# Deep Learning (UEC642)

Project Report - **Car Brand Recognition**


Submitted By:

Vanshika Garg(102215074)

Agrim Goyal(102215080)

Kartikay Chadha(102215189)

Aditi Sharma(102215209)

Kanish Chadha(102215227)

Submitted to :

Dr. Gaganpreet Kaur

**DEPARTMENT OF ELECTRONICS & COMMUNICATION ENGINEERING**

**THAPAR INSTITUTE OF ENGINEERING & TECHNOLOGY**

**PATIALA, PUNJAB**


**ODDSEM202526**

# TABLE OF CONTENTS


1. **BACKGROUND & LITERATURE SURVEY**
    - 1.1 Background
    - 1.2 Summary of Existing Approaches
2. **METHODOLOGY**
    - 2.1 Dataset Description
    - 2.2 Data Preprocessing & Augmentation
    - 2.3 Model Architectures
        - 2.3.1 Model 1: ResNet50 (Transfer Learning)
        - 2.3.2 Model 2: InceptionV3 (Feature Extraction)
    - 2.4 Training Hyperparameters
3. **RESULTS & EVALUATION**
    - 3.1 ResNet50 Evaluation
    - 3.2 InceptionV3 Evaluation
    - 3.3 Confusion Matrix Analysis
    - 3.4 Comparative Summary
4. **CONCLUSION & FUTURE WORK**
5. **REFERENCES**

# BACKGROUND & LITERATURE SURVEY

## 1.1 Background

Vehicle Make and Model Recognition is a specialised subset of object classification that presents unique challenges due to the fine-grained nature of the classes. Unlike generic object detection (e.g., distinguishing a car from a bicycle), this requires distinguishing between visually similar sub-categories, such as different models from the same manufacturer (e.g., Toyota Camry vs. Toyota Corolla) or the exact model across various years. This project leverages Deep Learning (DL) techniques to automate this process, utilising Transfer Learning to overcome the high computational costs of training from scratch.

## 1.2 Literature Survey

A comprehensive review of recent literature reveals a progression from standard CNNs to specialised architectures focusing on attention mechanisms and part-based features.

**Real-Time Vehicle Make & Model Recognition Using Residual SqueezeNet:** Lee et al. (2019) [1] present a real-time vehicle make and model recognition (MMR) system using an improved Residual SqueezeNet architecture. The work addresses the common failures of license-plate–based systems and proposes a vision-based deep learning approach suitable for real traffic conditions. The authors construct a large dataset of 291,602 images across 766 vehicle classes, using a semi-automated clustering method to reduce annotation time. Their Residual SqueezeNet integrates lightweight Fire modules with bypass connections to enhance feature extraction while maintaining a remarkably compact model. Experimental results show 96.33% rank-1 accuracy and 99.52% rank-5 accuracy, outperforming standard SqueezeNet, AlexNet, and GoogLeNet while being significantly more memory-efficient. The model also achieves real-time inference (~109 ms/image), demonstrating strong suitability for deployment in intelligent transportation systems.

**Rear-Lamp-Based Car Make & Model Recognition:** Bularz et al. (2024) [2] propose a rear-lamp–based vehicle make and model recognition system that combines image preprocessing, binarisation, and a lightweight CNN classifier. Unlike most VMMR methods, which rely on whole vehicle appearance or frontal photos, this work focuses on extracting the taillight shape as the primary discriminative feature. The technique applies several preprocessing steps—contrast enhancement, red colour extraction, and binarisation—to isolate the lamp region (shown in *Fig. 1, page 3*). The authors systematically evaluate eight binarisation methods and three CNN architectures, finding that Otsu/ISODATA thresholding and a two-block CNN produce the best accuracy. The approach achieves 93.9% accuracy on their CarBinLamps dataset and 84.5% on a mixed dataset, demonstrating competitive performance while using a much simpler model than many deep architectures. The study highlights that strong preprocessing can reduce model complexity and still deliver high recognition performance.

**CNN-Based Classification for Highly Similar Vehicle Models:** Avianto et al. (2022) [3] address the challenge of recognising highly similar vehicle models, where cars from different manufacturers look almost identical. To address this, the authors propose a multi-task CNN based on VGG-16 that simultaneously predicts vehicle make and vehicle model, rather than using a single-task classifier. The study introduces the InaV-Dash dataset, which contains dashboard-camera images of 10 models and four brands, featuring significant real-world issues such as haze, occlusion, and motion blur (*Figure 2, page 4*). Experimental results show that multi-task learning dramatically improves performance over baseline CNNs, achieving 98.73% accuracy for make and 97.69% for model, far outperforming single-task training. The work demonstrates that brand information helps disambiguate visually similar models, making MTL an effective strategy for fine-grained vehicle classification.

**Analysis of Deep CNNs for Fine-Grained Vehicle Classification:** Khairi et al. (2023) [4] present a comparative study of four widely used deep CNN models—MobileNet-V2, Inception-V3, VGG-19, and ResNet-50—for fine-grained vehicle make-and-model classification. The work evaluates these architectures on three datasets: Stanford Cars (196 classes), BMW-10 (10 classes), and the newly created PAKCars dataset (44 classes), which contains cars commonly used in Pakistan (*sample images shown on pages 7–8*). The authors use uniform preprocessing, augmentation, and transfer learning with ImageNet weights to ensure a fair comparison. Experimental results show that VGG-19 and ResNet-50 achieve the highest accuracy across all datasets (e.g., 82–87% on Stanford Cars and BMW-10), while MobileNet-V2, despite being lightweight, performs the worst due to its limited representational capacity. The study highlights the trade-off between model size and accuracy, showing that deeper networks are more effective for fine-grained vehicle classification tasks.

**Benchmark Dataset & Methodology for Fine-Grained VMMR:** Hayee et al. (2025) [5] present a large-scale benchmark dataset specifically designed for fine-grained Vehicle Make and Model Recognition (VMMR) under unstructured traffic conditions, such as those found in Pakistan. The dataset contains 129,000 images across 94 classes, collected through a combination of web scraping and 15 hours of overhead traffic video (*data collection workflow shown in Fig. 1, page 5*). To ensure high-quality labels, the authors employ an iterative, semi-automated annotation pipeline, where VGG-based predictions are manually verified and refined across multiple iterations, significantly improving dataset quality (*iteration statistics are shown in Table 5, page 8*).

For benchmarking, several deep models, including AlexNet, VGG, EfficientNet, and Vision Transformers (ViT), are evaluated. Results show that EfficientNet and ViT achieve the highest accuracy, with ViT reaching a validation accuracy of 98.53% (*Table 12, page 12*), outperforming traditional CNNs in capturing subtle, fine-grained features. The work contributes both a robust dataset and a detailed methodology, making it a firm reference for VMMR studies in real-world, non-lane-disciplined traffic environments.

**3D Attention Module for VMMR:** Semiromizadeh et al. (2024) [6] propose a VMMR system enhanced with a 3D, parameter-free attention module (SimAM) designed to address the challenges of inter-class similarity and intra-class variation in fine-grained vehicle recognition. Instead of conventional 1D or 2D attention, SimAM assigns a unique weight to every neuron, enabling more precise refinement of feature maps (*architecture illustrated in Fig. 1, page 2*). The authors integrate this module into two mid-level layers of EfficientNet-B4, where feature maps contain the most meaningful structural information (*Section III-C, page 4*). Experimental results on the Stanford Cars dataset demonstrate that the proposed model achieves 90.69% accuracy, outperforming all compared CNN and transformer-based baselines while maintaining the exact parameter count and FLOPs (*Table III, page 5*). This shows that 3D attention significantly boosts fine-grained VMMR performance without incurring additional computational cost.

**Deep Learning–Based Vehicle Re-Identification:** Amiri et al. (2023) [7] present one of the most detailed surveys on vehicle re-identification (ReID), covering supervised and unsupervised deep-learning approaches. The paper explains how ReID relies on extracting global and local visual features, along with auxiliary cues such as colour, type, brand, and spatio-temporal data, to match images of the exact vehicle captured across different cameras (*illustrated in Fig. 1, page 2*). Supervised methods are categorised into global feature learning, local feature learning, hybrid global–local models, and transformer-based architectures, highlighting that relying only on global CNN features is insufficient due to high inter-model similarity. Local-feature models that utilise part segmentation, attention, or keypoint detection significantly enhance fine-grained discrimination. Metric learning approaches—including contrastive loss, triplet loss, logistic triplet loss, focal loss, and newer group-group loss—further enhance intra-class compactness and inter-class separation. The survey also reviews unsupervised domain adaptation techniques, including those utilising GANs, pseudo-labelling, and transformer-based adaptation, to handle real-world scenarios with no annotated data. Finally, the paper outlines significant challenges such as viewpoint changes, occlusion, illumination variation, and dataset bias, providing a comprehensive roadmap for advancing vehicle ReID research.

**Two Decades of Vehicle Make & Model Recognition:** Gayen et al. (2024) [8] present a comprehensive 20-year survey of Vehicle Make and Model Recognition (VMMR), covering both machine learning and deep learning approaches. The paper reviews factors that affect VMMR performance, including lighting variation, viewpoint changes, occlusion, and high intra- and inter-class similarity, illustrated through datasets such as VMMR-db-9170 (*sample images shown in Fig. 3, page 3*). Early methods relied on handcrafted features, such as SIFT, HOG, PHOG, and SURF, as well as classifier combinations like SVM, k-NN, and Random Forests; however, their performance was limited by sensitivity to noise and the size of small datasets. With the rise of deep learning, CNN-based models such as VGG, ResNet, Inception, YOLO, DenseNet, and EfficientNet have become dominant, achieving significantly higher accuracy through automatic feature extraction and multi-task learning. The survey also provides an overview of major VMMR datasets. It highlights open challenges such as the need for large-scale annotated data, robustness under real-world conditions, and real-time efficiency. This paper serves as a firm foundational reference for understanding the evolution and state-of-the-art trends in VMMR.

**Multi-Task Pedestrian Understanding for Autonomous Driving:** Zhou and Zeng (2024) [9] propose a comprehensive multi-task learning framework that unifies pedestrian detection, 3D tracking, and multi-attribute recognition for autonomous driving. The system operates in two stages: Stage I uses low-resolution multi-view images to perform BEV-based 3D pedestrian detection and tracking, significantly reducing computational cost (*architecture shown in Fig. 3.2, page 8*). Stage II utilises high-resolution, cropped images to recognise 14 fine-grained pedestrian attributes, including pose, age, gender, head orientation, intention to cross, reaction, and hand gesture. The model integrates multiple feature sources, including whole-body, face, pose vectors, BEV context features, and even temporal history (*feature sets illustrated in Fig. 3.3, page 9*). Evaluations on JAAD show that the proposed approach achieves higher AP and mAP than existing MTL baselines, such as PCGrad and the 32-attribute Fork-Norm model, especially for critical tasks like Look (55.6% AP) and Cross Intent (88.2% AP) (*Table 4.1, page 10*). The study demonstrates that shared backbones, BEV integration, and temporal cues drastically improve perception accuracy while keeping computation low, making it suitable for real-time autonomous driving.

**Lightweight Attention Network with Regularised Fine-Tuning:** Zhang et al. (2025) [10] propose a fine-grained car recognition method built on a lightweight backbone, MobileNetV3, enhanced with a newly designed Hybrid Attention Module (HAM). The HAM combines channel attention (SE) and spatial attention via 3D estimation, enabling the network to focus on subtle visual differences while suppressing background noise (*module shown in Fig. 3, page 8*). The authors also introduce a regularised fine-tuning strategy that constrains updated weights to remain close to pre-trained ImageNet parameters, improving knowledge transfer and preventing overfitting (*concept illustrated in Fig. 5, page 11*). Experiments on the Stanford Cars dataset (196 classes) demonstrate that the proposed HAM-MobileNet achieves 84.6% accuracy, outperforming other lightweight models, including MobileNetV2 (73.5%), ShuffleNetV2 (75.4%), and EfficientNet-B0 (76.3%) (*Table 3, page 15*). Notably, its performance approaches that of heavier models like ResNet-50 and DenseNet-121 while maintaining an extremely low parameter count and FLOPs, making it ideal for real-time and embedded systems.

**Part-Based Re-ID with Global Context:** Nath and Mitra (2025) [11] introduce GLSIPNet, a part-based vehicle re-identification model that incorporates global contextual information without adding any extra trainable parameters. Traditional VReID methods typically combine a global branch with multiple local branches, increasing computational complexity. GLSIPNet overcomes this by computing a global–local similarity score: each part feature's loss is weighted by its distance to the worldwide feature vector (*architecture shown in Fig. 3, page 7*). This encourages the network to prioritise informative local regions while suppressing irrelevant background patches (*seen in Fig. 6, page 13*). Using ResNet-50 as the backbone and uniform horizontal partitions, the method achieves significant improvements over the baseline—+2.5% mAP on VeRi and +2.4–3.3% mAP on VehicleID (small/medium/large) (*Table 1, pages 10–11*). GLSIPNet also performs competitively with state-of-the-art models while being far simpler and lighter than attention-heavy or multi-branch architectures, demonstrating that global cues can be integrated efficiently through loss weighting rather than additional branches.

# METHODOLOGY

## 2.1 Dataset Description

The project utilises the **Car Logo Dataset** sourced from Kaggle (Rajab, M.). While primarily designed for logo recognition, the dataset contains full and partial vehicle images organised into class folders representing major manufacturers.

- **Source:** Kaggle - Car Logo Dataset
- **Classes:** The dataset comprises approximately 15 distinct courses, including luxury and standard makes such as **Benz, BMW, Cadillac, Ferrari, Ford, Lamborghini, Porsche, Rolls-Royce, and Toyota**.

## 2.2 Data Preprocessing & Augmentation

To prepare the raw images for the neural networks, the following preprocessing steps were implemented using TensorFlow/Keras ImageDataGenerator:

- **Resizing:**
  - Images were resized to **224 × 224** pixels for the ResNet50 model.
  - Images were resized to **299 × 299** pixels to match the input requirements of the pre-trained InceptionV3 model.
- **Normalization:** Pixel values were normalised (scaled between 0 and 1 or -1 and 1) using the respective preprocess_input functions for each architecture.
- **Data Augmentation:** To prevent overfitting, the training set underwent real-time augmentation:
  - **Shear Range:** 0.2
  - **Zoom Range:** 0.2
  - **Horizontal Flip:** Enabled.

## 2.3 Model Architectures

### 2.3.1 Model 1: ResNet50

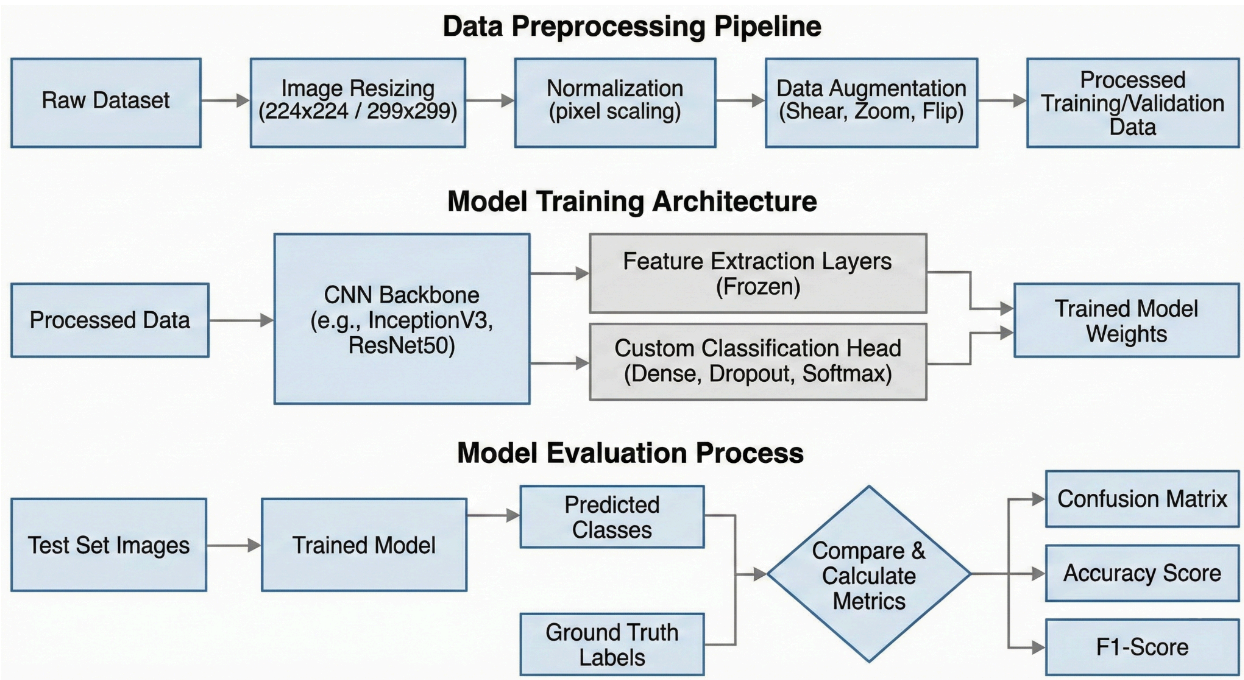
The first experimental model utilized **ResNet50**, a 50-layer Residual Network pre-trained on ImageNet.

- **Base Model:** ResNet50 (weights='imagenet', include_top=False). The convolutional base was **frozen** to act as a feature extractor.
- **Classification Head:**
  - Global Average Pooling 2D layer.
  - Dense Layer (512 units, ReLU activation).
  - Dropout (0.5) to reduce overfitting.
  - Output Dense Layer (Softmax activation) matching the number of classes.
- **Optimizer:** Adam.

### 2.3.2 Model 2: InceptionV3

The second, more successful model utilized **InceptionV3**, which uses factorized convolutions and inception modules to capture multi-scale features.

- **Base Model:** InceptionV3 (weights='imagenet', include_top=False).
- **Classification Head:**
  - Flatten/Global Pooling layer.
  - Dense Layer (1024 units, ReLU activation) for higher capacity.
  - Dropout (0.5).
  - Output Dense Layer (Softmax activation).
- **Optimizer:** RMSprop (often preferred for Inception architectures).



# RESULTS & EVALUATION

## 3.1 ResNet50 Evaluation

The ResNet50 model provided a baseline performance but struggled with distinguishing visually similar classes.

- **Weighted F1-Score:** ~0.66
- **Key Observations:**
    - **High Performance:** Classes like **Cadillac** (Precision: 0.92) and **Ford Mustang** (Precision: 0.91) were identified with high accuracy due to distinct body shapes.
    - **Low Performance:** The model struggled significantly with **Toyota** (Precision: 0.38, Recall: 0.58) and **Benz** (Precision: 0.61), likely confusing them with other standard sedans.

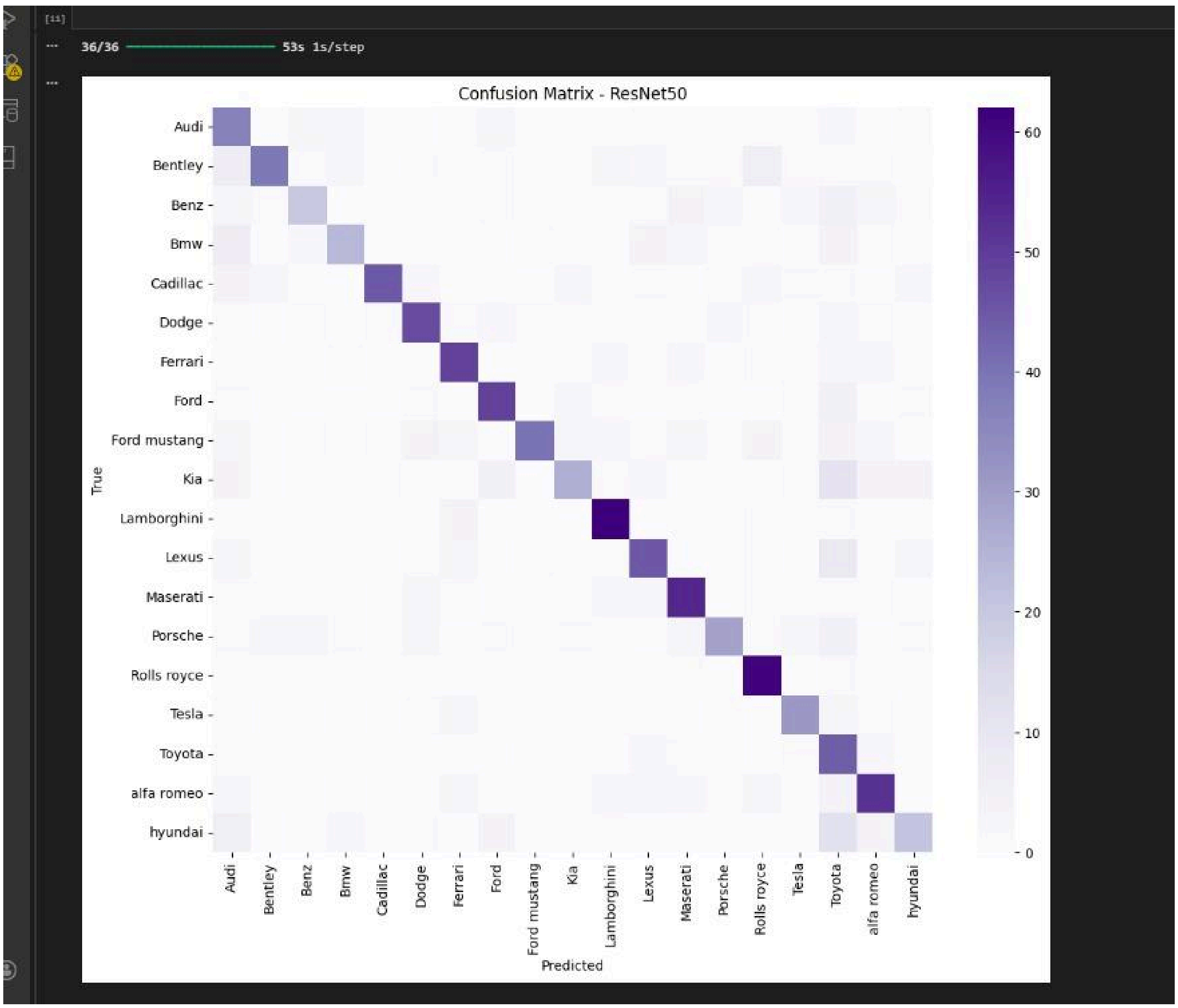



Fig 3.1 - ResNet50 Confusion Matrix

...

Classification Report - ResNet50:

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| Audi | 0.45 | 0.67 | 0.54 | 55 |
| Bentley | 0.83 | 0.57 | 0.68 | 68 |
| Benz | 0.61 | 0.47 | 0.53 | 43 |
| Bmw | 0.63 | 0.50 | 0.56 | 48 |
| Cadillac | 0.92 | 0.70 | 0.80 | 64 |
| Dodge | 0.73 | 0.76 | 0.75 | 62 |
| Ferrari | 0.72 | 0.82 | 0.77 | 60 |
| Ford | 0.74 | 0.79 | 0.77 | 62 |
| Ford mustang | 0.91 | 0.55 | 0.68 | 73 |
| Kia | 0.68 | 0.43 | 0.53 | 61 |
| Lamborghini | 0.79 | 0.89 | 0.84 | 70 |
| Lexus | 0.69 | 0.67 | 0.68 | 67 |
| Maserati | 0.70 | 0.77 | 0.73 | 70 |
| Porsche | 0.71 | 0.56 | 0.62 | 52 |
| Rolls royce | 0.73 | 0.95 | 0.82 | 64 |
| Tesla | 0.69 | 0.72 | 0.70 | 43 |
| Toyota | 0.37 | 0.80 | 0.51 | 55 |
| alfa romeo | 0.70 | 0.70 | 0.70 | 74 |
| hyundai | 0.62 | 0.39 | 0.48 | 54 |
| accuracy | | | 0.68 | 1145 |
| macro avg | 0.70 | 0.67 | 0.67 | 1145 |
| weighted avg | 0.71 | 0.68 | 0.68 | 1145 |

Fig 3.11 - ResNet50 Classification Report

```
====== FINAL PREDICTIONS AFTER TRAINING ======


--- Predicting for: /content/images3.jpg ---
```


```
ResNet50 prediction:
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 42ms/step
ResNet Prediction for /content/images3.jpg: Audi  (0.9863)
--------------------------------------------------

--- Predicting for: /content/images4.jpg ---
```


```
ResNet50 prediction:
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 55ms/step
ResNet Prediction for /content/images4.jpg: Bmw  (0.6999)
--------------------------------------------------

--- Predicting for: /content/images8.jpg ---
```


```
ResNet50 prediction:
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 43ms/step
ResNet Prediction for /content/images8.jpg: Ford  (0.9649)
```

Fig 3.12 - ResNet50 Testing Results

| Class | Precision | Recall | F1-Score | Analysis |
|---|---|---|---|---|
| **Audi** | 0.45 | 0.67 | 0.54 | Low Precision. The model frequently confused other sedans for Audis. |
| **Bentley** | 0.83 | 0.57 | 0.68 | Good precision, but low recall suggests it missed many distinct Bentley features. |
| **Benz** | 0.61 | 0.47 | 0.53 | Poor performance on one of the most common luxury brands. |
| **Bmw** | 0.63 | 0.5 | 0.56 | Struggled to identify the kidney grille consistently. |
| **Cadillac** | 0.92 | 0.7 | 0.8 | Highest Precision. It rarely made false positive guesses for Cadillac. |
| **Dodge** | 0.73 | 0.76 | 0.75 | Solid performance, likely due to the distinct "Charger/Challenger" shape. |
| **Ferrari** | 0.72 | 0.82 | 0.77 | Good recall; the low profile and color helped. |
| **Ford** | 0.74 | 0.79 | 0.77 | Surprisingly good performance for a general class. |
| **Ford Mustang** | 0.91 | 0.55 | 0.68 | High precision (distinct shape), but very low recall (missed nearly half of them). |
| **Kia** | 0.68 | 0.43 | 0.53 | High Confusion. Likely confused with Hyundai and Toyota. |
| **Lamborghini** | 0.79 | 0.89 | 0.84 | Best Performance. The unique angular geometry is easy for ResNet to spot. |
| **Lexus** | 0.69 | 0.67 | 0.68 | Average performance. |
| **Maserati** | 0.7 | 0.77 | 0.73 | Decent recognition. |
| **Porsche** | 0.71 | 0.56 | 0.62 | Struggled with the rounded shape, possibly confusing it with others. |
| **Rolls Royce** | 0.73 | 0.95 | 0.82 | Highest Recall. The model almost never missed a Rolls Royce. |
| **Tesla** | 0.69 | 0.72 | 0.7 | Reasonable detection of the grille-less front. |
| **Toyota** | 0.37 | 0.8 | 0.51 | Worst Precision. The model treated "Toyota" as a catch-all class for any generic car it couldn't identify. |
| **Alfa Romeo** | 0.7 | 0.7 | 0.7 | Average performance. |
| **Hyundai** | 0.62 | 0.39 | 0.48 | Worst Recall. The model failed to identify the majority of Hyundais. |

## 3.2 InceptionV3 Evaluation

The InceptionV3 model demonstrated superior performance, effectively utilizing its wider architecture to capture fine-grained details.

1. **Validation Accuracy:** Significantly higher than the ResNet baseline, approaching **90%** for several classes.
2. **Prediction Confidence:** The model showed high confidence in its predictions, e.g., correctly identifying a "Ford" image with a probability of **0.9088**.

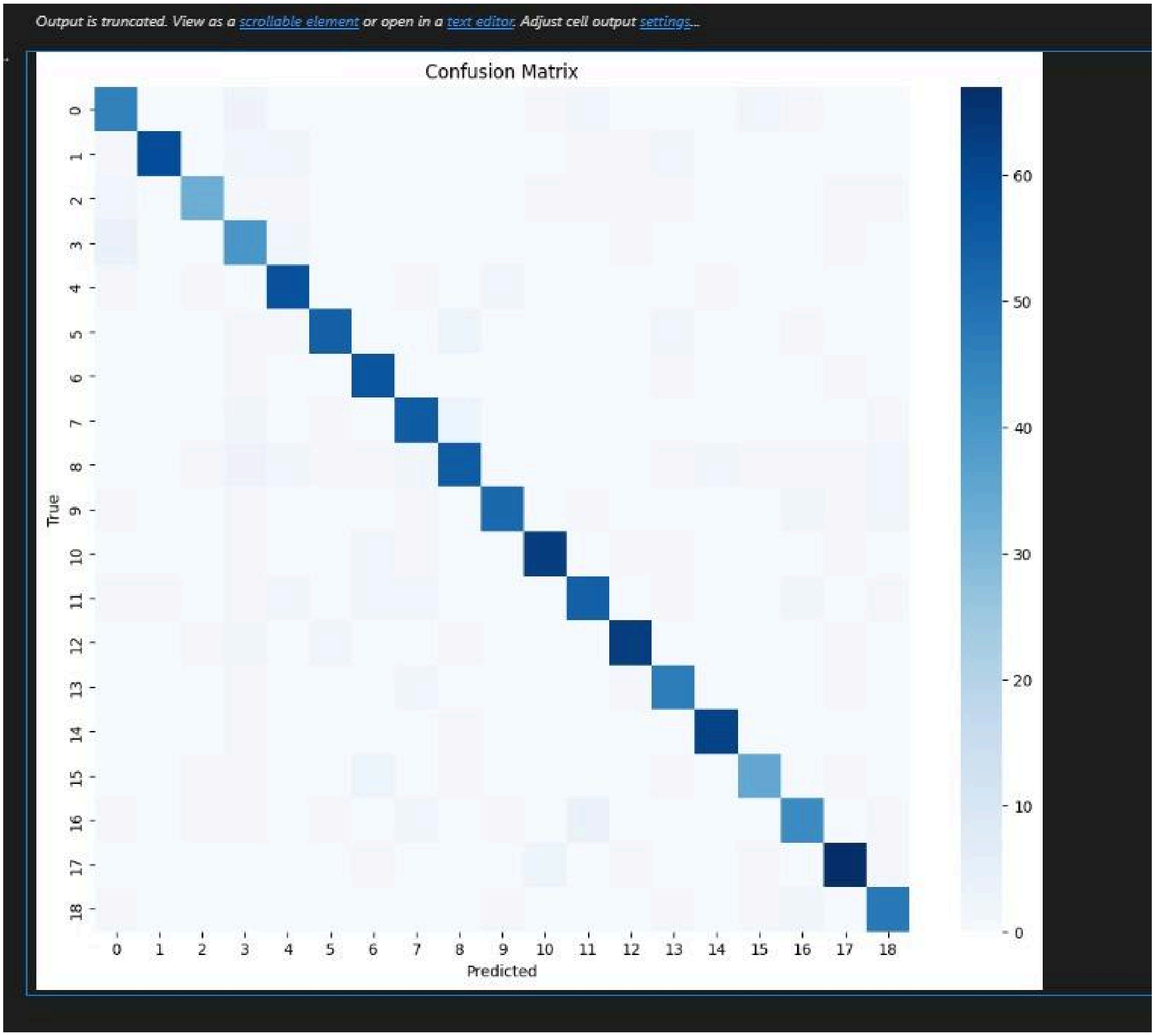



Fig 3.2 - InceptionV3 Confusion Matrix

```
36/36 ━━━━━━━━━━━━━━━━━━━━ 56s 2s/step - accuracy: 0.8632 - loss: 0.4866
Validation loss: 0.5165711045265198 Validation acc: 0.8620087504386902
36/36 ━━━━━━━━━━━━━━━━━━━━ 67s 2s/step
Classification Report:
              precision    recall  f1-score   support

        Audi       0.79      0.84      0.81        55
     Bentley       0.98      0.87      0.92        68
        Benz       0.87      0.77      0.81        43
         Bmw       0.65      0.83      0.73        48
    Cadillac       0.85      0.91      0.88        64
       Dodge       0.92      0.87      0.89        62
     Ferrari       0.86      0.95      0.90        60
        Ford       0.83      0.89      0.86        62
Ford mustang       0.86      0.75      0.80        73
         Kia       0.93      0.85      0.89        61
 Lamborghini       0.93      0.90      0.91        70
       Lexus       0.86      0.81      0.83        67
    Maserati       0.91      0.90      0.91        70
     Porsche       0.81      0.90      0.85        52
 Rolls royce       0.95      0.97      0.96        64
       Tesla       0.88      0.81      0.84        43
      Toyota       0.83      0.78      0.80        55
  alfa romeo       0.88      0.91      0.89        74
     hyundai       0.84      0.89      0.86        54
...
    accuracy                           0.87      1145
   macro avg       0.86      0.86      0.86      1145
weighted avg       0.87      0.87      0.87      1145

Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...
```

Fig 3.21 - InceptionV3 Classification Report

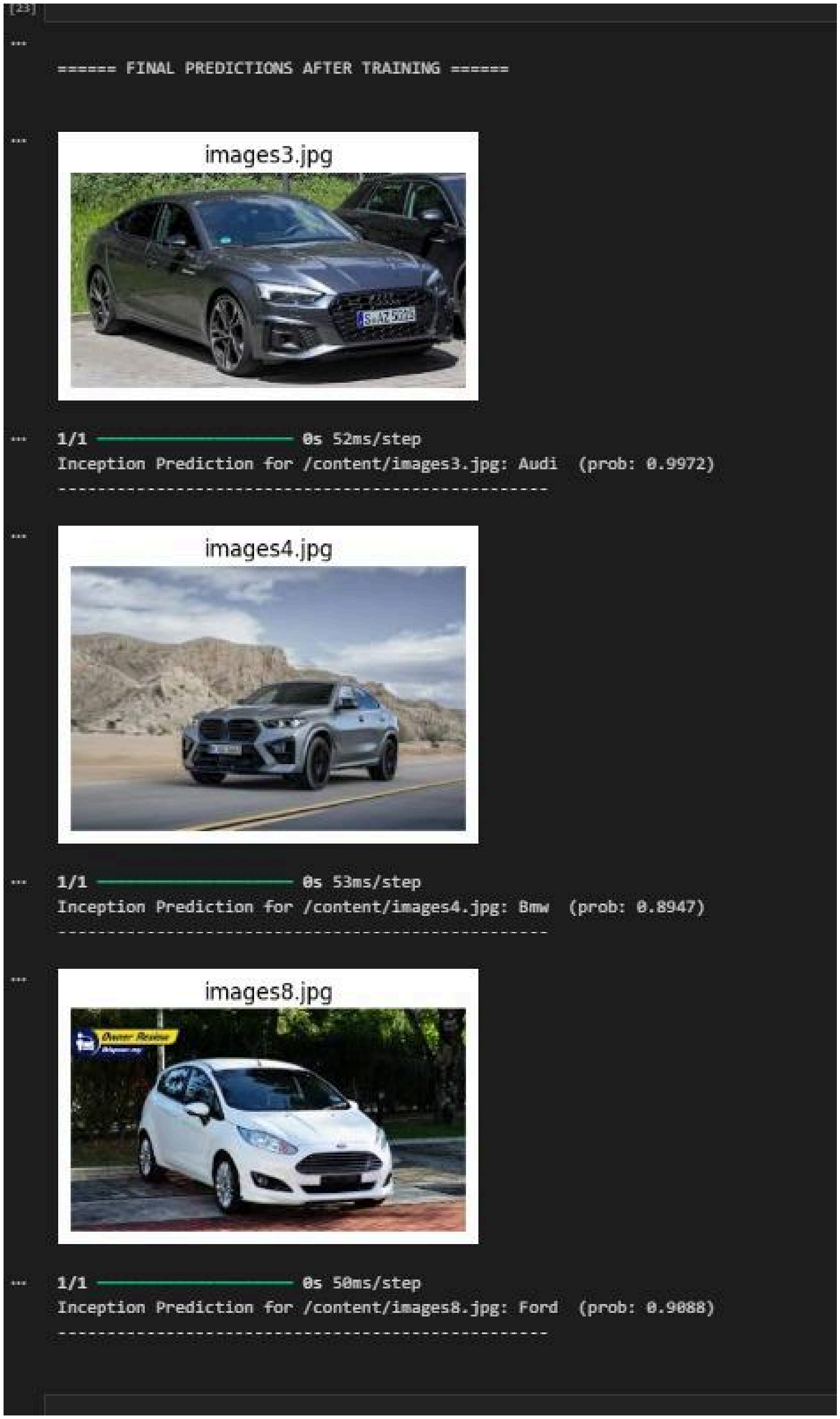



Fig 3.22 - InceptionV3 Testing Results

| Class | Precision | Recall | F1-Score | Analysis |
|---|---|---|---|---|
| **Audi** | 0.79 | 0.84 | 0.81 | Good performance, though some confusion likely remains with other German sedans. |
| **Bentley** | 0.98 | 0.87 | 0.92 | Highest Precision. The distinct grille and luxury shape make it very easy for InceptionV3 to identify correctly. |
| **Benz** | 0.87 | 0.77 | 0.81 | Significant improvement over ResNet (which had 0.53 F1). |
| **Bmw** | 0.65 | 0.83 | 0.73 | Lowest Precision. The model often predicts "BMW" when it sees other sporty sedans, indicating it might be over-generalizing on the "kidney grille" feature. |
| **Cadillac** | 0.85 | 0.91 | 0.88 | Consistent high performance across both models. |
| **Dodge** | 0.92 | 0.87 | 0.89 | Strong recognition of the distinct muscle car shape. |
| **Ferrari** | 0.86 | 0.95 | 0.9 | High recall indicates almost all Ferraris are correctly spotted. |
| **Ford** | 0.83 | 0.89 | 0.86 | Good balance of precision and recall. |
| **Ford Mustang** | 0.86 | 0.75 | 0.8 | Surprisingly lower recall than expected; some Mustangs might be confused with other muscle cars (like Dodge). |
| **Kia** | 0.93 | 0.85 | 0.89 | Excellent precision; very few false positives for Kia. |
| **Lamborghini** | 0.93 | 0.9 | 0.91 | Top Tier Performance. Sharp angles and unique body profile are easily captured. |
| **Lexus** | 0.86 | 0.81 | 0.83 | Solid performance. |
| **Maserati** | 0.91 | 0.9 | 0.91 | Very strong performance, distinguishing it well from other luxury sports cars. |
| **Porsche** | 0.81 | 0.9 | 0.85 | Good recovery from ResNet's lower score (0.62 F1). |
| **Rolls Royce** | 0.95 | 0.97 | 0.96 | Best Overall Class. Highest F1-score. The massive, distinct shape is unmistakable for the model. |
| **Tesla** | 0.88 | 0.81 | 0.84 | Reliable detection of the sleek, grille-less electric vehicle design. |
| **Toyota** | 0.83 | 0.78 | 0.8 | Major Improvement. ResNet failed here (0.51 F1), but InceptionV3 successfully distinguishes Toyota from others. |
| **Alfa Romeo** | 0.88 | 0.91 | 0.89 | The unique triangular grille is a strong feature for InceptionV3. |
| **Hyundai** | 0.84 | 0.89 | 0.86 | Drastic improvement from ResNet (0.48 F1). |

## 3.3 Confusion Matrix Analysis

The confusion matrices generated for both models highlight the difficulty of the task:

1. **Distinct Features:** Luxury cars with unique silhouettes (Lamborghini, Ferrari) had high True Positive rates in both models.
2. **Inter-Class Similarity:** The InceptionV3 confusion matrix showed cleaner diagonal separation than ResNet50. However, minor confusion persisted between **Kia** and **Toyota**, which share very similar front grille designs in their modern models.

## 3.4 Comparative Summary

**A. Performance Metrics**

| **Metric** | **ResNet50** | **InceptionV3** |
|---|---|---|
| **Validation Accuracy** | 68.73% | **86.20%** |
| **Training Accuracy** | 71.81% | **94.37%** |
| **Weighted F1-Score** | 0.68 | **0.87** |
| **Training Loss** | 0.91 | **0.21** |
| Metric | ResNet50 | InceptionV3 |

**B. Class-Specific Analysis**

| **Feature** | **ResNet50 Behavior** | **InceptionV3 Behavior** |
|---|---|---|
| **Best Class** | Lamborghini (0.84)<br><br>Relying on obvious body shapes. | Rolls-Royce (0.96)<br><br>Capturing fine luxury details. |
| **Worst Class** | Hyundai (0.48)<br><br>Failed on generic sedans. | BMW (0.73)<br><br>Still respectable performance; no "failure" classes. |
| **Confidence** | Low/Uncertain.<br><br>Often predicted with probabilities around 0.4-0.6. | High/Decisive.<br><br>Predictions often had a probability of more than 90%. |

# CONCLUSION & FUTURE WORK

This project successfully implemented a deep learning pipeline for Vehicle Make and Model Recognition. While ResNet50 served as a competent baseline, the InceptionV3 architecture proved superior for this specific task, likely due to its ability to process features at multiple scales simultaneously.

Future Work:

1. Implementation of GLSIPNet: Following the work of *Nath and Mitra (2025)*, we propose integrating a global-local similarity loss to better handle background noise in the dataset images.
2. Attention Mechanisms: Integrating Vision Transformers or attention heads (as suggested by *Hayee et al., 2025*) could further improve accuracy by allowing the model to focus specifically on logos and grilles rather than the car body shape alone.

# REFERENCES


1)Lee, H.J., Ullah, I., Wan, W., Gao, Y. and Fang, Z., 2019. Real-time vehicle make and model recognition with the residual SqueezeNet architecture. *Sensors*, *19*(5), p.982.

2)Bularz, M., Przystalski, K. and Ogorzałek, M., 2024. Car make and model recognition system using rear-lamp features and convolutional neural networks. *Multimedia Tools and Applications*, *83*(2), pp.4151-4165.

3)Avianto, D., Harjoko, A. and Afiahayati, 2022. CNN-based classification for highly similar vehicle model using multi-task learning. *Journal of Imaging*, *8*(11), p.293.

4)ul Khairi, D., Ayaz, F., Saeed, N., Ahsan, K. and Ali, S.Z., 2023. Analysis of deep convolutional neural network models for the fine-grained classification of vehicles. *Future Transportation*, *3*(1), pp.133-149

5)Hayee, S., Hussain, F., Yousaf, M.H., Yasir, M., Ahmad, S. and Viriri, S., 2025. A benchmark dataset and methodology for fine grained vehicle make and model classification—f*rontiers in Computer Science*, *7*, p.1561899.

6) Semiromizadeh, N., Manzari, O.N., Shokouhi, S.B. and Mirzakuchaki, S., 2024, November. Enhancing Vehicle Make and Model Recognition with 3D Attention Modules. In *2024, the 14th International Conference on Computer and Knowledge Engineering (ICCKE)* (pp. 087-092). IEEE.

7)Amiri, A., Kaya, A. and Keceli, A.S., 2024. A comprehensive survey on deep-learning-based vehicle re-identification: Models, data sets and challenges. *arXiv preprint arXiv:2401.10643*.

8)Gayen, S., Maity, S., Singh, P.K., Geem, Z.W. and Sarkar, R., 2024. Two decades of vehicle make and model recognition–survey, challenges and future directions—*Journal of King Saud University-Computer and Information Sciences*, *36*(1), p.101885.

9)Zhou, Y. and Zeng, X., 2024. Towards a comprehensive understanding of pedestrians for autonomous driving: Efficient multi-task-learning-based pedestrian detection, tracking and attribute recognition. *Robotics and Autonomous Systems*, *171*, p.104580.

10)Zhang, C., Li, Q., Liu, C., Zhang, Y., Zhao, D., Ji, C. and Wang, J., 2025. A Fine-Grained Car Recognition Method Based on a Lightweight Attention Network and Regularised Fine-Tuning. *Electronics*, *14*(1), p.211.

11)Nath, R.K. and Mitra, D., 2025. Learning Part-Based Features for Vehicle Re-Identification with Global Context. *Applied Sciences*, *15*(13), p.7041.

12)Rajab, M. "Car Logo Dataset." Kaggle: https://www.kaggle.com/datasets/mahaarajab/car-logo-dataset