OBJECT RECOGNITION

CSL-2050 PRML

Contents

Project Overview

PRIMARY GOAL

To understand standard machine learning techniques, analyse their performance, and applying various techniques to improve accuracy.
Through this exploration, we seek to gain insights into the challenges and limitations of using simple machine learning models for complex image classification tasks and identify areas for potential improvement.

DATASET

CIFAR10 image dataset with 60000 samples of 32x32 pixel images.

METHODS USED

- Classification models (KNN, tree based techniques, SVM, naive bayes)
- LDA, PCA
- Feature extraction from ResNet, HoG, hyperparameter tuning

Initial Results

Applied ML models like -

- **k-NN**: Explored k-NN with parameter optimization to determine the optimal number of neighbors.
- **Decision Tree**: Utilized decision tree-based methods for classification.
- **Random Forest**: Employed ensemble techniques like random forests for improved accuracy.
- **Naive Bayes**: Implemented Naive Bayes as a probabilistic classifier.
- **SVM** : powerful technique for linear and non linear decision boundaries, by using differenet kernel functions.

MODELS	ACCURACY	PRECISION	RECALL	F1-SCORE
KNN	0.339	0.430	0.339	0.326
Decision Tree	0.2673	0.267	0.2673	0.267
Random Forest	0.472	0.468	0.472	0.468
Naive Bayes	0.297	0.311	0.297	0.275
SVM	0.544	0.542	0.544	0.542
ANN	0.476	0.489	0.476	0.472

These results reveal the **poor** performance of classcial Machine learning models in Image Classification tasks.

REASONS for Failure

High Dimensionality

CIFAR-10 images have a high dimensionality (32x32x3 = 3072 features). KNN and Decision trees struggle in cases of high dimensional data.

Inability to capture spatial relationships

Pixel values in the features are correlated, which are not captured/learnt by these ML algorithms.

Independence Assumption

Naive Bayes assumes that features are independent, which is not the case for image data. Pixels in images are spatially correlated.

Overfitting

Decision trees can easily overfit to training data, especially with high-dimensional data like images

WHAT CAN BE DONE TO IMPROVE THE PERFORMANCE?

First Approach
Dimensionality Reduction

LDA

LDA is a supervised dimensionality reduction technique that aims to maximize the separation between multiple classes.

No of LDA components = 9
(i.e. n-1)

PCA

PCA is an unsupervised technique used for reducing the dimensionality of the data while preserving as much variance as possible.

No of PCA components = 217
(chosen by keeping 95% of variance)

Results

MODELS	Original ACCURACY	After LDA	After PCA
KNN	0.339	0.329	0.339
Decision Tree	0.2673	0.274	0.2673
Random Forest	0.472	0.371	0.472
Naive Bayes	0.297	0.366	0.297
SVM	0.544	0.544	0.544

Conclusions

- **Improvement**: LDA and PCA might have enhanced the feature separability, making the data more discriminative, which could have benefited the Naive Bayes classifier that assumes feature independence.

- **Decrease in Accuracy:**
    - Loss of Information: might have resulted in the loss of some discriminative information crucial for the Random Forest and SVM classifiers, affecting their performance negatively.
    - Complexity Reduction: Random Forest and SVM, being inherently capable of handling high-dimensional data, might not benefit significantly from dimensionality reduction and could even suffer due to reduced feature space.

Second Approach
Feature extraction from ResNet50

- ResNet, which stands for Residual Network, is a type of deep neural network architecture that introduced the concept of residual learning to CNN architectures to address the problem of vanishing gradients in very deep networks.

- Used the provided template for feature extraction from the last layer of a pre-trained ResNet-50 model.

- Why ResNet is good for IMAGES?
    - **Hierarchical Feature Learning**: ResNet's deep architecture allows it to learn hierarchical features from images, capturing both low-level and high-level features essential for image classification.
    - These deep architechtures make it relevant for tasks like image

Results after using features from ResNet50

MODELS	ACCURACY
KNN	0.841
Decision Tree	0.646
Random Forest	0.811
Naive Bayes	0.792
SVM	0.898
ANN	0.878

A significant improvement in accuracy was observed across all classifiers.

The reason is quite simple!

The features extracted from ResNet-50 are rich and high-dimensional, capturing intricate patterns and representations from the images. This solves the primary problem of capturing spatial correlations of the pixels.

Third Approach
Feature engineering (HoG)

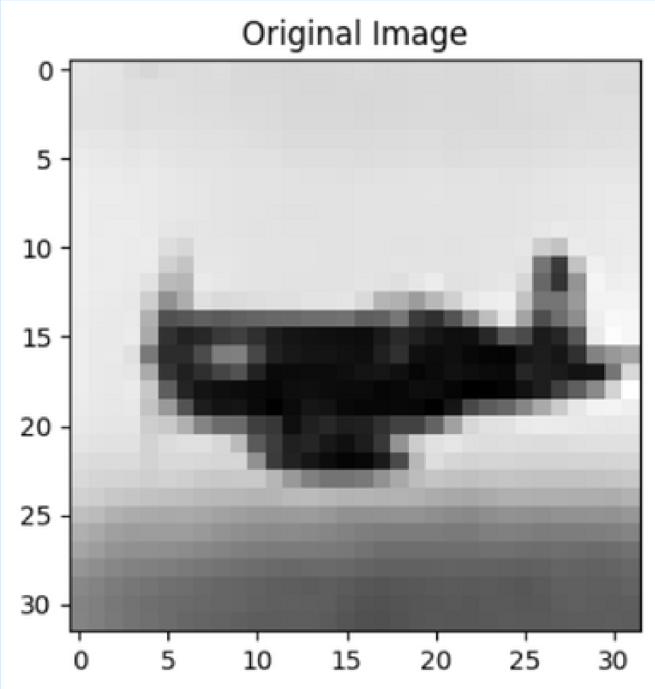
- HoG is a feature descriptor that captures the distribution of gradients (edge orientations) in an image. It is used primarily in computer vision tasks for object detection and image classification.
- We have used the template code provided for applying HoG.

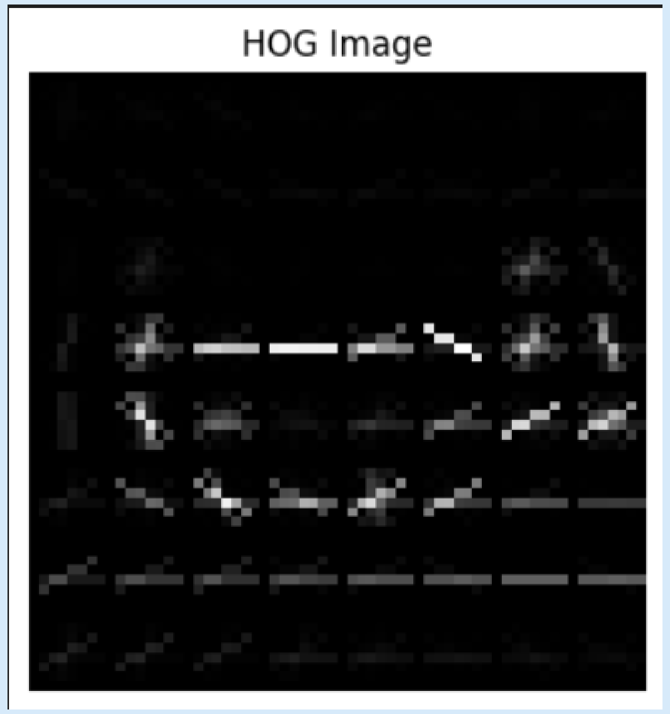
MODELS	KNN	DECISION TREE	RANDOM FOREST	NAIVE BAYES	SVM
ACCURACY	0.523	0.279	0.520	0.458	0.639

HoG

It is clear from the observation table that the classifiers trained on HoG features exhibited improved accuracy compared to models trained on raw pixel values, but not as high as those trained on ResNet features.

HoG captures texture and shape information but not the intricate details and relationships present in the data as effectively as a deep neural network

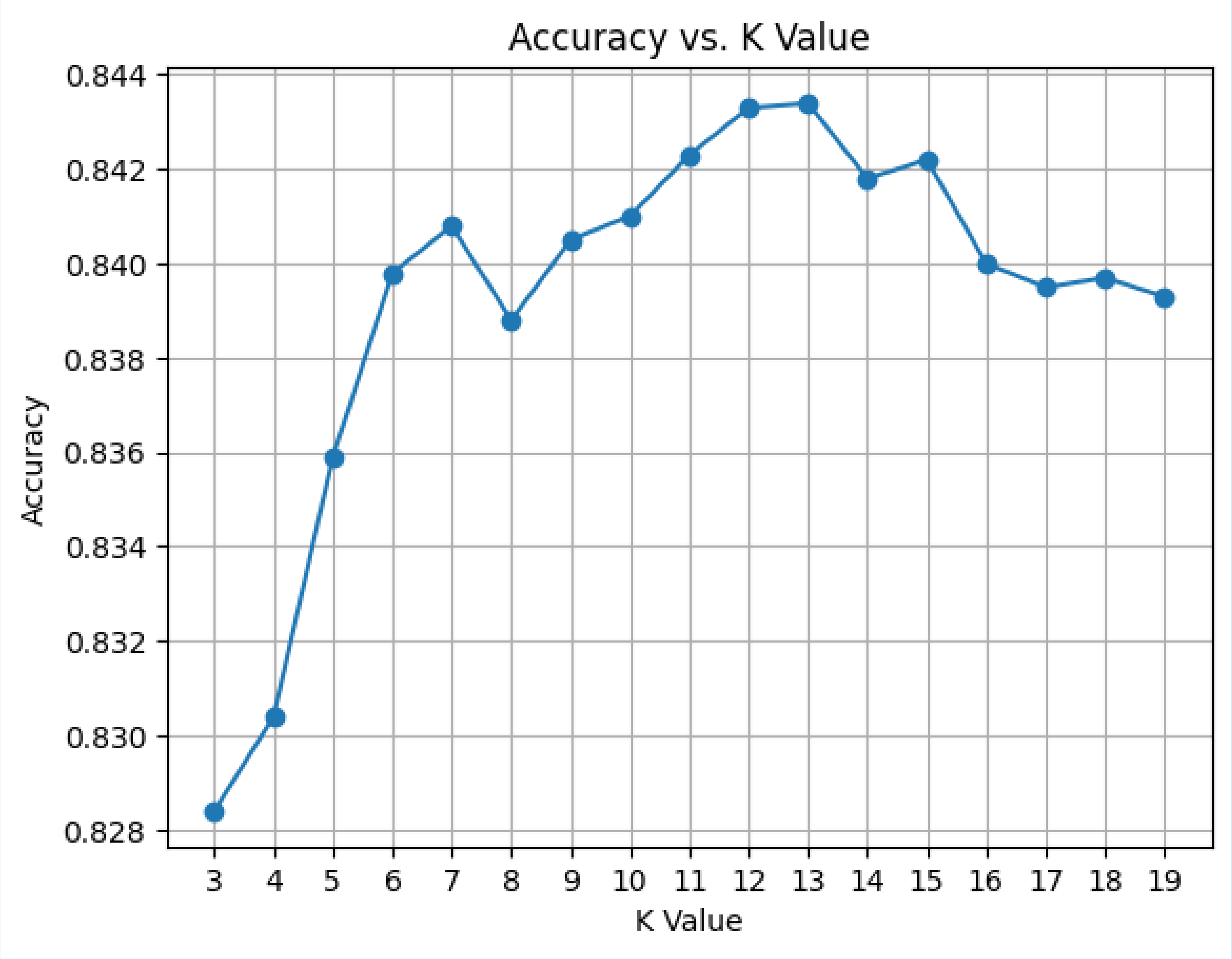




Final Step Hyperparameter Tuning

While applying a pre-trained model like ResNet-50 is a powerful approach for the CIFAR10 dataset, there's still room for improvement through hyperparameter tuning on the final classification layer.

- **K-Nearest Neighbors (KNN)-** By varying the number of neighbors (k), the accuracy of the KNN classifier was evaluated.



Best value of k = 13

- **Decision Tree -**
    - Two different criteria, "gini" and "entropy", were explored to split the nodes of the Decision Tree.
    - Best parameter found from the grid search and the values are
        - criteria = ‘entropy’
        - Max_depth = 10, min_samples_leaf = 10, and min_samples_split = 15
- **Naive Bayes -**
    - The **Multinomial** Naive Bayes classifier was optimized by tuning the alpha parameter (smoothing parameter) was obtained as 2.0
    - The **Gaussian** Naive Bayes classifier was optimized by tuning the var_smoothing parameter, which adds a fraction of the largest variance of all features to variances for calculation stability.
    - The best value for var_smoothing = 1e-09.
- **SVM -**
    - The SVM classifier was fine-tuned by exploring different combinations of hyperparameters such as C, gamma, and kernel.
    - Grid Search Cross Validation results
        - 'C'= 2
        - 'gamma' = 'scale'
        - 'kernel'= 'rbf'

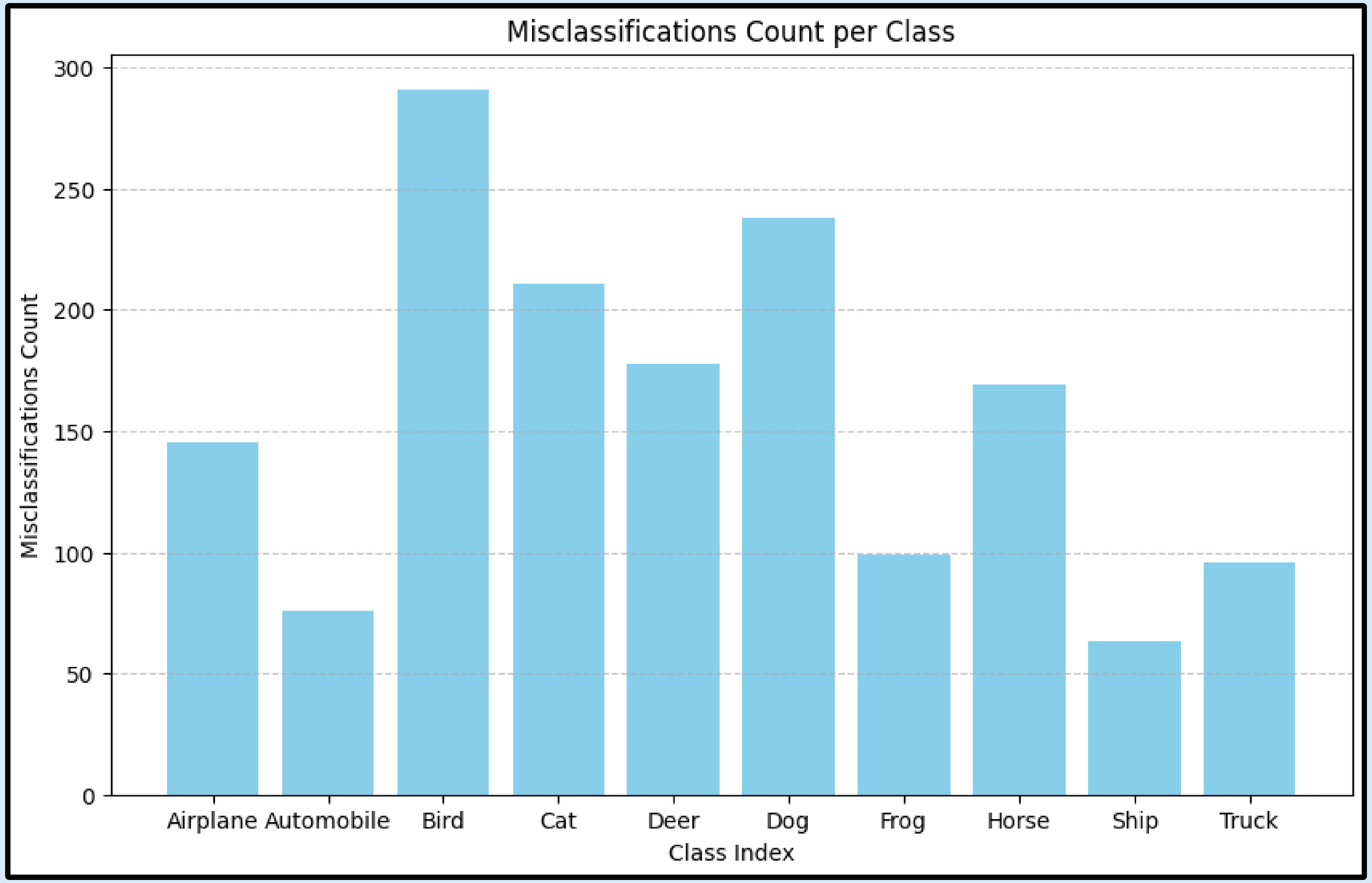
Models	ACCURACY
KNN	0.843
Decision Tree	0.672
Random Forest	0.820
Naive Bayes	0.795
SVM	0.896

Table of Final Accuracies

Hyperparameter tuning was helpful in further fine improvements of the performance of the classical models.

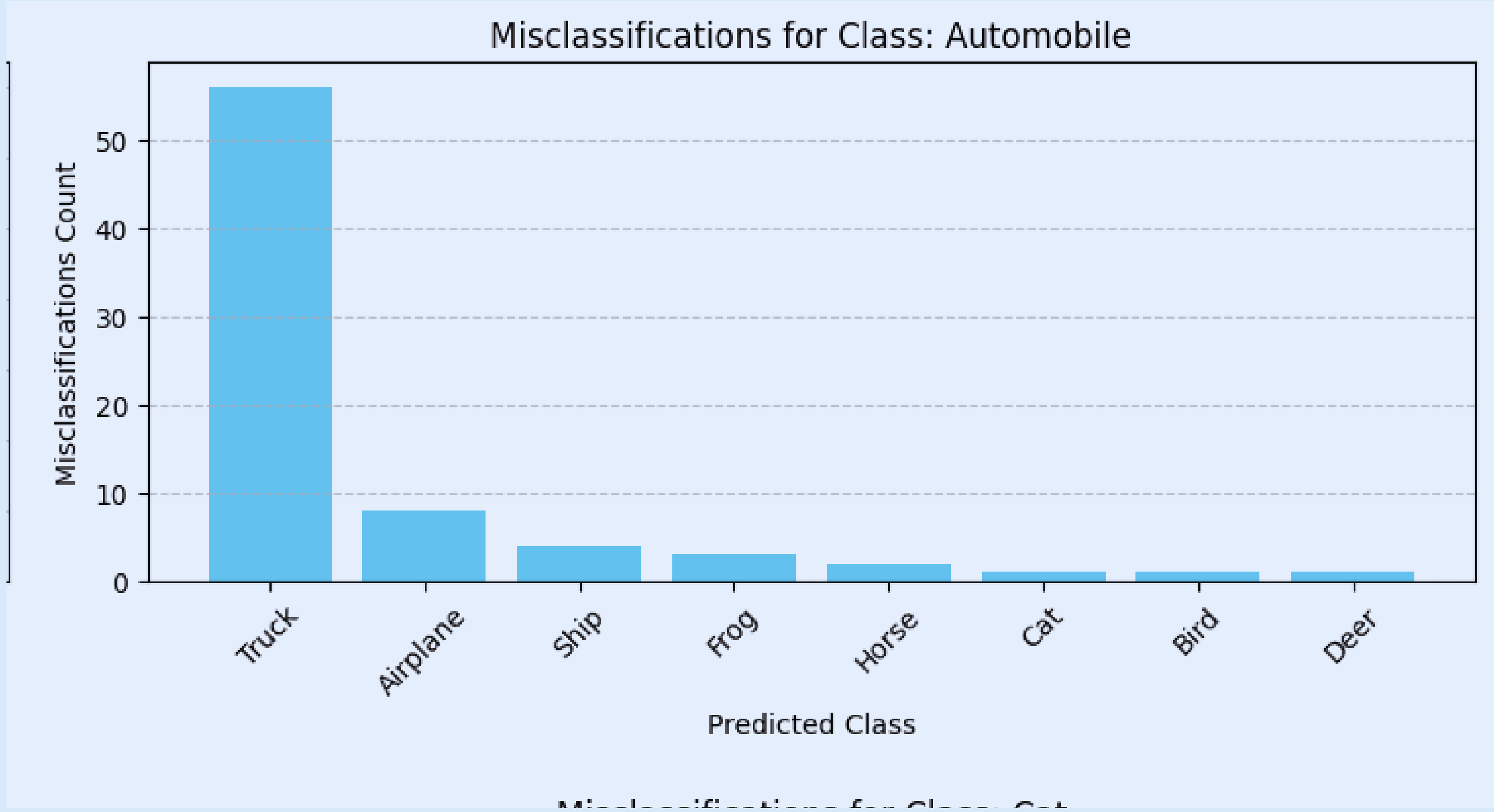
Misclassifications and Future Prospects

<u>Failure Case Analysis(using Misclassification)</u> - It provides insights into the specific classes that are often confused with each other by the models.



<u>Graph for KNN classifier</u>

Shows misclassifications over entire test set



Upon analysing for Automobile class, for the same KNN predictor, we can see that majority misclassifications are labeled 'TRUCK' which is quite reasonable.

Inferences from graphs of all models -

- Uniform misclassifications - Tree and forest methods
- Specific Confusions - Naive Bayes and SVM

FUTURE PROSPECTS

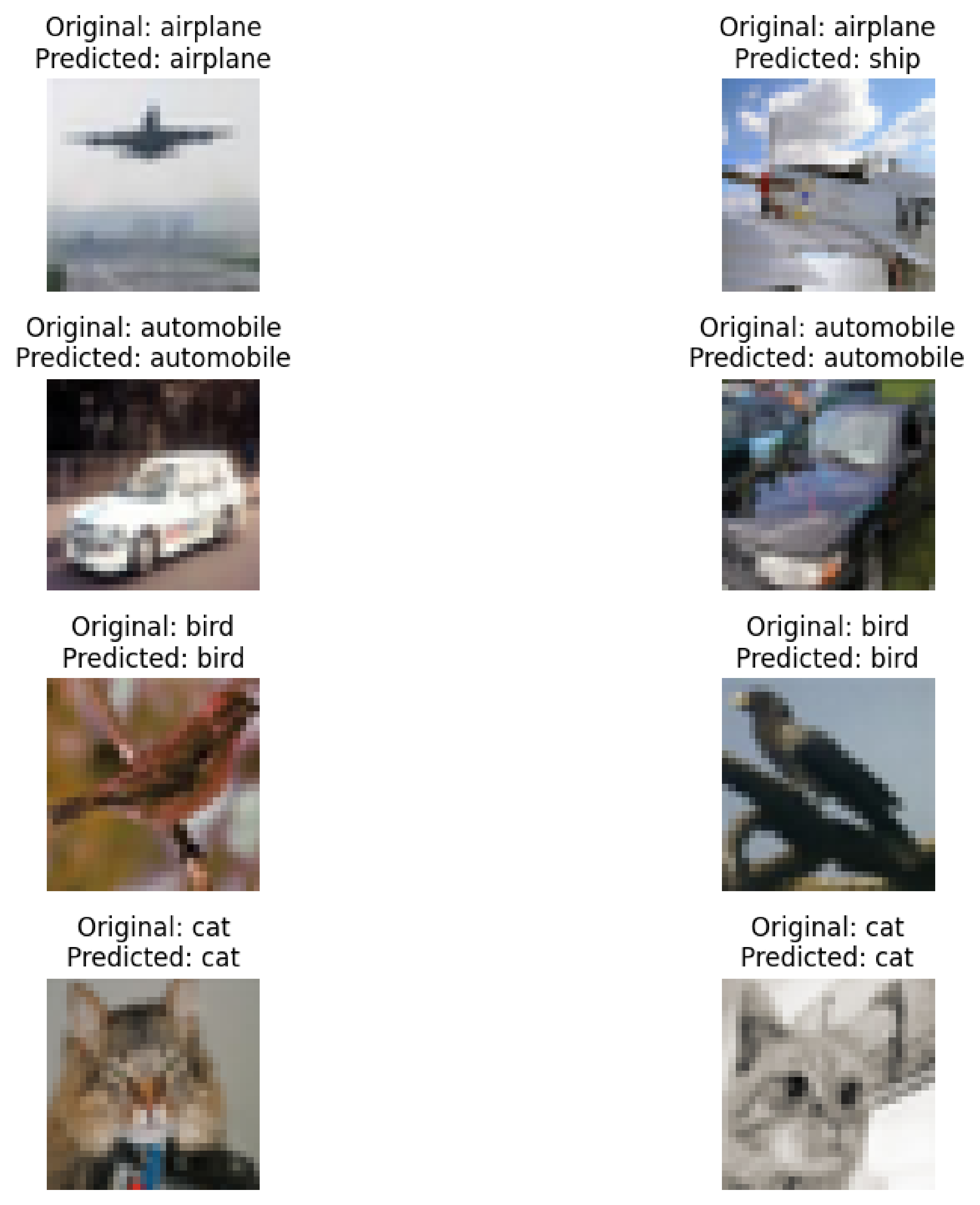
As we continue to explore the CIFAR-10 dataset, several avenues can be pursued to enhance model performance:

- **Advanced Feature Engineering**: Techniques like HoG and deep feature extraction (e.g., from pre-trained models like ResNet) have shown promise. Exploring more advanced feature engineering methods tailored to image data could yield significant improvements.
- **Ensemble Methods**: Combining the strengths of multiple models through ensemble methods can often lead to better generalization and reduced overfitting. Techniques like stacking or boosting could be explored.
- **Neural Network Architectures**: Experimenting with more complex neural network architectures, including convolutional neural networks (CNNs), can help in capturing intricate patterns in image data, potentially improving accuracy and reducing misclassifications..

Summary

We have explored various classical machine learning techniques for image classification task of CIFAR10 dataset.
We also looked into improvement methods like Dimensionality Reduction, using pretrained features from ResNet, feature engineering like HoG.

Thus, successful analysis has been conducted over these techniques, their pros and failure anaslysis.



KNN with feature extraction in action

Thank You!