

## **Deep Learning for Perception**

## **Project Report**

## 

## Fine-Grained Classification of Bone Marrow Blood Cells Using ResNet-50 CNN & Pretrained ViT Transformer

Team Members:  
Ali Raza 21K-4950

Muhammed Umer 21K-3219

Yahya Hussain 21K-4895

Section: BCS-8B

Course Instructor: Miss Sumaiyah Zahid

## **Objective** To develop and evaluate deep learning models capable of accurately classifying microscopic images of bone marrow blood cells into 14 fine-grained morphological classes. The goal is to assist pathologists in diagnosing hematological disorders such as leukemia by automating the cell-type classification process using advanced computer vision techniques.

## **Problem Statement** Bone marrow samples contain multiple types of cells, many of which appear visually similar under a microscope, making it hard even for trained specialists to classify them reliably. Mistakes or delays in diagnosis can lead to poor treatment outcomes.

## The main problem:

## Manual analysis is slow, error-prone, and not scalable.

## Some cell types are rare, leading to imbalanced data issues.

## Fine-grained differences are subtle, requiring highly capable models to distinguish between classes.

Hence, a system using deep learning, especially using CNNs and Transformers, can potentially automate this task with higher accuracy and reliability.

## **Methodology**

### **1. Dataset**

* The dataset comes from TCIA[1], specifically a medical collection of bone marrow blood smear images.
* It includes images labeled by experts into cell types such as promyelocytes, myeloblasts, lymphocytes, etc.

### **2. Preprocessing Steps**

* **Resizing** to 224x224 pixels to match the input size for ResNet-50 and ViT.
* **Normalization** so pixel values lie between 0 and 1 (or standardized).
* **Data augmentation** techniques like random rotations, flipping, and zooming help the model generalize better.

### **3. Model Architectures**

#### **ResNet-50 [2]**

* pretrained on bone marrow dataset.
* Final layer modified to output 14 classes

#### **ViT [3]**

#### Vision Transformer - Pretrained, using HuggingFace’s vit-base-patch16-224-in21k

* Fine-tuned on the dataset using PyTorch/Transformers integration

### **4. Training and Evaluation**

### Training with the Adam optimizer, cross-entropy loss (for multi-class classification).

### Metrics used:

### Accuracy: Overall correctness.

### Precision: Correct positive predictions vs. all positive predictions.

### Recall: Correct positive predictions vs. all actual positives.

### F1-Score: Harmonic mean of precision and recall.

## **Results**

## ***ResNet-50***

Test Accuracy: 68.36%  
Macro Avg F1: 0.68  
Notable Strengths: High recall in Eosinophils (EOS), Promyelocytes (PEB), and Plasma Cells (PLM)

Areas for Improvement: Misclassifications among similar myeloid precursors

#### **ViT**

Test Accuracy: 80.27%

Macro Avg F1: 0.80

Stronger performance across most classes compared to ResNet-50

Particularly accurate in identifying mature cells like Eosinophils, Plasma Cells, and Neutrophils

Demonstrated better generalization and class separation in confusion matrix analysis

## **References**

[1] Dataset: Helmholtz Imaging Platform, MLL, Fraunhofer – [Bone Marrow Cytomorphology Dataset](https://www.cancerimagingarchive.net/collection/bone-marrow-cytomorphology_mll_helmholtz_fraunhofer/)[2] He, K., Zhang, X., Ren, S., & Sun, J. (2016). *Deep Residual Learning for Image Recognition* (ResNet). [CVPR](https://arxiv.org/abs/1512.03385)[3] Dosovitskiy, A., et al. (2021). *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale*. [ViT](https://arxiv.org/abs/2010.11929)

[4]HuggingFace Transformers Library – https://huggingface.co/docs/transformers

[5]PyTorch Documentation –<https://pytorch.org>