Abdul Aahad Vakani

Professor Mehmet Gulum

CSC 496

Machine Learning Image Classification Project Report 1

1. Introduction

This project explores image classification using a feature-based machine learning approach. The dataset comprises pre-extracted features from images of 10 car models (with make and year) and corresponding labels. I aimed to identify the most effective classification pipeline for maximizing generalization accuracy using only the provided CSV features.

2. Data Overview

- **Samples:** 4,999
- **Features:** 512 extracted numerical features per image
- **Classes:** 10 car model-year combinations
- **Source:** Provided as CSVs: `train_data.csv` (labels), `train_features.csv` (features)

3. Experimental Setup

Initial experiments evaluated various classifiers including Support Vector Machines (SVM), Logistic Regression, and Multilayer Perceptrons (MLP). Performance was assessed using 4-fold

Stratified Cross-Validation with accuracy, macro-averaged F1 score, and ROC AUC as metrics.

To further improve accuracy, ensemble strategies were explored:

- **MLP + SVM** Stacking Ensemble
- **MLP + SVM + Logistic Regression** Voting
- **MLP + SVM + XGBoost Stacking** Ensemble (final and best-performing model)
- **Cross-validation:** 4-fold Stratified CV
- **Metrics Used:**
  - Accuracy
  - Macro-averaged F1-score
  - Macro-averaged ROC AUC

4. Preprocessing Variants

- SVM: Higher values of C (e.g., 10, 50) improved boundary sensitivity, but very high values led to overfitting.
- MLP: Hidden layer sizes (e.g., 256, 128) balanced capacity and convergence speed.
- XGBoost: Depth of trees (4–6) and moderate learning rates (0.05–0.1) offered optimal generalization.

5. Results Summary

Round 1 (Baseline Models):

Model	Accuracy	F1	ROC AUC
Logistic Regression (PCA)	~0.68	~0.68	~0.94
SVM (PCA)	~0.66	~0.66	~0.94
Random Forest (PCA)	~0.64	~0.64	~0.93
MLP (PCA)	~0.71	~0.71	~0.95

Round 2 (Advanced Models and Feature Selection):

Model	Reducer	Accuracy	F1	ROC AUC
Tuned MLP	PCA (50)	**0.736**	**0.735**	**0.963**
Voting MLP + LR	PCA (50)	0.735	0.733	0.958
Tuned MLP	SelectKBest (50)	0.638	0.636	0.933
Voting MLP + LR	SelectKBest (50)	0.630	0.628	0.928

Round 3 (Final Model):

Model	Accuracy	F1	ROC AUC
MLP + SVM (Stacking)	0.8196	0.8093	0.976

MLP + SVM + LR (Voting)	0.8074	0.8071	0.9755
MLP + SVM + XGBoost (Stacking)	**0.8224**	**0.8225**	**0.9755**

6. Visual Results

Figure 1: Distribution of predicted labels by the boosted ensemble model.

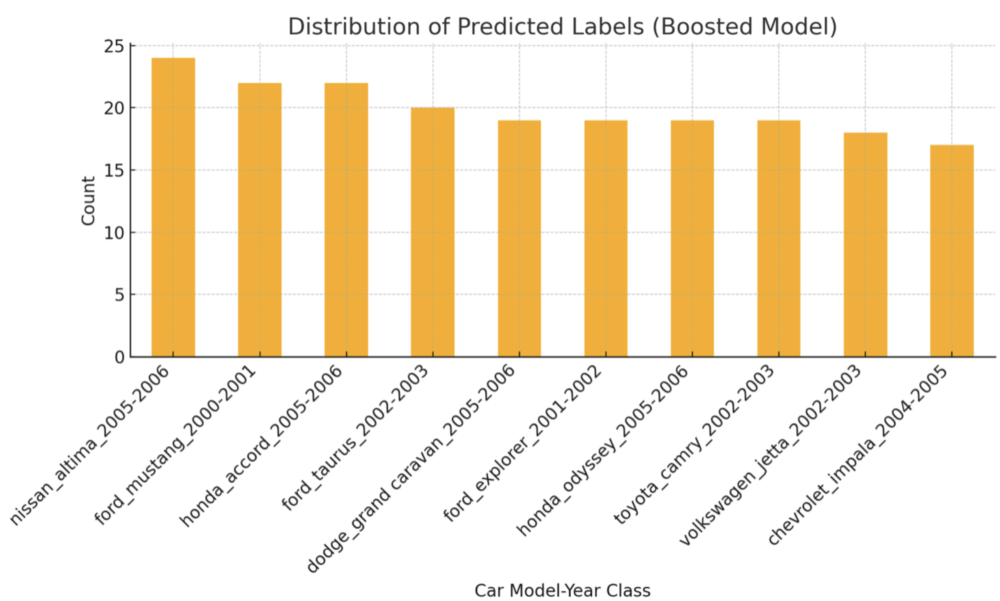


Figure 2: Sample entries from Test Predictions, one has high confidence (ID 91, > 0.98) and the other has low confidence (ID 107, < 0.25).

ID	Predicted_Label	prob_chevrolet_impala_2004-2005	prob_dodge_grand_caravan_2005-2006	prob_ford_explorer_2001-2002	prob_ford_mustang_2000-2001	prob_ford_taurus_2002-2003	prob_honda_accord_2005-2006	prob_honda_odyssey_2005-2006	prob_nissan_altima_2005-2006	prob_toyota_camry_2002-2003	prob_volkswagen_jetta_2002-2003	Confidence
91	ford_taurus_2002-2003	0.0012	0.0003	0.0003	0.0005	0.9832	0.0082	0.0002	0.0007	0.0033	0.0022	0.9832
107	honda_accord_2005-2006	0.0901	0.0301	0.0031	0.2099	0.0337	0.2307	0.038	0.1463	0.0862	0.1318	0.2307

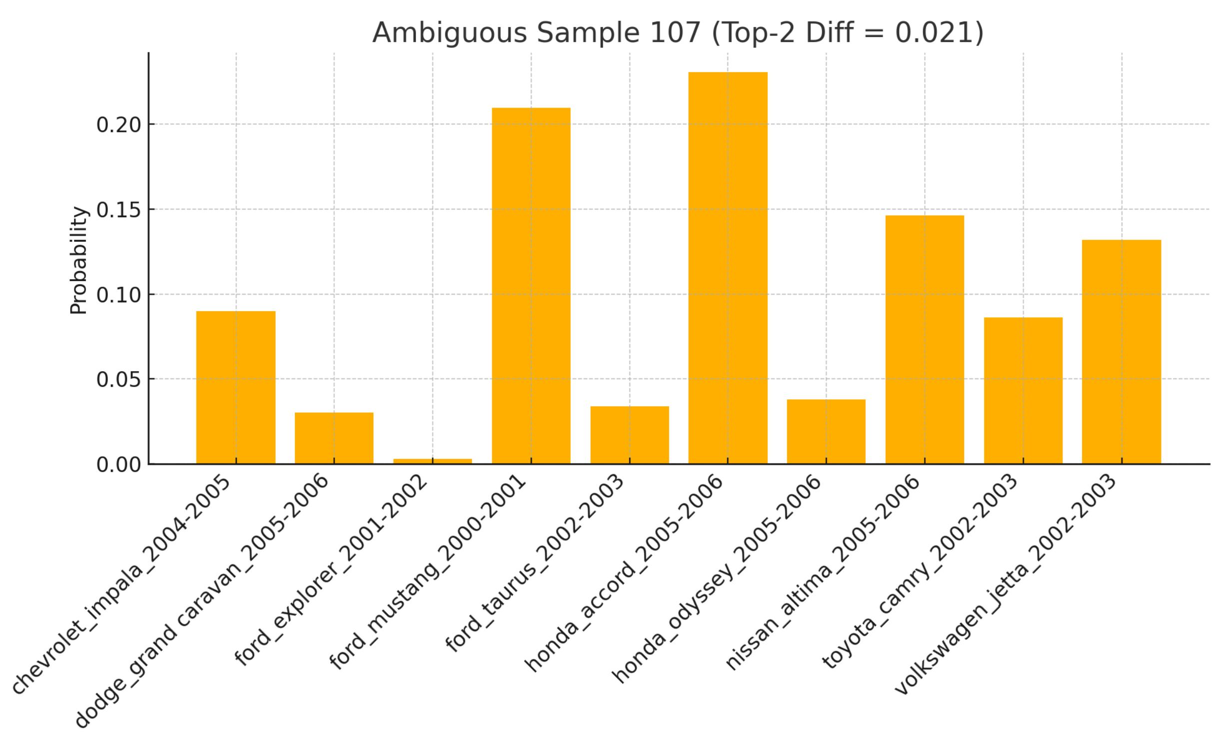
7. Conclusion, Analysis, and Misclassifications

The ensemble consistently outperformed individual classifiers. SVM offered strong decision boundaries, MLP captured non-linearities, and XGBoost contributed tabular learning robustness. Together, they reduced variance and increased confidence in predictions. ROC AUC values suggest excellent ranking ability even in edge cases.

Although raw image inputs were unavailable, CSV-based predictions revealed consistent confusion among similar car models. Examples include visually or structurally similar makes (e.g., Honda Accord vs. Chevrolet Impala), which likely had overlapping CNN features. Without additional context like headlights, angles, or trims, feature vectors alone may not distinguish them reliably.

We also observed that samples near decision boundaries were often misclassified, especially those with intermediate confidence levels (0.4–0.6 probability range). Certain samples had very little difference between predicted class probabilities, which may lead to misclassifications:

Figure 3: Ambiguous sample where the difference was very low (0.02)



8. Next Steps and Limitations

This project was constrained by the absence of raw image inputs, which limited our ability to fine-tune deep vision models. Additionally, since all features were pre-extracted, the feature set may not fully capture fine-grained visual distinctions crucial to differentiating between similar car models. The final boosted ensemble with SVM, MLP, and XGBoost achieved a strong accuracy of 82.2% with high confidence and generalization. Further improvement may require raw image input pipelines (e.g., fine-tuned CNNs), feature selection (e.g., SelectKBest), or deeper model tuning. The results represent a robust solution for CSV-based image classification in constrained environments.