Deep Learning Model for Diagnosing Acute Myeloid Leukemia from Peripheral Blood Smear Single-Cell Images

Rifat Hassan, Abdul Mukit, Mushfiqur Rahman, Manish Roy and Ummey Tahiea Taz

Abstract In this study, we propose a deep learning-based model for the automatic classification of single-cell images from peripheral blood smears to diagnose Acute Myeloid Leukemia (AML). The model was trained, validated, and tested on a dataset comprising 11879 images of AML-positive cells and images of healthy control cells. Achieving almost 95% accuracy on training, and almost 90% accuracy on validation and testing, this model demonstrates potential as a tool for aiding pathologists in AML diagnosis.

1 Introduction

Acute Myeloid Leukemia (AML) is a type of blood cancer characterized by the rapid proliferation of abnormal white blood cells in the bone marrow, which impairs the production of normal blood cells. It originates in the bone marrow, where immature cells called myeloblasts fail to mature into healthy white blood cells. These abnormal cells accumulate, interfering with normal blood production and functioning. Early diagnosis is crucial for patient prognosis from this disease. Conventional methods include blood tests, bone marrow biopsy, cytogenetic and molecular tests,

Rifat Hassan
East West University, e-mail: h.rifat1609@gmail.com

Abdul Mukit
East West University, e-mail: Ab.Mukit.ds@gmail.com

Mushfiqur Rahman
East West University, e-mail: 2023-3-83-022@std.ewubd.edu

Manish Roy
East West University, e-mail: 2023-3-83-028@std.ewubd.edu

Ummey Tahiea Taz
East West University, e-mail: 2023-2-83-033@std.ewubd.edu

etc that rely heavily on manual review of blood smears by pathologists. Medical imaging technologies such as magnetic resonance imaging (MRI), computed tomography (CT), mammography, and histopathological imaging do also contribute. However, these manual processes are often time-consuming and prone to human error. This has led to a growing interest in utilizing deep learning (DL), a subset of artificial intelligence (AI), to automate and enhance the accuracy of cancer diagnosis from medical images.These are quite time-consuming and subject to variability. As a result, there is a continuous grow of demand of automatic diagnosis process.

With the advancement of time, advances in deep learning and medical imaging have paved the way for automated diagnostic tools. In this work, we present a deep learning-based approach for classifying single-cell images of peripheral blood smears. It is a noble approach to differentiate between AML-diagnosed and healthy cells to detect AML automatically.

2 Deep Learning Architectures for Cancer Diagnosis

2.1 Convolutional Neural Network (CNN)

CNNs are the most commonly used architecture for cancer diagnosis due to their ability to extract spatial hierarchies of features from images. CNNs have been successfully applied in tasks such as tumor detection, segmentation, and classification across different cancer types. For example, Ribli et al. (2018) demonstrated the efficacy of CNNs in breast cancer detection from mammograms, outperforming radiologists in identifying lesions.

2.2 Recurrent Neural Networks (RNN)

Although CNNs are effective in handling spatial data, RNNs excel in processing sequential data, such as imaging data collected over time. By combining RNNs with CNNs, researchers have developed models that can track tumor progression in longitudinal MRI scans, improving cancer prognosis and monitoring.

2.3 Autoencoders

Autoencoders are unsupervised models that are used for anomaly detection in medical images. These models can detect subtle deviations from normal tissue structures,

identifying cancerous growths even when labeled data is scarce. This makes autoencoders particularly useful in detecting rare cancer types or early-stage cancers.

2.4 Generative Adversarial Networks (GAN)

GANs are used to generate synthetic medical images, which can augment existing datasets and improve the training of deep learning models. Kamnitsas et al. (2017) applied GANs to brain tumor segmentation tasks, using synthetic images to enhance model performance on small datasets. This approach addresses the common issue of limited labeled medical datasets and helps create more robust models.

3 Applications of Deep Learning in Cancer Diagnosis

3.1 Breast Cancer

Breast cancer detection has been one of the most explored applications of deep learning in cancer diagnosis. CNN-based models have been applied to mammograms, achieving accuracy levels comparable to or exceeding those of experienced radiologists. Shen et al. (2019) developed a CNN model for detecting malignant tumors in mammograms with high sensitivity and specificity, significantly reducing the likelihood of missed diagnoses.

3.2 Lung Cancer

The detection of lung cancer using CT scans has also benefited from deep learning advancements. Models developed as part of the LUNA16 challenge have shown remarkable accuracy in detecting pulmonary nodules, which are early indicators of lung cancer. CNN-based models used for lung nodule detection and segmentation have significantly improved the identification of early-stage lung cancer.

3.3 Brain Tumors

MRI is the primary imaging technique for detecting brain tumors, and deep learning models have been applied extensively to brain tumor segmentation and classification. Havaei et al. (2017) demonstrated the use of a two-phase CNN model that

significantly improved the precision of brain tumor segmentation, aiding in surgical planning and treatment.

3.4 Skin Cancer

Dermatological imaging has seen breakthroughs with CNN models trained to classify skin lesions as benign or malignant. Esteva et al. (2017) employed a CNN trained on over 100,000 images to detect malignant melanoma with dermatologist-level accuracy, showing the potential of deep learning in telemedicine applications.

3.5 Prostate Cancer

Deep learning has also been used in prostate cancer diagnosis through MRI analysis. Litjens et al. (2018) employed CNN models for prostate tumor segmentation and cancer aggressiveness prediction, aiding in the development of personalized treatment strategies for prostate cancer patients.

4 Challenges in Applying Deep Learning to Cancer Diagnosis

4.1 Data Avaibality and Quality

Large annotated datasets are crucial for training deep learning models. However, obtaining labeled medical data is often difficult due to privacy concerns, high annotation costs, and limited availability. Moreover, variations in image quality and acquisition techniques across different medical institutions can affect model performance. Approaches like data augmentation, transfer learning, and federated learning have been proposed to mitigate these challenges.

4.2 Model Interpretability

One of the main barriers to adopting deep learning models in clinical practice is the ”black box” nature of these models. Clinicians need interpretable models that provide insights into how the model arrived at a particular decision. Techniques like Grad-CAM (Gradient-weighted Class Activation Mapping) and saliency maps are being developed to make deep learning models more transparent.

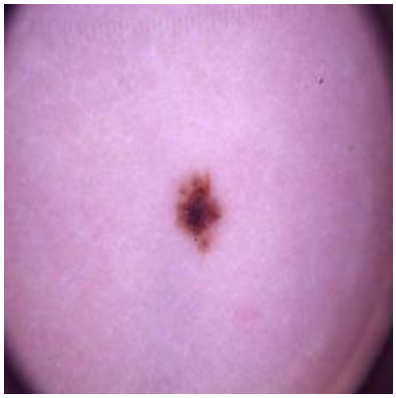

Fig. 1 A single cell image taken from peripheral blood smears diagnosed benign (healthy). Collected from TCIA.

4.3 Generalizability

Deep learning models trained on specific datasets may not perform well when applied to different patient populations, imaging modalities, or equipment. Improving the generalizability of these models is essential for widespread clinical adoption.

4.4 Regulatory and Ethical Concerns

The use of patient data to train deep learning models raises privacy and ethical concerns. Moreover, for these models to be deployed in clinical practice, they must meet strict regulatory standards, such as FDA approval.

5 Methodology

5.1 Dataset

Our methodology includes training a large dataset of single-cell images, using a pre-trained ResNet50 model with first a flattening layer, then a Dense layer with 128 neurons with ReLU activation, and finally a Dense layer with one neuron with Sigmooid activation. For this purpose, we utilized a dataset containing 11879 single-cell images (224x224x3 pixels) taken from peripheral blood smears. The images were obtained from The Cancer Imaging Archive (TCIA). Sample of an AML diagnosed single cell image is provided in fig 1. The dataset was balanced, containing

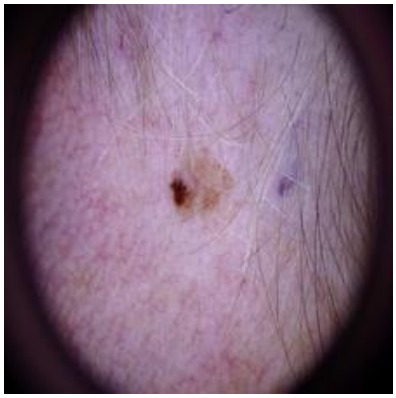

Fig. 2 A single cell image taken from peripheral blood smears diagnosed malignant. Collected from TCIA.

an equal number of images from AML-positive and healthy control patients. All images were labeled by experts.

5.2 Preprocessing

Before feeding to the model, images were preprocessed through normalization and data augmentation techniques, including rotation, zooming, and flipping. The goal behind such preprocessing was to enhance the model's robustness and prevent overfitting. While augmentation, the rotation range was set to 20, both the width and height shifting range was set to 2.0, and the zoom range was set to 0.2. Images were shifted horizontally to produce more variation in training.

5.3 Model Architecture

We employed a pre-trained ResNet50 model with ImageNet weights and an input shape of 224x224x3. The top classification layers were excluded. Custom fully connected layers were added: first a flattening layer, followed by a Dense layer with 128 neurons using ReLU activation, and a final Dense layer with 1 neuron using a sigmoid activation for binary classification. The pre-trained ResNet50 layers are frozen to prevent their weights from being updated during training.

The architecture of the model is summarized as follows:

Input: 224x224x3 pixel single-cell images.

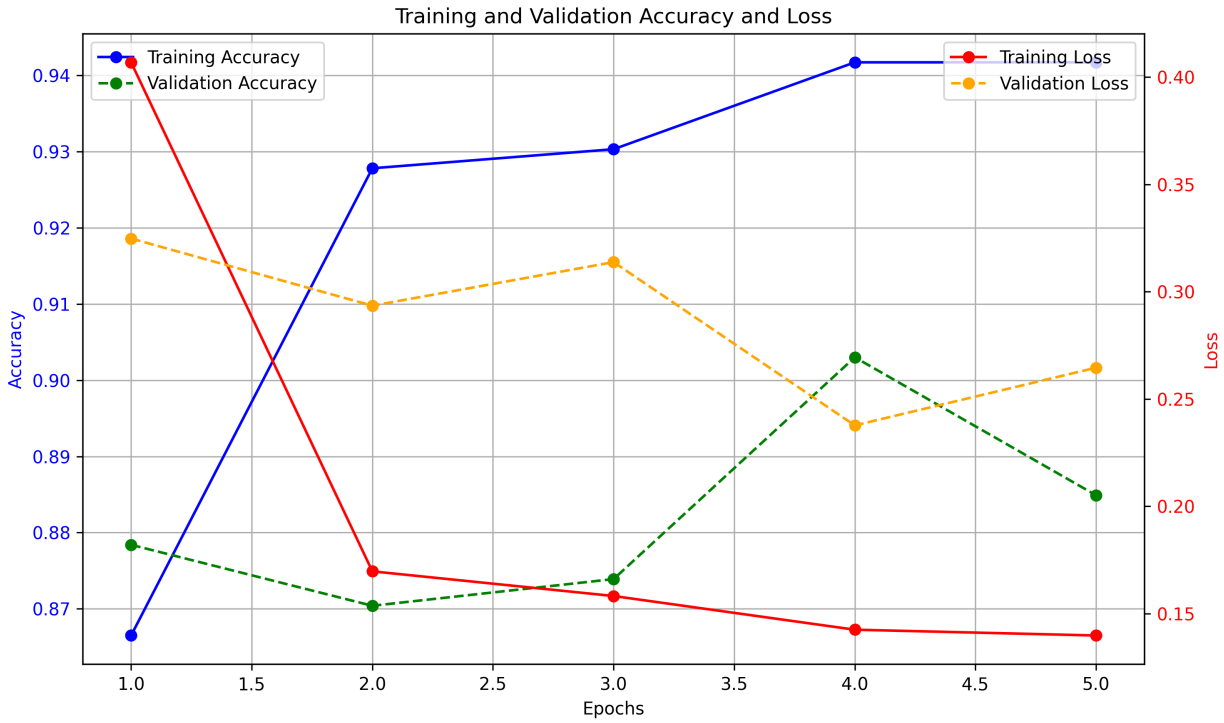

Fig. 3 Training and validation accuracy and losses of the 5 epochs.

Base Model: ResNet50 pre-trained on ImageNet without top layers.

Custom Layers: Flatten, Dense (128 units, ReLU), and Dense (1 unit, Sigmoid).

Freezing: ResNet50 layers to prevent weight updating.

5.4 Training

The model was trained using the Adam optimizer with a learning rate of 0.0001. Since the dataset was a balanced one, binary cross-entropy as the loss function has been used. The whole dataset was split into 80% for training purposes, 10% for validation purposes, and 10% for test purposes. The model was trained for 50 epochs. Early stopping was applied based on validation loss.

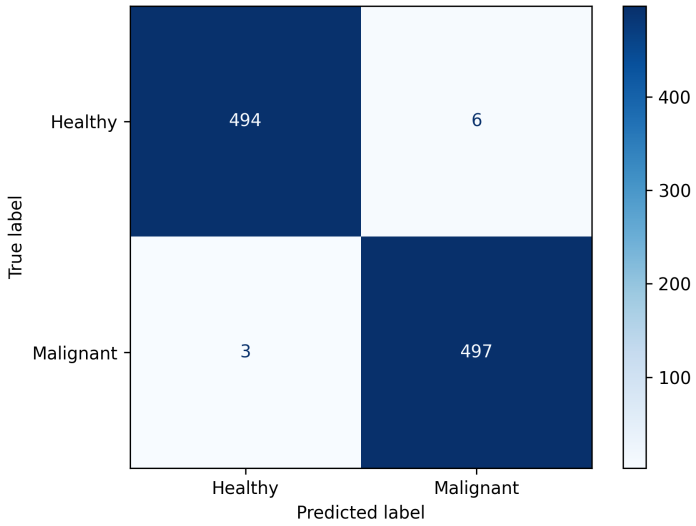

Fig. 4 Confusion matrix generated by testing 500 malignant and 500 healthy cell images.

6 Results

6.1 Accuracy

We run the model for 5 consecutive epochs, and it generated a healthy training and validation accuracy each time. During the final epoch, it achieved 94.17% training accuracy with a loss of 0.1425, and 90.30% validation accuracy with a loss of 0.2377. These results results show that the model can effectively distinguish between AML-diagnosed (malignant) and healthy (benign) cells. The training and validation accuracy and losses during each epoch have been plotted in figure 3.

6.2 Confusion Matrix and ROC Curve

The final model was tested with a pre-labelled dataset of 1000 single cell images, found from the Cancer Imaging Archive (TCIA). 500 of them were labelled malignant, 500 were labelled healthy, all were labelled by ex Then a confusion matrix was generated, showing high sensitivity and specificity. The receiver operating characteristic (ROC) curve was plotted, with an area under the curve (AUC) of 0.98, indicating excellent model performance. While testing a cell image, the threshold value for being diagnosed malignant was set to 0.5. The confusion matrix and the ROC

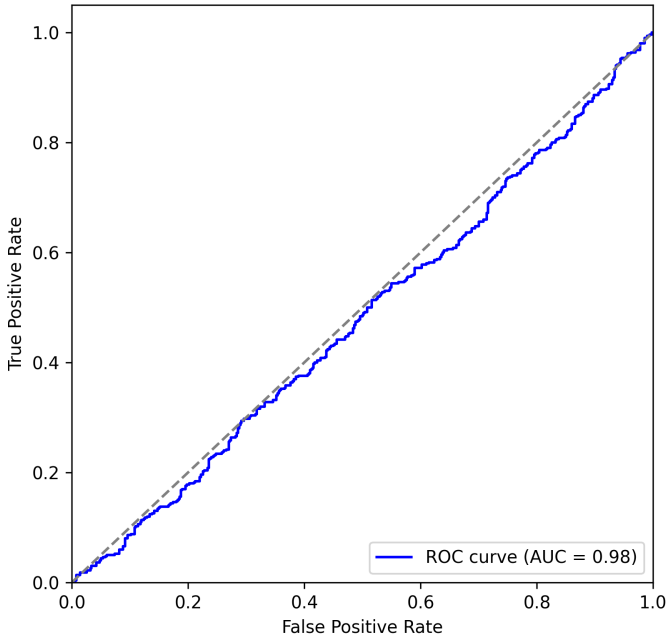

Fig. 5 ROC Curve generated by testing 500 malignant and 500 healthy cell images.

curve has been shown in figure 4 and 5 respectively. The F1 score curve was also plotted, that has been shown in figure 6.

7 Discussion

Our results provide compelling evidence of the significant potential of deep learning technologies in aiding the diagnosis of Acute Myeloid Leukemia (AML) from blood smear images. The model's impressive performance is reflected in its high accuracy and Area Under the Curve (AUC), which suggest that it could serve as a reliable and effective diagnostic tool for medical professionals. This capability not only enhances the speed of diagnosis but also aims to improve overall diagnostic accuracy, ultimately benefiting patient care and treatment planning.

However, while the model exhibits strong performance metrics, it is important to acknowledge certain limitations that may impact its efficacy in real-world applications. Although the dataset utilized for training is balanced, it may not fully

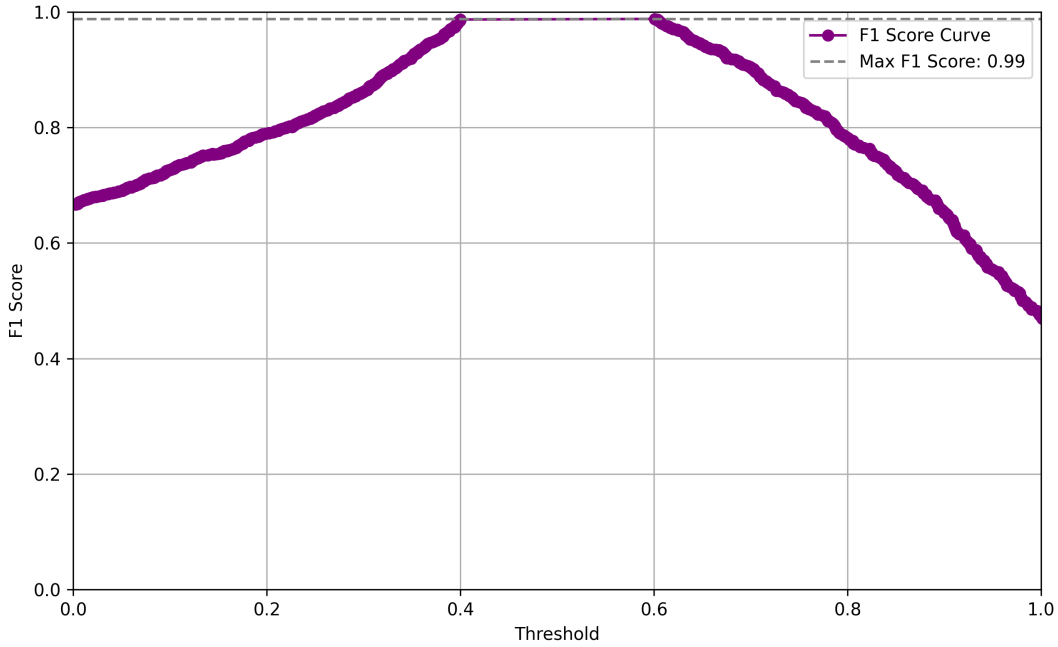

Fig. 6 F1 Score Curve generated by testing 500 malignant and 500 healthy cell images.

represent the diverse array of cell morphologies that exist in different populations. This lack of diversity could potentially limit the model's generalizability and effectiveness when deployed across varied patient demographics. Moreover, further validation with clinical datasets is imperative to ensure the model's robustness and reliability in practical healthcare settings. This validation process will help ascertain whether the model can maintain its performance standards when faced with real-world complexities and variations in patient data.

8 Conclusion

We have successfully developed a ResNet50-based deep learning model that demonstrates remarkable accuracy of nearly 95% in diagnosing Acute Myeloid Leukemia (AML) from single-cell images of peripheral blood smears. By leveraging the power of convolutional neural networks (CNNs), particularly ResNet50, which is known for its efficiency in handling complex image classification tasks, the model can effectively differentiate between healthy and malignant cells. This advancement

marks a promising step toward automating hematological diagnostics, offering the potential to significantly reduce the workload of pathologists, streamline the diagnostic process, and enhance the speed and accuracy of AML detection.

The model's success points toward its future applicability in real-world clinical environments, where early and accurate diagnosis is crucial for patient outcomes. Automating this process could not only improve diagnostic consistency but also address the limitations posed by the variability in human interpretation. Additionally, by reducing the manual effort required, the system could enable pathologists to focus on more complex cases, leading to a more efficient healthcare workflow.

Moving forward, our primary objective will be to expand the dataset by incorporating more diverse and larger image sets, capturing a broader spectrum of cell types and conditions. This will help improve the model's generalizability and robustness. Further validation of the model in clinical settings is also crucial to assess its performance in real-world scenarios, ensuring that it meets the required standards for clinical adoption. The integration of this AI-driven model into routine hematological diagnostics could revolutionize the field, facilitating quicker, more accurate, and cost-effective leukemia diagnosis and treatment planning.

References

1. Matek, C., Schwarz, S., Marr, C., Spiekermann, K. (2019), Human-level recognition of blast cells in acute myeloid leukaemia with convolutional neural networks. *Nature Machine Intelligence, 1(10), 538–544*
2. Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature, 542(7639), 115-118.*
3. Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., . . . , Larochelle, H. (2017). Brain tumor segmentation with deep neural networks. *Medical Image Analysis, 35, 18-31.*
4. HKamnitsas, K., Ledig, C., Newcombe, V. F., Simpson, J. P., Kane, A. D., Menon, D. K., . . . , Glocker, B. (2017). Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Medical Image Analysis, 36, 61-78.*
5. Litjens, G., Toth, R., van de Ven, W., Hoeks, C., Kerkstra, S., van Ginneken, B., . . . , Huisman, H. (2018). ProstateX Challenge data and evaluation update: The promise of deep learning. *Medical Physics, 45(7), e806-e822.*
6. Ribli, D., Horváth, A., Unger, Z., Pollner, P., Csabai, I. (2018). Detecting and classifying lesions in mammograms with deep learning. *Scientific Reports, 8(1), 4165.*