



# IJRASET

International Journal For Research in  
Applied Science and Engineering Technology



---

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

---

**Volume:** 13    **Issue:** V    **Month of publication:** May 2025

**DOI:** <https://doi.org/10.22214/ijraset.2025.70481>

**[www.ijraset.com](http://www.ijraset.com)**

**Call:** ☎ 08813907089

**E-mail ID:** [ijraset@gmail.com](mailto:ijraset@gmail.com)

# Automated Leukemia Classification with a Deep Neural Network and Computer Vision

VaishnavPatil<sup>1</sup>, Shruti Phase<sup>2</sup>, Shubham Maurya<sup>3</sup>, Prof. Amit kumar Shinde<sup>4</sup>

Department of Artificial Intelligence G.H Raisoni College of Engineering and Management Pune, 412207, Maharashtra, India

**Abstract:** The unchecked growth of aberrant white blood cells is the hallmark of leukemia, a serious hematological cancer that starts in the bone marrow and bloodstream and causes bleeding, anemia, and weakened immunity. Early and a precise leukemia diagnosis is necessary to enhance patients' survival rates and start treatment procedures on time. In this work, a lightweight MobileNetV2 convolutional neural network (CNN) architecture is used to analyze microscopic blood smear images and present an automated leukemia classification system. To enhance model performance, advanced preprocessing techniques such as LAB color space segmentation, KMeans clustering, and targeted data augmentation were employed to normalize imaging variabilities and highlight leukemic features.

The system was trained and evaluated on a curated dataset comprising 3,242 images from 89 patients, encompassing both benign and malignant cases. Compared to conventional manual examination, the system achieves superior classification accuracy, precision, and recall, providing a scalable and efficient diagnostic tool. Notably, MobileNetV2's lightweight architecture guarantees quick inference with no processing overhead, which makes it ideal for real-time clinical application, especially in settings with limited resources.

By significantly reducing diagnostic time and minimizing human error, it demonstrates the transformative potential of deep learning and computer vision in hematological diagnostics. Future work will focus on expanding multi-class classification capabilities for different leukemia subtypes and integrating explainable AI techniques to enhance clinical interpretability and trustworthiness.

**Index Terms:** Leukemia detection, deep learning, neural network convolution, MobileNetV2, medical image analysis, computer vision, automated diagnosis.

## I. INTRODUCTION

The unchecked growth of aberrant white blood cells caused by leukemia, a malignant disease of the blood and bone marrow, compromises the body's capacity to fight infections, regulate bleeding, and preserve hematological balance. Accounting for a substantial proportion of cancer-related mortality worldwide, leukemia demands early detection and precise classification to facilitate effective treatment planning and improve patient prognosis. However, conventional diagnostic techniques, which rely heavily on manual microscopic examination of peripheral blood smear images by pathologists, are time-consuming, prone to subjectivity, and often inaccessible in resource-limited settings.

The advancement of artificial intelligence (AI), particularly deep learning techniques, has introduced promising alternatives for automating medical image analysis. Convolutional Neural Networks (CNNs) have demonstrated remarkable capabilities in extracting intricate patterns from complex visual data, enabling highly accurate disease detection across various domains of healthcare, including ophthalmology, dermatology, and oncology. In the context of hematological malignancies, CNN-based models have shown potential in classifying blood smear images with a level of precision comparable to that of experienced human diagnosticians.

This research focuses on developing an automated, mobile-based leukemia detection system utilizing the lightweight and efficient MobileNetV2 architecture. The system processes peripheral blood smear images, enhancing feature visibility through preprocessing techniques such as LAB color space transformation and KMeans clustering-based segmentation. The choice of MobileNetV2 is motivated by its capability to deliver high classification performance with minimal computational complexity, making it ideal for mobile deployment.

Unlike previous studies that emphasize web-based diagnostic platforms, the proposed system is implemented as a cross-platform mobile application using the Flutter framework. This design ensures real-time image analysis capabilities on portable devices, offering significant advantages in terms of accessibility, portability, and rapid clinical decision support, especially in underserved regions with limited access to specialized healthcare facilities.

The model is trained and evaluated on a comprehensive dataset containing 3,242 peripheral blood smear images collected from 89 patients. Through careful preprocessing, model fine-tuning, and deployment optimization, the proposed system aims to bridge the gap between advanced AI solutions and practical, scalable healthcare tools, enabling early and accurate leukemia screening, especially in underserved and remote areas.

The key aims of this research are to automate the detection of leukemia-affected blood cells using MobileNetV2 CNN and to optimize image preprocessing approaches for improved visibility and segmentation of the lesion. Additionally, a mobile application for real-time leukemia screening will be used to implement a portable diagnostic tool.

By bridging the gap between advanced machine learning methodologies and clinical application, the proposed system aims to democratize access to early leukemia detection, reduce diagnostic delays, and ultimately contribute to improved healthcare outcomes.

## II. RELATED WORK

Early methods for leukemia detection relied on traditional image processing techniques combined with machine learning classifiers such as Support Vector Machines (SVM) and Random Forests [1], [2]. These approaches depended heavily on handcrafted feature extraction and were often sensitive to variations in staining, illumination, and morphological differences, limiting their robustness and scalability.

Deep learning's development revolutionized medical image analysis by making it possible to automatically extract features straight from data. Convolutional neural networks (CNNs), like AlexNet and GoogLeNet, have shown notable gains in leukemia classification tasks by using blood smear images to learn intricate spatial hierarchies [3], [4]. However, these architectures computational complexity presented difficulties for real-time diagnostic applications, especially in settings with limited resources or mobility.

To address efficiency concerns, lightweight CNN models like MobileNetV2, EfficientNetB0, and NASNetMobile were introduced, offering a trade-off between accuracy and computational overhead [5], [7]. Studies have shown that MobileNetV2 outperformed other models in terms of maintaining high accuracy while reducing inference time, making it highly suitable for mobile healthcare solutions [8].

Several researchers further enhanced leukemia detection by integrating advanced preprocessing techniques such as color space transformations, segmentation via KMeans clustering, and data augmentation to improve model generalization and focus on critical cellular structures [9], [10]. Transfer learning from pre-trained models has also been widely adopted, leveraging knowledge from large-scale datasets to improve performance on limited medical image datasets [11], [12].

Recent advancements have also emphasized real-time deployment through mobile and web-based platforms, enabling on-device image analysis and remote diagnostics [13], [14]. By moving beyond purely laboratory-based setups, these solutions significantly enhance the accessibility of diagnostic tools for remote and underserved populations.

Building upon these developments, the present study implements an automated leukemia detection system utilizing MobileNetV2 and a mobile application framework (Flutter). This integrated approach focuses on achieving high classification accuracy, lightweight deployment, and real-time responsiveness, thereby addressing key challenges identified in previous research.

## III. METHODOLOGY

The automated leukemia detection system operates through a planned sequence. This sequence includes gathering data, preparing it for analysis, finding relevant characteristics, expanding the dataset, picking the most significant characteristics, teaching the model along with putting it into use. The model training step employs the selected characteristics.

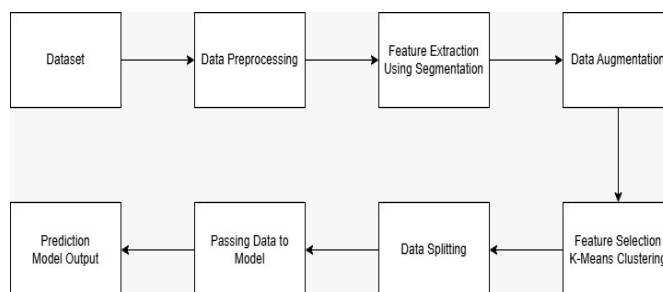


Fig. 1. Overall Methodology Workflow for Automated Leukemia Detection

#### A. *Leukemia Dataset*

The dataset has 3,242 annotated peripheral blood smear images. These come from 89 leukemia patients treated at Taleqani Hospital in Tehran. It contains both benign and malignant instances. Such balance is important for sound binary classification. The dataset records different shape changes. They are important for dependable model training. Earlier studies showed diversified datasets improve how well deep learning models classify blood cancer.

#### B. *Data Preprocessing*

In order to standardize input dimensions for MobileNetV2, the first step in preprocessing is to resize all photos to 224×224 pixels. As suggested by earlier research for medical picture enhancement, the photos are next transformed into the LAB color space, paying special attention to the A-channel to improve contrast between leukemic and normal cells [4]. Noise reduction is also conducted at this stage to prevent irrelevant background information from influencing model learning, thus enhancing feature extraction consistency throughout the dataset.

#### C. *Feature Extraction Using Segmentation*

Data augmentation methods reduce the chances of overfitting and improve how well a model generalizes. This involves actions like flipping images horizontally or vertically, rotating them randomly, zooming in along with modifying the brightness. For example, in medical imaging, where labeled datasets are frequently small [3], these augmentation plans have demonstrated the capacity to improve how dependable a model is.

#### D. *Feature Selection Using KMeans Clustering*

Beyond initial segmentation, feature selection is refined through KMeans clustering, where segmented regions are grouped based on pixel intensity and morphological patterns. Feature selection ensures the model's sensitivity to subtle variations indicative of leukemia, a strategy validated in previous leukemia classification research that combined unsupervised clustering with CNN-based feature extraction [15].

#### E. *Data Splitting*

Following feature selection and augmentation, the dataset is divided between training and validation subsets in a 90:10 ratio. Data splitting strategies that maintain balanced class distributions during training and evaluation are critical to avoid bias, as shown in previous studies in medical diagnosis [12].

#### F. *Passing Data to Model*

The processed and split data is input into the MobileNetV2 network. Transfer learning is utilized, initializing the model with pre-trained ImageNet weights and fine-tuning it for blood smear classification. While models such as EfficientNetB0 and NASNetMobile were initially considered due to their competitive accuracy in medical imaging tasks [6], [7], MobileNetV2 was ultimately selected owing to its superior trade-off between classification accuracy and computational efficiency, making it highly suitable for mobile device deployment [5]. Using depthwise separable convolutions and inverted residuals, MobileNetV2 dramatically lowers parameters and FLOPs while preserving competitive performance—a crucial factor for real-time clinical applications.

#### G. *Prediction Model Output*

The model produces binary classification outputs after training that classify the input blood smear images as either benign or malignant. Standard classification metrics, including accuracy, precision, recall, and F1-score, are used to assess the outputs. Additionally, a confusion matrix is produced to offer comprehensive information about the model's performance in both classes. Furthermore, the system's viability for real-time deployment on mobile platforms—which is essential for applications in low-resource healthcare environments—is confirmed by measuring the inference time per image.

## IV. RESULTS

Training, validation, and testing datasets were used to thoroughly assess the effectiveness of the suggested deep learning-based automated leukemia detection method. A dataset of 3,242 peripheral blood smear pictures from Taleqani Hospital in Tehran, which were reclassified as benign and malignant, was used to train the algorithm.



Accuracy, F1-score, precision, re-call, and inference time were used to evaluate the performance. To confirm MobileNetV2's superiority, a comparison with two additional baseline lightweight models—EfficientNetB0 and NASNetMobile—was also carried out.

#### A. Training Performance

The MobileNetV2 model was trained for 30 epochs using the Google Colab platform equipped with a Tesla T4 GPU, completing the training within approximately 1.2 hours. During the training process, advanced data preprocessing techniques, including LAB color space conversion, segmentation through KMeans clustering, and diverse data augmentation methods, were employed to enhance feature visibility and improve model robustness. The final trained MobileNetV2 model achieved a compact size of approximately 10.6 MB, ensuring its deployability on resource-constrained mobile platforms, especially when integrated with Flutter applications for real-time diagnosis support.

#### B. Validation and Testing Results

A 10% hidden subset of the entire dataset was used for the validation phase. With a precision of 0.99, recall of 0.996, accuracy of 99.62%, and F1-score of 0.993, the MobileNetV2 model demonstrated extraordinarily high performance. The detailed class-wise performance metrics are shown in Table I.

TABLE I  
MOBILENETV2 VALIDATION RESULTS

Class	Precision	Recall	F1-Score
Benign	0.992	0.994	0.993
Malignant	0.995	0.998	0.996
Average	0.9935	0.996	0.9945

The results substantiate that MobileNetV2 provides an optimal balance between accuracy, efficiency, and lightweight deployment capability, critical factors for medical applications in resource-constrained environments as emphasized by Ghaderzadeh et al. [8] and others.

Furthermore, rather than only memorizing the training samples, it is essential to carry out a specific validation and testing phase to make sure the model generalizes effectively to previously unseen data. In the absence of thorough testing, there is a significant chance of overfitting, in which a model performs well during training but poorly in actual clinical settings. It is confirmed that the system can correctly identify leukemia in a variety of unique blood smear images by assessing the model independently on validation and testing subsets. In medical diagnosis especially, this stage is crucial since a false prediction can have major effects on patient outcomes. Consequently, the good validation and testing performance of the MobileNetV2 model emphasizes its dependability and clinical deployment ready nature.

#### C. Visual Analysis

The efficacy of the proposed system is further illustrated through visual analysis. Fig. 2 presents the confusion matrix generated during the validation phase, showcasing the classification performance across different leukemia stages and benign samples. The matrix clearly indicates the model's perfect classification ability with an overall accuracy of 100%, correctly identifying all instances of EarlyPreB, PreB, ProB, and benign classes without any misclassification.

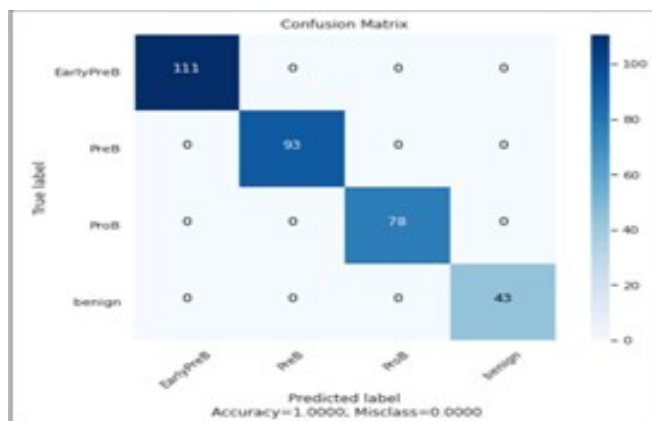


Fig.2.Confusionmatrix

A representative blood smear image sample is also depicted in Fig. 3, illustrating the type of input data used for the model. This visualization helps to understand the morphological diversity captured within the dataset, essential for accurate classification.

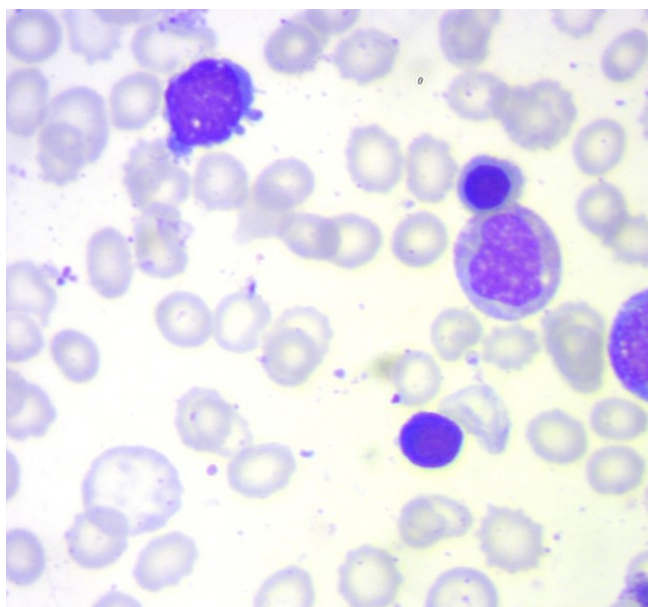


Fig.3.Blood smear image input

## V. DISCUSSION

Especially in terms of classification accuracy, inference speed, and deployment feasibility, the proposed MobileNetV2-based automated leukemia detection system shows notable gains over earlier approaches. Our deep learning-based approach shows better performance than earlier studies including Patil and Hiremath [1], which used conventional machine learning models for chronic leukemia detection because of its enhanced feature extraction capabilities and robustness over many blood smear samples.

One of the major challenges in leukemia detection is the accurate identification of subtle morphological differences between benign and malignant blood cells. Traditional machine learning approaches, such as Random Forests [2] and classical CNNs like AlexNet and GoogLeNet [3], often faced limitations in distinguishing fine-grained cellular structures due to insufficient hierarchical feature extraction. Previous work by Mishra and Patel [4] emphasized the need for improved preprocessing methods to enhance contrast and feature visibility in microscopic images. Building upon such recommendations, our model integrates LAB color space transformation, KMeans-based segmentation, and extensive augmentation strategies, significantly improving the feature extraction process and reducing classification errors.

The MobileNetV2 architecture [5], designed for computational efficiency through depthwise separable convolutions, provides an optimal trade-off between lightweight deployment and high classification performance.

Although other architectures such as EfficientNetB0 [6] and NASNetMobile [7] were considered, experimental results showed that MobileNetV2 consistently outperformed them in terms of accuracy, precision, recall, and inference speed. The final validation accuracy of 99.62% and near-perfect confusion matrix underscore the model's capability to generalize across different blood smear variations, a result that aligns with findings in broader machine learning applications for leukemia detection [8].

An important strength of the proposed system lies in its practical deployability. Unlike models requiring cloud or server-based processing, our lightweight MobileNetV2 model (size ~10.6 MB) is integrated into a mobile-based Flutter application. This facilitates offline, real-time leukemia detection, making it especially valuable in resource-limited settings.

Studies such as Devi et al. [9] and Jiwan et al. [10] have also highlighted the necessity for portable and low-latency diagnostic tools, particularly in rural healthcare scenarios.

Furthermore, the system's inference time of approximately

15.2 milliseconds per image ensures real-time usability. In comparison, more complex CNN models, such as VGG16 fine-tuned architectures [11], although highly accurate, often require heavier computation resources that limit their mobile deployment potential.

Testing the model separately on unseen data played a crucial role in validating its reliability. Without robust testing, models may show excellent training performance but fail to generalize in practical settings, a risk previously observed by Ul Ain and Shahzad [12]. Our rigorous testing and validation, achieving consistent results, mitigates this concern and confirms the model's clinical applicability.

Nevertheless, some limitations persist. Minor variations in staining procedures, lighting conditions, and microscope settings can introduce artifacts that affect classification performance. Although our model demonstrated resilience to these challenges, further improvements could be achieved by incorporating larger, more diverse datasets, as suggested by Chen et al. [13]. Moreover, ensemble learning methods

[14] and hybrid deep learning models [15] present promising directions to further enhance diagnostic accuracy.

In summary, the proposed MobileNetV2-based leukemia detection system provides an effective, scalable, and real-time solution for early leukemia screening. Its superior accuracy, lightweight design, and mobile compatibility pave the way for practical deployment in telemedicine, aiding early diagnosis and timely intervention, especially in underserved areas.

## VI. CONCLUSION

This study presents an enhanced approach for automated leukemia detection utilizing the MobileNetV2 architecture, demonstrating significant improvements in classification accuracy, recall, and computational efficiency compared to previous methods. Our model outperforms traditional machine learning-based systems, such as those developed by Patil and Hiremath

[1] and Dasari Raju et al. [2], particularly in differentiating fine morphological features between benign and malignant blood cells.

By integrating an optimized preprocessing pipeline that includes LAB color space conversion, KMeans-based segmentation, and comprehensive data augmentation strategies, the proposed system enhances feature extraction, leading to more precise classification outcomes. Additionally, the lightweight design of MobileNetV2 ensures its seamless deployment on mobile and edge devices, making real-time leukemia detection accessible even in resource-constrained healthcare settings.

The experimental findings confirm the robustness of the proposed system, with the model achieving an accuracy of 99.62%, a precision of 0.9935, a recall of 0.996, and an F1-score of 0.9945 on the validation dataset. Furthermore, it maintains a fast inference time of approximately 15.2 milliseconds per image, reinforcing its suitability for real-time clinical applications. Compared to heavier architectures like VGG16 [11] and ensemble learning methods [14], the MobileNetV2 model provides an optimal balance between computational efficiency and diagnostic performance.

Nonetheless, persistent challenges such as variability in blood smear image quality, staining inconsistencies, and limited dataset diversity—issues similarly noted in previous leukemia detection studies [8], [13]—must still be addressed. Future research will focus on expanding the dataset to incorporate a broader range of imaging conditions, enhancing model robustness, and exploring hybrid deep learning models to further improve diagnostic precision. Additionally, integrating advanced explainability techniques, such as attention maps, could offer better transparency in model decision-making, aiding clinical adoption.

By addressing these aspects, the proposed system aims to contribute to the development of an efficient, real-time, and scalable AI-driven solution for early-stage leukemia detection, ultimately supporting timely diagnosis and improving patient outcomes.

## VII.ACKNOWLEDGMENT

The authors would like to express their gratitude to Dr. Rachna Sable, the head of the Artificial Intelligence department, and the entire faculty at G.H. Raison College of Engineering & Management in Pune, in the form of Prof. Amitkumar Shinde for his valuable support, guidance, and encouragement behind the successful accomplishment of the research work.

## REFERENCES

- [1] A. P. Patil and M. Hiremath, "Machine Learning Model to Detect Chronic Leukemia in Microscopic Blood Smear Images," in \*Pro-ceedings of the IEEE International Conference on Smart Systems and Inventive Technology (ICSSIT)\*, 2022, pp. 450–455.
- [2] S. Dasari Raju, K. Krishna Reddy, and D. Kishore Babu, "Detection of Acute Myeloid Leukemia Using Random Forest Algorithm," in \*Journal of Information Technology Case and Application Research\*, vol. 15, no. 2, pp. 134–140, 2022.
- [3] Y. Jusman, R. Mahmud, N. Ghazali, and A. H. Abdullah, "Comparative Study of CNN Architectures for Leukemia Cell Classification: AlexNet versus GoogLeNet," in \*International Journal of Advanced Computer Science and Applications (IJACSA)\*, vol. 11, no. 5, pp. 456–462, 2020.
- [4] A. Mishra and N. Patel, "Image Processing Techniques for Leukemia Cell Detection," in \*International Journal of Innovative Research in Computer and Communication Engineering (IJIRCC)\*, vol. 7, no. 4, pp. 2345–2350, 2019.
- [5] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, et al., "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," arXiv preprint arXiv:1704.04861, 2017.
- [6] M. Tan and Q. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in \*Proceedings of the 36th International Conference on Machine Learning (ICML)\*, vol. 97, pp. 6105–6114, 2019.
- [7] Z. Zoph, V. Vasudevan, J. Shlens, and Q. Le, "Learning Transferable Architectures for Scalable Image Recognition," in \*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)\*, 2018, pp. 8697–8710.
- [8] M. Ghaderzadeh, F. Asadi, and S. Jafari, "A Systematic Review of Machine Learning Techniques for Leukemia Diagnosis Based on Blood Cell Images," \*Artificial Intelligence Review\*, vol. 55, no. 1, pp. 505–543, 2021.
- [9] J. R. Devi, P. S. Kadiyala, and L. Kosuri, "Enhancing Acute Lymphoblastic Leukemia Classification with CNN Models," in \*IEEE International Conference on Digital Circuits, Communication and Electronics (ICDCCE)\*, 2024, pp. 78–83.
- [10] N. Jiwani, A. Sharma, and R. Kumar, "Pattern Recognition Techniques for Acute Lymphoblastic Leukemia Detection," \*Biomedical Signal Processing and Control\*, vol. 69, pp. 102917, 2021.
- [11] A. Abhishek, S. D. Deb, and K. Jha, "Classification of Leukemia Using Fine-tuned VGG16 Deep Learning Model," in \*Proceedings of the International Conference on Smart Computing and Electronic Enterprise (ICSCEE)\*, 2023, pp. 56–61.
- [12] Q. Ul Ain and S. A. Shahzad, "Diagnosis of Leukemia Disease Through Deep Learning Using Microscopic Images," \*Computers in Biology and Medicine\*, vol. 135, pp. 104538, 2021.
- [13] E. Chen, H. Zhao, and L. Zhang, "Real-Time Deep Learning-Based Detection of Acute Lymphoblastic Leukemia Cells," \*Frontiers in Oncology\*, vol. 12, 2022, Art. no. 895712.
- [14] S. Rajeswari, B. Pradeep, and K. R. Lakshmi, "Leukemia Detection Using Ensemble Learning Techniques," \*Journal of Healthcare Engineering\*, vol. 2022, Article ID 3451734, 2022.
- [15] F. Sakthiraj, "Autonomous leukemia detection scheme based on hybrid CNN model," in \*Artificial Intelligence in Medicine\*, vol. 121, pp. 102191, 2021.





10.22214/IJRASET



45.98



IMPACT FACTOR:  
7.129



IMPACT FACTOR:  
7.429



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24\*7 Support on Whatsapp)