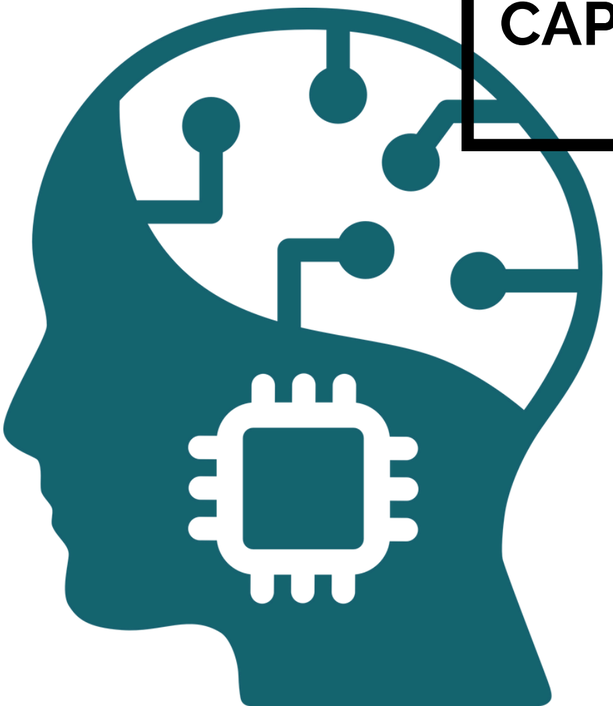
MIT | Academy of Engineering

(An Autonomous Institute Affiliated to Savitribai Phule Pune University)

ENCODER- DECODER NETWORKS FOR IMAGE CAPTIONING AND SEGMENTATION OF ANAEMIC RBCS



Presented by

Samruddhi Bhujbal - 202201070011

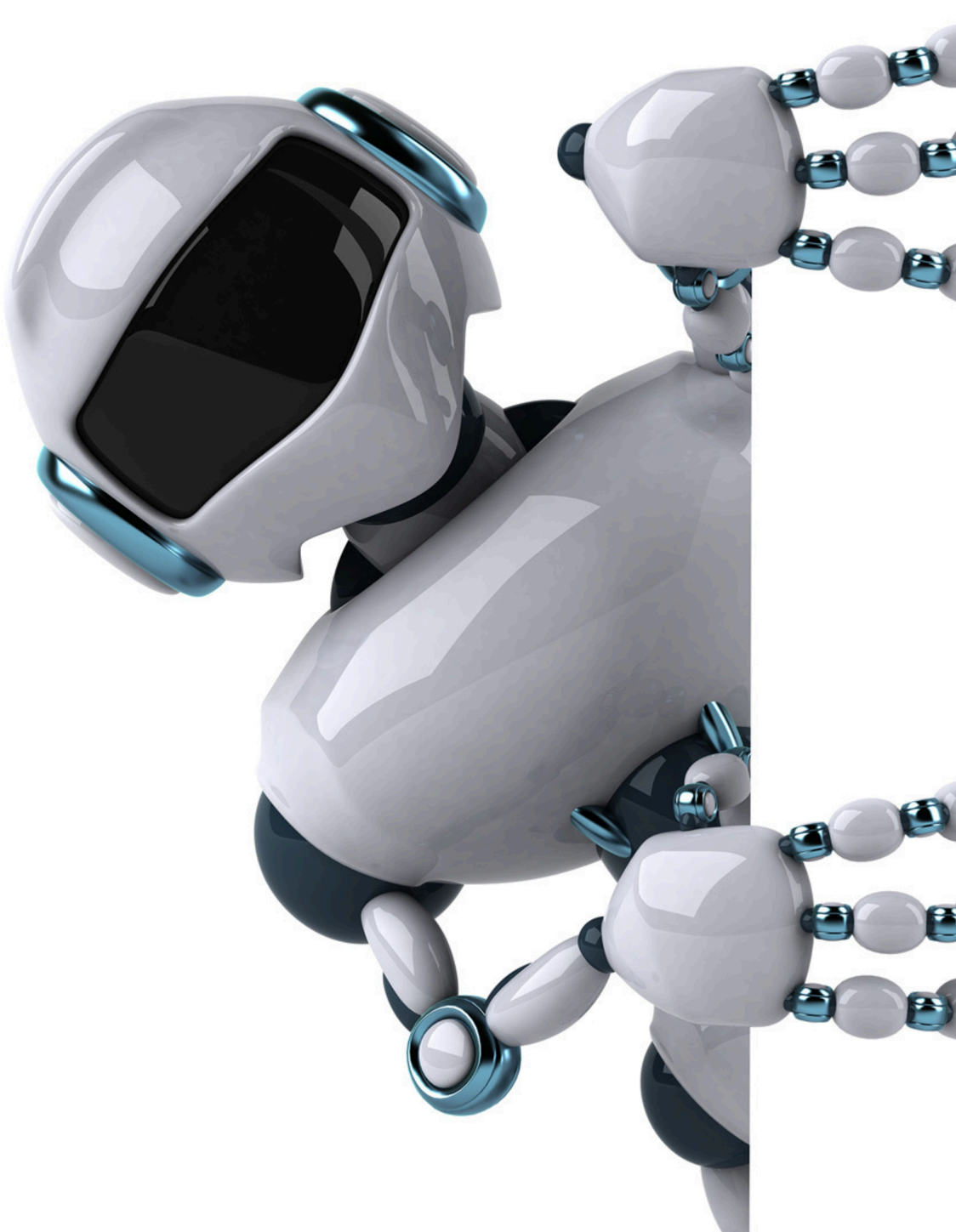
Snehal Bankar- 202201070020

Srushti Ghadge- 202201090125

Course Code- 2311332L

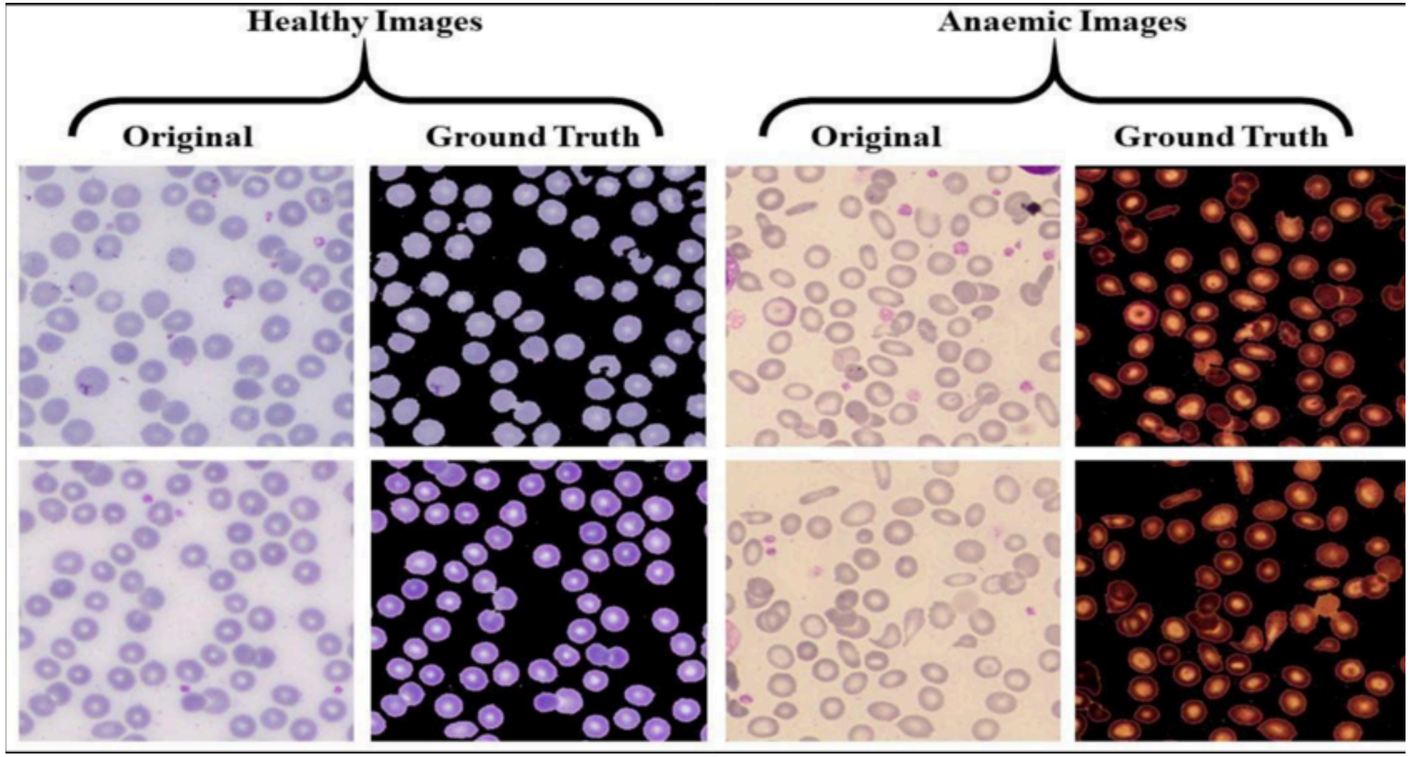
Guided by - Dr. Sunita Barve

Index



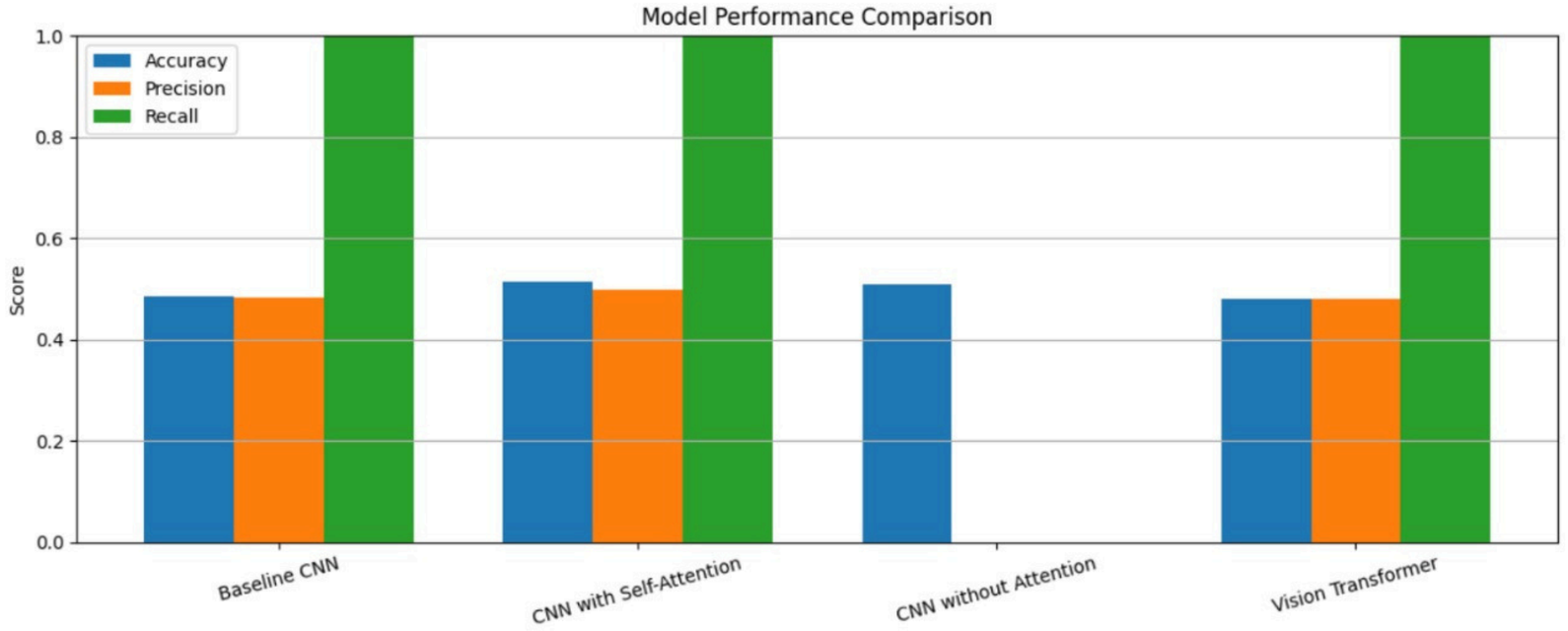
INTRODUCTION

Anaemia is a condition marked by a lack of healthy red blood cells, often diagnosed through microscopic analysis of blood smears. Manual examination is time-consuming and prone to error. This project proposes a deep learning-based semantic segmentation approach using a multilevel encoder-decoder network to automatically identify anaemic RBCs, aiming to improve diagnostic accuracy and efficiency.



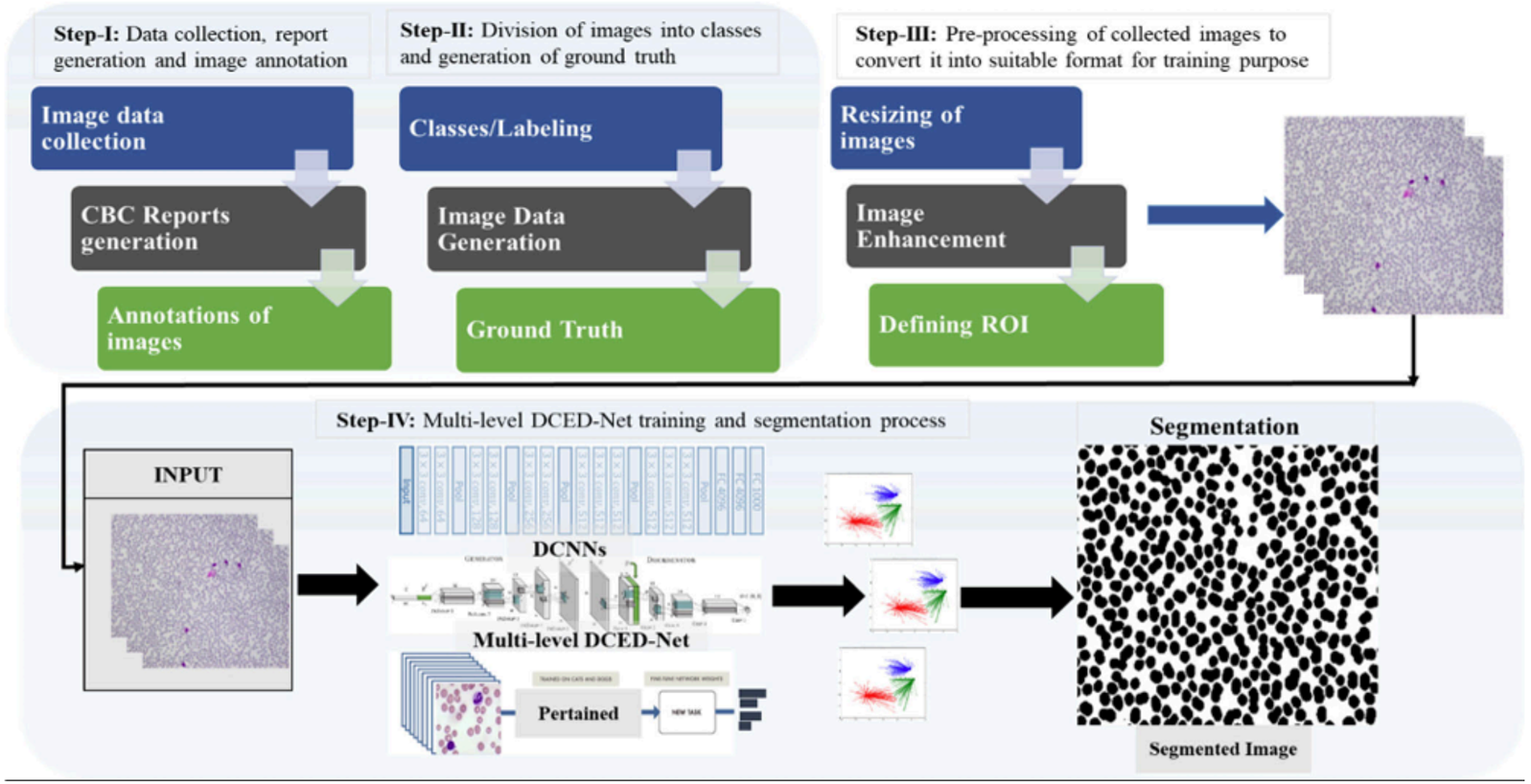
Objectives

1.Implement 3 models:
No Attention
Bahdanau Attention
Self-Attention (Transformer)
2. Train on RBC CELL dataset
3.Compare performance using BLEU & ROUGE

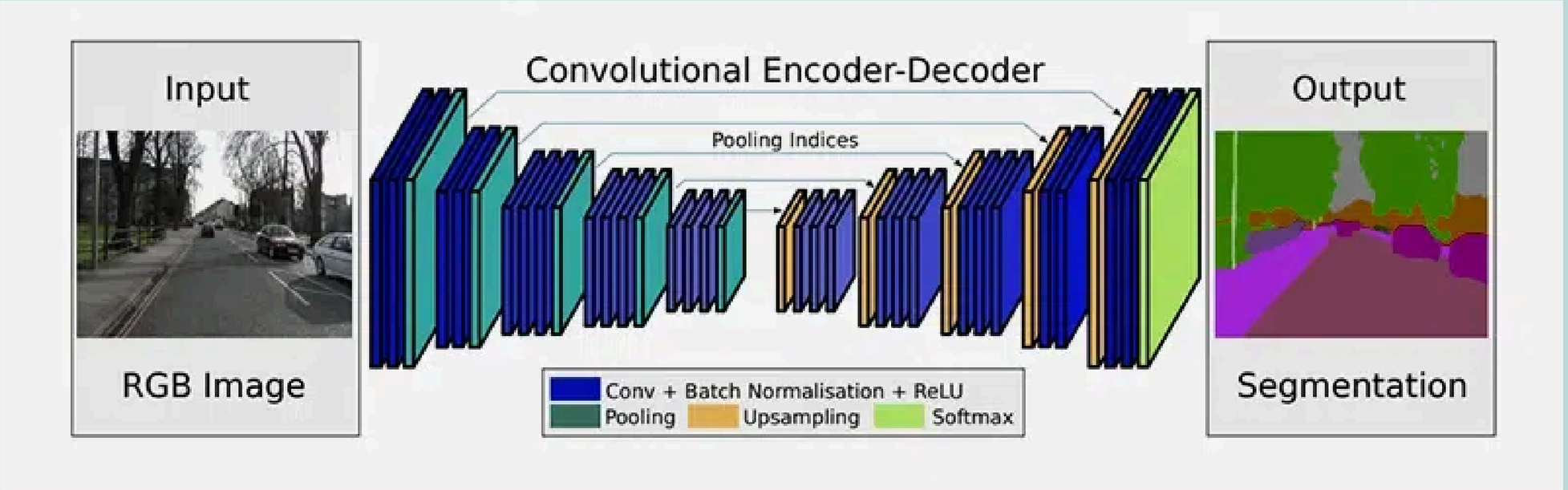


Summary

The base paper focuses on the semantic segmentation of RBCs in microscopic images to automate anaemia detection. It introduces a deep learning model based on an encoder-decoder CNN framework. The model segments individual RBCs and classifies them as anaemic or normal based on shape, size, and color features.



Architecture



- **Encoder**: Extracts features using convolution and pooling layers, reducing image size.
- **Decoder**: Upsamples features to original image size using transposed convolutions.
- **Skip Connections**: Link encoder and decoder layers to preserve spatial details.
- **Output:** Pixel-wise segmentation map classifying each pixel (e.g., anaemic or normal RBC).

Dataset Description

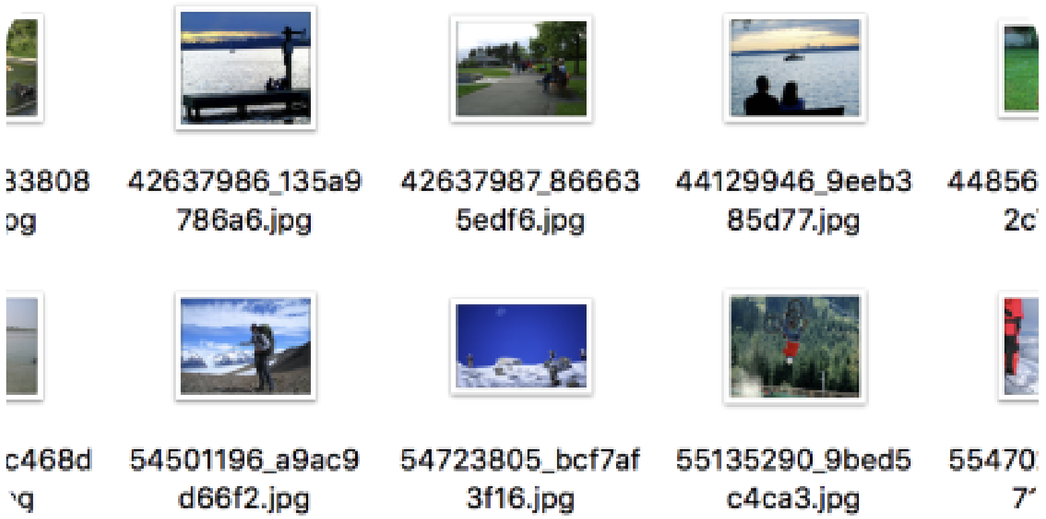
Purpose: Designed for anemia diagnosis using RBC (Red Blood Cell) images.

Classes: Two categories – Anemia and Normal.

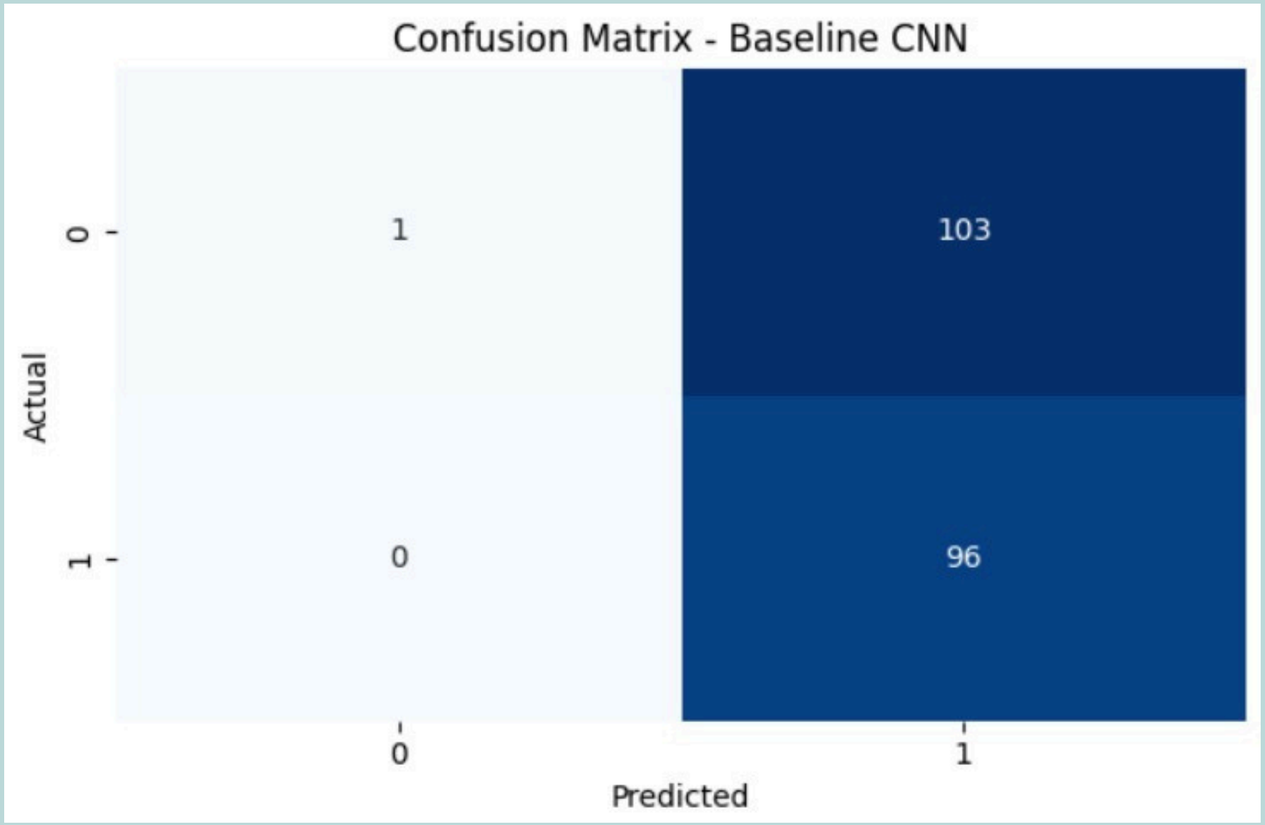
Data Type: Microscopic images of RBCs in .jpg format.

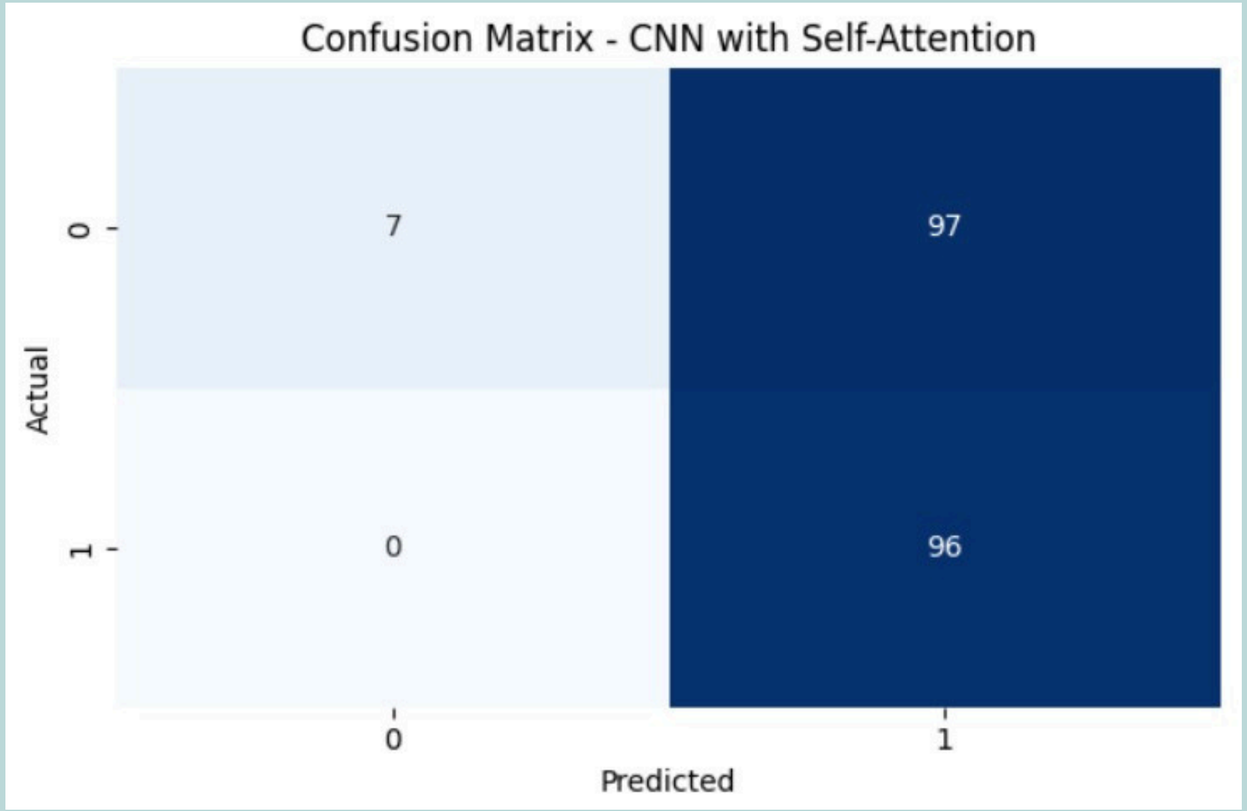
Usage: Ideal for binary image classification tasks in medical imaging.

Labels: Inferred from folder names (Anemia/, Normal/).

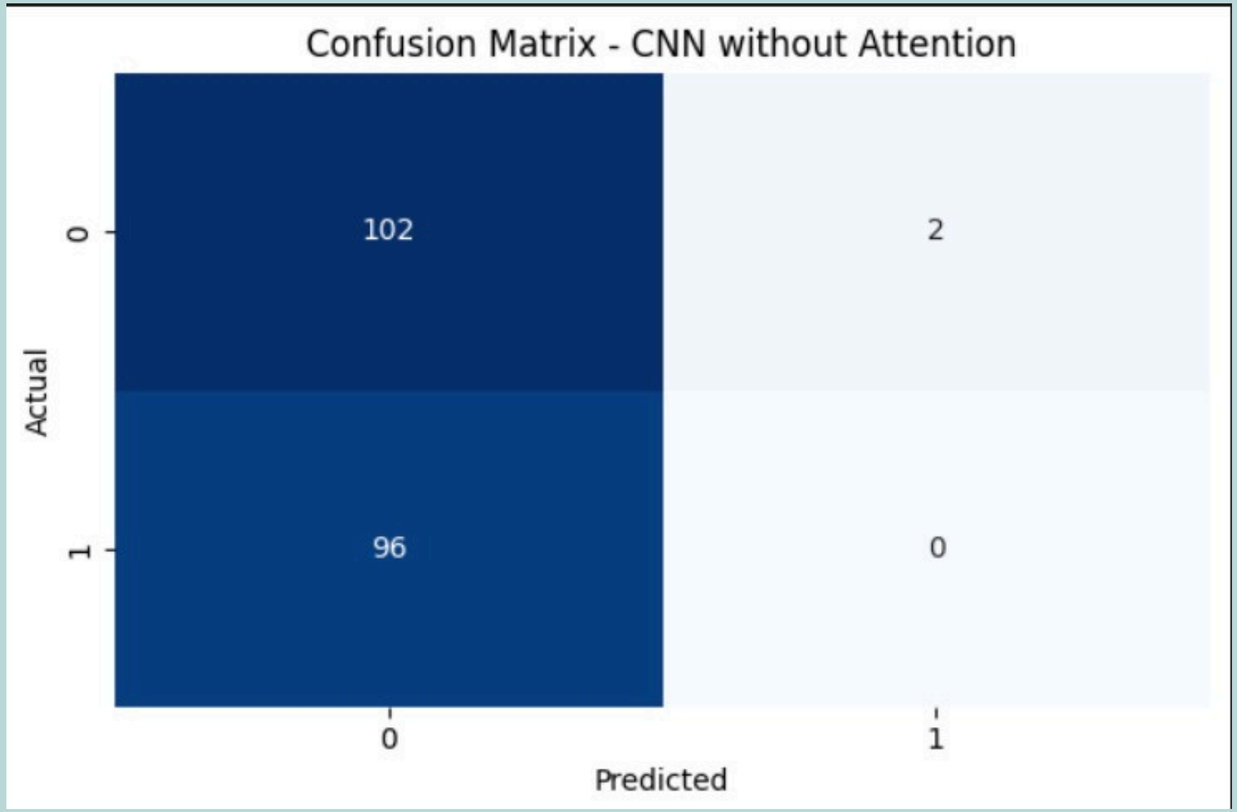


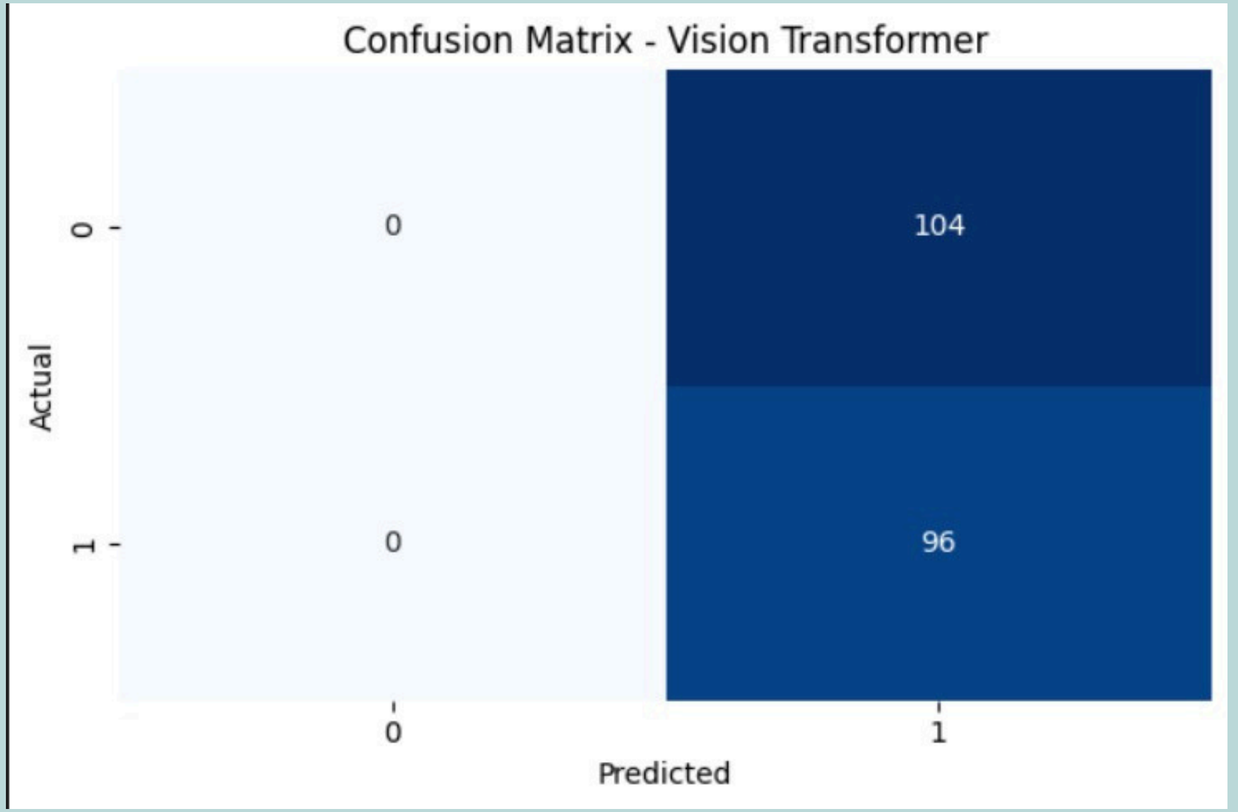
Performance Comparision





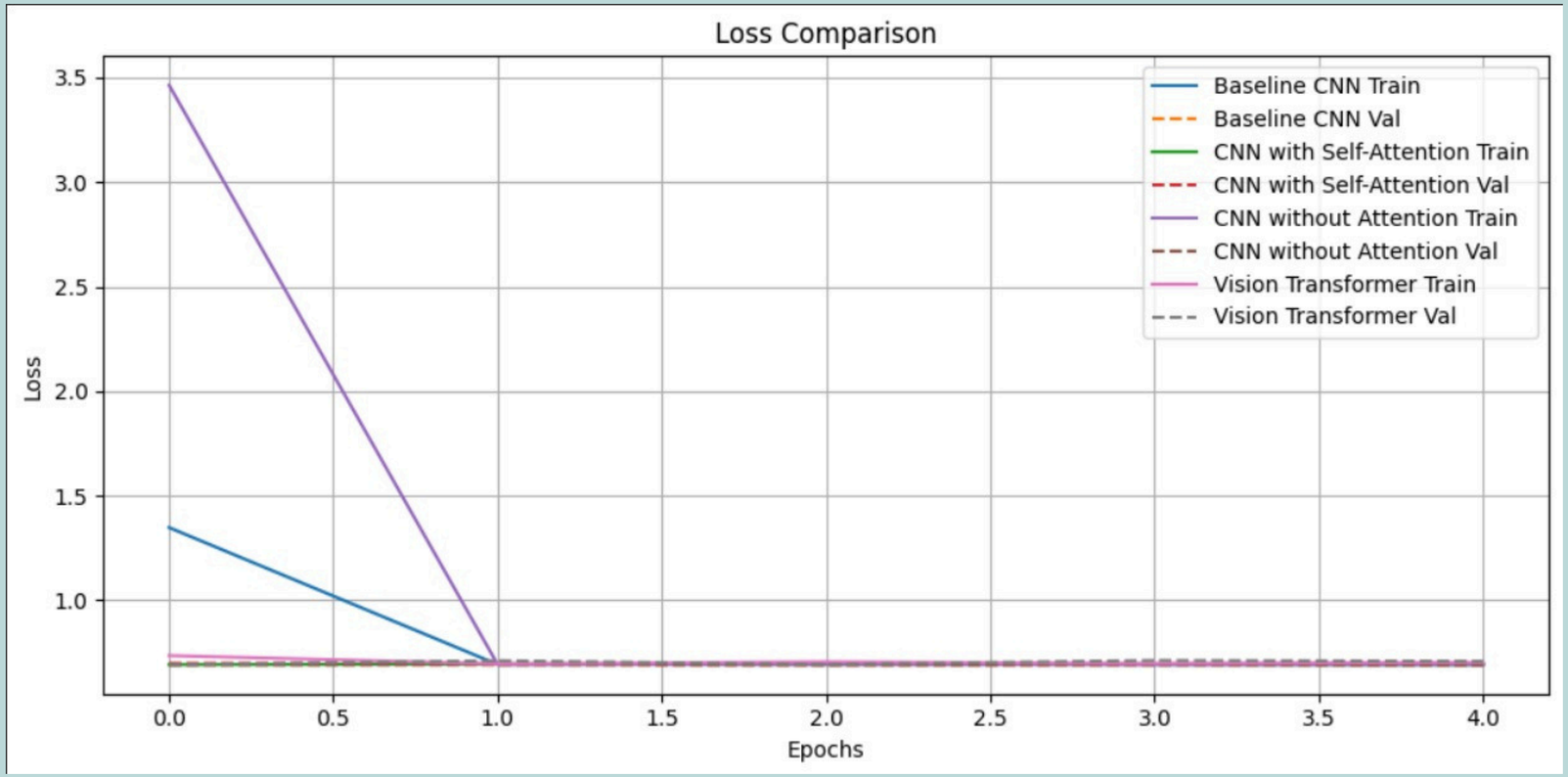
Performance Comparision





Graphs

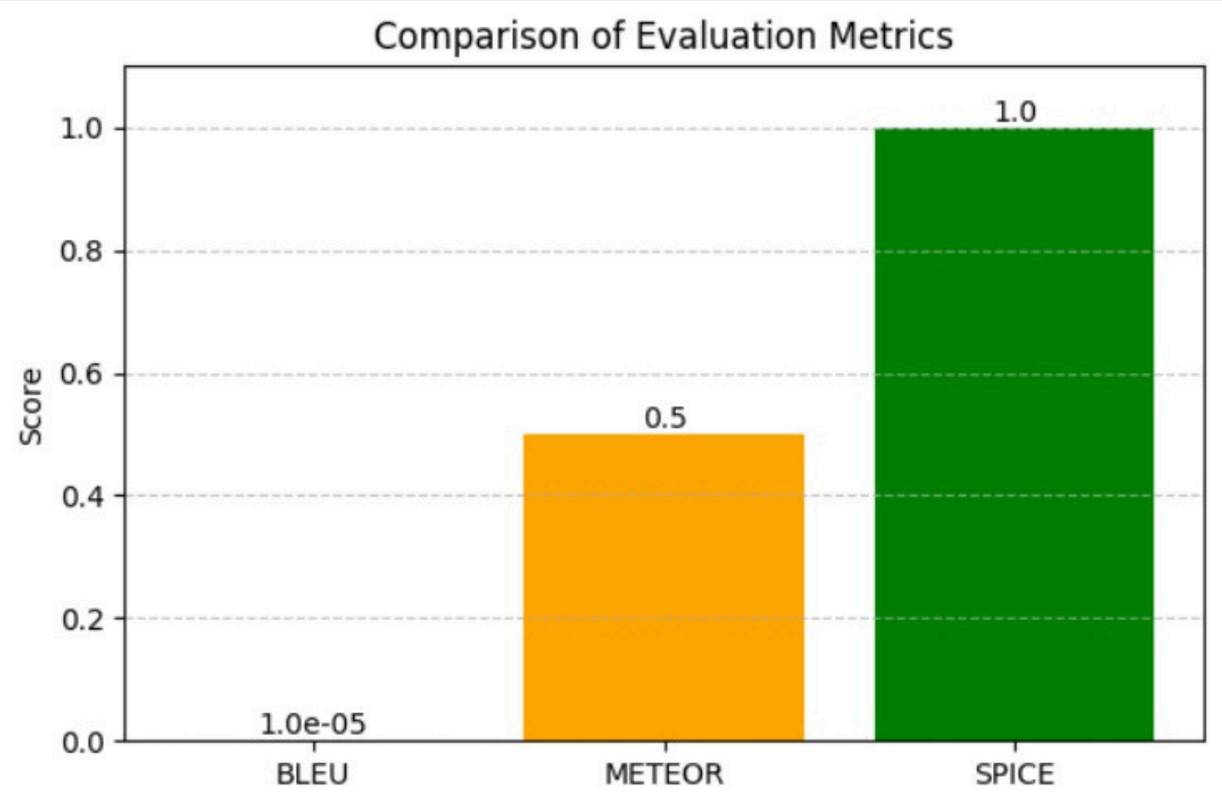
Training curves



Conclusion from Graph:

- Vision Transformer and CNN with Self-Attention outperform other models in terms of training stability and low loss.
- CNN without attention is the least stable initially.
- The attention mechanism significantly enhances learning efficiency and reduces loss early in training.

Comparison of Evaluation Metrics



Segmentation: Evaluated with IoU, F1 Score, Pixel Accuracy
Captioning: Evaluated with BLEU, ROUGE Transformer-based models performed best in capturing context and generating relevant outputs Efficient, accurate models suitable for clinical applications

Conclusion and Future work

CONCLUSION:

Encoder-decoder models efficiently segment anaemic RBCs and generate relevant image captions Transformer and Attention models offer higher accuracy and relevance

FUTURE WORK:

- Incorporate larger datasets
- Deploy hybrid models (CNN + Transformer)
- Enable real-time medical diagnostics on edge devices

Thank You!