Hematology in Global Health: Blood Cell Classification of microscopic data

by Alex Kroumov

Introduction

Blood research is critical in identification of diseases that are burdens to high-income countries (HICs) like leukemia, but it is also crucial in diagnosing diseases in low and middle income countries (LMICs), where there are fewer physicians per patient. This study will focus on applying an AI model to Acute Leukemia (AL) diagnosis in order to expedite differential complete blood counting in assessing disease state progression. Acute Leukemia (AL) is used as the model disease, but the results from the study apply to other blood diseases including anemia, malaria, and hematologic symptoms that may be indicative of broader underlying conditions.

Global Burden of Leukemia

In HICs, the prevalence rate of leukemia is 30 per 100,000 people; in LMICs, this rate is closer to 10 per 100,000 people (Huang & Zhang, 2025). Additionally, the WHO estimates 36.81 physicians per 10,000 people in the US, an HIC, while LMICs have much lower physician concentrations, such as India with 7.23 physicians per 10,000 people or Uganda with 1.91 physicians per 10,000 people (WHO 2020). Granted, this accounts for all physicians but this metric generally represents the availability of healthcare workers. Therefore, despite there being more cases in HICs, LMICs have less access to care, indicating a universal need for a low-cost method of identifying and tracking blood cell types that increases diagnosis rates for HICs and access to care for LMICs.

Current Landscape and Potential for Deep Learning

Currently, in blood research, hematologists meticulously spend time looking through microscopes to identify different types of cells in blood: infection-fighting neutrophils, allergy-battling eosinophils, rarer basophils, and more. This process is called the differential blood count, and it is critical in providing medical diagnosis of infection or a condition such as leukemia.

The advent of deep learning allows us to accelerate the creation of a complete blood count (CBC). For example, in leukemia, the differential CBC indicates an abnormal distribution of white blood cells, with a “predominance of blasts and a decrease in mature neutrophils and lymphocytes” (Tripathi 2025). The proposed project seeks to reduce the burden of counting blood cells in hematology settings, thus increasing diagnosis rates in HICs and increasing healthcare availability in LMICs. The CNN model detailed below distinguishes between eight different types of blood cells from microscopic images, paving the way for faster, more efficient, and widely accessible diagnostic tools. This report outlines the methodology and the results of this analysis.

Methods

Dataset

The analysis uses the BloodMNIST dataset which is a part of a larger collection called MedMNIST v2 to help verify and validate machine learning model performance on medical imaging tasks (Acevedo et al., 2020; Yang et al., 2023). BloodMNIST contains 17,092 labeled images of blood cells from healthy individuals. Each image is labeled as one of eight types: basophil, eosinophil, erythroblast, immature granulocyte, lymphocyte, monocyte, neutrophil, or platelet. They already split the source dataset with a ratio of 7:1:2 into training, validation and testing sets. The source images with resolution 3 × 360 × 363 pixels are center-cropped into 3 × 200 × 200, and then resized into 3 × 28 × 28.

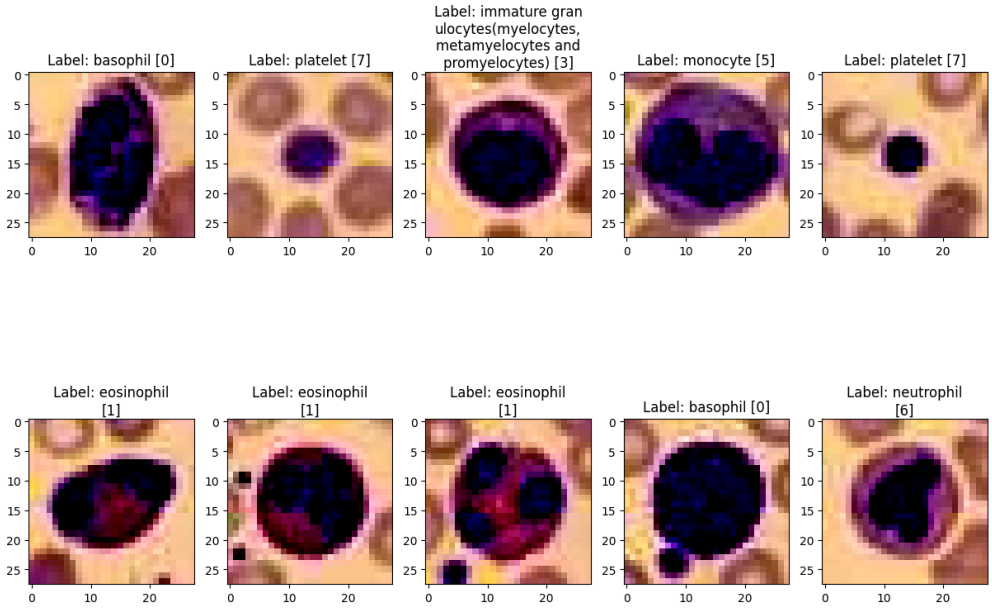

Fig 1. Diagram of the first 10 training samples with labels.

Architecture

The lightweight CNN was built with 5 convolutional layers. Each layer consists of three processes: 1) convolution of the input image with a 3 × 3 kernel and 1 unit of padding to preserve spatial dimensions, 2) batch normalization to ensure stability, and 3) a ReLU nonlinear activation function to encourage learning. Layers 2 and 5 contain a max_pooling parameter which takes the maximum value in a 2 × 2 square pixel area to the next layer in order to reduce spatial dimensions. Finally, the outputs to the convolutional layers are put into the fully connected layer, which takes each patch and passes it through a multi-layer perceptron. Below, I elaborate on these steps

Convolutional Layers: These layers find the patterns in the data from low-level to high-level vision tasks. The first few layers might learn to detect simple things like edges, curves, and color spots. Deeper layers combine these simple patterns to recognize more complex features, like the specific texture of a cell's nucleus or the color of its cytoplasm.

ReLU (Rectified Linear Unit): This is a piecewise function whose output is the input if the input is positive, and zero if the input is negative. This nonlinearity has been found to be effective in training a model by forcing it to

fit to the training data. If it were just a linear function, it would be as though the entire network were just one linear operation, which is not conducive to learning.

Max Pooling Layers: These layers are needed for downsampling the image, effectively summarizing the most important features found by the convolutional layers. This makes the process more efficient and helps the model generalize better.

Fully Connected Layers: After passing through all the pattern-detection layers, the data should be in a form that is easier for a machine to learn. The information is flattened, meaning each convolved patch is fed as an input into the perceptron. This is where classifications occur. Based on all the features it has detected, the model calculates the probability that the image belongs to each of the eight cell categories and makes its final prediction.

Training Parameters

The model processed the training images in batches of size 128, making a prediction for each one. A Cross-Entropy Loss function was used to measure error in predictions compared to the label, in standard fashion for supervised learning. The goal of the model is to minimize the loss.

To minimize loss, a Stochastic Gradient Descent (SGD) function was used. After each batch, it looks at the error and slightly adjusts the CNN's internal parameters in a direction that should reduce error by next time. The new weight vector equals the old weight vector minus a learning rate scalar times a unit directional gradient. This process was repeated for 10 "epochs," meaning the model looked at the entire training dataset ten times. To handle the massive number of calculations, this entire process was run on Google Colab’s T4 GPU, which gave results in less than four minutes.

Results

After 10 epochs of training, the model was tested. The model achieved a final test accuracy of approximately 94%. This means that when exposed to 100 new images, it correctly classified 94 of them. Considering there were 8 classes, this is an encouraging result.

To get a deeper understanding of its performance, the true positives and false negatives were identified through a confusion matrix. Confusion matrices give a detailed breakdown of the model's successes and failures. The model almost never mistook a platelet for a lymphocyte. This high level of accuracy demonstrates that the CNN

successfully learned the subtle visual differences between the various blood cells.

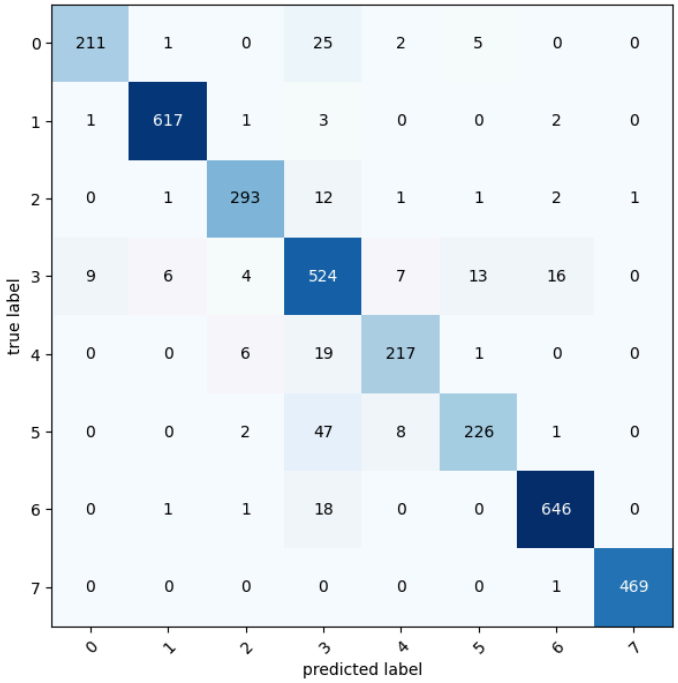

Fig 2. Confusion Matrix indicating model successes in accurately classifying between types of blood cells on the testing data. Misclassifications occurred

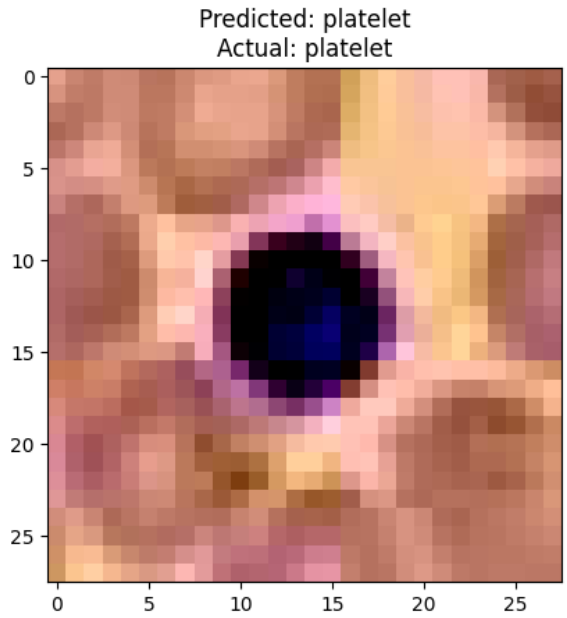

Fig 3. Model showing how the predicted value matches the actual value.

Concluding Remarks

In conclusion, this project successfully demonstrated that a Convolutional Neural Network can be trained to classify different types of blood cells with a high degree of accuracy even with 28 × 28 pixel resolution.

This technology alone is useful for research, but it shines in the global health space. Typical blood tests take a sample of blood and smear on a microscope slide. Then, trained technicians count the number and proportion of

each class of blood cells to create a CBC with a differential. This laborious process takes 15 minutes per slide. With this CNN, this bottleneck can be reduced to an entire dataset in five minutes.

Thus, AI-powered tools could one day act as assistants to hematologists, performing initial screenings and flagging abnormal cells, thereby increasing the speed and efficiency of diagnoses. In under-resourced areas with limited access to specialists, such a system could revolutionize access to care. This experiment is a promising glimpse into a future where artificial intelligence and medicine work hand-in-hand to improve patient outcomes.

References

Acevedo, A., Merino, A., Alférez, S., Molina, Á., Boldú, L., & Rodellar, J. (2020). A dataset of microscopic peripheral blood cell images for development of automatic recognition systems. Data in brief, 30, 105474. https://doi.org/10.1016/j.dib.2020.105474

Huang, P., Zhang, J. Global leukemia burden and trends: a comprehensive analysis of temporal and spatial variations from 1990—2021 using GBD (Global Burden of Disease) data. BMC Public Health 25, 262 (2025). https://doi.org/10.1186/s12889-025-21428-w

Sharma, R., Jani, C. Mapping incidence and mortality of leukemia and its subtypes in 21 world regions in last three decades and projections to 2030. Ann Hematol 101, 1523–1534 (2022). https://doi.org/10.1007/s00277-022-04843-6

Tripathi AK, Chuda R. Laboratory Evaluation of Acute Leukemia. [Updated 2025 Jan 5]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025 Jan-. Available from: https://www.ncbi.nlm.nih.gov/sites/books/NBK611988/

Yang, J., Shi, R., Wei, D. et al. MedMNIST v2 - A large-scale lightweight benchmark for 2D and 3D biomedical image classification. Sci Data 10, 41 (2023). https://doi.org/10.1038/s41597-022-01721-8