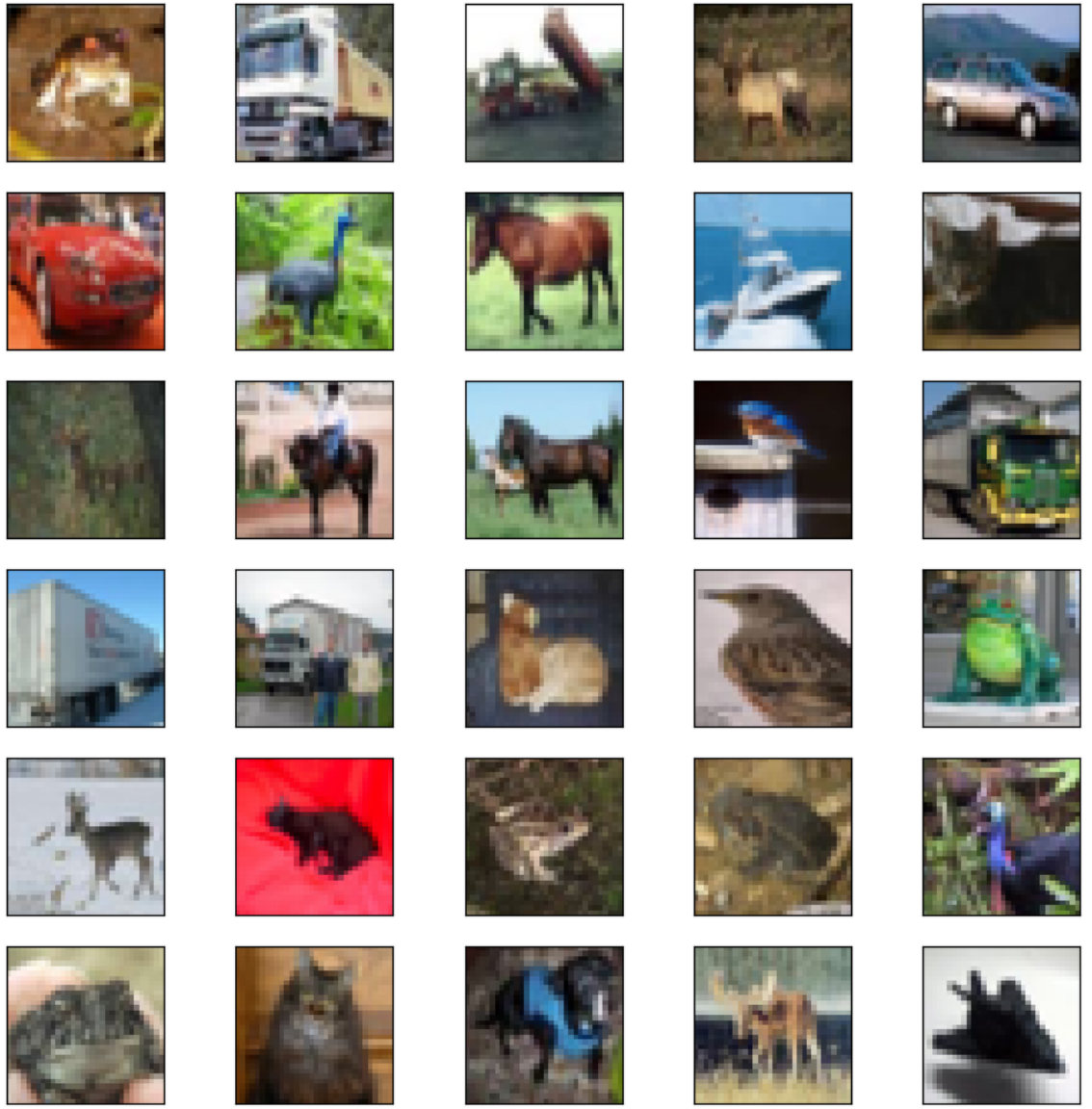
CIFAR-10 DATASET ML CLASSIFICATION STUDY

TOPIC HEADINGS

- ABOUT DATASET
- BACKGROUND
- DATA PREPROCESSING
- MODEL ARCHITECTURE
- RESULTS
- ANALYSIS

ABOUT DATASET

The CIFAR-10 dataset is a widely used benchmark for image classification tasks. It consists of 60,000 color images of size 32x32 pixels, belonging to 10 different classes such as airplanes, cars, birds, cats, deer, dogs, frogs, horses, ships, and trucks.



BACKGROUND

We will use various machine learning models to evaluate the classification performance of the 10-class CIFAR-10 dataset.

The models to be used are:

- Random Forests (RandomForest)
- K-Nearest Neighbors (KNN)
- Decision Tree
- Convolutional Neural Network (CNN)

The evaluation metrics will be:

- Classification Accuracy
- Precision
- Recall
- F1-Score

The goal is to determine the best performing model.

This comprehensive analysis will reveal the most suitable machine learning model for the CIFAR-10 dataset.

DATA PREPROCESSING

The CIFAR-10 data set consists of 50,000 training and 10,000 test images.

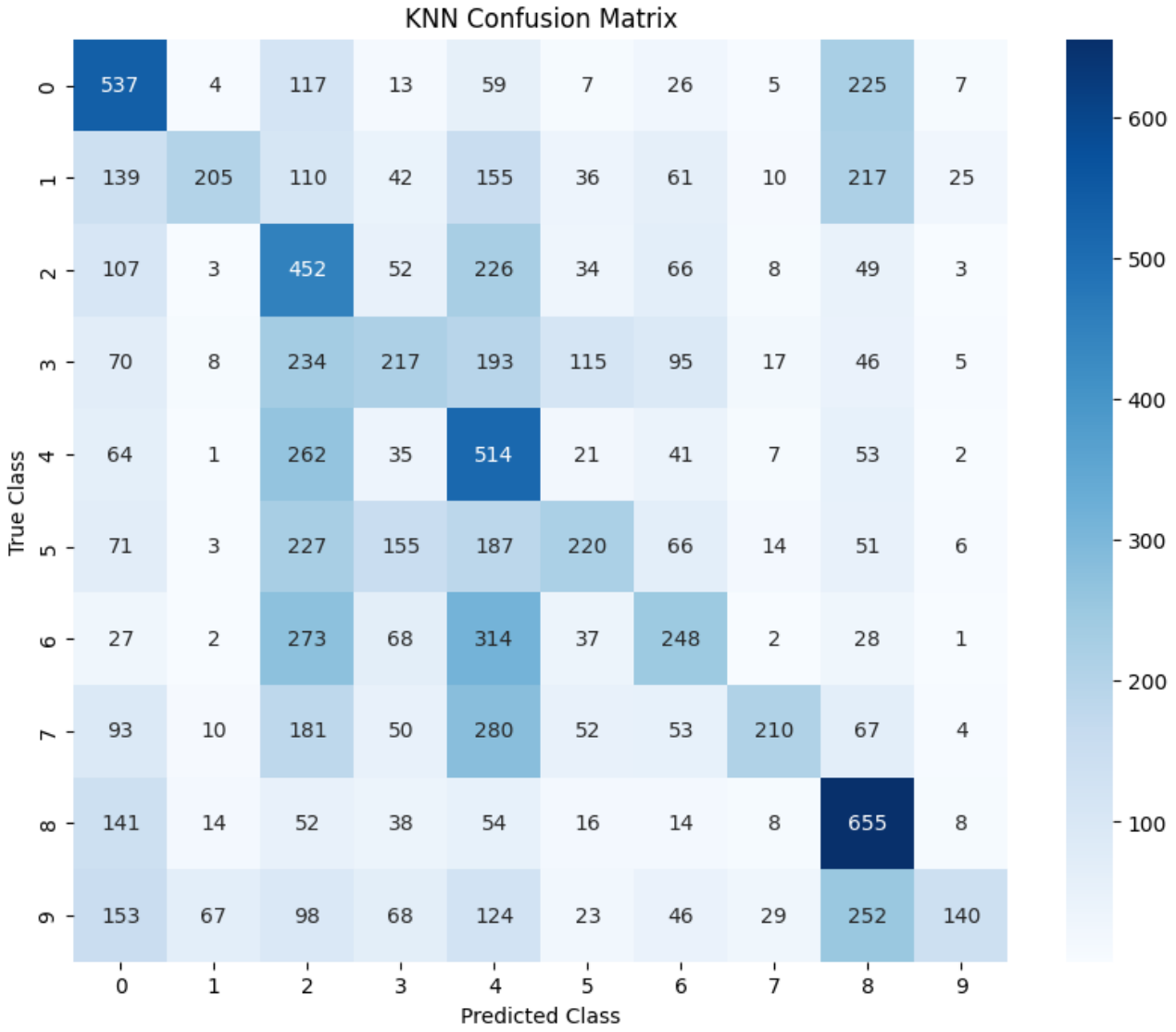
The images in the data set contain RGB pixel values ranging from 0-255.

To improve the training performance, the images have been normalized to the range of 0-1.

MODEL ARCHITECTURE

I used traditional machine learning algorithms such as KNeighborsClassifier, RandomForestClassifier, and DecisionTreeClassifier on the Cifar10 data set. Additionally, I used a high-performing Convolutional Neural Network (CNN) model.

The CNN model consists of sequential convolutional, pooling, and fully-connected layers. The initial layers extract image features using 3x3 filters, while the final layers perform the classification task.

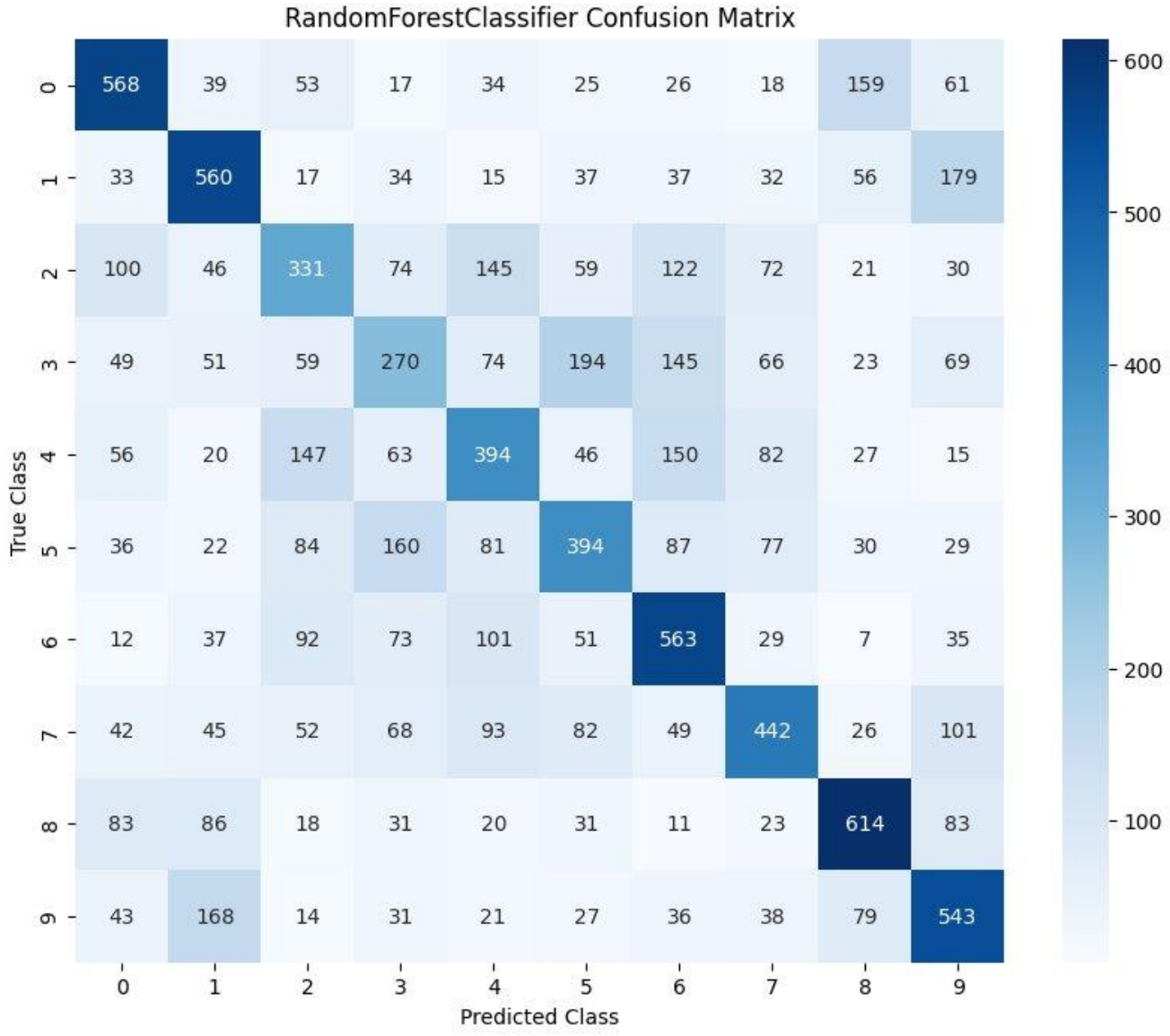


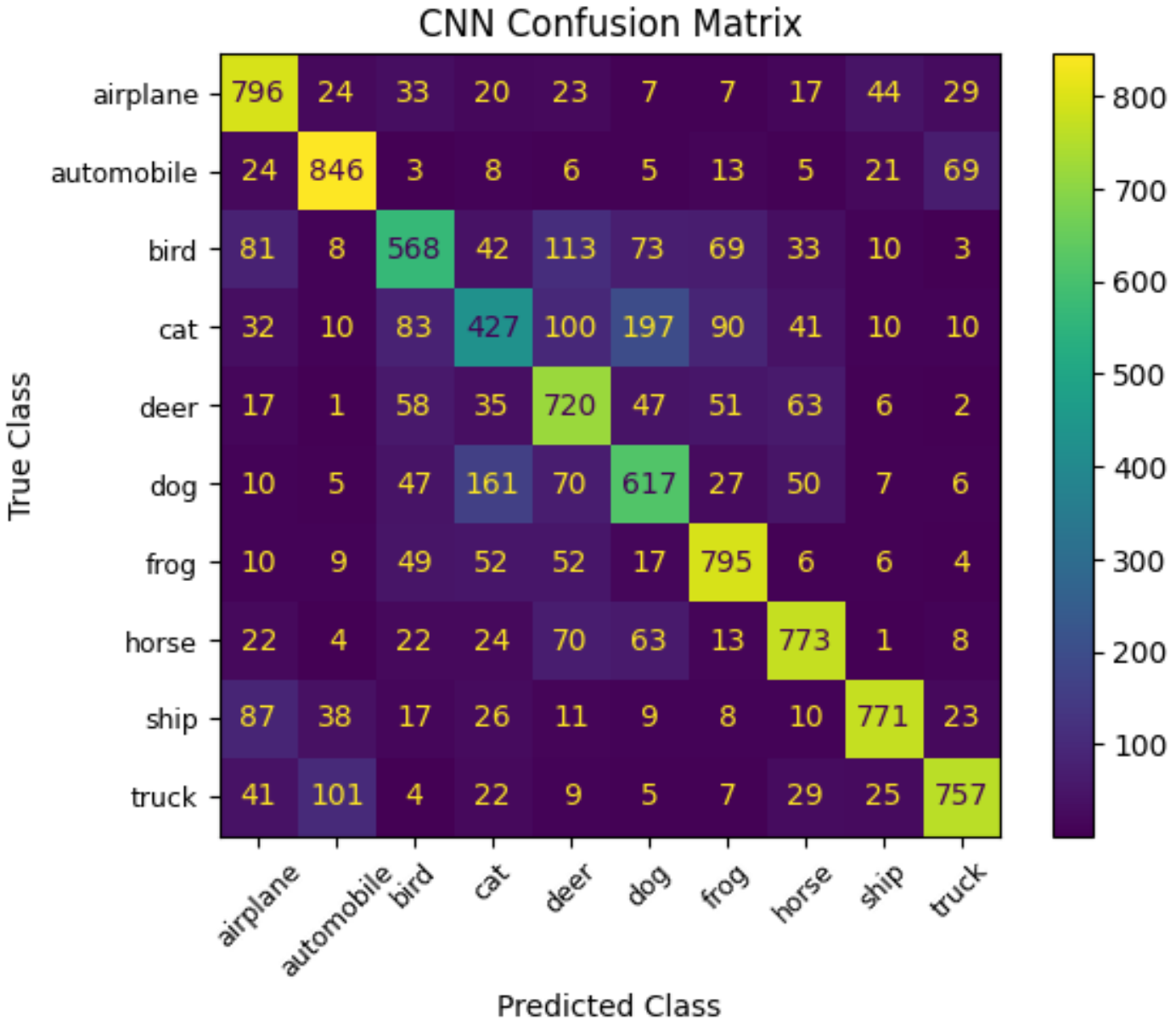
KNN RESULTS

- Accuracy: 0.3398
- F1 Score: 0.32601
- Recall: 0.3398
- Precision: 0.430426

RANDOM FOREST CLASSIFIER RESULTS

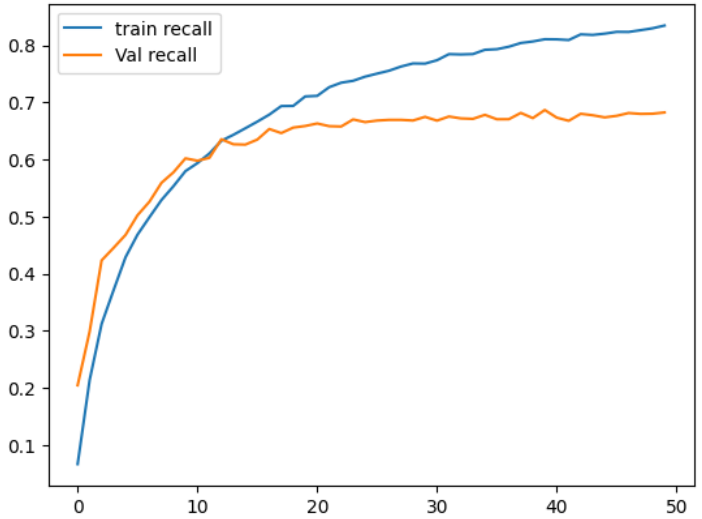
- Accuracy: 0.4679
- F1 Score: 0.4640
- Recall: 0.4679
- Precision: 0.4633





CNN RESULTS

- Accuracy: 0.8710
- F1 Score: 0.8715
- Recall: 0.8346
- Precision: 0.9130



ANALYSIS

The results I obtained from the traditional machine learning algorithms are as follows:

- KNeighborsClassifier: Performance was quite low
- RandomForestClassifier: Performance was quite low
- DecisionTreeClassifier: Performance was quite low

In contrast, the performance of the Convolutional Neural Network (CNN) model was considerably high. In conclusion, compared to the traditional machine learning algorithms, the performance of the CNN model on the Cifar10 data set was significantly superior.