International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 21 No. 3 (1) 2025

ijerst.editor@gmail.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Vol. 21, No. 3(1), 2025

Research Paper

IMPROVING WHITE BLOOD CELL CLASSIFICATION WITH MINIMAL LABELS: A SEMI-SUPERVISED LEARNING FRAMEWORK

Zainab Uzma[1], Rubeena Afsar[2]
[1]PG Scholar, Department of CSE, Shadan Women's College of Engineering and Technology, Hyderabad,
zainabuzmass95@gmail.com
[2]Asst. Professor, Department of CSE, Shadan Women's College of Engineering and Technology
swcetrubeenaafsar@gmail.com

ABSTRACT

The diagnosis of numerous blood disorders depends on the accurate classification of white blood cell subtypes. Manually designed features are frequently needed for traditional computer vision techniques, which can be time-consuming and performance-limiting. On the other hand, machine learning techniques provide higher accuracy but usually require large labeled datasets, which are expensive and difficult to collect. This work presents a semi-supervised learning strategy designed specifically for the classification of white blood cells. Through the utilization of a limited quantity of labeled data and a more extensive collection of unlabeled data, the model is able to recognize and classify various subtypes of white blood cells straight from microscopic pictures. This approach improves classification performance by leveraging the data's natural structure and patterns rather than depending just on pre-established features.The suggested method was assessed on a dataset of artificial images that depicted different subtypes of white blood cells. The results indicate promising accuracy in differentiating across cell types, indicating potential uses in clinical diagnostics. White blood cell analysis in medical settings can be automated and made more efficient with this scalable method that reduces the need for manually labeled data while retaining excellent classification accuracy.

Received: 16-7-2025 Accepted: 23-8-2025 Published: 30-8-2025

1. INTRODUCTION

White blood cell (WBC) classification accuracy is essential for the diagnosis of leukemia and anemia, among other blood-related conditions. Manually constructed features and domain knowledge are key components of traditional WBC classification techniques, which can be laborious and error-prone. The volume and complexity of medical data are frequently too much for these approaches to handle. By learning directly from data and achieving high levels of accuracy, machine learning techniques—especially those that make use of Xception—offer a potent substitute. But in most cases, these methods need huge labeled datasets, which are costly and time-consuming to produce. This project suggests a semi-supervised learning method to overcome this difficulty by enhancing WBC classification performance by combining a limited quantity of labeled data with a larger pool of unlabeled data.

OBJECTIVE

Creating an architecture for Xception that makes efficient use of a growing collection of unlabeled photographs as well as a small group of identified images. Using information gleaned from the data, train the model to correctly differentiate between various WBC kinds. measuring the model's effectiveness in terms of generalization to new data and classification accuracy. In clinical contexts where labeled data is scarce and unlabeled data is more accessible; this approach's viability is demonstrated.

2.1 PROBLEM STATEMENT

White blood cell (WBC) subtype classification accuracy is essential for the diagnosis of a number of blood diseases. Manually built features, which take a long time to build and can restrict performance, are a major component of traditional computer vision techniques. Although machine learning techniques enhance accuracy, they usually necessitate substantial quantities of labeled data, which are expensive and challenging to acquire in medical settings. As a result, a scalable and effective approach that can reliably identify WBC subtypes with less labeled data while maintaining diagnostic performance is desperately needed.

2.2 Existing System

Convolutional neural networks, or CNNs, are a family of deep learning techniques that are mostly utilized for computer vision and image processing tasks. CNNs, which are modeled after the visual cortex of animals, are made to automatically and adaptively learn feature spatial hierarchies from input images.

They are now the foundation of many contemporary computer vision applications, such as segmentation,

object detection, and image categorization. To provide non-linearity to the model and help it learn intricate patterns, an activation function—typically ReLU—is used. By decreasing the feature maps' width and height, pooling layers preserve the most crucial data while lowering the computational burden and overfitting risk. Max pooling and average pooling are two popular forms of pooling.

Disadvantage of Existing System

- Significant computational resources are needed for CNNs.
- Large volumes of labeled data are usually needed for CNN training in order to attain excellent performance.
- The overfitting.

2.3 Proposed System

"Extreme Inception," or "Xception," is a deep learning model architecture that was first presented by Keras founder François Chollet. "Xception: Deep Learning with Depth wise Separable Convolutions" was the 2017 paper in which it was introduced. The Inception architecture, which is the foundation of Xception, is renowned for its effectiveness and superior performance in picture categorization tasks. Xception uses depth-wise separable convolutions to reduce computational cost and improve performance over conventional convolutional neural networks.

With fewer parameters and lower computing demands, Xception is a noteworthy example of a memory-efficient program. This makes it appropriate for operating on memory-constrained devices like embedded systems and cell phones.

Advantages of Proposed System

- Decreased Complexity of Computation
- Improved Performance.
- Improved Extraction of Features.

2. RELATED WORKS

Because white blood cells (WBCs) are essential for diagnosing hematological disorders, their classification has long been a subject of study. The majority of early methods relied on conventional image processing methods, which involved manually extracting handmade elements from tiny images, such as shape, color, and texture. Following that, classification techniques such decision trees, support vector machines, and k-nearest neighbors (KNN) were applied. Despite offering a starting point for WBC classification, these methods had poor generalizability and a strong reliance on domain knowledge.

In WBC image classification problems, convolutional neural networks (CNNs) have demonstrated notable advancements since the introduction of deep learning. Models like VGGNet, ResNet, and Inception are notable examples of models that have been refined on labeled WBC datasets and have achieved great accuracy. But in the medical field, these approaches usually need for a lot of labeled data, which is expensive and time-consuming.

In order to tackle the problem of scarce labeled data, semi-supervised learning (SSL) approaches have been investigated in recent research. Promising outcomes have been obtained when applying techniques like pseudo-labeling, graph-based learning, and consistency regularization (e.g., Mean Teacher) to medical picture analysis. SSL has demonstrated promise in integrating tiny labeled datasets with a larger unlabeled data pool, which can significantly enhance model generalization and lower annotation costs.

3. METHODOLOGY OF PROJECT

The goal of the suggested methodology is to use semi-supervised learning (SSL) to accurately classify different subtypes of white blood cells (WBCs). The technique is intended to minimize reliance on human annotated data while preserving good classification performance by combining a small selection of labeled WBC images with a larger pool of unlabeled photos.

Data Gathering:

This is how raw data is gathered from different sources. Databases, polls, experiments, and internet sources are some of the possible sources of the data. In order to answer the research question or address the current issue, it is necessary to gather sufficient pertinent data.

Examine the information:

Finding patterns, trends, and insights in the gathered data is what this entails. To comprehend the aspects and structure of the data, analysis can be done by exploratory data analysis approaches, statistical methods, or visualizations.

Data Purification:

The process of fixing or eliminating errors, inconsistencies, and inaccuracies from a dataset is known as data cleaning. To make sure the dataset is accurate and usable, this may entail addressing missing numbers, fixing errors, eliminating duplication, and standardizing data formats.

Splitting Data:

The dataset is separated into several subsets in this step, usually a training set and a testing set. While the testing set is used to assess the model's performance, the training set is used to train the machine learning model. Model parameters are occasionally adjusted using a validation set as well.

Using the Algorithm on the Data:

In order to create a model, machine learning algorithms are applied to the training data. This stage entails choosing the right algorithm (such as neural networks or decision trees) and applying it to the training data in order to identify patterns and relationships.

Model Performance:
Using the testing set, you assess the model's performance after training it. This entails assessing how accurately the model forecasts or categorizes fresh data. F1 score, accuracy, precision, recall, and area under the curve (AUC) are common performance indicators that vary based on the type of problem (classification, regression, etc.).

The outcome:
The results of your model's performance are interpreted and presented in this last stage. The outcomes ought to demonstrate how successfully the model has fulfilled the project's goals, and any new information gleaned from the research is condensed. At this point, recommendations or conclusions based on the findings are frequently given.

4. ALGORITHM USED IN PROJECT

One kind of artificial neural network made especially to analyze structured grid data, like pictures, is called a convolutional neural network (CNN). Convolutional layers are used to identify patterns and spatial hierarchies in the data.Usually, an image in the form of a multi-dimensional array serves as the CNN's input. A collection of filters, sometimes known as kernels, are applied to the input image by these layers. In order to create a feature map, each filter performs element-wise multiplications on the image while adding up the results. Detecting local patterns like edges, textures, and forms is the responsibility of convolutional layers.

5. DATA FLOW DIAGRAM

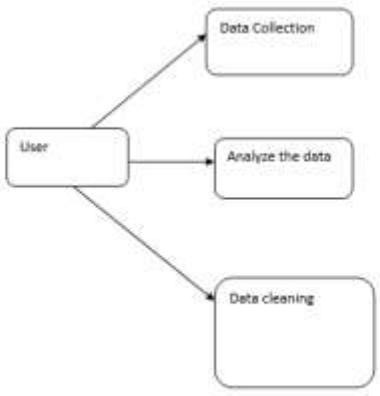
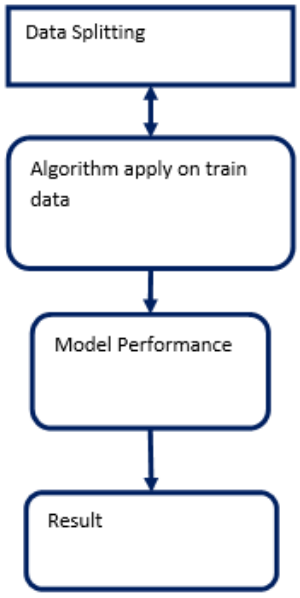


Fig: 6 Flow Diagram



6. SYSTEM ARCHITECTURE

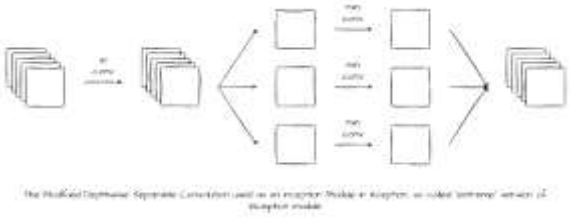
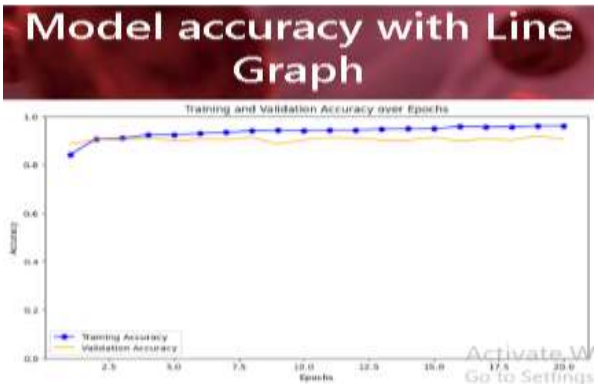
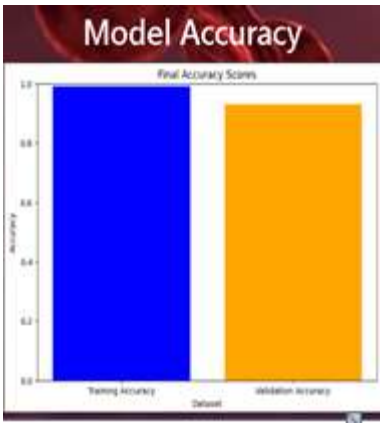


Fig: 7 SYSTEM ARCHITECTURE OF PROJECT

7. RESULTS

The proposed Xception model outperforms the existing CNN by providing higher accuracy, faster training, and lower computational cost. Its use of depthwise separable convolutions reduces memory usage and improves feature extraction, making it more efficient and suitable for mobile/embedded devices. In contrast, CNNs require more resources, large datasets, and face higher risk of overfitting.





8. FUTURE ENHANCEMENT

The integration of transfer learning and multi-modal data could be the focus of future research to improve the white blood cell image classification model. Particularly when there is a lack of labeled data, the model's performance can be improved by using pre-trained models on extensive datasets and optimizing them for white blood cell classification. Future developments are anticipated to significantly improve the model's efficacy and application, including transfer learning and multi-modal data integration.

9. CONCLUSION

This study presents a sophisticated white blood cell categorization technique that uses blood smear images and a novel semi-supervised learning framework in conjunction with convolutional neural networks. The suggested solution combines a mean instructor model with an advanced feature extraction network, as well as an improved pseudo-label generating technique and a more sophisticated sampling plan. Together, these components optimize classification accuracy while reducing the requirement for large amounts of labeled data. The outcomes reveal that this method not only greatly enhances classification performance but also lessens the amount of human annotation required, indicating its potential for real-world uses in medical research and diagnostics.

REFERENCES:

[1]. X. Li, Y. Cao, and Y. Wang, ‘‘A robust classification method for five types of leukocytes in peripheral blood based on mean-shift clustering,’’ J. Biomed. Eng., vol. 35, no. 5, pp. 761–766, 2018.

[2]. I. Radosavovic, P. Dollár, R. Girshick, G. Gkioxari, and K. He, ‘‘Data distillation: Towards omni-supervised learning,’’ in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 4119–4128.

[3]. S. Sanei and T. K. Lee, ‘‘Cell recognition based on PCA and Bayesian classification,’’ in Proc. 4th Int. Symp., Nara, Japan, 2003, pp. 239–243.

[4]. B. C. Ko, J. W. Gim, and J. Y. Nam, ‘‘Cell image classification based on ensemble features and random forest,’’ Electron. Lett., vol. 47, no. 11, p. 638, 2011.

[5]. S. H. Rezatofighi and H. Soltanian-Zadeh, ‘‘Automatic recognition of five types of white blood cells in peripheral blood,’’ Computerized Med. Imag. Graph., vol. 35, no. 4, pp. 333–343, Jun. 2011.

[6]. O. Sarrafzadeh, H. Rabbani, A. Talebi, and H. U. Banaem, ‘‘Selection of the best features for leukocytes classification in blood smear microscopic images,’’ in Proc. SPIE, San Diego, CA, USA, Mar. 2014, pp. 159–166.

[7]. D. Gupta, J. Arora, U. Agrawal, A. Khanna, and V. H. C. de Albuquerque, ‘‘Optimized binary bat algorithm for classification of white blood cells,’’ Measurement, vol. 143, pp. 180–190, Sep. 2019.

[8]. E. Abdullah and M. K. Turan, ‘‘Classifying white blood cells using machine learning algorithms,’’ Int. J. Eng. Res. Dev., vol. 11, pp. 141–152, Jun. 2019.

[9]. M. Alruwaili, ‘‘An intelligent medical imaging approach for various blood structure classifications,’’ Complexity, vol. 2021, pp. 1–10, May 2021.

[10]. A. N. Nithyaa, R. P. Kumar, G. M. Gokul, and G. Aananthi, ‘‘MATLAB based potent algorithm for WBC cancer detection and classification,’’ Biomed. Pharmacol. J., vol. 14, no. 4, pp. 2277–2284, Dec. 2021.

[11]. K. Simonyan and A. Zisserman, ‘‘Very deep convolutional networks for large-scale image recognition,’’ 2014, arXiv:1409.1556.

[12]. K. He, X. Zhang, S. Ren, and J. Sun, ‘‘Deep residual learning for image recognition,’’ in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 770–778.

[13]. M. Jiang, L. Cheng, F. Qin, L. Du, and M. Zhang, ‘‘White blood cells classification with deep convolutional neural networks,’’ Int. J. Pattern Recognit. Artif. Intell., vol. 32, no. 9, Sep. 2018, Art. no. 1857006.

[14]. K. Almezhghwi and S. Serte, ‘‘Improved classification of white blood cells with the generative adversarial network and deep convolutional neural network,’’ Comput. Intell. Neurosci., vol. 2020, pp. 1–12, Jul. 2020.

[15]. A. I. Shahin, Y. Guo, K. M. Amin, and A. A. Sharawi, ‘‘White blood cells identification system based on convolutional deep neural learning networks,’’ Comput. Methods Programs Biomed., vol. 168, pp. 69–80, Jan. 2019.

[16]. Y. Liu, Y. Fu, and P. Chen, ‘‘WBCaps: A capsule architecture-based classification model designed for white blood cells identification,’’ in Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC), Jul. 2019, pp. 7027–7030.

[17]. X. Yao, K. Sun, X. Bu, C. Zhao, and Y. Jin, ‘‘Classification of white blood cells using weighted optimized deformable convolutional neural networks,’’ Artif. Cells, Nanomedicine, Biotechnol., vol. 49, no. 1, pp. 147–155, Jan. 2021.
[18]. B. Sheng, M. Zhou, M. Hu, Q. Li, L. Sun, and Y. Wen, ‘‘A blood cell dataset for lymphoma classification using faster R-CNN,’’ Biotechnol. Biotechnological Equip., vol. 34, no. 1, pp. 413–420, Jan. 2020.
[19]. A. M. Patil, M. D. Patil, and G. K. Birajdar, ‘‘White blood cells image classification using deep learning with canonical correlation analysis,’’ IRBM, vol. 42, no. 5, pp. 378–389, Oct. 2021.
[20]. F. Özyurt, ‘‘A fused CNN model for WBC detection with MRMR feature selection and extreme learning machine,’’ Soft Comput., vol. 24, no. 11, pp. 8163–8172, Jun. 2020.
[21]. M. Toğaçar, B. Ergen, and Z. Cömert, ‘‘Classification of white blood cells using deep features obtained from convolutional neural network models based on the combination of feature selection methods,’’ Appl. Soft Comput., vol. 97, Dec. 2020, Art. no. 106810.
[22]. A. Acevedo, S. Alférez, A. Merino, L. Puigví, and J. Rodellar, ‘‘Recognition of peripheral blood cell images using convolutional neural networks,’’ Comput. Methods Programs Biomed., vol. 180, Oct. 2019, Art. no. 105020.
[23]. M. Sharma, A. Bhave, and R. R. Janghel, ‘‘White blood cell classification using convolutional neural network,’’ in Soft Computing and Signal Processing. Berlin, Germany: Springer, 2019, pp. 135–143.
[24]. Q. Huang, W. Li, B. Zhang, Q. Li, R. Tao, and N. H. Lovell, ‘‘Blood cell classification based on hyperspectral imaging with modulated Gabor and CNN,’’ IEEE J. Biomed. Health Informat., vol. 24, no. 1, pp. 160–170, Jan. 2020.
[25]. R. B. Hegde, K. Prasad, H. Hebbar, and B. M. K. Singh, ‘‘Feature extraction using traditional image processing and convolutional neural network methods to classify white blood cells: A study,’’ Australas. Phys. Eng. Sci. Med., vol. 42, no. 2, pp. 627–638, Jun. 2019.
[26]. R. Ahmad, M. Awais, N. Kausar, and T. Akram, ‘‘White blood cells classification using entropy-controlled deep features optimization,’’ *Diagnostics*, vol. 13, no. 3, p. 352, Jan. 2023.
[27]. L. Boldú, A. Merino, A. Acevedo, A. Molina, and J. Rodellar, ‘‘A deep learning model (ALNet) for the diagnosis of acute leukaemia lineage using peripheral blood cell images,’’ *Comput. Methods Programs Biomed.*, vol. 202, Apr. 2021, Art. no. 105999.
[28]. B. Sen, A. Ganesh, A. Bhan, and S. Dixit, ‘‘Deep learning based diagnosis of sickle cell anemia in human RBC,’’ in *Proc. 2nd Int. Conf. Intell. Eng. Manage. (ICIEM)*, Apr. 2021, pp. 526–529.
[29]. S. S. R. Bairaboina and S. R. Battula, ‘‘Ghost-ResNeXt: An effective deep learning based on mature and immature WBC classification,’’ *Appl. Sci.*, vol. 13, no. 6, p. 4054, Mar. 2023.
[30]. J. L. D. Resendiz, V. Ponomaryov, R. R. Reyes, and S. Sadovnychiy, ‘‘Explainable CAD system for classification of acute lymphoblastic leukemia based on a robust white blood cell segmentation,’’ *Cancers*, vol. 15, no. 13, p. 3376, Jun. 2023.
[31]. C. Li and Y. Liu, ‘‘Improved generalization of white blood cell classification by learnable illumination intensity invariant layer,’’ *IEEE Signal Process. Lett.*, vol. 31, pp. 176–180, 2024.
[32]. T. A. Elhassan, M. S. Mohd Rahim, M. H. Siti Zaiton, T. T. Swee, T. A. Alhaj, A. Ali, and M. Aljurf, ‘‘Classification of atypical white blood cells in acute myeloid leukemia using a two-stage hybrid model based on deep convolutional autoencoder and deep convolutional neural network,’’ *Diagnostics*, vol. 13, no. 2, p. 196, Jan. 2023.
[33]. M. Yildirim and A. Çinar, ‘‘Classification of white blood cells by deep learning methods for diagnosing disease,’’ *Revue d’Intelligence Artificielle*, vol. 33, no. 5, pp. 335–340, Nov. 2019.
[34]. M. M. Alam and M. T. Islam, ‘‘Machine learning approach of automatic identification and counting of blood cells,’’ *Healthcare Technol. Lett.*, vol. 6, no. 4, pp. 103–108, Aug. 2019.
[35]. Q. Wang, S. Bi, M. Sun, Y. Wang, D. Wang, and S. Yang, ‘‘Deep learning approach to peripheral leukocyte recognition,’’ *PLoS ONE*, vol. 14, no. 6, Jun. 2019, Art. no. e0218808.
[36]. G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, ‘‘Densely connected convolutional networks,’’ in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 2261–2269.
[37]. S. Sharma, S. Gupta, and D. Gupta, ‘‘Deep learning models for the automatic classification of white blood cells,’’ *Comput. Intell. Neurosci.*, vol. 2022, Jan. 2022, Art. no. 7384131.
[38]. D. Baby, S. J. Devaraj, and A. Raj, ‘‘Leukocyte classification based on transfer learning of VGG16 features by K-nearest neighbor classifier,’’ in *Proc. 3rd Int. Conf. Signal Process. Commun. (ICPSC)*, Coimbatore, India, May 2021, pp. 252–256.
[39]. K. A. Fathy, H. K. Yaseen, M. T. Abou-Kreisha, and K. A. ElDahshan, ‘‘A novel meta-heuristic optimization algorithm in white blood cells classification,’’ *Comput., Mater. Continua*, vol. 75, no. 1, pp. 1527–1545, 2023.
[40]. T. Miyato, S.-I. Maeda, M. Koyama, and S. Ishii, ‘‘Virtual adversarial training: A regularization method for supervised and semi-supervised learning,’’ *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 8, pp. 1979–1993, Aug. 2019.

[41]. X. Fu, Y. Sheng, and H. Li, ‘‘A semi-supervised encoder generative adversarial networks,’’ *Acta Automatica Sinica*, vol. 46, no. 3, pp. 531–539, 2020.
[42]. A. Abdelwahab, A. Afifi, and M. Salama, ‘‘An integrated active deep learning approach for image classification from unlabeled data with minimal supervision,’’ *Electronics*, vol. 13, no. 1, p. 169, Dec. 2023.
[43]. X. Li, X. Wang, X. Chen, Y. Lu, H. Fu, and Y. C. Wu, ‘‘Unlabeled data selection for active learning in image classification,’’ *Sci. Rep.*, vol. 14, no. 1, p. 424, Jan. 2024.
[44]. A. Tarvainen and H. Valpola, ‘‘Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results,’’ in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–10.
[45]. I. Loshchilov and F. Hutter, ‘‘SGDR: Stochastic gradient descent with warm restarts,’’ 2016, arXiv:1608.03983.
[46]. I. Sutskever, J. Martens, and G. Dahl, ‘‘On the importance of initialization and momentum in deep learning,’’ in *Proc. Int. Conf. Mach. Learn.*