Chinese Sign Language Classification

Nada Abdellatef, Hadeer Mamdouh, Khaled Elsaka, Mostafa Nofal
University of Ottawa

Introduction

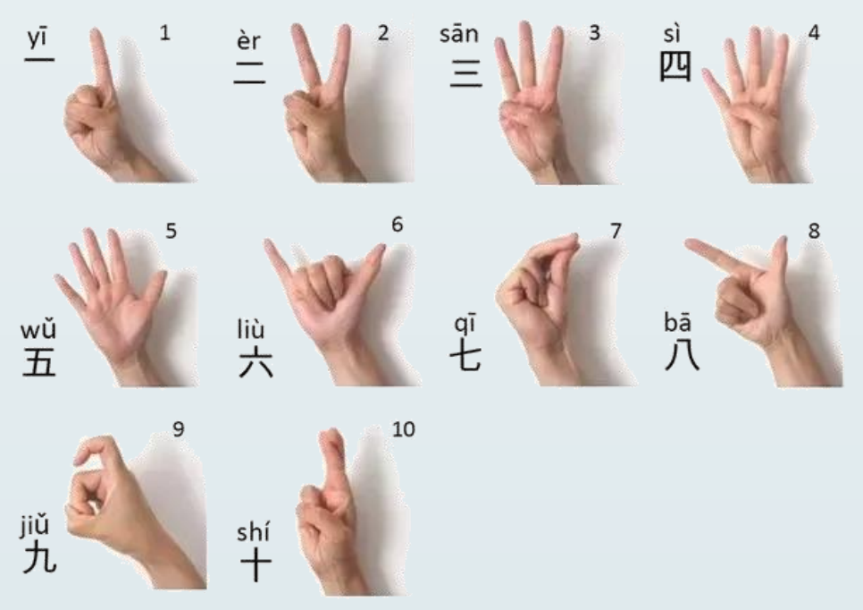
Nowadays, about 5% of the population suffers from the issue of hearing loss in the world. There is also a large number of hearing-impaired people in China, number of which 27.9 million. Sign language is the easiest form of communication for people who have trouble hearing,. Sign language recognition is an important and difficult task when it comes to converting sign language into text.

Problem Formulation

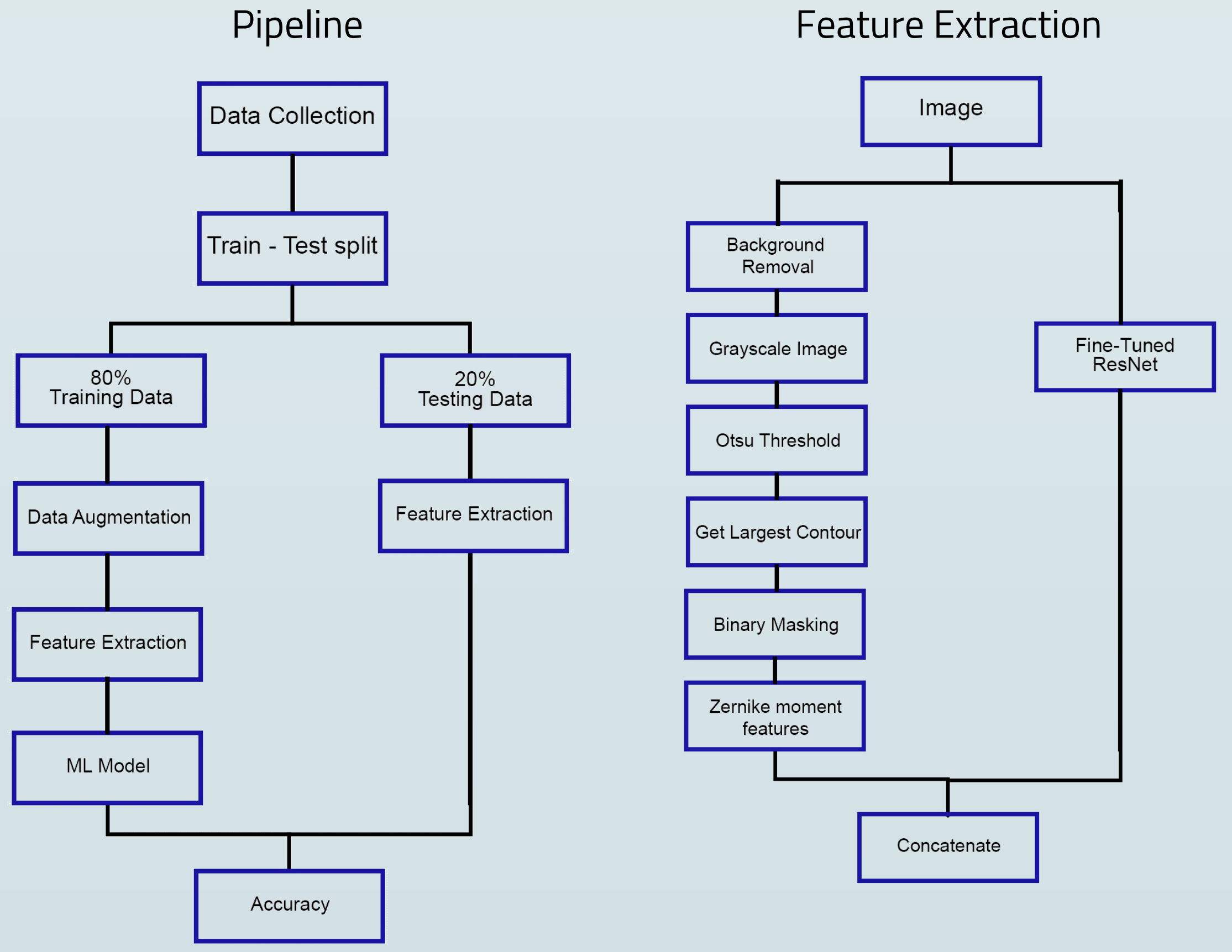
The current sign language recognition methods are broadly divided into two types: the traditional machine learning method using image features and the deep neural network-based method. Other one uses machine learning techniques to perform classification based on the image features of the interested area. The former uses traditional image segmentation algorithms to segment hand shapes from sign language images or video frames of sign language video.

Dataset

This dataset consists of 500 RGB images of Chinese numbers in finger sign language. The Chinese numbers finger sign language consists of 10 signs in accordance with the state-issued universal sign language standard. Figure 1 demonstrates categories of Chinese finger sign language intercepted from sample images. These samples were preprocessed and normalized to 128 × 128 background-removed images. Our experiment was executed with this private dataset including 500 images. Among them, 400 images were used for training, and the rest were used for testing.



Methodology



Methodology (cont.)

Split the images into training and testing 80:20 respectively. Then, Using transfer learning to fine-tune the pre-trained ResNet model by some hidden layers for the classification task compute the accuracy of the test data to represent our baseline. Combining different feature engineering techniques:

1- We know that the convolution layers act as feature extractor, so we collect the features from the last convolution layer, then adding 1024 (1*1) convolution layer to reduce the dimensions and by applying global average pooling we got 1024 feature vector as output.

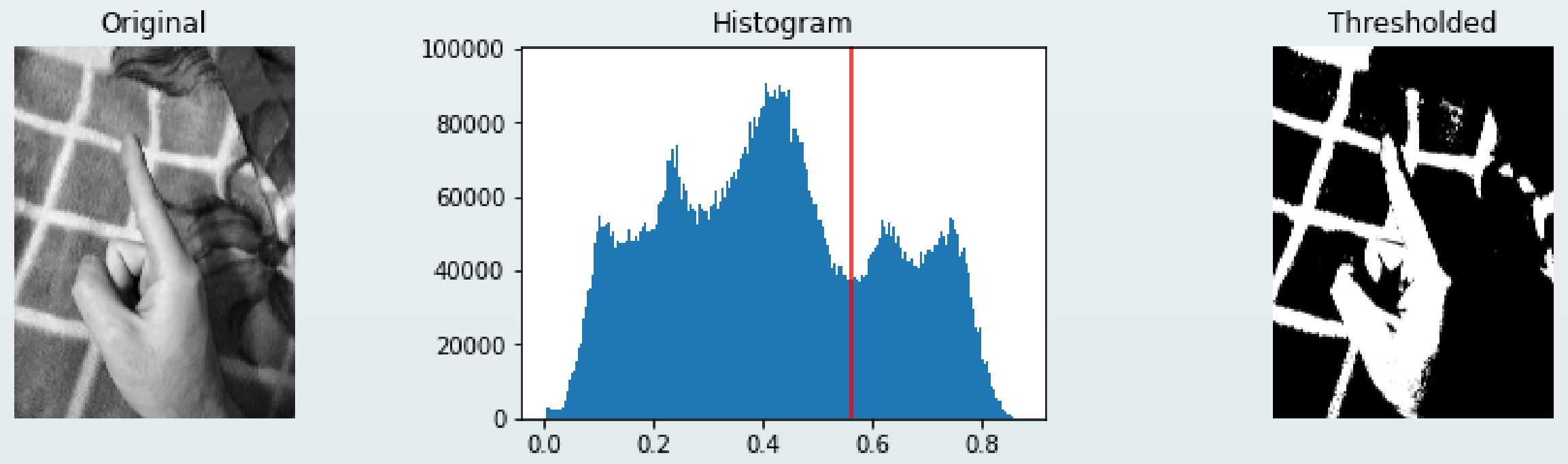
2- Using The Zernike moments after applying a binary mask on the images for hand pixels to be 1 and 0 otherwise. Zernike moment requires 2 important parameters the first one is the radius and the other one is the degree of polynomial. The radius is for a circle surrounds the target object and the radius to represent the shape.

After concatenating the two methods we feed them into ML models which are linear SVM, RBF kernel SVM, Random Forest, MLP, Naïve Bayes, and XGBoost. Selecting the champion model based on the average between the training and testing accuracy and comparing the results with the baseline performance.

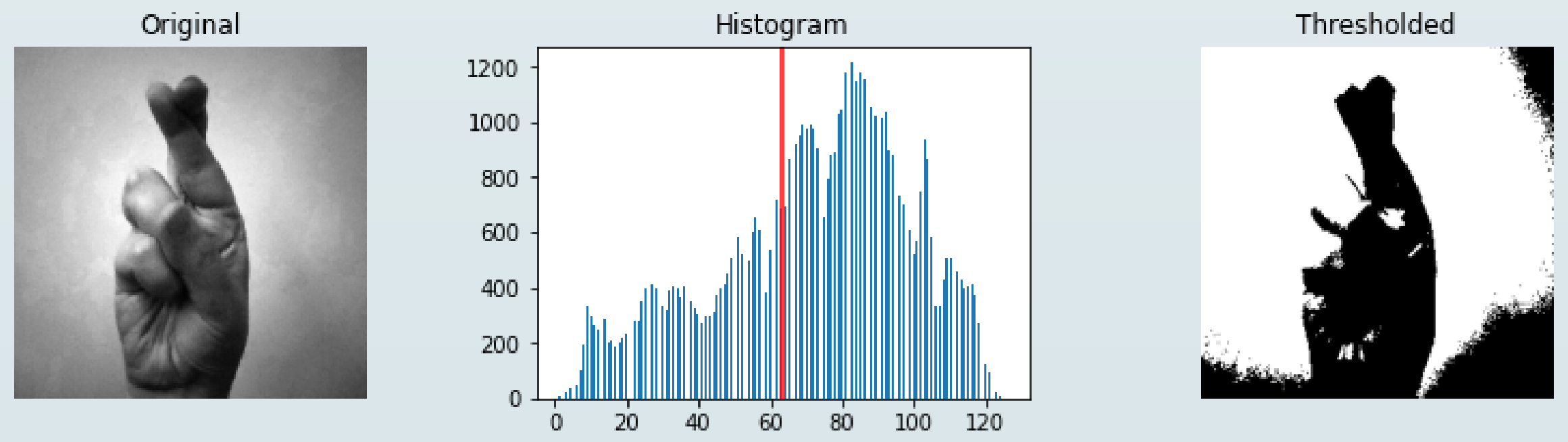
Challenges

Selecting the threshold for the binary images (applied thresholding using Otsu):

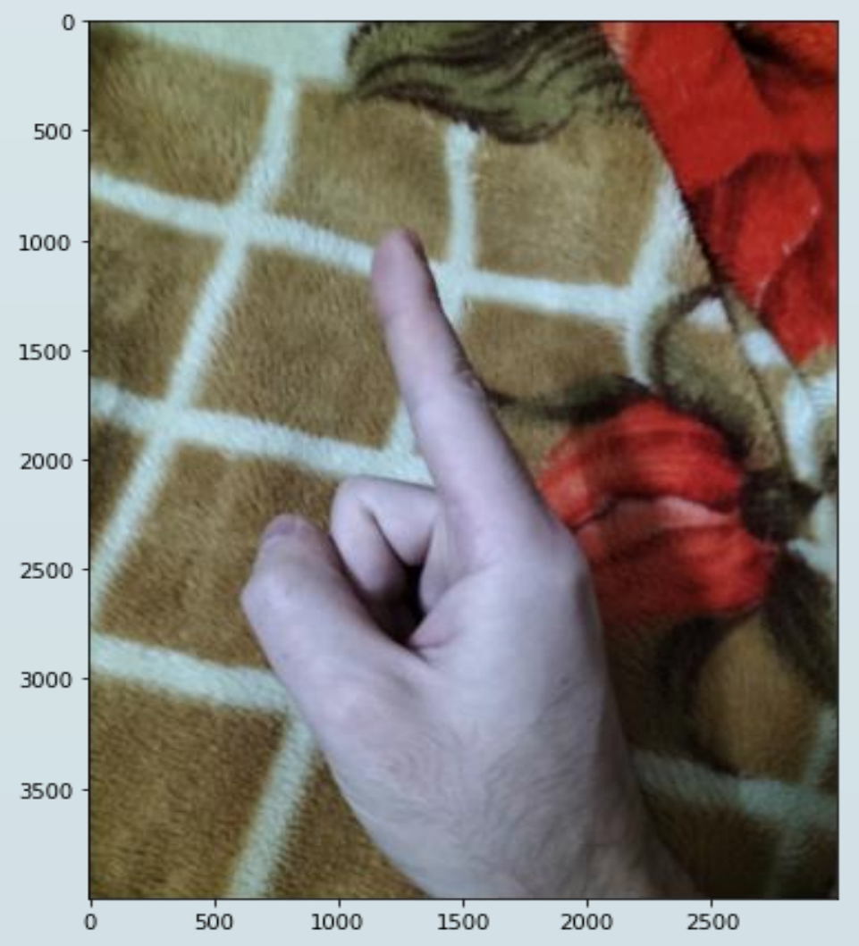
- Noisy images.



- Different lightening

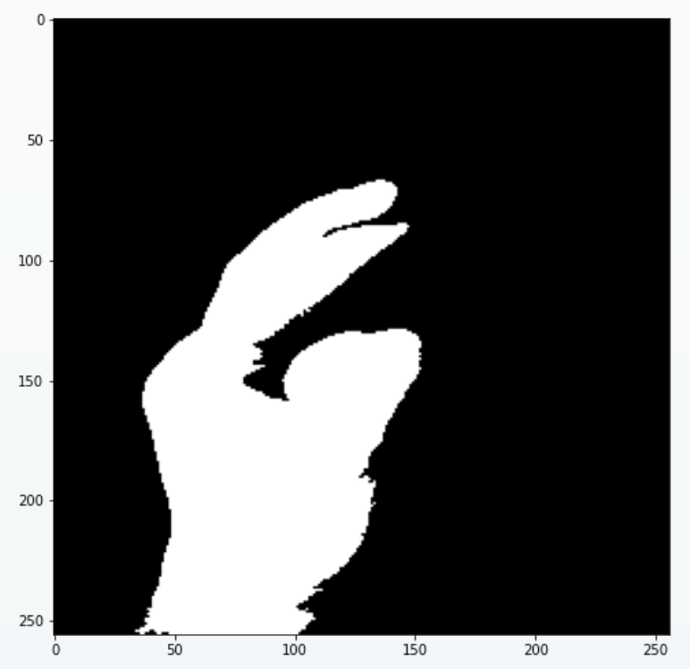


- Solving this problem by the rembg library to remove the background from the images.



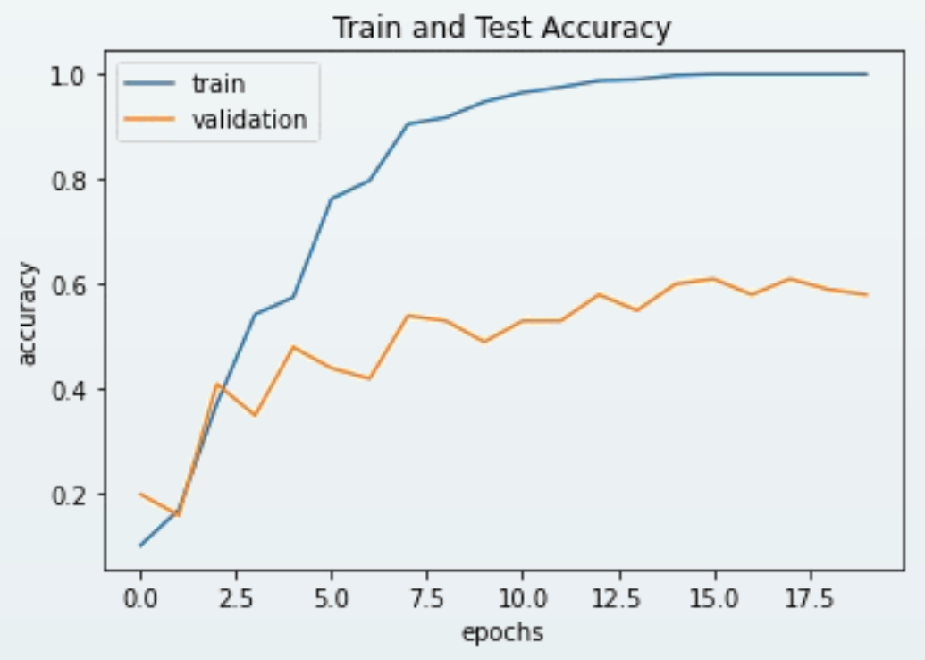
Challenges (cont.)

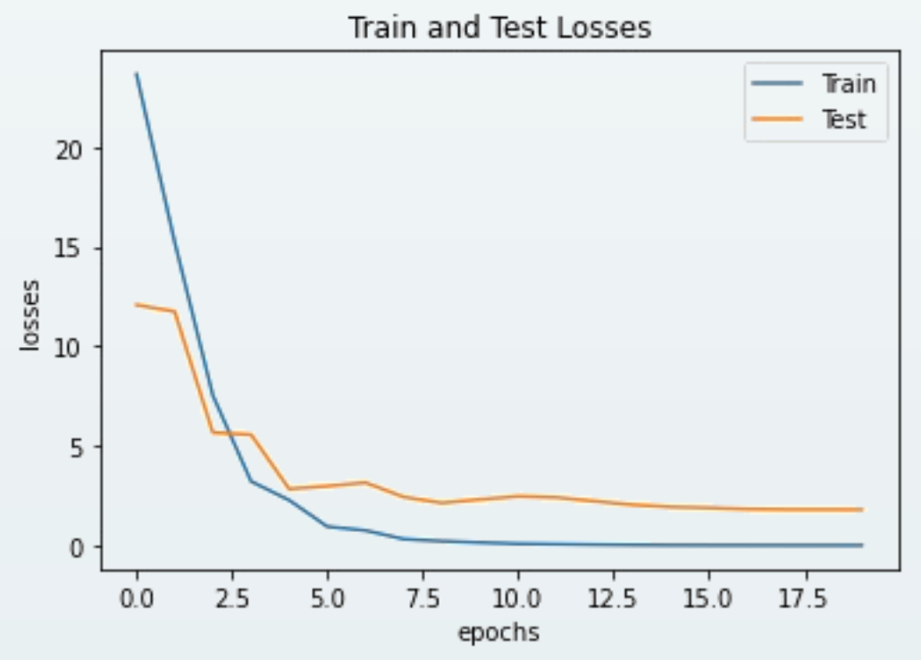
- But it doesn't do well, if the size of the image is small and the image is not clear and noisy. After removing the background it's much easier to prepare the images for the Zernike.

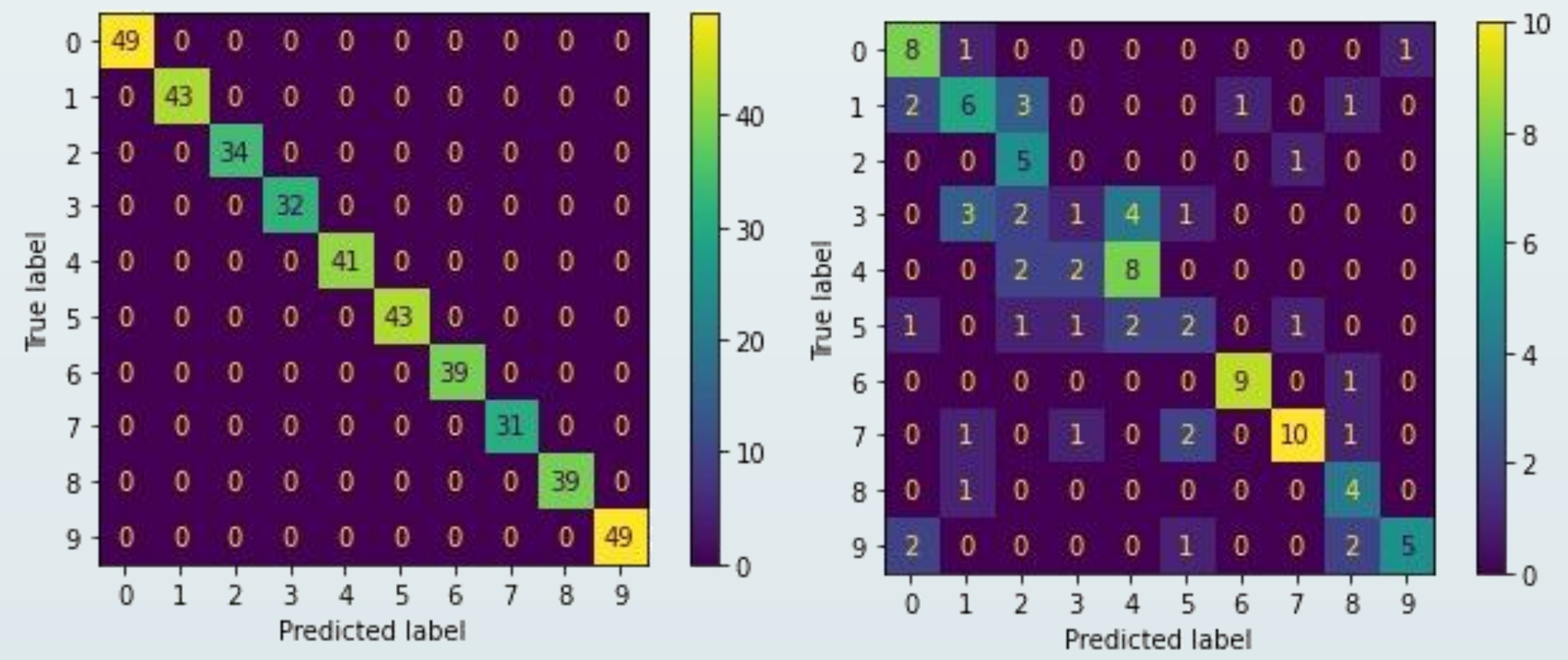


Evaluation

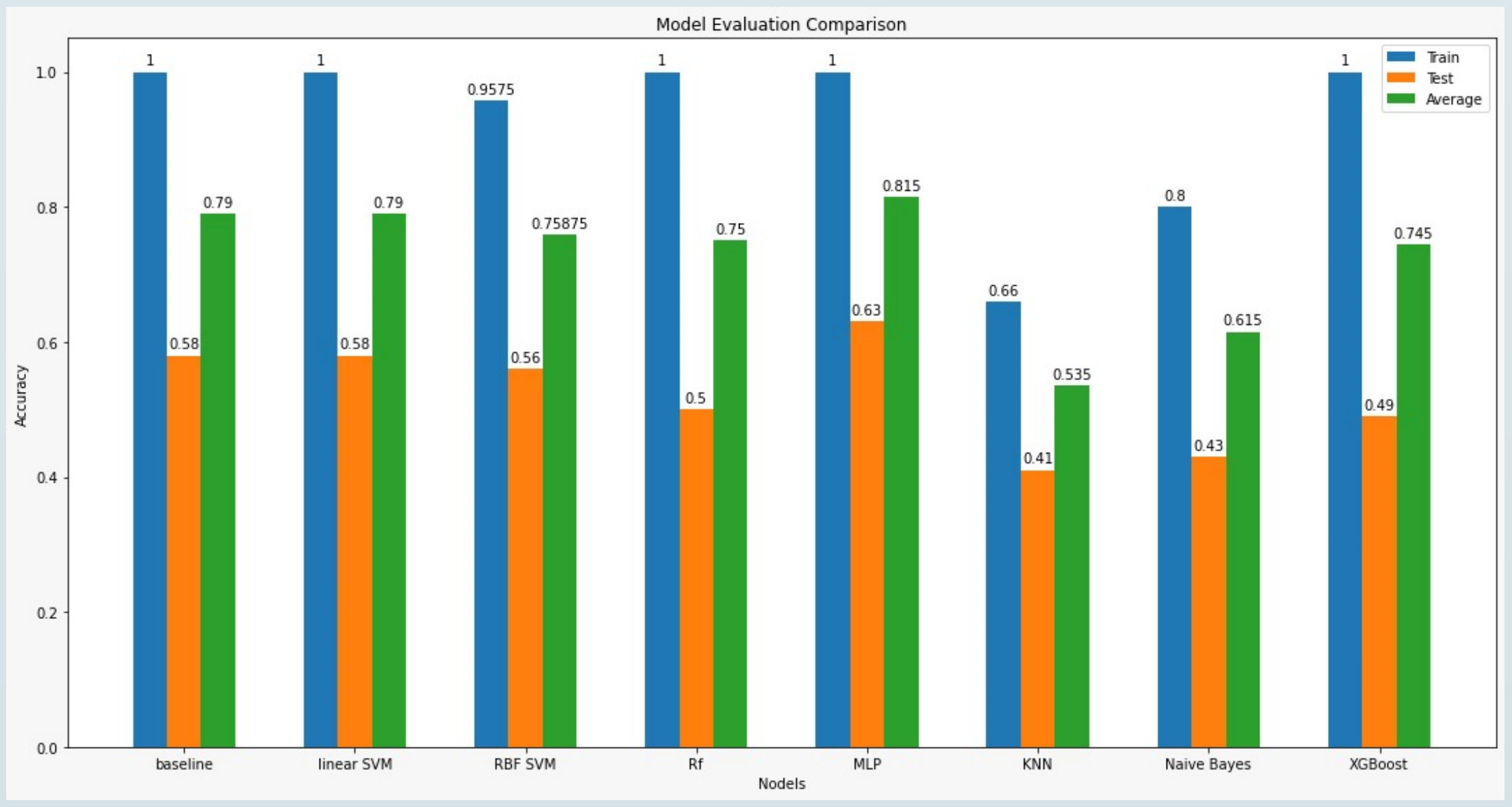
- Here are the baseline losses and accuracies







- Accuracy Comparison of the Models



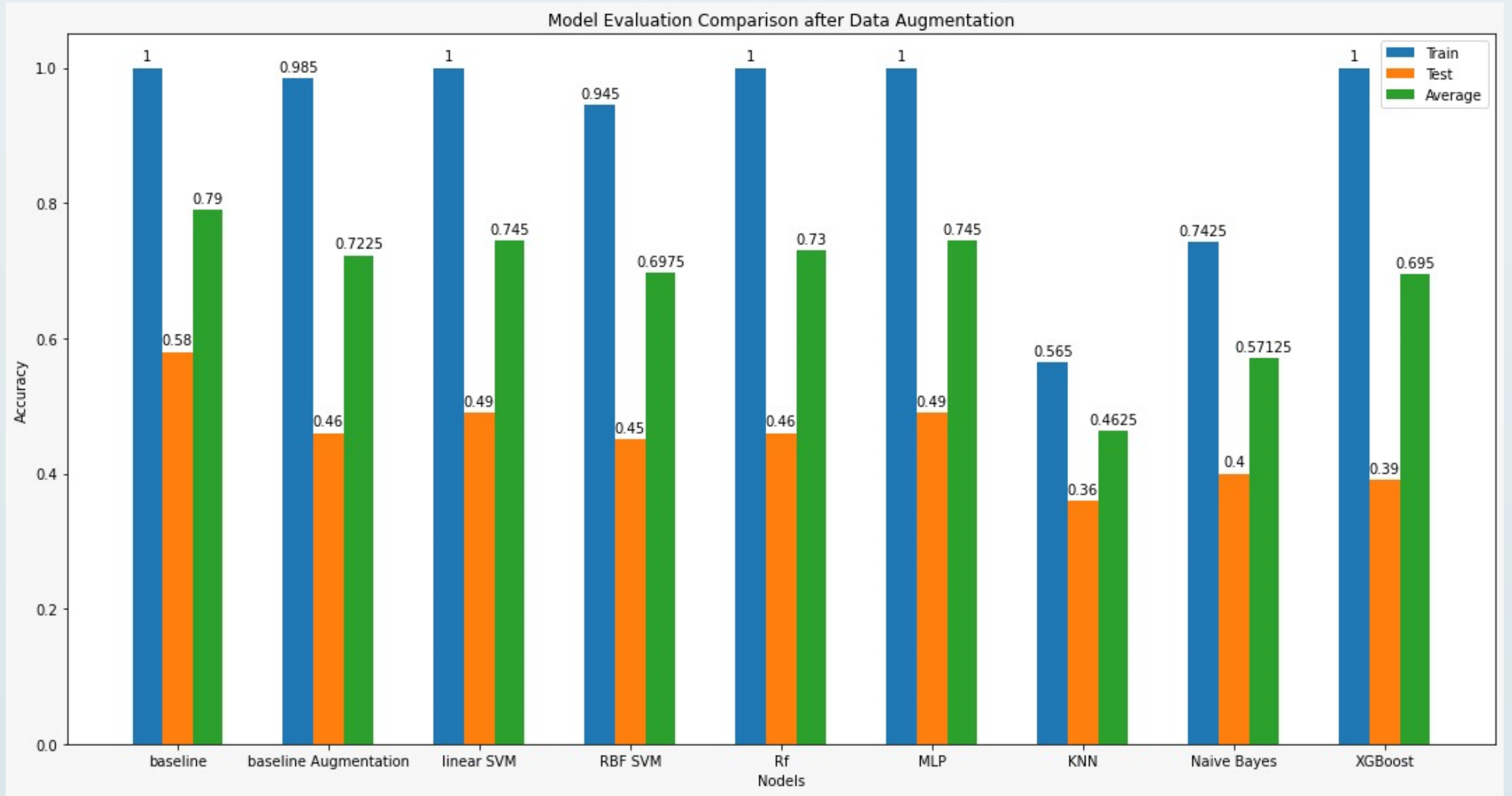
-Model-base, Linear-SVM, Random Forest, MLP, and XG-Boost have the same training accuracy rates as MLP, which is our champion model is with the highest accuracy of 63% accuracy compared with the other models

Evaluation (cont.)

- We tuned the hyperparameters and found that the accuracy reached 64% and the best parameters are as follows:
- ('activation', 'logistic')
- ('alpha', 0.003)
- ('learning_rate', constant')
- ('learning_rate_init', 0.0001)
- ('max_iter', 500)

- Photos sample visualization after data augmentation

- Accuracy Comparison of the Models with data augmentation



Conclusion

Combining the two feature engineering techniques from the ResNet pre-trained model and Zernike moments results in improving performance over the baseline with a training accuracy of 99% and testing accuracy of 64% for the tuned best model MLP. The model is overfitting and that is because the training data is very few. Although we've tried data augmentation, the results didn't meet our expectations so, we plan to increase the size of the images to have better results from the rembg library for background removal. Also, building the MLP using Keras instead of Sklearn to have more flexibility and solve the overfitting problem.