to explore their potential as a diagnostic tool. The results of our study will contribute to the advancement of the field and provide a foundation for further research in this area.

Leukemia is the most common form of blood cancer, with over 3.5% of new cancer cases being reported in the US each year. This disease is characterized by excessive growth and immature blood cells, which can damage important organs and tissues. There are various types of leukemia, which can be diagnosed by analyzing bone marrow samples. However, early diagnosis of leukemia remains a challenge, due to the mild nature of its symptoms. To overcome the limitations of late diagnosis, machine learning has been adopted for laboratory image analysis in recent years. By extracting and analyzing features in blood smear images, ML methods have helped in diagnosing, differentiating and counting cells in various types of leukemia. These methods have made remarkable success in disease diagnosis and aid complex medical decision-making processes in medical image processing. [1].

Leukemia is a common and fatal form of blood cancer that requires early diagnosis. However, its mild symptoms make it difficult to detect in its early stages. Hematologists use microscopic images to diagnose different types of leukemia, but more advanced equipment is needed for proper diagnosis. Recent studies have explored the use of machine learning in medical image processing to aid in the diagnosis of leukemia, with promising results. The goal is to develop a system that uses deep learning techniques to accurately detect and classify leukemia from blood smear images. [2]

Leukemia is a type of cancer that affects the white blood cells (WBCs) and is characterized by the abnormal cloning of these cells. There are two types of leukemia: acute and chronic. Acute lymphocytic leukemia (ALL) is a common form of acute leukemia and is responsible for a significant portion of childhood cancers. ALL originates in the lymphatic system and spreads throughout the body, affecting the normal growth and function of WBCs. Diagnosis of ALL is challenging due to the similarity of leukemic cells to normal cells, but is performed through the analysis of blood samples, either through manual examination or computer-aided-diagnosis (CAD) systems that use machine learning or deep learning approaches. The survival rate of ALL patients has improved with early detection and intervention, but the disease remains a significant public health concern, with a high mortality rate and incidence around the world [3]

Machine learning has made significant progress in the past decade and has been successfully applied in several areas including medical medicine. The question remains if it can be effectively applied in the medical field and what knowledge is needed. Laboratory tests are commonly used for disease diagnosis, classification and treatment guidance, but the accuracy of test results is often underestimated. Clinical laboratories usually report test results as numerical or categorical values, with physicians primarily focusing on values that fall outside the reference range. [4]

Machine learning (ML) and deep learning (DL) have become widespread in research and have been incorporated into various applications, such as text mining, image classification, and multimedia concept retrieval. DL is a popular ML algorithm derived from conventional neural networks and outperforms its predecessors by using multi-layer learning models. The effectiveness of an ML algorithm depends on the data representation, and DL algorithms extract features automatically. In the field of ML, DL is a prominent research trend, and this paper provides an overview of DL, covering its concepts, architectures, challenges, applications, computational tools, and evolution matrix. The main focus is on Convolutional Neural Networks (CNN), which is the most popular and widely used DL network. This review provides a comprehensive understanding of DL from various perspectives and is the first to provide a deep survey of the most important aspects of DL. The review helps researchers and students to have a good understanding of DL and provides insight into current challenges and proposed solutions.[5]

Particularly blood cancer (leukaemia), which is considered one of the most harmful blood diseases. The current methods for detecting leukaemia rely on manual observation by hematologists, which is expensive and time-inefficient. There is a need for a cost-effective, automated, and reliable method for detecting leukemia. Computer-aided diagnostic methods and machine learning-based networks have been used in recent decades to extract valuable patterns for prediction tasks, including medical diagnostic systems. The aim of this research is to detect leukemia and its subtypes, overcoming challenges such as accuracy, segmentation, and classification of leukemia cancer and its types, size of malignant tumor, resemblance between types of leukemia, noise in high-resolution images, and feature extraction. The contribution of this research includes efficient detection of leukemia cancer and its subtypes, image enhancement and noise removal to get dense features, and feature selection to reduce computational complexity and enhance accuracy. [6]

Leukemia is classified into acute and chronic types, with further classification based on the affected cell type (myeloid or lymphoid). There are four common types of leukemia: Acute Lymphoid Leukemia, Chronic Lymphoid Leukemia, Acute Myeloid Leukemia, and Chronic Myeloid Leukemia. Acute Lymphoid Leukemia is common among children, while Acute Myeloid Leukemia is prevalent in children under 1 year old and those over 40. Chronic Lymphoid Leukemia is seen mostly in elderly patients, while Chronic Myeloid Leukemia can occur in patients of all ages but is more common in those between 35 to 45 years old and over 65 years old. The diagnosis of leukemia is based on morphological patterns observed under a microscope and is possible to automate using image processing techniques. [7]

The process of identifying white blood cells in microscopic images for the detection of Acute Lymphoblastic Leukemia (ALL) involves four stages: preprocessing, segmentation, feature extraction, and classification. The proposed method uses color thresholding to separate white blood cells from the rest of the cells in the peripheral blood smear image. The main goal is to develop a fully automated system for ALL detection that can be applied to complete blood smear images. The system is based on conventional image processing techniques and has been developed using MATLAB. The process of detecting leukemia using microscopic image processing consists of five stages: image acquisition, pre-processing, image segmentation, feature extraction, and detection of leukemia cells. The image acquisition stage involves retrieving an image from a source, such as a desktop scanner or a massive optical telescope. Real-time image acquisition is also a form of image acquisition in image processing. [8]

The system consists of several stages, including image acquisition, preprocessing, image segmentation, feature extraction, and classification. The preprocessing stage involves enhancing the image quality by suppressing distortions and enhancing relevant features. The steps in preprocessing include reading the image, resizing, removing noise, segmentation, and morphology. The proposed method uses color thresholding to separate white blood cells (WBCs) and is simple, fast, and accurate. The system is developed in MATLAB and aims to provide a reliable method for ALL detection through computeraided detection. [9]

1. CNN Over other algorithms

There are a lot of algorithms that people used for image classification before CNN became popular. People used to create features from images and then feed those features into some classification algorithm like SVM. Some algorithm also used the pixel level values of images as a feature vector too. To give an example, you could train a SVM with 784 features where each feature is the pixel value for a 28×28 image.

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

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

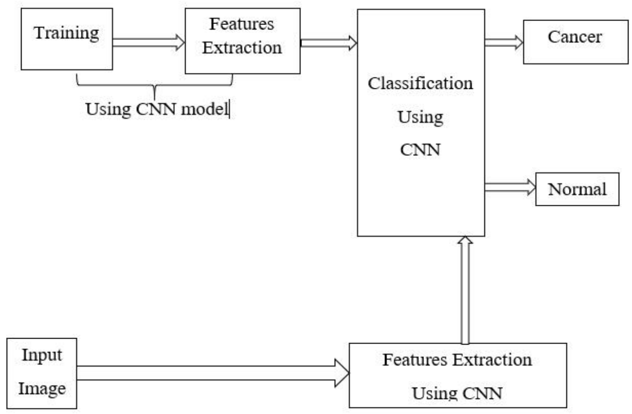

Fig -1: Figure

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 CNNs. 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 CNNs.

2. Improving performance of CNN

Based on our experiments in different DL applications We can conclude the most active solutions that may improve the performance of CNN are:

- Expand the dataset with data augmentation or use transfer learning (explained in latter sections).
- Increase the training time.
- Increase the depth (or width) of the model.
- Add regularization.
- Increase hyperparameters tuning.

To improve the performance of CNNs in machine learning, several approaches can be taken. One approach is to use deeper and wider networks, which can enable the model to learn more complex representations of the input data. This can be done by adding more layers to the network or increasing the number of neurons in each layer.

Another approach is to use more sophisticated architectures, such as ResNet, Inception, and DenseNet, which have been designed to address specific challenges in computer vision tasks. These architectures can incorporate techniques such as residual connections, inception modules, and dense connectivity, which can improve the performance of the model by allowing it to learn more effective representations of the data.

Data augmentation techniques can also be used to improve the performance of CNNs. This involves artificially increasing the size of the training set by applying transformations, such as rotations, scaling, and flips, to the images. This can help the model to learn robust features that are invariant to transformations, leading to better generalization performance on new data.

In addition, transfer learning, where a pre-trained CNN is fine-tuned on a specific task, can also lead to improved performance. This is because pre-trained models have already learned generic features from large datasets that can be useful for other tasks. Fine-tuning the model on a specific task can then adapt the model to the task-specific characteristics, leading to better performance.

In conclusion, there are many approaches that can be taken to improve the performance of CNNs in machine learning. By using deeper and wider networks, sophisticated architectures, data augmentation techniques, and transfer learning, researchers and practitioners can continue to push the boundaries of what is possible with CNNs.

RESEARCH CHALLANGES

Availability of high-quality datasets: Many studies have been limited by the availability of high-quality datasets, particularly for less common types of leukemia. Developing large, diverse datasets that accurately represent different populations and subtypes of leukemia is essential for advancing research in this field.

Robustness of models: While many machine learning models have shown high accuracy in detecting leukemia, it is important to ensure that they are robust and able to generalize to new data. Testing models on different datasets and evaluating their performance under different conditions can help improve the reliability and robustness of these models.

Integration of deep learning techniques: While deep learning techniques have shown promise in improving accuracy and efficiency in medical image analysis, there is a need for more research on how to effectively integrate these techniques with traditional machine learning methods. This includes exploring transfer learning approaches and developing new architectures optimized for medical image analysis.

Interpretability and explainability of models: As machine learning models become increasingly complex, it is important to ensure that they are interpretable and explainable to clinicians and patients. Developing methods for explaining model predictions and identifying features that contribute to classification decisions can improve the transparency and trustworthiness of these models.

Real-world deployment and usability: While many machine learning models have shown high accuracy in controlled settings, there is a need for more research on how to effectively deploy these models in real-world clinical settings. This includes developing user-friendly interfaces and ensuring that models are compatible with existing clinical workflows and infrastructure.

3. CONCLUSIONS

Different categories of classification techniques and computing methodologies applied for cancer diagnosis using images are presented in this paper. The categorization of these techniques with their advantages and disadvantages are discussed. This work agrees with other researcher's findings such that texture feature analysis is dependent on the quality of the images of cancerous cells, and statistics modeling could be inaccurate in some specific situations. Each method own some sort of advantages and disadvantages. Diagnostic results of cancer using computer-aided detection methods are depends on the type of images and imaging techniques. The images play foremost roles in determining the results of a cancer diagnosis. In this work, we reviewed various statistical and machine learning methodologies that perform analysis with a texture feature of images and different data pre-processing techniques. We presented a comparative analysis of various techniques based on their performance. This paper conclude based on our review results that a suitable selection of single or combination of machine learning / soft computing algorithms depends upon the data set, capable to yield results with the accuracy of more or equal to 95% on the earlier detection of cancer.

REFERENCES

[1] Machine Learning in Detection and Classification of Leukemia Using Smear Blood Images: A Systematic Review, Pengwei Wang, Published 25 Jun 2021 by Scientific Programming link – https://www.hindawi.com/journals/sp/2021/9933481/

(2) Ishfaque Qamar Khilji1 , Kamonashish Saha2 , Jushan Amin Shonon3 , Muhammad Iqbal Hossain4 Department of Computer Science and Engineering BRAC University Dhaka, Bangladesh Application of Homomorphic Encryption on Neural Network in Prediction of Acute Lymphoid Leukemia International Journal of Advanced Computer Science and Applications, Vol. 11, No. 6, 2020 link-https://www.researchgate.net/profile/Ishfaque-Khilji/publication/342650881_Application_of_Homomorphic_Encryption_on_Neural_Network_in_Prediction_of_Acute_Lymphoid_Leukemia/links/5f11630e299bf1e548bd4ca9/Application-of-Homomorphic-Encryption-on-Neural-Network-in-Prediction-of-Acute-Lymphoid-Leukemia.pdf

(3) Ijraset Journal For Research in Applied Science and Engineering Technology Author Siddhesh Kamthe Publish Date : 2022-04-25 link-https://www.ijraset.com/research-paper/detection-and-classification-of-leukemia-using-deep-learning

(4) Academic Editor: Nouman Ali A Deep Learning Framework for Leukemia Cancer Detection inMicroscopic Blood Samples Using Squeeze andExcitation Learning January 2022 link=https://www.researchgate.net/publication/358253652_A_Deep_Learning_Framework_for_Leukemia_Cancer_Detection_in_Microscopic_Blood_Samples_Using_Squeeze_and_Excitation_Learning

(5) International Research Journal of Engineering and Technology (IRJET) A SURVEY ON BLOOD DISEASE DETECTION USING MACHINE LEARNINGSenthil Kumar V1, Ragul S S2, Ramya Shree3, Agustus A4 Volume: 09 Issue: 05 | May 2022 link - https://www.irjet.net/archives/V9/i5/IRJET-V9I5464.pdf

(6) Laith Alzubaidi, Jinglan Zhang, Amjad J. Humaidi, Ayad Al-Dujaili, Ye Duan, Omran Al-Shamma, J. Santamaría, Mohammed A. Fadhel, Muthana Al-Amidie & Laith Farhan Springerope Review of deep learning: concepts, CNN architectures, challenges, applications, future directions Published: 31 March 2021 link – https://journalofbigdata.springeropen.com/articles/10.1186/s40537-021-00444-8

[7] Raheel Baig 1,2, Abdur Rehman 3 , Abdullah Almuhaimeed 4,* and Abdulkareem Alzahrani 5 and Hafiz Tayyab Rauf 6,*Detecting Malignant Leukemia Cells Using Microscopic Blood Smear Images: A Deep Learning Approach Published: 21 June 2022 link – https://www.mdpi.com/2076-3417/12/13/6317

[8] Accurate Machine-Learning-Based classification of Leukemia from Blood Smear Images Kokeb Dese a, Hakkins Raj a, Gelan Ayana b, Tilahun Yemane c, Wondimagegn Adissu c, Janarthanan Krishnamoorthy PhD (Associate professor) a, Timothy Kwa PhD (Associate Professor) Volume 21, Issue 11, November 2021 link - https://www.sciencedirect.com/science/article/abs/pii/S2152265021002706

[9] LEUKEMIA DISEASE DETECTION AND CLASSIFICATION USING MACHINE LEARNING APPROCHES K.SAIKIRAN A.PASYANTHI PADMA MALINI M.SATYA SAI DURGA PRASAD A.SURYA VENKATA RAMYA Sangivalasa, bheemili mandal, visakhapatnam dist.(A.P)2020-2021 link – https://ece.anits.edu.in/Project%20Reports%202020-21/Sec-B/B-16.pdf

[10] Machine Learning in Detection and Classification of Leukemia Using Smear Blood Images approved by Iran National Committee for Ethics in Biomedical Research with Approval ID IR.SBMU.RETECH.REC.1399.735. Published 25 Jun 2021

[11] Blood Cancer Detection using Machine Learning According to to the All Authors of this Research paper which approved by in link -ICECA 2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA)

[12] Leukemia Detection Using Machine and Deep Learning Through Microscopic Images First Online: 07 July 2022 School of Clinical Medicine, Zhengzhou University, Zhengzhou, Henan, PR China Noor Ayesha

[13] Accurate Machine-Learning-Based classification of Leukemia from Blood Smear Images Hematology and immunohematology course team, School of Medical Laboratory Sciences, Jimma University, Jimma, Ethiopia November 202

[14] School of Computing and Information Sciences, Florida International University, University Park, USA Vaibhav Rupapara Published: 19 January 2022

[15] E. Fathi, M. J. Rezaee, R. Tavakkoli-Moghaddam, A. Alizadeh, and A. Montazer, "Design of an integrated model for diagnosis and classification of pediatric acute leukemia using machine learning," Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine, vol. 234, no. 10, pp. 1051–1069, 2020.

[16] F. E. Al-Tahhan, M. E. Fares, A. A. Sakr, and D. A. Aladle, "Accurate automatic detection of acute lymphatic leukemia using a refined simple classification," Microscopy Research & Technique, vol. 83, no. 10, pp. 1178–1189, 2020.

[17] I. Q. Khilji, K. Saha, J. A. Shonon, and M. I. Hossain, "Application of homomorphic encryption on neural network in prediction of acute lymphoid leukemia," International Journal of Advanced Computer Science and Applications (IJACSA), vol. 11, no. 6, 2020.

[18] M. Sharif, J. Amin, A. Siddiqa et al., "Recognition of different types of leukocytes using YOLOv2 and optimized bag-of-features," Emerging Deep Learning Theories and Methods for Biomedical Engineering, vol. 8, pp. 167448–167459, 2020.

[19] S. Praveena and S. P. Singh, "Sparse-FCM and deep convolutional neural network for the segmentation and classification of acute lymphoblastic leukaemia," Biomedical Engineering/Biomedizinische Technik, vol. 65, no. 6, pp. 759–773, 2020.