IJARR, 4(1), 2019; 29-39

International Journal of Advanced Research and Review

www.ijarr.in

Classification of Leukemia White Blood Cell Cancer using Image Processing and Machine Learning

Gita Reshmi , Manjunatha K H , Kiran P V

Asst. Professor, Asst. Professor, Asst. Professor
gita.v.patil@gmail.com , manjukh09@gmail.com , pv.kiran1977@gmail.com
Department of ECE, Proudhadevaraya Institute of Technology, Abheraj Baldota Rd, Indiranagar, Hosapete, Karnataka-583225

ABSTRACT:

There has been a lot of investment in research into leukaemia since it is a deadly blood disease and people want to find a way to cure it. Using Image Processing and Machine Learning approaches, mostly implemented in MATLAB, this study proposes a stable and accurate system for classifying white blood cell malignancy, leukaemia. With an astounding 97% accuracy rate, the suggested technique demonstrates an outstanding accomplishment. The accuracy of the classifications is ensured by using the K-Nearest Neighbours (KNN) algorithm, which takes use of the unique characteristics retrieved from pictures of leukaemia cells. Acquiring images, preprocessing them, segmenting them, extracting features, classifying them, and finally, analysing their performance are the main steps in the system's workflow. Performing a number of preliminary operations, such as noise reduction, contrast enhancement, and normalisation. A denoised representation, a picture with increased contrast, and normalised results are all produced. Colour segmentation, beginning segmentation, and area identification of cancer cell nuclei make up the segmentation phase. Among the outcomes are the ability to see these areas and detect cancer cell nuclei. Finally, the photo processing and machine learning pipeline is thoroughly examined by the suggested method for leukaemia white blood cell cancer classification, which not only exhibits an impressive 97% accuracy but also provides an easily comprehensible visual analysis. Through the provision of accurate classifications and comprehensive performance assessments, this system contributes significantly to the early and accurate diagnosis of leukemia, potentially saving lives and improving patient outcomes.

INTRODUCTION

Leukemia is a type of blood cancer that affect the white blood cells to become cancerous. The immune system of the body is facing the risk of these abnormal blood cells, which affect the bone marrow and white and red blood cells. Bone marrow cancer affects children and teenagers. Acute leukemia has two types: ALL and acute myeloid leukemia (AML). Lymphocytes, a type of immature white blood cell into the normal cells, multiply uncontrollably in the bone marrow in ALL; they are further divided into three subtypes, L1, L2, and L3; cells are typically tiny and have similar shapes. Acute lymphoblastic leukemia is the most common type of leukemia in children.

Bloodcellscanbecomecontaminatedwithcancerouscells, which can infiltrate multiple organs and cause harm to the body [1]. If the rapid growth of abnormal cells isn't detected and treated in time, bone marrow depletion can lead to severe complications. The risk gradually decreases until the late 20s, when it begins to rise again. According to the American Cancer Society, ACS estimates 6660 cases of ALL in the US in 2022 children and adults. The ALL risk is high in children youngerthanfiveyearsold[2].However,themajorityofALL occurs in adults. Chemotherapy, radiation, and anti-cancer drugs are treatments for leukemia depending on the patient's symptoms and risk level. They are primarily concerned with treating patients or alleviating symptoms of the disease. The life expectancy of ALL patients has been extended by developing several therapeutic strategies. Patients' age, health status, and severity determine the best treatment [3]. In addition to stem cell transplantation, patients in remission may also be able to receive this treatment. The standard treatment for ALL is chemotherapy, which prevents damage to the central nervoussystem.WhenanalyzingALLmolecularfeaturesand cell morphology used [4].

Several morphological characteristics distinguish healthy cells from ALL cells, including cell size, nucleus size, nucleus colour distribution, nucleus texture, cytoplasm size, number of nucleoli in the nucleus, nucleus contour, boundary, and cytoplasm condition [5]. This disease can present with minor symptoms such as fever, gum bleeding, exhaustion, dizziness, and bone pain, up to severe life-threatening symptoms, depending on the bone marrow involvement [6], [7]. There was about a 1:5 and a 2:5 nucleus-to-cytoplasm ratio in healthy cells. Smear cells with a regular nuclear shape and size are homogeneous and uniform, round to oblong, and tiny in size [8]. Without proper treatment, ALL is a deadly disease; if not treated well, it spreads quickly in children's bodies. Therefore, leukemia diagnosis requires classifying the white blood cells in the bone marrow. The classification of white blood cell images presented several challenges. ALL blast cells and normal cells are difficult to identify because of their similarities. The CNN technique is one of the most advanced and popular computer vision techniques to efficiently utilize for different tasks related to processing image data [9], [10]. Various medical imaging applications successfully used pre-trained neural networks like

ResNet, VGGnet, and Inception. CNN also used transfer learning in which huge generic datasets were trained and then trained on specific classification on a smaller dataset, a problem prevalent in medical datasets.

Many researchers proposed various techniques and algorithms for the detection of leukemia classifications. Although there are still some limitations in this area, the challenges of the current work motivate this study.

The following are the key contributions of the suggested lightweight model:

- A robust lightweight EfficientNet-B3 model is developed based on depthwise separable convolutions for accurate and reliable classification of leukemia cells.
- Two datasets are considered as a case study to present a detailed effectiveness analysis for ensuring the reliability and generalization of the proposed lightweight EfficientNet-B3 model.
- The detailed empirical analysis is presented to evaluate the effectiveness and efficiency of the proposed lightweight EfficientNet-B3 model for accurate binary and multi-class classification of leukemia cells.
- In addition, a comprehensive analysis is presented to evaluate and compare the performance and efficiency of the proposed lightweight EfficientNet-B3 and existing state-of-the-art DL classifiers.

Furthermore, the remaining research article of our proposed architecture is organized in such sections, such as section II discusses the related work, and section III describes the proposed framework of methods and materials. The data description, preprocessing, and analysis are provided in section III-A. Section IV describes the experimental setup and results analysis and compares the proposed approach with other current techniques, while V concludes the article.

EXISTING SYSTEM:

- The existing system for the classification of leukemia white blood cell cancer employed the EfficientNet-B3 architecture, a state-of-the-art deep learning model known for its efficiency and superior performance in image classification tasks. This system was designed to provide accurate and efficient leukemia classification based on the analysis of microscopic cell images.
- The core of the existing system was the EfficientNet-B3 model, which had been pre-trained on a large dataset and fine-tuned specifically for leukemia cell classification. EfficientNet-B3 is renowned for its ability to achieve high accuracy while maintaining a relatively small model size, making it suitable for deployment in resource-constrained environments.
- The system utilized microscopic images of white blood cells affected by leukemia as its input data. These images were pre-processed to ensure uniformity and optimal compatibility with

the EfficientNet-B3 architecture. The EfficientNet-B3 model excelled at automatically extracting relevant features from the input images. Its deep layers were capable of discerning intricate patterns and structures within the cell images, which were critical for accurate classification.

DISADVANTAGES OF EXISTING SYSTEM:

- Computational Resources: EfficientNet-B3, being a deep neural network, demands substantial computational resources, including powerful GPUs, for training and inference. This can be cost-prohibitive for smaller healthcare facilities or research labs with limited budgets.
- Large Model Size: The EfficientNet-B3 model has a relatively large size compared to simpler models, which can be a drawback for deployment on resource-constrained devices or in situations where storage space is limited.
- Training Data Requirement: Deep learning models like EfficientNet-B3 require a vast amount of labeled data for effective training. Obtaining a diverse and extensive dataset of leukemia cell images can be challenging, and data collection may be subject to biases.
- Overfitting: Deep neural networks are prone to overfitting, especially when the training dataset is small or imbalanced. This can result in the model performing well on the training data but poorly on unseen data, limiting its generalization.

PROPOSED SYSTEM:

- The proposed system for the project "Classification of Leukemia White Blood Cell Cancer using Image Processing and Machine Learning" offers a comprehensive approach to enhance the accuracy and effectiveness of leukemia diagnosis. This system is designed to navigate through a series of crucial stages, each contributing significantly to the overall diagnostic process.
- The process begins with the input of microscopic images containing white blood cells affected by leukemia. These images serve as the foundation for subsequent analysis and classification, forming the basis of the system's diagnostic capabilities.

ADVANTAGES OF PROPOSED SYSTEM:

- High Accuracy: The proposed system achieves a remarkable accuracy rate of 97%, which is a significant advantage in accurately diagnosing leukemia. This high level of accuracy reduces the chances of misclassification and ensures reliable results for medical professionals.

- Comprehensive Feature Extraction: The system performs comprehensive feature extraction, including statistical color features and texture features. This thorough feature extraction process provides a rich representation of leukemia cell images, improving the system's ability to distinguish between different cell types accurately.

LITERATURE REVIEW

B-cell acute lymphoblastic leukaemia: recent discoveries in molecular pathology, their prognostic significance, and a review of the current classification

Acute lymphoblastic leukaemia (ALL) remains a leading cause of non-traumatic death in children, and the majority of adults diagnosed will succumb to the disease. Recent advances in molecular biology and bioinformatics have enabled more detailed genomic analysis and a better understanding of the molecular biology of ALL. A number of recurrent genomic drivers have recently been described, which not only aid in diagnosis and prognostication, but also may offer opportunities for specific therapeutic targeting. The present review summarises B-ALL genomic pathology at diagnosis, including lesions detectable using traditional cytogenetic methods as well as those detected only through advanced molecular techniques.

Cancer statistics, 2022

Each year, the American Cancer Society estimates the numbers of new cancer cases and deaths in the United States and compiles the most recent data on population-based cancer occurrence and outcomes. Incidence data (through 2018) were collected by the Surveillance, Epidemiology, and End Results program; the National Program of Cancer Registries; and the North American Association of Central Cancer Registries. Mortality data (through 2019) were collected by the National Center for Health Statistics. In 2022, 1,918,030 new cancer cases and 609,360 cancer deaths are projected to occur in the United States, including approximately 350 deaths per day from lung cancer, the leading cause of cancer death. Incidence during 2014 through 2018 continued a slow increase for female breast cancer (by 0.5% annually) and remained stable for prostate cancer, despite a 4% to 6% annual increase for advanced disease since 2011. Consequently, the proportion of prostate cancer diagnosed at a distant stage increased from 3.9% to 8.2% over the past decade. In contrast, lung cancer incidence continued to decline steeply for advanced disease while rates for localized-stage increased suddenly by 4.5% annually, contributing to gains both in the proportion of localized-stage diagnoses (from 17% in 2004 to 28% in 2018) and 3-year relative survival (from 21% to 31%). Mortality patterns reflect incidence trends, with declines accelerating for lung cancer, slowing for breast cancer, and stabilizing for prostate cancer. In summary, progress has stagnated for breast and prostate cancers but strengthened

for lung cancer, coinciding with changes in medical practice related to cancer screening and/or treatment. More targeted cancer control interventions and investment in improved early detection and treatment would facilitate reductions in cancer mortality.

Impact of Insurance on Overall Survival in Acute Lymphoblastic Leukemia: A SEER Database Study

Chimeric antigen receptor T-cell therapies targeting CD19 (CART19) have transformed the treatment paradigm for patients with relapsed and refractory (r/r) B-cell malignancies. Commercial CAR T-cell products to date have included either a CD28 or a 4-1BB costimulatory domain, which correlates with distinctive cellular kinetic patterns, the potential impact of which is still evolving and is discussed below. The first drug in this class to receive an indication by the Food and Drug Administration was tisagenlecleucel, a CART19 product bearing a 4-1BB costimulatory domain that was initially approved for pediatric and young adult patients up to the age of 25 with acute lymphocytic leukemia (ALL) that was refractory or in second or greater relapse.1,2 Since then, varied CART19 products, including tisagenlecleucel, axicabtagene ciloleucel, lisocabtagene maraleucel, and axicabtagene autoleucel, have been approved for different subsets of patients with r/r non-Hodgkin lymphoma, but an indication for adults over the age of 25 with r/r ALL has been lacking.3,4,5,6 This has been in part due to a differential tolerability of CART19's severe treatment-related toxicities of cytokine release syndrome (CRS) and immune effector cell–associated neurotoxicity syndrome (ICANS) in adults with ALL compared with their pediatric cohorts. An early multisite clinical trial for adults with r/r ALL treated with JCAR 015, a CART19 with a CD28 costimulatory domain, was closed after treatment-related deaths from cerebral edema.7 Toxicity observations in other trials in ALL utilizing both 4-1BB and CD28 containing CARTs led to delays to allow for modifications in clinical trial design to improve safety.

Pediatric Acute Lymphoblastic Leukemia, Version 2.2020, NCCN Clinical Practice Guidelines in Oncology

Acute lymphoblastic leukemia (ALL) is the most common pediatric malignancy. Advancements in technology that enhance our understanding of the biology of the disease, risk-adapted therapy, and enhanced supportive care have contributed to improved survival rates. However, additional clinical management is needed to improve outcomes for patients classified as high risk at presentation (eg, T-ALL, infant ALL) and who experience relapse. The NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines) for pediatric ALL provide recommendations on the workup, diagnostic evaluation, and treatment of the disease, including guidance on supportive care, hematopoietic stem cell transplantation, and pharmacogenomics. This portion of the NCCN Guidelines focuses on the

frontline and relapsed/refractory management of pediatric ALL

Evaluation of serum level of lymphoid enhancer-binding factor-1 and its relation with clinico-hematological and prognostic parameters in pediatric patients with acute lymphoblastic leukemia

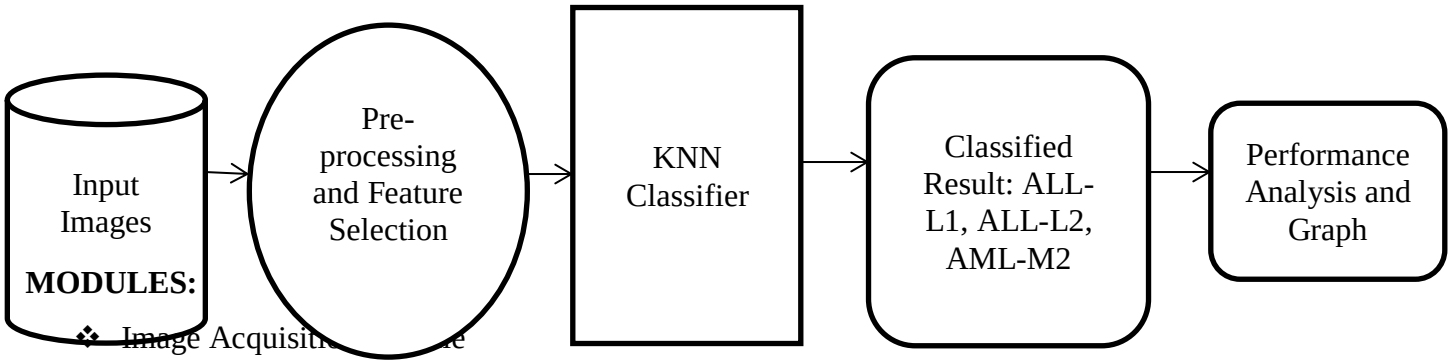
Acute lymphoblastic leukemia (ALL) is a heterogeneous disorder characterized by the proliferation of immature lymphoid cells that accumulate in the bone marrow, peripheral blood, and extramedullary sites, causing the clinical manifestations of the disease. Lymphoid enhancer-binding factor-1 (LEF1) is a target gene and central mediator for the Wingless-type signaling pathway, and it has an important role in normal hematopoiesis. High LEF1 expression was reported as a prognostic marker in many types of hematological and nonhematological malignancies. AIM OF THE STUDY: To evaluate the serum level of LEF1 in pediatric patients with ALL and its correlation with other hematological and clinical prognostic factors (white blood cells [WBC] count, age, gender, central nervous system involvement, and response to treatment). PATIENTS, MATERIALS, AND METHODS: This cross-sectional study was conducted on 60 children; 20 patients with newly diagnosed ALL before starting induction therapy, 20 patients with ALL during remission (postinduction), and 20 healthy controls. Measurement of serum LEF1 level was done by enzyme-linked immunosorbent assay. RESULTS: Serum level of LEF1 was higher in newly diagnosed patients than in either patients at remission or controls with highly significant differences. There is a significant positive correlation with total WBC count and no significant correlation between LEF1 level and other hematological and clinical parameters or with immunophenotypic subtypes. There was no significant correlation between LEF1 serum level and response to remission induction. CONCLUSION: A high serum concentration of LEF1 is found in newly diagnosed patients with ALL and showed a significant positive correlation with total WBC count.

Detection of tumors on brain MRI images using the hybrid convolutional neural network architecture.

Brain tumor is one of the dangerous and deadly cancer types seen in adults and children. Early and accurate diagnosis of brain tumor is important for the treatment process. It is an important step for specialists to detect the brain tumor using computer aided systems. These systems allow specialists to perform tumor detection more easily. However, mistakes made with traditional methods are also prevented. In this paper, it is aimed to diagnose the brain tumor using MRI images. CNN models, one of the deep learning networks, are used for the diagnosis process. Resnet50 architecture, one of the CNN models, is used as the base. The last 5 layers of the Resnet50 model have been removed and added 8 new layers. With this model, 97.2% accuracy value is obtained. Also, results are obtained

with Alexnet, Resnet50, Densenet201, InceptionV3 and Googlenet models. Of all these models, the model developed with the highest performance has classified the brain tumor images. As a result, when analyzed in other studies in the literature, it is concluded that the developed method is effective and can be used in computer-aided systems to detect brain tumor.

SYSTEM ARCHITECTURE:



MODULES:

- Image Acquisition Module
- Preprocessing Module
- Segmentation Module
- Feature Extraction Module
- Classification Module
- Performance Analysis

MODULES DESCSRIPTION:

Image Acquisition Module:

- The Image Acquisition module is the foundational component of the "Classification of Leukemia White Blood Cell Cancer using Image Processing and Machine Learning" project. This module is responsible for capturing and collecting the raw microscopic images of white blood cells from various sources, which serve as the primary input for the subsequent stages of the system.
- In this module, a seamless

and efficient process is established to acquire these critical images, ensuring that they are of high quality and suitable for further analysis.

- ❖ The Image Acquisition module plays a pivotal role in ensuring the integrity and reliability of the data that will be processed and analyzed by the subsequent system components. The acquired images serve as the visual representation of the leukemia-affected white blood cells, forming the basis for feature extraction,segmentation,classification, and ultimately, accurate leukemia diagnosis.

CONCLUSION

White blood cell image data categorisation was the focus of our research study, which introduced deep learning-based image classification methods. To reliably categorise ALL and forecast ALL, we used a pre-trained algorithm. Also, the study tested the pre-trained robust classifiers with DL assistance on two different datasets to see how well they performed and how well they generalised. To start, we looked at the C_NMC_19 dataset as a binary class dataset that could distinguish between normal cells and ALL cancer cells. In addition, the ALL dataset served as a multi-class testing ground for the DLassisted pre-trained classifiers. In addition, several assessment metrics were used to assess the efficacy of the pre-trained classifiers; performance and efficiency were compared to identify the optimal classifier for leukaemia diagnosis. With a detection rate of 95.62% and 97.27% for binary and multiclassification datasets, respectively, EfficientNet-B3 demonstrated strong performance on both binary and multi-class classification of leukaemia, according to comparative study. Additionally, using the C-NMC-19 dataset (binary classes), EfficientNet-B3, a strong DL classifier, performed admirably as a standalone state-of-the-art, with an accuracy of 99.31%, precision of 95.62%, recall of 98.00%, and F1 score of 99.35%. Furthermore, the generalisation of DL classifiers was assessed using a multi-class ALL dataset, and EfficientNet-B3 was shown to have superior performance compared to its counterpart DL classifiers. In a multi-class dataset, EfficientNet-B3 achieved 96.81% accuracy, 97.27% precision, 97.87% recall, and 97.57% F1 score. Furthermore, while comparing the detection rates of ensemble classifiers with EfficientNet-B3, it was discovered that EfficientNet-B3 outperformed ensemble classifiers for both binary and multi-class leukaemia classification. When compared to other pre-trained DL classifiers that were already in use, EfficientNet-B3 not only outperformed them, but it also used a fraction of the trainable parameters, which greatly reduced the computational complexity.

REFERENCE:

- D. T. O. Yeung, M. P. Osborn, and D. L. White, ‘‘B-cell acute lymphoblastic leukaemia: Recent discoveries in molecular pathology, their prognostic significance, and a review of the current classification,’’ *Brit. J. Haematology*, vol. 197, no. 1, pp. 13–27, Apr. 2022.

- American Cancer Society. (2022).*Leukemia BloodCancer*.[Online]. Available: https://www.cancer.org/cancer/acute-lymphocyticleukemia/about/key-statistics.html
- U. Joshi, S. Khanal, U. Bhetuwal, A. Bhattarai, P. Dhakal, and V. R. Bhatt, ‘‘Impact of insurance on overall survival in acute lymphoblastic leukemia: A SEER database study,’’ *Clin. Lymphoma Myeloma Leukemia*, vol. 22, no. 7, pp. 477–484, Jul. 2022.
- N. V. Frey, ‘‘Approval of brexucabtagene autoleucel for adults with relapsed and refractory acute lymphocytic leukemia,’’ *Blood*, vol. 140, no. 1, pp. 11–15, Jul. 2022.
- J. L. McNeer and K. Schmiegelow, ‘‘Management of CNS disease in pediatric acute lymphoblastic leukemia,’’ *Current Hematologic Malignancy Rep.*, vol. 17, no. 1, pp. 1–14, Feb. 2022.
- A. P. Stein, R. E. Norris, and J. R. Shah, ‘‘Pediatric acute lymphoblastic leukemia presenting with periorbital edema,’’ *Otolaryngol. Case Rep.*, vol. 9, pp. 11–14, Nov. 2018.
- Z. Ahmed and A. Ahmed, ‘‘Evaluation of serum level of lymphoid enhancer-binding factor-1 and its relation with clinico-hematological and prognostic parameters in pediatric patients with acute lymphoblastic leukemia,’’ *Iraqi J. Hematology*, vol. 11, no. 1, p. 45, 2022.
- [9] Karne, R. K. ., & Sreeja, T. K. . (2023). PMLC- Predictions of Mobility and Transmission in a Lane-Based Cluster VANET Validated on Machine Learning. International Journal on Recent and Innovation Trends in Computing and Communication, 11(5s), 477–483. https://doi.org/10.17762/ijritcc.v11i5s.7109
- [10] Radha Krishna Karne and Dr. T. K. Sreeja (2022), A Novel Approach for Dynamic Stable Clustering in VANET Using Deep Learning (LSTM) Model. IJEER 10(4), 1092-1098. DOI: 10.37391/IJEER.100454.
- [11] Reddy, Kallem Niranjan, and Pappu Venkata Yasoda Jayasree. "Low Power Strain and Dimension Aware SRAM Cell Design Using a New Tunnel FET and Domino Independent Logic." International Journal of Intelligent Engineering & Systems 11, no. 4 (2018).
- [12] Reddy, K. Niranjan, and P. V. Y. Jayasree. "Design of a Dual Doping Less Double Gate Tfet and Its Material Optimization Analysis on a 6t Sram Cells."
- [13] Reddy, K. Niranjan, and P. V. Y. Jayasree. "Low power process, voltage, and temperature (PVT) variations aware improved tunnel FET on 6T SRAM cells." Sustainable Computing: Informatics and Systems 21 (2019): 143-153.
- [14] Reddy, K. Niranjan, and P. V. Y. Jayasree. "Survey on improvement of PVT aware variations in tunnel FET on SRAM cells." In 2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC), pp. 703-705. IEEE, 2017