ARABIC FONT RECOGNITION SYSTEM

Team 4 – CMPS450

CMPS451 – Big Data

Team Members

Ahmed Emad Reda	1190180	Hla Hany Mohamed	1190344
Nada Tarek	1180504	Ziad Ahmed Hamed	4200002

Problem Description

Objective: Develop a system to classify Arabic paragraphs into four font categories based on input images.

Challenges: Characterizing unique font features; handling variations in writing styles, spacing, and noise.

Font Code	Font Name
0	Scheherazade New
1	Marhey
2	Lemonada
3	IBM Plex Sans Arabic

Project Pipeline

Data Loading & Splitting → Preprocessing → Feature Extraction → Feature Selection → Model Development → Hyperparameter Tuning → Prediction

Preprocessing Module

Median Filter

Image Sharpening

Conversion to Binary

Image Deskewing

Image Resizing

Feature Extraction & Selection Modules

HOG

SIFT

Standardization

PCA

Unsuccessful Trials

Edge Direction Matrix (EDM):

- **Features from EDM1 & EDM2:**
  - ✓ Edges Direction
  - ✓ Homogeneity
  - ✓ Pixel Regularity
  - ✓ Edges Regularity

- Accuracy obtained: 85%

Unsuccessful Trials

Segmentation technique (inspired by Variance Threshold by scikit-learn)

After segmentation
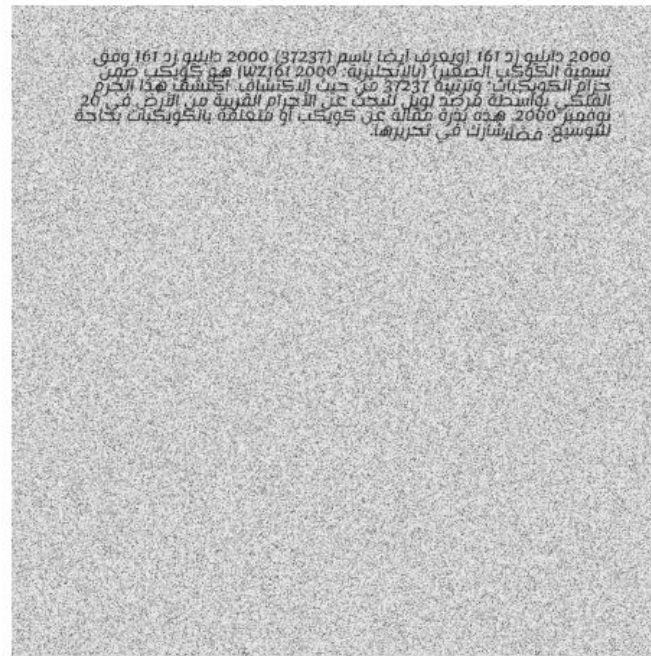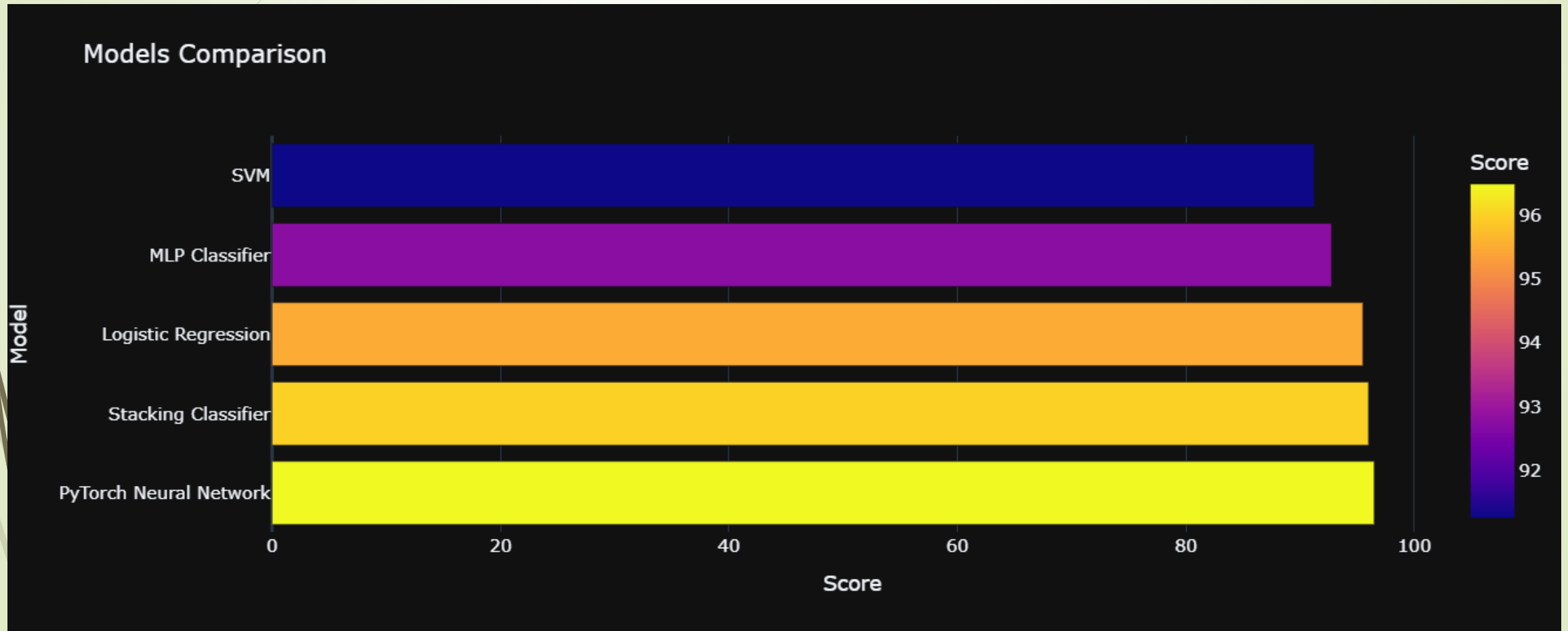
Before segmentation
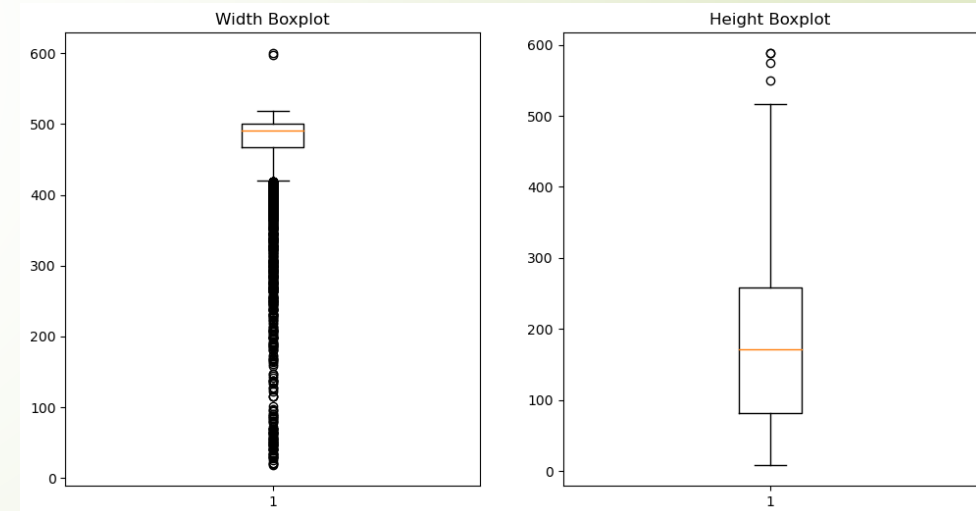
Dimensions of images after segmentation

Performance Analysis

Model Development

PyTorch Neural Network

Model Architecture Summary:

- Input layer processes feature vectors.

- 2 hidden layers learn complex patterns with ReLU activation.

- Output layer generates class probabilities with softmax activation.

Accuracy obtained: 96.5%

Workload Distribution

Team member	Tasks
Ahmed Emad	Preprocessing, Model Development, API
Hla Hany	EDM, Model Development, Deployment
Nada Tarek	HOG, SIFT, PCA
Ziad Ahmed	Preprocessing, Model Development, API

Thank you!