NAVTAC AI Training Program

Machine Learning & Data Science

Mid-Term Project Report

Title: Image Classification Using Support Vector Machine (SVM) for Helmet and Non-Helmet Detection

Submitted by: Sajid Ali

Instructor: Gulshan Yasmeen

Submission Date: October 19, 2025

Table of Contents

1. Project Proposal

Title

Image Classification Using Support Vector Machine (SVM) for Helmet and Non-Helmet Detection

Problem Statement

Road safety is one of the most critical challenges in developing countries where motorbike riders often neglect helmet usage. This project aims to detect whether a person is wearing a helmet or not using machine learning techniques applied to image data.

Objectives

The objectives of this project are to:
- Develop an image classification model that distinguishes between helmet and non-helmet images.
- Apply feature extraction using a pre-trained deep learning model and use SVM for final classification.
- Evaluate the performance of the model and discuss its practical applicability.

Dataset Description

A custom dataset was created manually by downloading images of people wearing helmets and not wearing helmets from various websites. The dataset was organized into separate folders for training, validation, and testing. Each class (Helmet and Non-Helmet) contained an approximately equal number of images to maintain balance.

2. Data Mining and Exploration

Initial exploration involved organizing and loading image data using Keras' ImageDataGenerator. Basic checks were performed to verify the image counts, class distributions, and dimensions. No missing values were found since images were well-organized in folders.

3. Data Preprocessing

Data augmentation was applied to increase the diversity of training samples and prevent overfitting. The following transformations were used:

- Rotation range: 20 degrees
- Width and height shifts: 0.2
- Shear and zoom range: 0.2
- Horizontal flipping

All images were rescaled by a factor of 1/255 to normalize pixel values. The dataset was then split into training, validation, and testing subsets for model development.

4. Data Visualization

Sample visualizations of augmented images were examined to ensure transformations maintained class integrity. Histograms of class distributions were generated to confirm balanced data across Helmet and Non-Helmet categories.

5. Model Development

Feature extraction was performed using the MobileNetV2 model, pre-trained on the ImageNet dataset. The convolutional base of MobileNetV2 was used to extract deep features from the image data, which were then flattened. These extracted features were passed to a Support Vector Machine (SVM) classifier implemented using scikit-learn.

The SVM model was built using a pipeline that included StandardScaler for normalization, followed by SVC for classification. This hybrid approach combined the representational power of deep learning with the robustness of classical ML.

6. Model Evaluation

The model's performance was evaluated using classification metrics and a confusion matrix. The results were as follows:

- Test Accuracy: 81.18%
- Validation Accuracy: 80.49%
- Precision (Helmet): 0.83
- Recall (Helmet): 0.79
- Precision (Non-Helmet): 0.80
- Recall (Non-Helmet): 0.83
- Overall F1-score: 0.81

The confusion matrix indicates that the model correctly identified most Helmet and Non-Helmet instances with a balanced performance across classes.

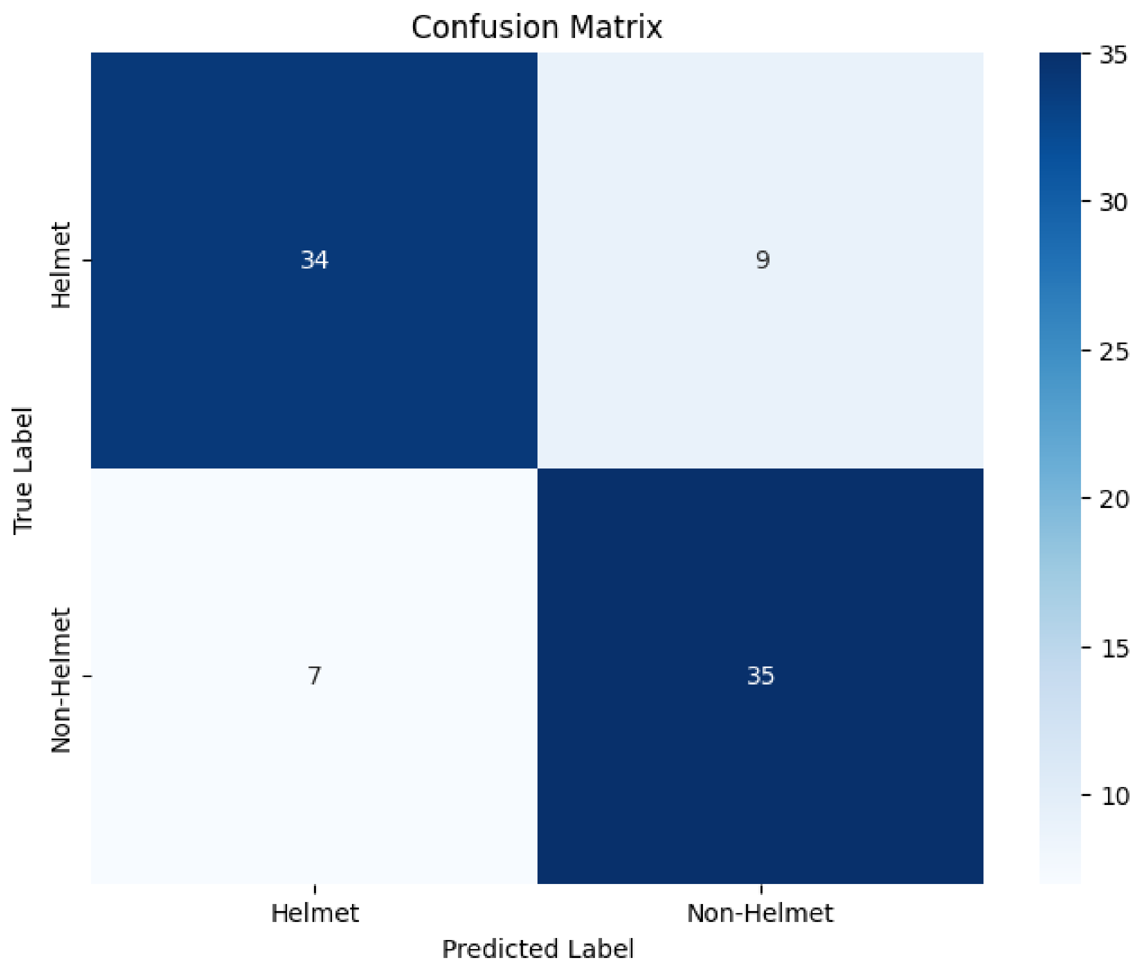

Figure 1: Confusion matrix showing classification performance for Helmet and Non-Helmet images.

7. Conclusion and Recommendations

The SVM-based image classification model successfully distinguished between Helmet and Non-Helmet images with an accuracy of over 81%. The approach of using MobileNetV2 feature extraction combined with SVM proved effective for this small-scale dataset. Future improvements may include training on a larger dataset, fine-tuning the CNN layers, and experimenting with alternative classifiers such as Random Forest or XGBoost.

8. References

- Chollet, F. (2017). Deep Learning with Python. Manning Publications.
- Pedregosa et al. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research.
- TensorFlow and Keras Documentation.
- Image sources: Public domain and open-access websites for educational use.