

Pattern Recognition: CSE-5367-001

Team Members: -

- Satya Shah (1002161494)
- Varad Nair (1002161475)
- Harshita Dhingra (1002052823)

REPORT

1. Introduction This report presents the classification of the USPS dataset using K-Nearest Neighbors (KNN) and Naive Bayes classifiers. The goal was to achieve over 90% accuracy on the test set. *Feature extraction techniques such as Histogram of Oriented Gradients (HOG) were applied to improve classification performance.*

2. Dataset and Preprocessing

- The USPS dataset was loaded using the h5py library.
- The dataset consists of grayscale images of handwritten digits (16x16 pixels).
- **Histogram of Oriented Gradients (HOG)** was applied to extract **meaningful features** for classification.

3. Feature Extraction using HOG Transform

- The HOG (Histogram of Oriented Gradients) transform was used to extract edge and texture features from digit images' array.
- This method enhances feature representation by capturing gradient orientations in localized portions of the image.
- HOG features significantly improved the classifier's performance by providing robust feature descriptors.

4. Classification Models

K-Nearest Neighbors (KNN):

- Implemented using KNeighborsClassifier from sklearn.neighbors.

- Experimented with different values of k to optimize performance.
- Evaluated using 10-fold cross-validation.

Naive Bayes Classifier:

- Implemented using GaussianNB from sklearn.naive_bayes.
- Applied to the feature-extracted dataset.
- Optuna was used for hyperparameter tuning of var_smoothing parameter of Naïve Bayes..

5. Performance Metrics

Cross-Validation Results (10-Fold):

- KNN achieved an average accuracy above **96.84%** for an optimal **k=6**.
- Naive Bayes achieved an average accuracy of **93.58%** for 10 folds .

Confusion Matrix:

- Generated for both classifiers to analyze misclassifications.
- Higher accuracy was observed in digit recognition for KNN compared to Naive Bayes.

Precision and Recall:

- Precision and recall scores were computed to measure class-wise performance.
- KNN achieved higher precision and recall, indicating better classification capability.

6. Impact of K Value in KNN

- Different values of k were tested to find the optimal choice.
- As the value of k increased the accuracy was decreasing.
- The best performance was observed for **k = 6**.

7. Accuracy on Test data

- KNN'S accuracy on test data was **94.72%**.
- Naïve Bayes accuracy on test data was **91.23%**.

8. Conclusion

- KNN outperformed Naive Bayes in digit classification on the USPS dataset.
- **HOG-based feature extraction** contributed to achieving high accuracy.
- Future work can explore additional classifiers and ensemble methods to further enhance performance.