Feature Fusion Ensemble Architecture with Active Learning for Microscopic Blood Smear Analysis

Jeevan Jamakayala

Machine Learning Engineer

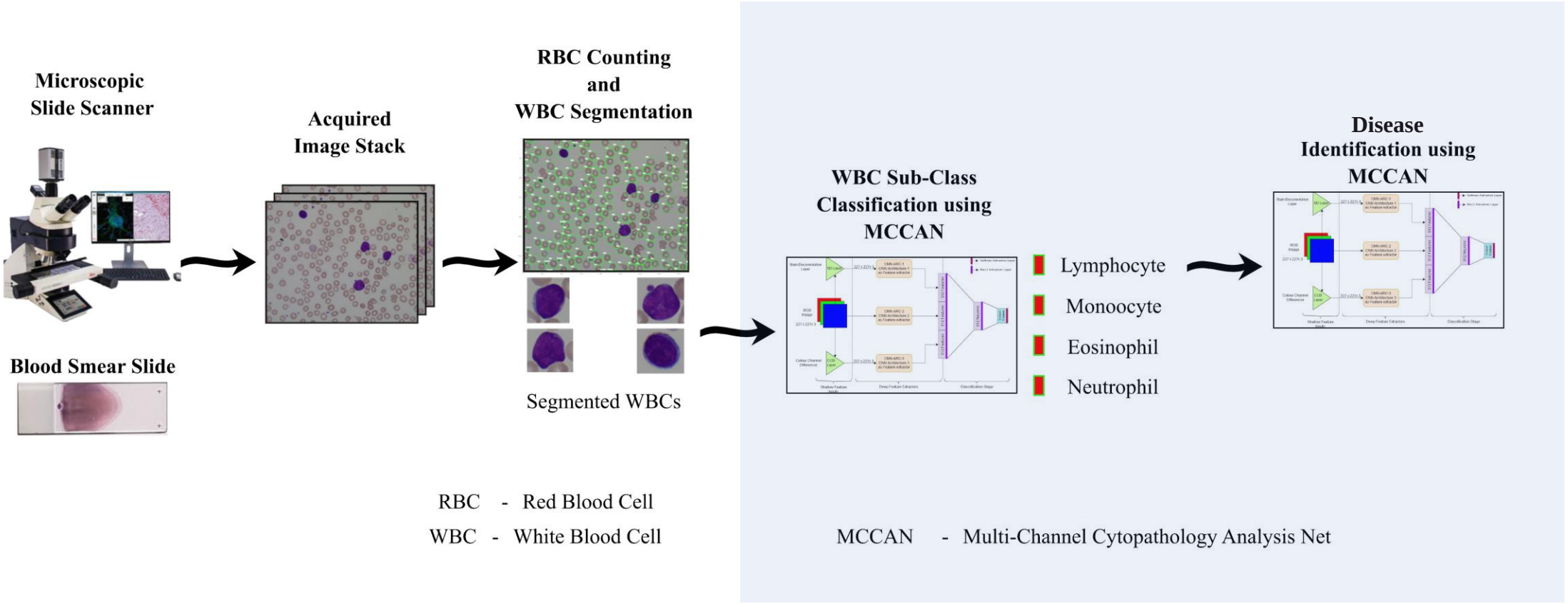
Exafluence Inc - India

Dr. Rama Krishna Sai Gorthi

Department of Electrical Engineering

Indian Institute of Technology Tirupati (IITT) - India

Overview of our Work



RBC - Red Blood Cell

WBC - White Blood Cell

MCCAN - Multi-Channel Cytopathology Analysis Net

Introduction

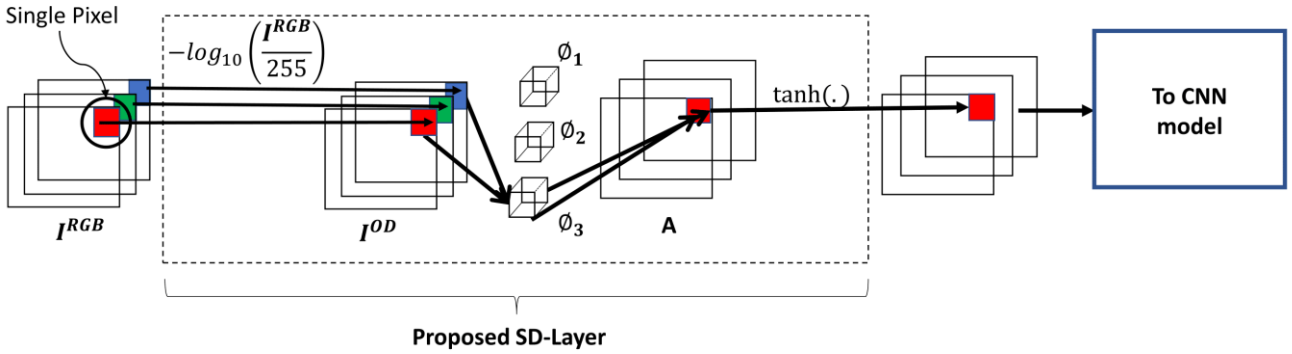
- We propose a lightweight convolutional Neural Network (CNN) model called **Multi-channel Cytopathology Analysis Network (MCCAN)** for cell classification or disease identification from the blood smear images after conventional cell segmentation
- We employ Active Learning to train the model to overcome the labeling cost and to produce better accuracies with less data samples
- Following are two main problem statement we tried to address using our architecture:
  - Classification of White Blood Cells (WBC) into sub-classes
  - Identification of malignant B-cells from normal pre-cursor B-cells

Related Work

- There has been a lot of efforts for WBC classification & disease identification as well to come up with novel methods which provide high accuracy
- In many approaches, features are being extracted by using different image processing techniques, which are used to train the neural networks as a classifier for identifying malign cells.
- In other classes of methods, with the help of Transfer learning, the pretrained deep networks are being employed to achieve higher accuracy. However, the lack of sufficient data for training is a key bottleneck for the evolution of dedicated deep networks for this application
- Effective feature extraction strategies from RGB cell images, before applying the transfer learning is also an active area of research to combat this limited training data

Related Work

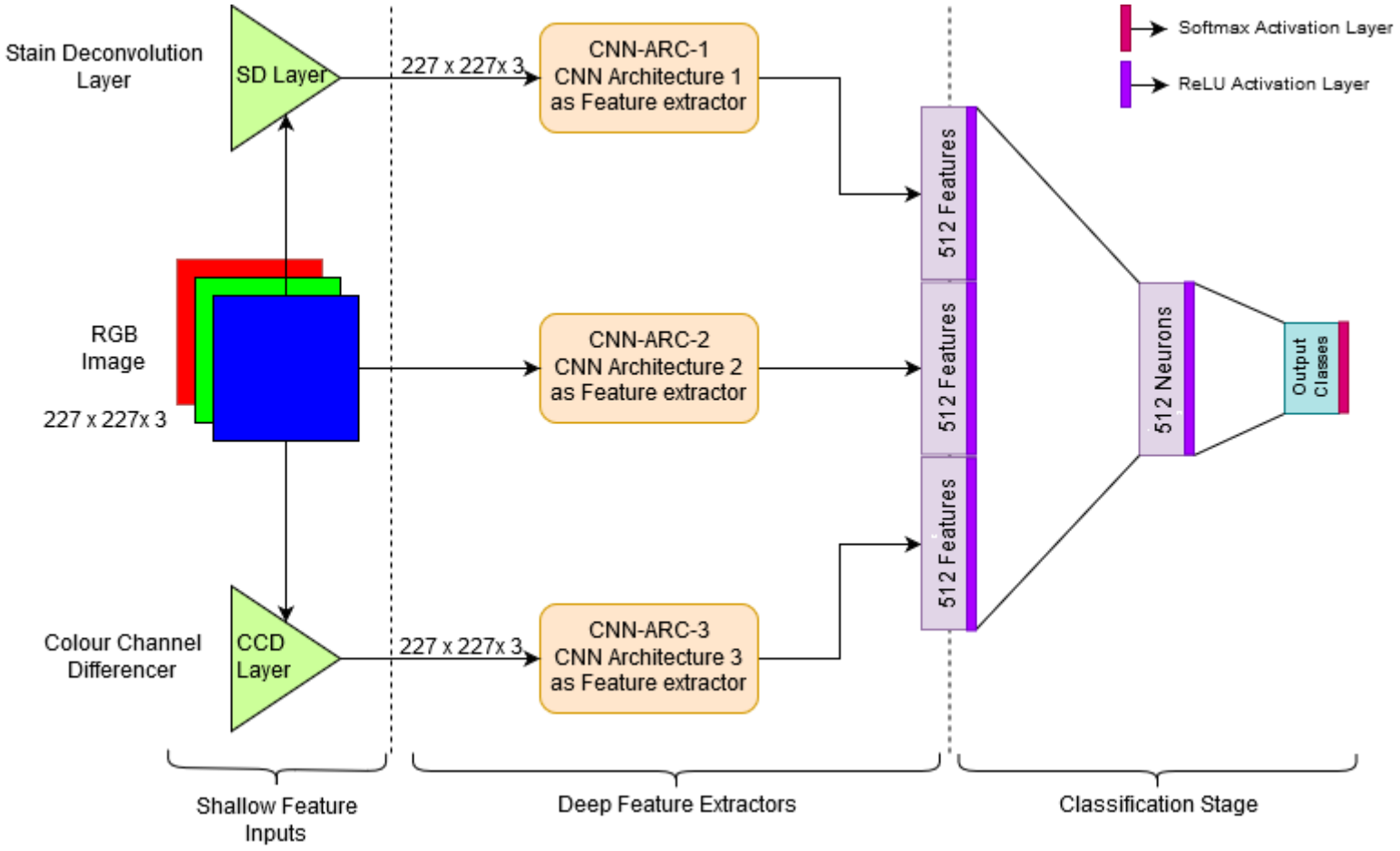
- In "***SD-layer: Stain deconvolutional layer for CNNs in medical microscopic imaging***" paper at MICCAI 2017, conversion of RGB images into Optical Density space using stain deconvolution layer (SD-Layer) then training on Alex net to extract high-level feature extraction & classification is proposed
- However, to combat less sample-based learning problems Active Learning is developed
- With the help of Active Learning, one could eliminate labeling more similar data samples, reducing the cost of labeling and effort



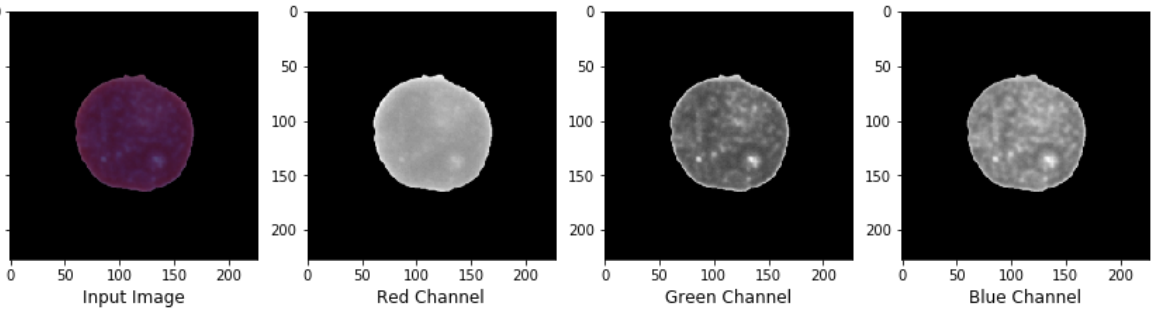
Proposed Approach

- There are two important steps in the proposed approach
  - Multi-channel Cytopathology Analysis Net (MCCAN) is a light-weight CNN model with three shallow feature input layers. Our aim is to build a **generalized simple yet effective architecture**
  - Use of Active learning based training of the proposed MCCAN architecture

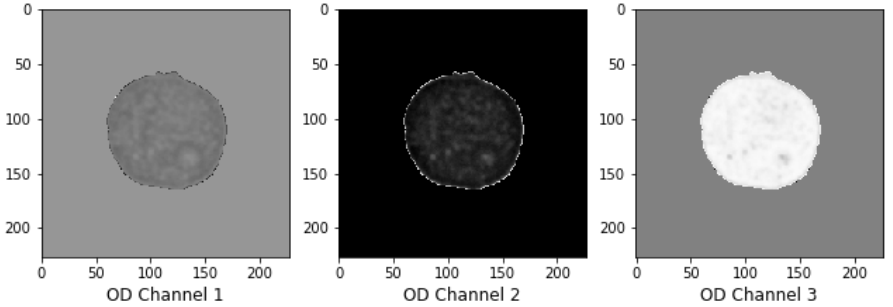
Proposed Approach - MCCAN



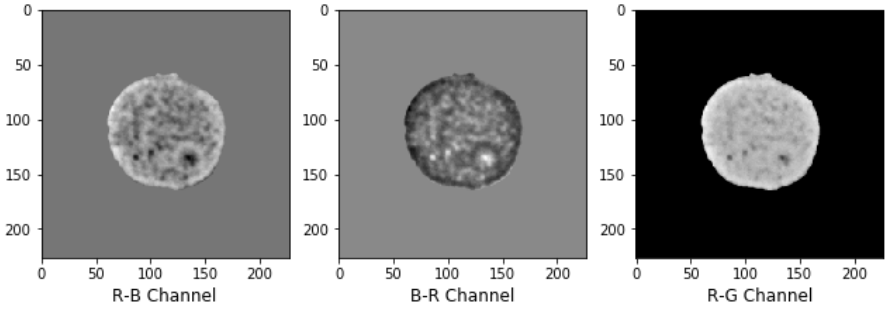
Proposed Approach – Shallow Feature Input Layer



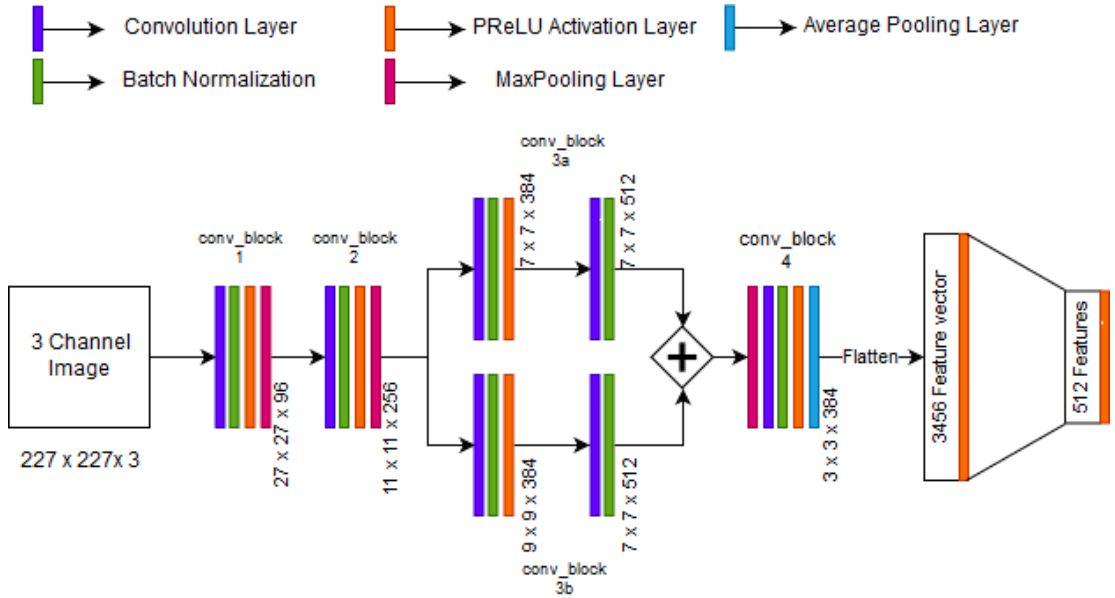
Stain Deconvolution (SD) Layer Output



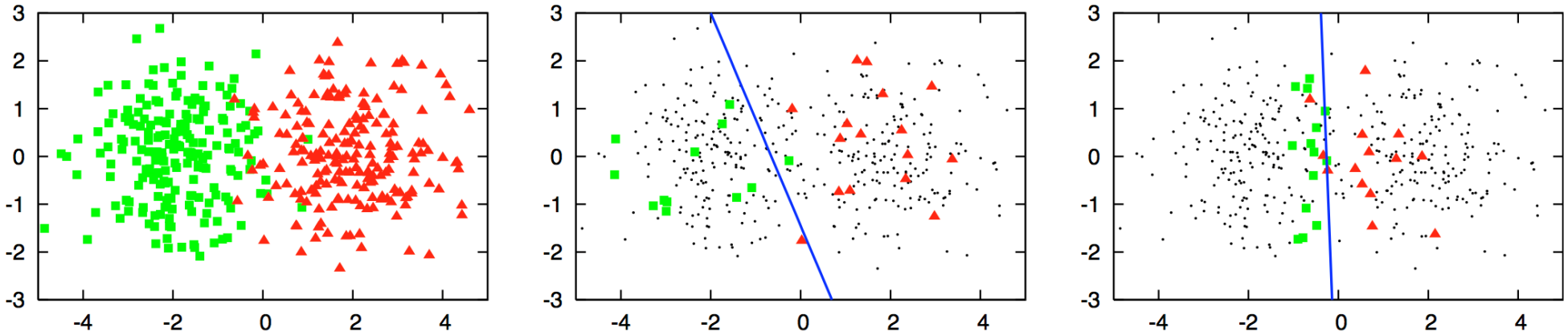
Colour Channel Differencer (CCD) Layer Output

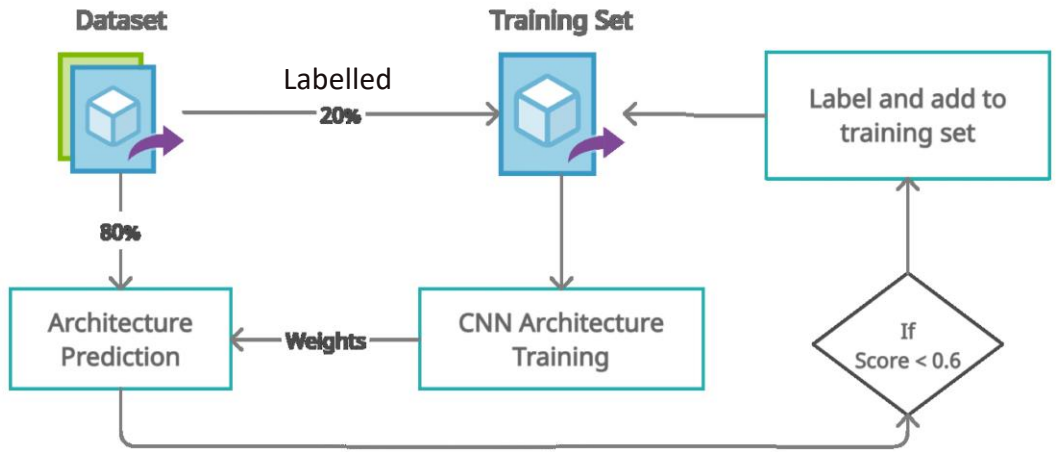


Proposed Approach – CNN Feature Extractors



Proposed Approach – Active Learning





Results – WBC Sub-class Classification

- The base model (CNN-ARC-2) produced an accuracy of 86.87% after 20 epochs of training. While, the same base model with the implementation of Active learning achieved an accuracy of 89.84% with just 28% of the whole dataset
- The proposed architecture (MCCAN), with all 3-stages, when trained through Active learning for 130 epochs and 30% of data results in a much higher accuracy of 96.34% in WBC sub-class classification than ResNet v152 with fine-tuning the last layers

Architecture	Epochs	% Train Data	Accuracy
ResNet v1 152	3000	100	87.46%
MCCAN	130	31.70	96.34%

Results – WBC Disease Classification

- In comparison with Multi-stream Inception ResNetV2, our ensemble-based approach can achieve higher accuracy of 85.80% with small networks and active learning while operating on different shallow features

Architecture	Accuracy
Multi stream Inception ResNet v2	83.35%
MCCAN	**85.80%**

- Performance of our proposed architecture with Active learning

No. of Training Samples	% Train Data	Test Accuracy
6000	8.77	75.06%
30438	44.51	82.91%
34471	**50.40**	**85.80%**
36734	53.72	84.21%

Conclusions

- The proposed Multi-channel Cytopathology Analysis Net (MCCAN) along with Active learning comes in handy to analyze stained microscopic images
- It leverages information from different domain-specific features from RGB space and Optical Density space, to analyze the microscopic images better
- Active learning, in this framework, helps to achieve higher accuracies with less labeled data and hence provides a cost-effective solution to blood smear analysis
- The generality and compactness enable its users to perform multiple classification tasks in Cytopathology with the same architecture

THANK YOU