

# Finding Optimized Machine Learning Model For Recognizing English Handwritten Digit

Nowfel Mashnoor Amir Faruk

Rajshahi University of Engineering and Technology

June 22, 2018

# Outline

- Search previous works on this topic
- Pick two classifier for comparison
- Preparing Datasets for both training and testing
- Find Optimal Hyper parameters for the classifiers
- Train the classifiers using Scikit Learn Library
- Evaluate the classifiers
- Exporting the scores of the classifiers

# Introduction

Handwritten Digit Recognition has been very successful in recent years. A lot of research and studies has been done in recent years on it like Devnagari Handwritten Character Recognition. Handwritten digit recognition technique is used in various fields like PDA, bank cheque, handwritten fields in form etc. There are many classifiers for recognizing Handwritten Digit. In our research we are going to compare among Random Forest and Artificial Neural Network. We will chose the best classifier among them based on their accuracy on testing set.

# Objective

Compare the performance and accuracy of Artificial Neural Network and Random Forest classifier on MNIST Data Set and evaluate their scores.

# Literature Review

A comparison study has been already done where *Base Linear Classifier*, *Baseline Nearest Neighbor Classifier*, *Large Fully Connected Multi-Layer Neural Network*, *Tangent Distance Classifier(TDC)*, *LeNet 4 With KNN*, *Optimal Margin Classifier* are compared among.<sup>1</sup> But it doesn't include basic ANN and Random Forest. In our work, we have basically compared this two.

---

<sup>1</sup>Yann LeCun et al. "Learning algorithms for classification: A comparison on handwritten digit recognition". In: *Neural networks: the statistical mechanics perspective* 261 (1995), p. 276.

# Methodology

- Collected 28x28 Grey Scale Converted MNIST Data From a Website
- Processing Data
  - ▶ Extracted 60000 Training Data
  - ▶ Extracted 10000 Testing Data
- Made Instance of Classifier from SKLearn
- Fed the training data to the classifiers

# Data



Figure: MNIST Dataset Sample Images

# Result Analysis

Table: Mean Accuracy

Artificial Neural Netwrok	Random Forest
0.875087508750875	0.9478947894789479

$$CM = \begin{bmatrix} 965 & 0 & 1 & 1 & 1 & 2 & 3 & 1 & 5 & 0 \\ 0 & 1118 & 3 & 2 & 1 & 2 & 2 & 1 & 6 & 0 \\ 6 & 4 & 984 & 6 & 6 & 1 & 4 & 12 & 5 & 4 \\ 1 & 3 & 11 & 958 & 1 & 13 & 0 & 7 & 10 & 6 \\ 5 & 2 & 4 & 0 & 937 & 2 & 4 & 4 & 5 & 19 \\ 6 & 1 & 2 & 26 & 5 & 836 & 4 & 1 & 4 & 7 \\ 12 & 6 & 0 & 0 & 4 & 11 & 920 & 1 & 4 & 0 \\ 5 & 6 & 25 & 9 & 4 & 0 & 0 & 962 & 3 & 14 \\ 14 & 1 & 13 & 29 & 10 & 21 & 8 & 6 & 863 & 9 \end{bmatrix}$$

Random Forest Confusion Matrix



## Result Analysis

$$CM = \begin{bmatrix} 899 & 0 & 43 & 2 & 0 & 23 & 8 & 1 & 2 & 1 \\ 0 & 1096 & 2 & 2 & 0 & 1 & 1 & 1 & 30 & 2 \\ 13 & 3 & 857 & 3 & 8 & 43 & 25 & 7 & 53 & 20 \\ 1 & 5 & 9 & 900 & 0 & 55 & 0 & 7 & 33 & 0 \\ 0 & 0 & 10 & 0 & 875 & 0 & 19 & 4 & 5 & 69 \\ 21 & 4 & 52 & 142 & 0 & 626 & 5 & 0 & 38 & 4 \\ 18 & 1 & 38 & 0 & 20 & 3 & 871 & 0 & 5 & 2 \\ 1 & 31 & 3 & 1 & 2 & 0 & 1 & 924 & 19 & 46 \\ 4 & 14 & 37 & 18 & 6 & 13 & 0 & 7 & 860 & 15 \\ 1 & 6 & 8 & 5 & 66 & 14 & 2 & 40 & 25 & 842 \end{bmatrix}$$

Confusion Matrix for Artificial Neural Network

# Future Work

- Improve hyper parameters
- Try other variants of Neural Network (CNN, RNN etc)
- Compare Other Classifiers

# Conclusion

So from our we can say that Random Forest Classifier performed better than ANN. But there could be better hyper parameters for which ANN would outperform RF.