Convolutional Neural Network for Acute Lymphoblastic Leukemia Classification

Instituto Tecnológico y de Estudios Superiores de Monterrey

Author: *Gamaliel Marines*, gamaliel_projects@hotmail.com

Abstract

Implementation of a *Convolutional Neural Network* (CNN) to detect different stages of *Acute Lymphoblastic Leukemia (ALL)* and *classify* them correspondingly. The data set consists of *3256* PBS *images* from 89 suspected of ALL patients. This dataset is divided into four classes: *benign, early, pre,* and *pro*. The result of the implementation is a *CNN model* using *transfer learning ResNet50*, reaching an *accuracy* of 97% in test cases, 99% in validation cases, and 98% in training cases.

Impact

Leukemia is one of the most lethal forms of cancer, particularly affecting children and young adults. The implementation of a Convolutional Neural Network (CNN) for the detection and classification of Acute Lymphoblastic Leukemia (ALL) stages has the potential to significantly enhance diagnostic accuracy and speed. Early and accurate detection is crucial in improving patient outcomes, as timely intervention can lead to better treatment responses and increased survival rates.

Improved Diagnostic Accuracy: The CNN model demonstrated great accuracy in test cases and validation cases. By automating the classification process, healthcare providers can reduce misdiagnosis and ensure that patients receive appropriate treatment more swiftly.

Reduction in Diagnostic Time: The use of CNNs can drastically reduce the time required for diagnosis. Traditional methods may take longer, while a trained CNN can analyze and classify images in a matter of seconds. This rapid processing capability is particularly beneficial in emergency situations where prompt diagnosis is essential for effective treatment.

Scalability and Accessibility: The deployment of CNNs in clinical settings can enhance the scalability of leukemia detection. With an increasing number of patients requiring evaluation, automated systems can handle larger volumes of data without compromising accuracy. This can be particularly advantageous in regions with limited access to specialized medical professionals, making advanced diagnostic tools more accessible to underserved populations.

Introduction

Acute Lymphoblastic Leukemia (ALL) is a rapidly progressing cancer of the blood and bone marrow that predominantly affects children but can also occur in adults. Early and accurate diagnosis of ALL is critical, as timely intervention significantly improves survival rates and patient outcomes. Traditional diagnostic methods rely heavily on manual examination of Peripheral Blood Smear (PBS) images by hematologists, a process that is both time-consuming and prone to human error due to the visual similarities between leukemic and normal cells.

Advances in deep learning, particularly Convolutional Neural Networks (CNNs), have revolutionized image-based diagnostics by

enabling automatic feature extraction and classification with high accuracy. CNNs are especially well-suited for medical imaging tasks, as they can identify subtle patterns and anomalies that may be overlooked by the human eye.

This study presents the implementation of a CNN model using transfer learning with ResNet50 to detect and classify different stages of ALL. The model is trained on a dataset comprising 3,256 PBS images from 89 patients suspected of having ALL, categorized into four classes: benign, early, pre, and pro.

By leveraging the power of deep learning, the proposed system aims to support medical professionals in diagnosing ALL more accurately and efficiently. The CNN model achieved 97% accuracy on the test set, 99% on validation set, and 98% on training set, demonstrating its potential as a reliable diagnostic tool in clinical practice.

Understanding the Dataset

In the context of machine learning and deep learning, a dataset refers to a structured collection of data that is used to train, validate, and test models. In image classification tasks such as medical diagnostics, a dataset typically comprises labeled images that represent various classes or conditions to be recognized by the model. The quality, diversity, and representativeness of the dataset are crucial for building a model that performs reliably in real-world scenarios.

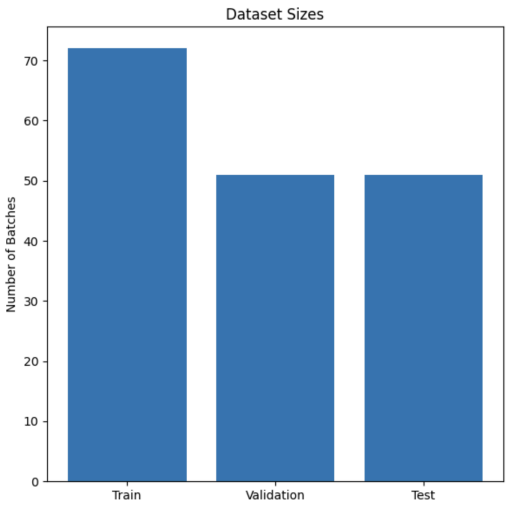
A high-quality dataset is characterized by the following attributes:

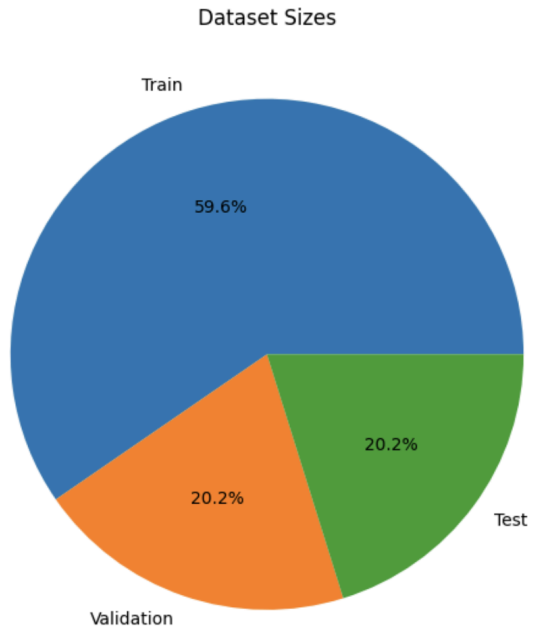
1. Clear labeling of samples by qualified experts.
2. Balanced representation of all target classes to prevent bias.
3. Sufficient sample size to allow the model to learn meaningful patterns.
4. High-resolution images that preserve important visual features.
5. Consistency in imaging conditions, such as lighting and magnification, to reduce noise.

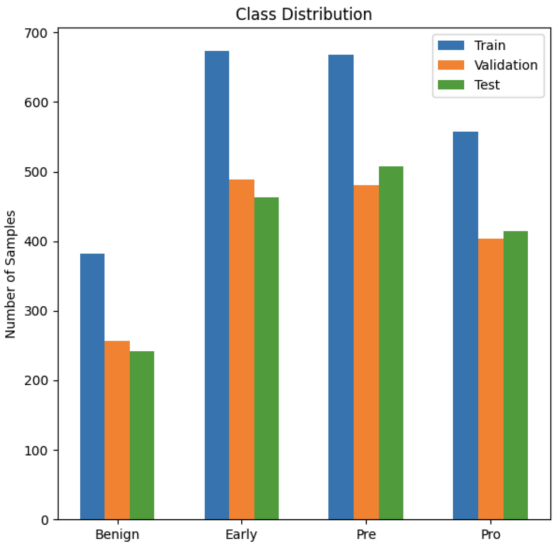
For this study, I use the publicly available Acute Lymphoblastic Leukemia (ALL) image dataset prepared at the bone marrow laboratory of *Taleghani Hospital* in Tehran, Iran. This dataset includes 3,256 peripheral blood smear (PBS) images collected from 89 individuals, including 25 healthy (benign) cases and 64 patients diagnosed with ALL. The ALL class is further categorized into three malignant subtypes: *Early Pre-B*, *Pre-B*, and *Pro-B* ALL. All images were captured using a Zeiss microscope camera at 100x magnification and saved in JPG format, ensuring visual quality.

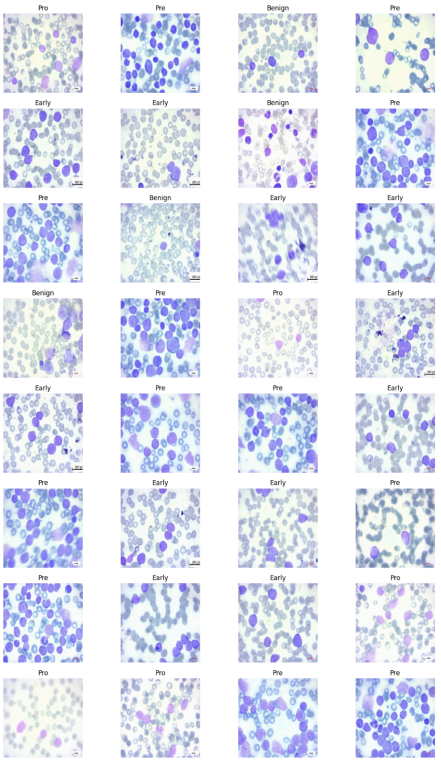
The labeling of cell types and subtypes was done by a specialist using flow cytometry, a gold standard in hematological diagnostics, which adds to the reliability of the dataset. Furthermore, the dataset includes both original and segmented images, with segmentation performed using color thresholding in the HSV color space, providing additional data for model training or augmentation.

Given its professional preparation, expert labeling, and public availability, this dataset is well-suited for developing and evaluating a Convolutional Neural Network (CNN) model for the detection and classification of ALL and its subtypes.









Dataset Source:
Mehrad Aria et al., "Acute Lymphoblastic Leukemia (ALL) image dataset." Kaggle (2021). DOI: 10.34740/KAGGLE/DSV/2175623

Publication Reference:
Ghaderzadeh, M. et al. (2022). A fast and efficient CNN model for B-ALL diagnosis and its subtypes classification using peripheral blood smear images. Int J Intell Syst. 37: 5113–5133. doi:10.1002/int.22753

Related Work and Research

The application of deep learning in medical image analysis has gained significant momentum in recent years, particularly in hematological disorders such as Acute Lymphoblastic Leukemia (ALL). Manual examination of peripheral blood smear (PBS) images remains the standard diagnostic procedure in many clinical laboratories, but it is often prone to errors due to the visual similarity between benign and malignant cells, variability in staining, and subjectivity in human interpretation.

To address these limitations, researchers have proposed various machine learning and deep learning approaches to automate the detection and classification of leukemic cells. Early works often relied on handcrafted features such as texture, shape, and color histograms extracted from segmented cells, which were then fed into classifiers like Support Vector Machines (SVM), Decision Trees, or Random Forests. While these approaches showed promise, their performance was limited by the quality of feature extraction and their inability to generalize across datasets [1].

The introduction of Convolutional Neural Networks (CNNs) marked a turning point in this field. CNNs are capable of learning hierarchical features directly from raw image data, thereby eliminating the need for manual feature engineering. Several studies have successfully applied CNN architectures to classify PBS images, demonstrating superior performance over traditional methods. For instance, Ghaderzadeh et al. proposed a fast and efficient CNN model for the diagnosis of B-ALL and its subtypes using PBS images, achieving high accuracy and significantly reducing processing time [2]. Their work also laid the foundation for the publicly available dataset used in this study.

Other notable contributions include works that employed transfer learning techniques using pre-trained models such as VGG16, InceptionV3, and ResNet50. These models, originally trained on large-scale datasets like ImageNet, can be fine-tuned on medical imaging data to leverage their robust feature extraction capabilities while requiring fewer training samples [3], [4]. Transfer learning has proven particularly effective in domains like medical diagnostics where acquiring large annotated datasets is often challenging.

Additionally, researchers have explored segmentation-based preprocessing to improve classification results. Techniques such as color thresholding in HSV color space, morphological operations, and contour extraction help isolate leukocytes from the background, thereby reducing noise and improving model focus. Some studies combine segmentation with attention mechanisms or ensemble learning to boost classification accuracy [5].

Despite these advancements, challenges remain in achieving consistent performance across diverse datasets and clinical settings. Variations in microscope type, staining techniques, and image acquisition parameters can affect model generalizability. Therefore, continuous validation with real-world clinical data and interpretability of model predictions remain crucial areas of ongoing research.

The current study builds upon this body of work by implementing a CNN model based on ResNet50 using transfer learning to classify benign and malignant PBS images, as well as the subtypes of B-ALL. By leveraging a high-quality, well-labeled dataset and focusing on the early identification of leukemic subtypes, this research contributes to the development of reliable and accessible AI-assisted diagnostic tools.

Implementation of the model

For the implementation of the classification model, a pre-trained ResNet50 architecture was used as the base. This model, originally trained on a large dataset, was imported without its top classification layers and with preloaded custom weights. The input shape was set to $224 \times 224 \times 3$ to match the resolution of the processed images. To prevent overfitting and preserve learned features, the base model's weights were frozen during training.

On top of the ResNet50 base, a custom classification head was added. This consisted of a flattening layer, a dropout layer with a rate

of 0.5 to reduce overfitting, and a final dense layer with four output units and a softmax activation function, reflecting the multi-class nature of the classification task. The model was compiled using the Adam optimizer and the categorical cross-entropy loss function, with accuracy as the evaluation metric.

To optimize training performance, two callbacks were incorporated: ModelCheckpoint, which saved the best model during training, and EarlyStopping, which halted training early if the validation loss did not improve. The model was trained for three epochs using a defined training and validation dataset, and its performance was monitored throughout the training process.

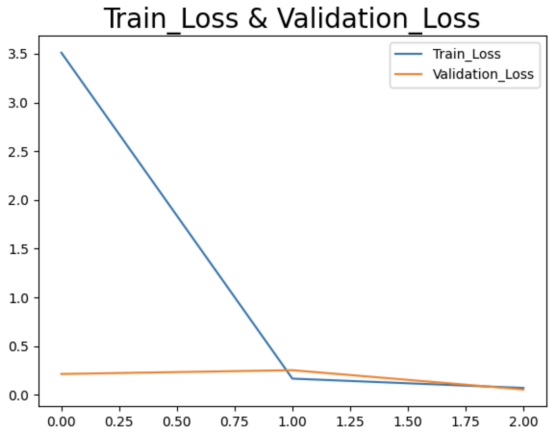
Layer (type)	Output Shape	Param #
resnet50 (Functional)	(None, 7, 7, 2048)	23,587,712
flatten (Flatten)	(None, 100352)	0
dropout (Dropout)	(None, 100352)	0
dense (Dense)	(None, 4)	401,412

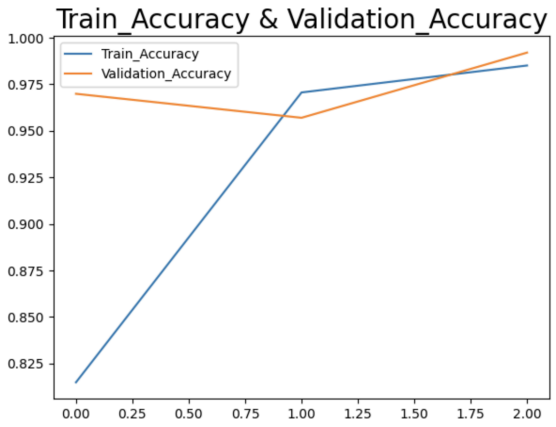
```
Epoch 1/3
72/72 ━━━━━━━━━━━━━━━━━━━━ 274s 4s/step - accuracy: 0.6686 - loss: 3.8103 - val_accuracy: 0.9343 - val_loss: 0.4046
Epoch 2/3
72/72 ━━━━━━━━━━━━━━━━━━━━ 239s 3s/step - accuracy: 0.9718 - loss: 0.2245 - val_accuracy: 0.9859 - val_loss: 0.0896
Epoch 3/3
72/72 ━━━━━━━━━━━━━━━━━━━━ 230s 3s/step - accuracy: 0.9911 - loss: 0.0534 - val_accuracy: 0.9705 - val_loss: 0.1388
```

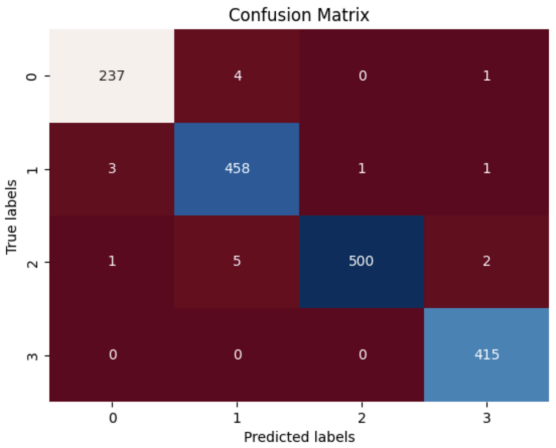
Results

blah blah

blah







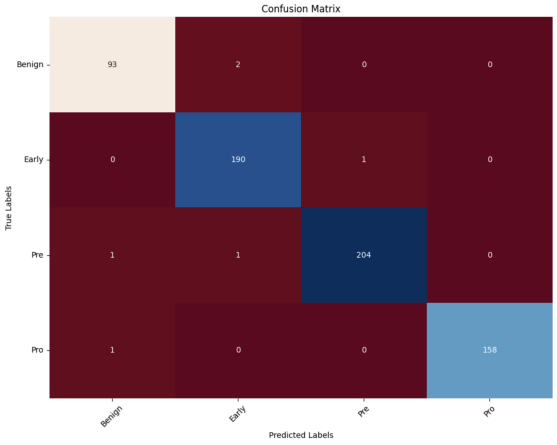
class_id	precision	recall	f1-score	support
0				
1				
2				
3				

Classification Report is :

	precision	recall	f1-score	support
0	0.97	0.99	0.98	95
1	0.99	0.99	0.99	191
2	1.00	0.99	0.99	206
3	1.00	0.99	1.00	159
accuracy			0.99	651
macro avg	0.99	0.99	0.99	651
weighted avg	0.99	0.99	0.99	651

x

After doing modifications:



Classification Report is :

	precision	recall	f1-score	support
0	0.98	0.98	0.98	95
1	0.98	0.99	0.99	191
2	1.00	0.99	0.99	206
3	1.00	0.99	1.00	159
accuracy			0.99	651
macro avg	0.99	0.99	0.99	651
weighted avg	0.99	0.99	0.99	651

Discussion

blah blah blah

Future Work

A logical next step is to train the model using a different dataset that includes images with varying characteristics. This would enhance the robustness of the training data and further improve the model's accuracy and precision. Expanding the dataset to incorporate more diverse samples can also help the model generalize better to new, unseen data.

Once the model has been retrained and validated with a broader dataset, the next phase should involve real-world testing in clinical settings. Specifically, the model should be evaluated using live images from laboratories and its results compared against expert diagnoses. This would provide valuable insights into the model's reliability in practical environments.

Based on these results, further refinement may be necessary—either through fine-tuning or retraining—to address any discrepancies or performance issues. If the model demonstrates consistent accuracy and alignment with expert assessments, it can then be considered for deployment in under-resourced or remote healthcare settings. In such environments, the model could serve as a valuable diagnostic aid, helping bridge the gap in medical expertise and improving early detection of Acute Lymphoblastic Leukemia (ALL)

References

[1] S. Rehman, M. A. Khan, M. S. Anwar, and S. W. Baik, "A hybrid deep learning framework for the detection of acute lymphoblastic leukemia using blood smear images," *Microscopy Research and Technique*, vol. 84, no. 12, pp. 2902–2912, 2021. doi: 10.1002/jemt.23814

[2] M. Ghaderzadeh, M. Aria, A. Hosseini, F. Asadi, D. Bashash, and H. Abolghasemi, "A fast and efficient CNN model for B-ALL diagnosis and its subtypes classification using peripheral blood smear images," *International Journal of Intelligent Systems*, vol. 37, no. 7, pp. 5113–5133, 2022. doi: 10.1002/int.22753

[3] H. P. Nguyen, H. T. Le, D. D. Nguyen, and H. X. Nguyen, "Transfer learning with deep convolutional neural networks for cancer cell classification," *Proceedings of the 2019*

IEEE-RIVF International Conference on Computing and Communication Technologies, Danang, Vietnam, 2019. doi: 10.1109/RIVF46745.2019.8968010

[4] A. Spanhol, L. Oliveira, C. Petitjean, and L. Heutte, “Breast cancer histopathological image classification using Convolutional Neural Networks,” *2016 International Joint Conference on Neural Networks (IJCNN)*, Vancouver, BC, Canada, 2016, pp. 2560–2567. doi: 10.1109/IJCNN.2016.7727519

[5] T. M. Abbas, A. M. Sarhan, and H. AlZubaidi, “Improved segmentation technique for leukemia detection using color thresholding and morphological operations,” *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 10, no. 2, pp. 540–546, 2019. doi: 10.14569/IJACSA.2019.0100268