Original Article

Access this article online

Quick Response Code:

Website: https://journals.lww.com/ijhm

DOI: 10.4103/ijh.ijh_110_24

Optimizing deep learning for accurate blood cell classification: A study on stain normalization and fine-tuning techniques

Mohammed Tareq Mutar, Jaffar Nouri Alalsaidissa[1], Mustafa Majid Hameed[2], Ali Almothaffar[2]

Abstract:

BACKGROUND: Deep learning's role in blood film screening is expanding, with recent advancements including algorithms for the automated detection of sickle cell anemia, malaria, and leukemia using smartphone images.

OBJECTIVES: This study aims to build the artificial intelligence (AI) models and assess their performance in classifying blood film cells as normal or abnormal.

MATERIALS AND METHODS: The dataset included 171,374 images from 961 patients which were classified by experts. These images were resized, denoized, normalized, augmented, and classified into two categories, normal and abnormal cells. Two stain normalization techniques were used in this study; Reinhard and Mackenko techniques. The data were split into training and testing sets with a ratio of (8:2). The model was built through transfer learning by using the pretrained model Inception-Resnet v2 as a backbone. Three different fine-tuning techniques were tested in this study. The training was done using Python with Keras library on Google Colab for 10 epochs. The model was tested for accurately classifying individual blood cells whether normal or abnormal and evaluated using accuracy and area under receiver operator characteristic curve.

RESULTS: The counts of the three most common cell types were as follows: Segmented neutrophils: 29,424; erythroblasts: 27,395; and lymphocytes: 26,242. The Reinhard stain normalization had better accuracy than Mackenko, the best AI model achieved the highest accuracy of 96.7%%, the area under the curve (AUC) of 99.87%, while the second technique achieved an accuracy of 91.46% and an AUC of 97.23% in classifying normal from abnormal cells.

CONCLUSION: In conclusion, AI can effectively classify the blood cells as either normal or abnormal, yielding accurate results in a time-effective manner, especially with the use of transfer learning of pretrained models and fine-tuning. In this study, Inception-Resnet V2 showed good accuracy in differentiating normal from abnormal cells.

Keywords:

Deep learning, fine-tuning, malignant cells, stain normalization

Departments of Surgery, [1]Pathology and Forensic Medicine and [2]Medicine, College of Medicine, University of Baghdad, Baghdad, Iraq

Address for correspondence:
Dr. Mohammed Tareq Mutar,
Department of Surgery,
College of Medicine,
University of Baghdad,
Baghdad, Iraq.
E-mail: mohammed.tareq1600c@comed.uobaghdad.edu.iq

Submission: 23-09-2024
Revised: 04-12-2024
Accepted: 04-01-2025
Published: 07-03-2025

Introduction

Artificial neural networks, artificial intelligence (AI) algorithms that simulate human neural tissue in terms of structure and functionality, are being increasingly utilized in multiple medical diagnostic and predictive aspects. They can learn the relationship between the inputs and outputs by extracting features and using them for decision-making.[1,2] The role of AI in medicine is rapidly expanding. It is believed that AI will be used by every type of clinician, ranging from paramedics to

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: WKHLRPMedknow_reprints@wolterskluwer.com

How to cite this article: Mutar MT, Alalsaidissa JN, Hameed MM, Almothaffar A. Optimizing deep learning for accurate blood cell classification: A study on stain normalization and fine-tuning techniques. Iraqi J Hematol 2025;14:60-5.

© 2025 Iraqi Journal of Hematology | Published by Wolters Kluwer - Medknow

specialty doctors, as it can assist healthcare providers in various ways, such as interpreting radiology imaging,[3-5] pathology slides,[6,7] retinal images,[8,9] skin lesions,[10,11] electrocardiograms,[12] and endoscopy.[13,14] The reasons for expanding the role of AI in medicine are related to increased data.[15]

Hematological neoplasms contribute significantly to global morbidity and mortality.[16] According to the global burden of cancer, leukemia ranked 15th as the most commonly diagnosed cancer and 11th as the leading cause of cancer-related death globally.[17,18] Early detection of hematological malignancies is essential as it enhances the outcome. The diagnosis of these diseases is complex and based on the multiple diagnostic modalities, such as cytology, histology, genetics, immunophenotyping, and imaging.[19] For detecting blood malignancies, blood smears or bone marrow examination is required and usually examined by hematologists under the microscope.[20]

The role of deep learning in the blood film screening of various pathologies is expanding. de Haan *et al*. developed a deep-learning algorithm for automated sickle cell anemia screening in smartphone-based microscopy. They utilized two deep networks, one for enhancing phone-captured images and the second one for segmentation and classification.[21] Fuhad *et al*. developed an algorithm based on deep neural networks for the automated detection of malaria-infected cells using smartphone-captured images.[22] A recent systematic review by Ghaderzadeh *et al*. evaluated machine learning in the diagnosis and classification of leukemia in blood films and revealed that deep learning was utilized in 37% of the studies, with increasing frequency in the last few years. They used different pretrained models and transfer learning in their works.[23]

This study aims to build AI models and assess their performance accuracy in classifying blood film cells as normal or abnormal cells.

Materials and Methods

The dataset used in this study comprised bone marrow cytological preparations from 961 patients at Munich Leukemia Laboratory with various hematological diseases. The bone marrow slides were stained using Giemsa stain and cropped into single-cell images that experts classified into 22 categories.

The data was then preprocessed by resizing the images to 180 × 180 pixels due to limited resources. Next, the images were denoized using a median filter with a kernel size of 3, followed by stain normalization and data augmentation. Images were classified into two categories based on whether the cell could be seen normally in a blood film or not, as determined by the hematologists' opinion of the class to be included in each category. The data were split into training and testing sets with an 8:2 ratio.

Stain normalization techniques are robust methods for color correction that involve a heterogeneous group of unprocessed images as standards for mapping various target images.

There are three types of color normalization methods, including global color normalization techniques like Reinhard, color normalization after stain separation based on supervised techniques, and color normalization after stain separation based on unsupervised techniques.[24]

Two preprocessing stain normalization functions were applied to all images including:

A. Separating eosin-like and the hematoxylin-like components based on the principle component analysis suggested by Macenko *et al*. method[25,26]
B. Matching the mean and standard deviation of each channel between the image and the target using linear transforms based on the work of Reinhard *et al*.,[27] Figure 1 shows the two stain normalization technique applied on a blast cell.

Convolutional neural networks (CNNs) have a large number of parameters and typically require a large amount of data for training. Even with small CNNs, this problem is applied. Various data augmentation techniques are used to overcome this limitation.[28] Data augmentation is applied by creating a set of new images using variations of the original images. The increase in data aims to reduce CNN overfitting and improve the generalization of the trained model.[29]

A CNN model was built through the transfer learning by using the pretrained model on the ImageNet dataset with fine-tuning. The pretrained Inception-Resnet v2 model was adopted as a backbone, which is built based on two deep learning models, Inception and Resnet. This represents a modification in inception models by introducing residual blocks that add the output of the inception block to the input layer by using a 1 × 1

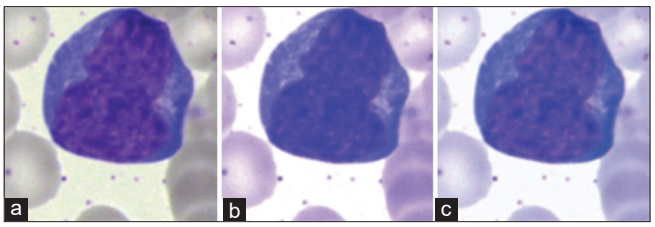

Figure 1: (a) The original image, (b) Mackenko normalized, (c) Reinhard normalized

convolution layer to match the depth. The output of this model is (8 × 8 × 1536). This is followed by global average pooling and two dense layers, and a classifier sigmoid. The models were built using Python v3.1 and Keras library with 25GB RAM provided by Google Colaboratory.

Fine-tuning the model involved implementing three different techniques. The first method entailed freezing the layers and only training the two dense and classifier layers. The second method involved training the classifier in the first step and gradually unfreezing the layers, starting with the first 10% and moving on to the next 10% until the training was complete. In the third method, the model was first trained to classify the 22 different cell categories by gradually unfreezing the first 10% of layers and then the next 10% until all layers were trained, to improve the model's ability to extract the features of blood cells. After completing the training process, the model was fine-tuned to classify cells as either normal or abnormal by retraining the sigmoid classifier while freezing all other layers of the model. The study was approved by the ethical committee of the Pathology Department College of medicine, University of Baghdad in February 11, 2024.

The three models were trained on Google Colab for 10 epochs, the training parameters are shown in Table 1.

All the statistical analyses were done using Python V3.1, the data were summarized with count and proportion, and the models were evaluated using accuracy and area under the curve (AUC).

Accuracy = (TP + TN)/(FP + FN + TP + TN), where TP = True Positives, TN = True Negatives, FP = False Positives, FN = False Negatives.

Results

The age range of included patients was 18.1–92.2 years, with a median age of 69.3 years and a mean age of 65.6 years. The cohort included 575 males (59.8%) and 385 females (40.1%).[27] There were 171,374 images with 83726 cells considered normal and the remainder were abnormal, the summary of the cell types and their count is shown in Table 2.

These models achieved different levels of accuracy, with Reinhard stain normalization had the highest accuracy during testing and was used as the standard normalization technique. The heat map for classifying a blast cell is shown in Figure 2.

The accuracy and AUC are presented in Figure 3 and Table 3, the 3rd fine-tuning technique achieved the highest accuracy and AUC 96.7 and 99.87, respectively, while the first technique had the least accuracy of 88.7%. Figure 4 displays the confusion matrix for the third fine-tuning technique using Mackenko and Reinhard stain normalization methods.

Table 1: Training parameters for the three algorithms in Google Colab

	Parameters / Specifications
Learning rate	0.001
Optimizer	Adam
Loss function	Binary crossentropy
Classifier	Sigmoid
Epochs	10
Augmentation techniques	Flipping and rotation

Table 2: Dataset cell types and their count for the used dataset

Cell type	Count	Considered normal or abnormal
Segmented neutrophils	29,424	Normal
Erythroblasts	27,395	Abnormal
Lymphocytes	26,242	Normal
Artefacts	19,630	Abnormal
Promyelocytes	11,994	Abnormal
Blasts	11,973	Abnormal
Band neutrophils	9968	Normal
Plasma cells	7629	Normal
Myelocytes	6557	Abnormal
Eosinophils	5883	Normal
Monocytes	4040	Normal
Not identifiable	3538	Abnormal
Metamyelocytes	3055	Abnormal
Proerythroblasts	2740	Abnormal
Other cells	294	Abnormal
Basophils	441	Normal
Hairy cells	409	Abnormal
Immature lymphocytes	65	Abnormal
Faggot cells	47	Abnormal
Smudge cells	42	Abnormal
Abnormal eosinophils	8	Abnormal

Table 3: The accuracy and area under the curve of three fine-tuning techniques using Reinhard normalization

	Accuracy (%)	AUC (%)
The first technique	88.76	94.22
The second technique	91.46	97.23
The third technique	96.7	99.87

AUC=Area under the curve

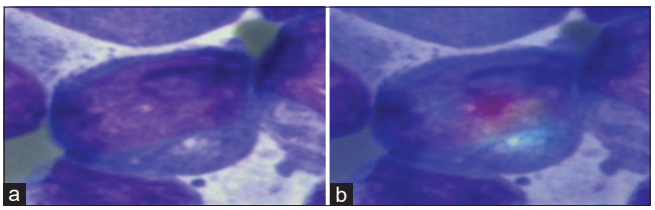

Figure 2: Heat map for a blast cell. (a) Blast cell, (b) Heatmap of the cell

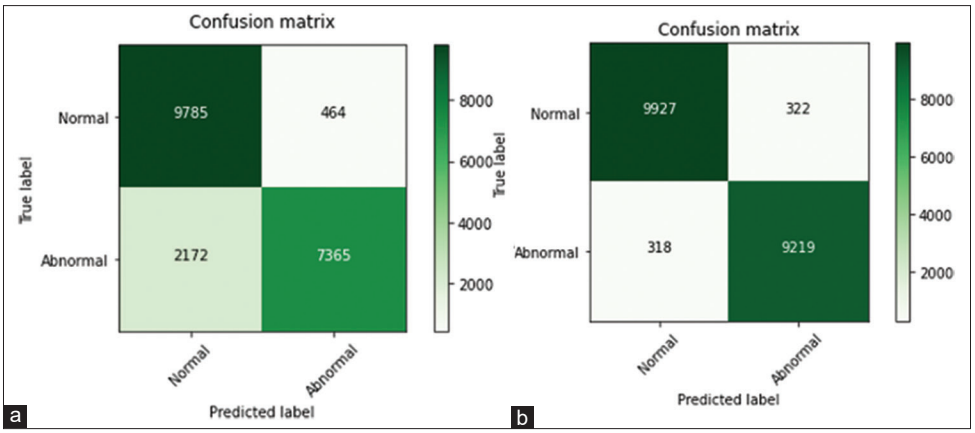

Figure 3: Receiver operator characteristics curve for the two stain normalization. (a) Mackenko Normalization, (b) Reinhard Normalization. ROC = Receiver operator characteristics

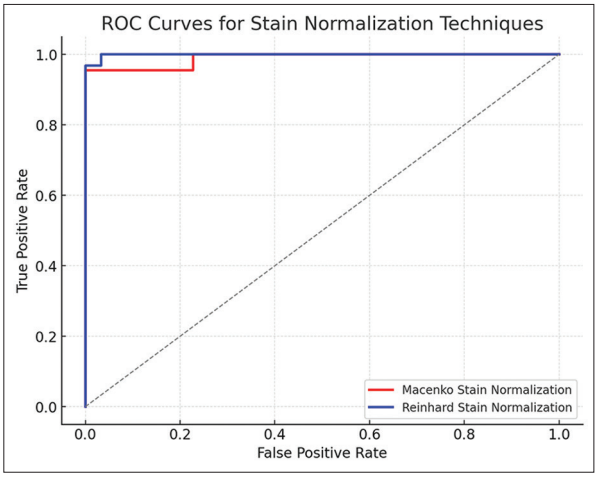

Figure 4: Confusion matrix for the third fine-tuning technique

Discussion

AI has become a modern-day innovation that has demonstrated its effectiveness in both medical and nonmedical fields and has been widely applied in various medical domains.[30,31] Machine learning, as an automated pathological diagnosis, is increasingly used in the field of pathology.[32] When diagnosing leukemia, it is typical to undertake a series of procedures that involve clinically oriented history, thorough physical examination, and laboratory investigation directed towards identifying abnormal cells in peripheral and/or bone marrow blood samples. Delay in the diagnosis is associated with a delay in treatment and poor prognosis.[33] Therefore, this study was designed to detect abnormal white blood cells and differentiate them from normal ones using a relatively accurate and quick method.

To evaluate the performance of the model presented in this study, it was compared with similar research that aimed to achieve the same objective. Boldú *et al*.[34] used an ALNet deep learning-based network to diagnose the acute leukemia lineage, and the achieved sensitivity, specificity, and precision values were 100%, 92.3%, and 93.7%, respectively, for myeloid leukemia. They obtained a sensitivity of 89% and specificity and precision values of 100% for lymphoid leukemia. Shafique and Tehsin[35] used deep CNN to detect acute lymphoblastic leukemia automatically and classify its subtypes, achieving 99.50% accuracy for leukemia detection and 96.06% accuracy for subtype classification. In addition, Hegde *et al*.[36] proposed a deep-learning approach for the classification of white blood cells in peripheral blood smear images. A systematic review by Salah *et al*.[37] concluded that using machine learning in hematology, particularly in hematological malignancy, improves diagnostic accuracy and increases clinical care efficiency in workflow and cost. Thus, the application of machine learning in such areas could have significant benefits since manual blood review is widely used not only in the hematology department.[38]

As with any deep learning project, the major limitations of the model presented in this paper are the need for a larger amount of data to achieve better performance and a larger and more complicated neural net, which increases the risk of overfitting. Furthermore, limited resources restricted the image size to smaller sizes, resulting in the loss of image details. Technical errors, such as mislabeled data, also have an impact on the model's performance.[39] These limitations must be addressed in future research.

Conclusion

In conclusion, AI can effectively screen the blood films for the presence of abnormal cells, yielding accurate results in a time-effective manner, especially with the use of transfer learning of pretrained models and fine-tuning.

In this study, Inception-Resnet V2 showed good accuracy in detecting abnormal cells.

Acknowledgment

We sincerely thank all the patients who participated in this study and those who made this dataset available. Their contributions are invaluable and greatly appreciated.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

References

1. Chung CC, Chan L, Bamodu OA, Hong CT, Chiu HW. Artificial neural network based prediction of postthrombolysis intracerebral hemorrhage and death. Sci Rep 2020;10:20501.
2. Topol EJ. High-performance medicine: The convergence of human and artificial intelligence. Nat Med 2019;25:44-56.
3. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJ. Artificial intelligence in radiology. Nat Rev Cancer 2018;18: 500-10.
4. Zhou LQ, Wang JY, Yu SY, Wu GG, Wei Q, Deng YB, *et al.* Artificial intelligence in medical imaging of the liver. World J Gastroenterol 2019;25:672-82.
5. Gore JC. Artificial intelligence in medical imaging. Magn Reson Imaging 2020;68:A1-4.
6. Chang HY, Jung CK, Woo JI, Lee S, Cho J, Kim SW, *et al*. Artificial intelligence in pathology. J Pathol Transl Med 2019;53:1-12.
7. Janowczyk A, Madabhushi A. Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. J Pathol Inform 2016;7:29.
8. Lin M, Bao G, Sang X, Wu Y. Recent advanced deep learning architectures for retinal fluid segmentation on optical coherence tomography images. Sensors (Basel) 2022;22:3055.
9. Liu X, Cao J, Fu T, Pan Z, Hu W, Zhang K, *et al*. Semi-supervised automatic segmentation of layer and fluid region in retinal optical coherence tomography images using adversarial learning. IEEE Access 2018;7:3046-61.
10. Takiddin A, Schneider J, Yang Y, Abd-Alrazaq A, Househ M. Artificial intelligence for skin cancer detection: Scoping review. J Med Internet Res 2021;23:e22934.
11. Jaworek-Korjakowska J, Kłeczek P. Automatic classification of specific melanocytic lesions using artificial intelligence. Biomed Res Int 2016;2016:8934242.
12. Aziz S, Ahmed S, Alouini MS. ECG-based machine-learning algorithms for heartbeat classification. Sci Rep 2021;11:18738.
13. Min JK, Kwak MS, Cha JM. Overview of deep learning in gastrointestinal endoscopy. Gut Liver 2019;13:388-93.
14. Wu J, Chen J, Cai J. Application of artificial intelligence in gastrointestinal endoscopy. J Clin Gastroenterol 2021;55: 110-20.
15. May M. Eight ways machine learning is assisting medicine. Nat Med 2021;27:2-3.
16. The Lancet Haematology. The global burden of haematological diseases. Lancet Haematol 2018;5:e1.
17. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018;68:394-424.
18. Bispo JA, Pinheiro PS, Kobetz EK. Epidemiology and etiology of leukemia and lymphoma. Cold Spring Harb Perspect Med 2020;10:a034819.
19. Papakonstantinou E, Dragoumani K, Efthimiadou A, Palaiogeorgou AM, Pierouli K, Mitsis T, *et al*. Haematological malignancies implications during the times of the COVID-19 pandemic. Oncol Lett 2021;22:856.
20. Rehman A, Abbas N, Saba T, Rahman SI, Mehmood Z, Kolivand H. Classification of acute lymphoblastic leukemia using deep learning. Microsc Res Tech 2018;81:1310-7.
21. de Haan K, Ceylan Koydemir H, Rivenson Y, Tseng D, Van Dyne E, Bakic L, *et al*. Automated screening of sickle cells using a smartphone-based microscope and deep learning. NPJ Digit Med 2020;3:76.
22. Fuhad KM, Tuba JF, Sarker MR, Momen S, Mohammed N, Rahman T. Deep learning based automatic malaria parasite detection from blood smear and its smartphone based application. Diagnostics (Basel) 2020;10:329.
23. Ghaderzadeh M, Asadi F, Hosseini A, Bashash D, Abolghasemi H, Roshanpour A. Machine learning in detection and classification of leukemia using smear blood images: a systematic review. Sci Program. 2021;2021:1-14.
24. Roy S, Kumar Jain A, Lal S, Kini J. A study about color normalization methods for histopathology images. Micron 2018;114:42-61.
25. Matek C, Krappe S, Münzenmayer C, Haferlach T, Marr C. An expert-annotated dataset of bone marrow cytology in hematologic malignancies. Cancer Imaging Arch 2021. [doi: 10.7937/TCIA. AXH3-T579].
26. Macenko M, Niethammer M, Marron JS, Borland D, Woosley JT, Guan X, *et al*. A method for normalizing histology slides for quantitative analysis. In: 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro. Boston, MA, USA. Piscataway (NJ). IEEE; 2009. p. 1107-10.
27. Reinhard E, Ashikhmin M, Gooch B, Shirley P. Color transfer between images. IEEE Comput Graph Appl 2001;21:34-41.
28. Perez L, Wang J. The effectiveness of data augmentation in image classification using deep learning. arXiv preprint arXiv:1712.04621. 2017.
29. Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. J Big Data 2019;6:60.
30. Sekhar A, Biswas S, Hazra R, Sunaniya AK, Mukherjee A, Yang L. Brain tumor classification using fine-tuned GoogLeNet features and machine learning algorithms: IoMT enabled CAD system. IEEE J Biomed Health Inform 2022;26:983-91.
31. Liu Y, Jain A, Eng C, Way DH, Lee K, Bui P, *et al*. A deep learning system for differential diagnosis of skin diseases. Nat Med 2020;26:900-8.
32. Dimitriou N, Arandjelović O, Caie PD. Deep learning for whole slide image analysis: An overview. Front Med (Lausanne) 2019;6:264.
33. Okten IN, Sezen BA, Gunaydin UM, Koca S, Ozturk T, Ozder HT, *et al*. Factors associated with delayed diagnosis and treatment in patients with malignant diseases. J Clin Oncol 2018;36:e22141.
34. Boldú L, Merino A, Acevedo A, Molina A, Rodellar J. A deep learning model (ALNet) for the diagnosis of acute leukaemia lineage using peripheral blood cell images. Comput Methods Programs Biomed 2021;202:105999.
35. Shafique S, Tehsin S. Acute lymphoblastic leukemia detection and classification of its subtypes using pretrained deep convolutional neural networks. Technol Cancer Res Treat 2018;17:1533033818802789.
36. Hegde RB, Prasad K, Hebbar H, Singh BM. Feature extraction

using traditional image processing and convolutional neural network methods to classify white blood cells: A study. Australas Phys Eng Sci Med 2019;42:627-38.

37. Salah HT, Muhsen IN, Salama ME, Owaidah T, Hashmi SK. Machine learning applications in the diagnosis of leukemia: Current trends and future directions. Int J Lab Hematol 2019;41:717-25.
38. Leung E, Johnston A, Olsen B, Chang H, Martin T, Wozniak M, *et al*. Laboratory practices for manual blood film review: Results of an IQMH patterns of practice survey. Int J Lab Hematol 2021;43:184-90.
39. Debs P, Fayad LM. The promise and limitations of artificial intelligence in musculoskeletal imaging. Front Radiol 2023;3:1242902.