Prepared by group 8

Knowledge Discovery and Data Mining

Multi-object Classification in Autonomus Driving

Section 513-C

Team Members

Bhushan Asati
20030940

Suyash Madhavi
20030933

Rujuta Dabake
20033905

Anirudh Sharma
20034386

Introduction and Problem Statement

- Objective:

  Developing a robust classification model for objects in autonomous driving environments.
- Categories: Focusing on two broad categories: Human and Vehicle.
- Significance:
  - Supports real-time decision-making in traffic scenarios.
  - Enhances the safety and efficiency of autonomous vehicles.

Dataset Description

- **Dataset Used**
  - KITTI 3D Object Detection Dataset.
- **Multimodal Data:**
  - Combines sensor data from images and LiDAR.
- **Original Categories:**
  - Pedestrian, Cyclist, Car, Truck, Van, Tram, Miscellaneous.
- **Merged Categories into two main classes:**
  - Human: Pedestrian, Cyclist.
  - Vehicle: Car, Truck, Van, Tram.
- **Dataset Size:**
  - Over 7,481 labeled frames with detailed annotations, including 3D bounding boxes.
  - Size of dataset - 30 GB.

Visual Representation

Data Distribution

Data Preprocessing

- Extract and preprocess images from the KITTI dataset.
- Normalize images to the range [0, 1].
- Map object labels into the defined categories (Human and Vehicle).

Models Used

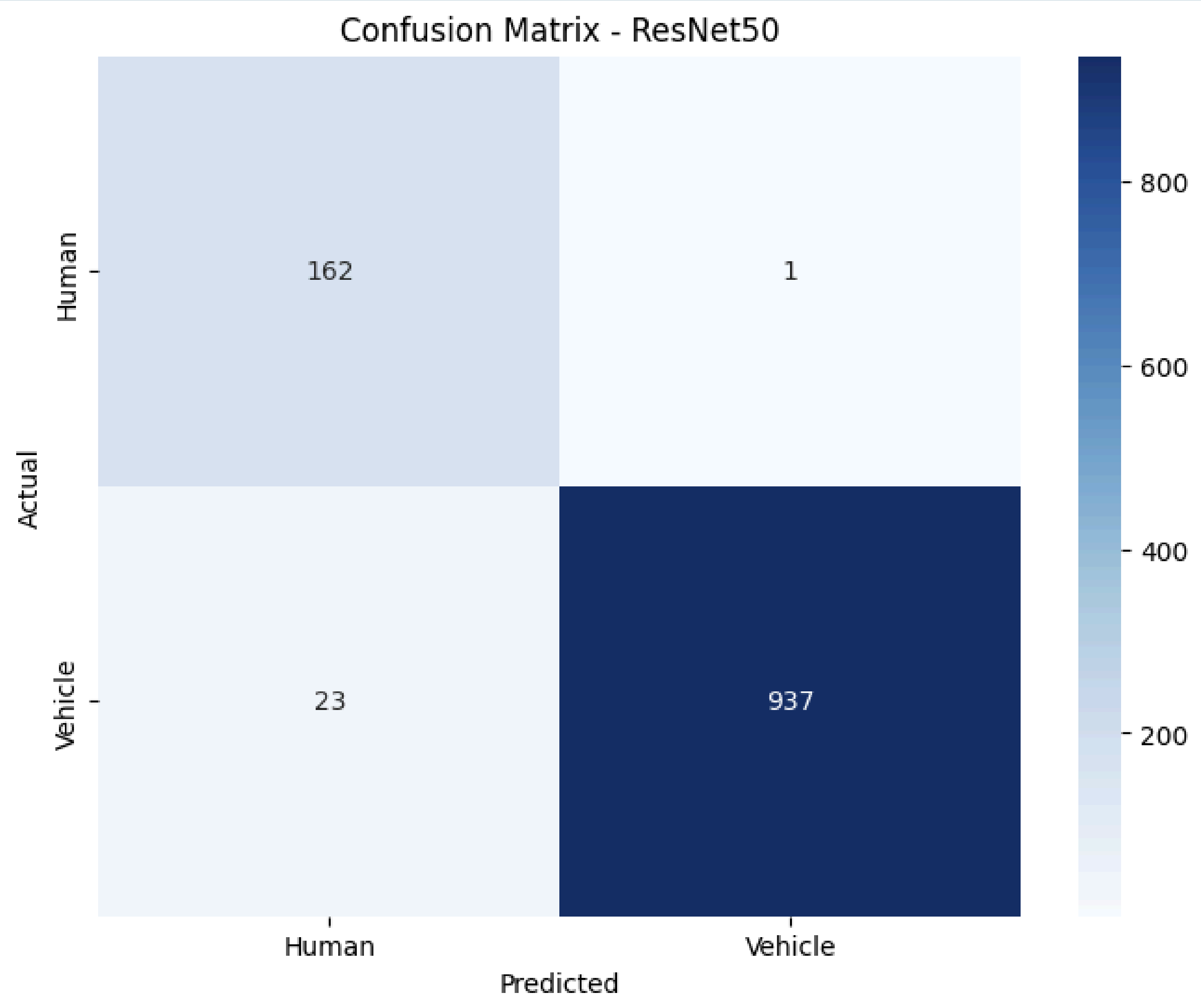
- ResNet50
- EfficientNetB0
- MobileNetV2
- Inception-V3
- DenseNet-121
- Faster R-CNN
- XGBoost
- Random Forest

ResNet50

- **Why Chosen**
  - Residual connections prevent vanishing gradients in deep networks.
  - Proven reliability for feature extraction in diverse datasets.
- **Architecture**
  - Consists of residual blocks with skip connections.
  - Pre-trained on ImageNet for transfer learning.
- **Fine-Tuning**
  - Fully connected layer modified for Human/Vehicle classification.
  - Batch size: 32, Learning rate: 0.001, Optimizer: SGD.
  - Normalized input images for stable training.
- **Applications**
  - Image recognition, medical diagnostics, and object tracking.

Classification Report for ResNet50:

	precision	recall	f1-score	support
Human	0.88	0.99	0.93	163
Vehicle	1.00	0.98	0.99	960
accuracy			0.98	1123
macro avg	0.94	0.98	0.96	1123
weighted avg	0.98	0.98	0.98	1123



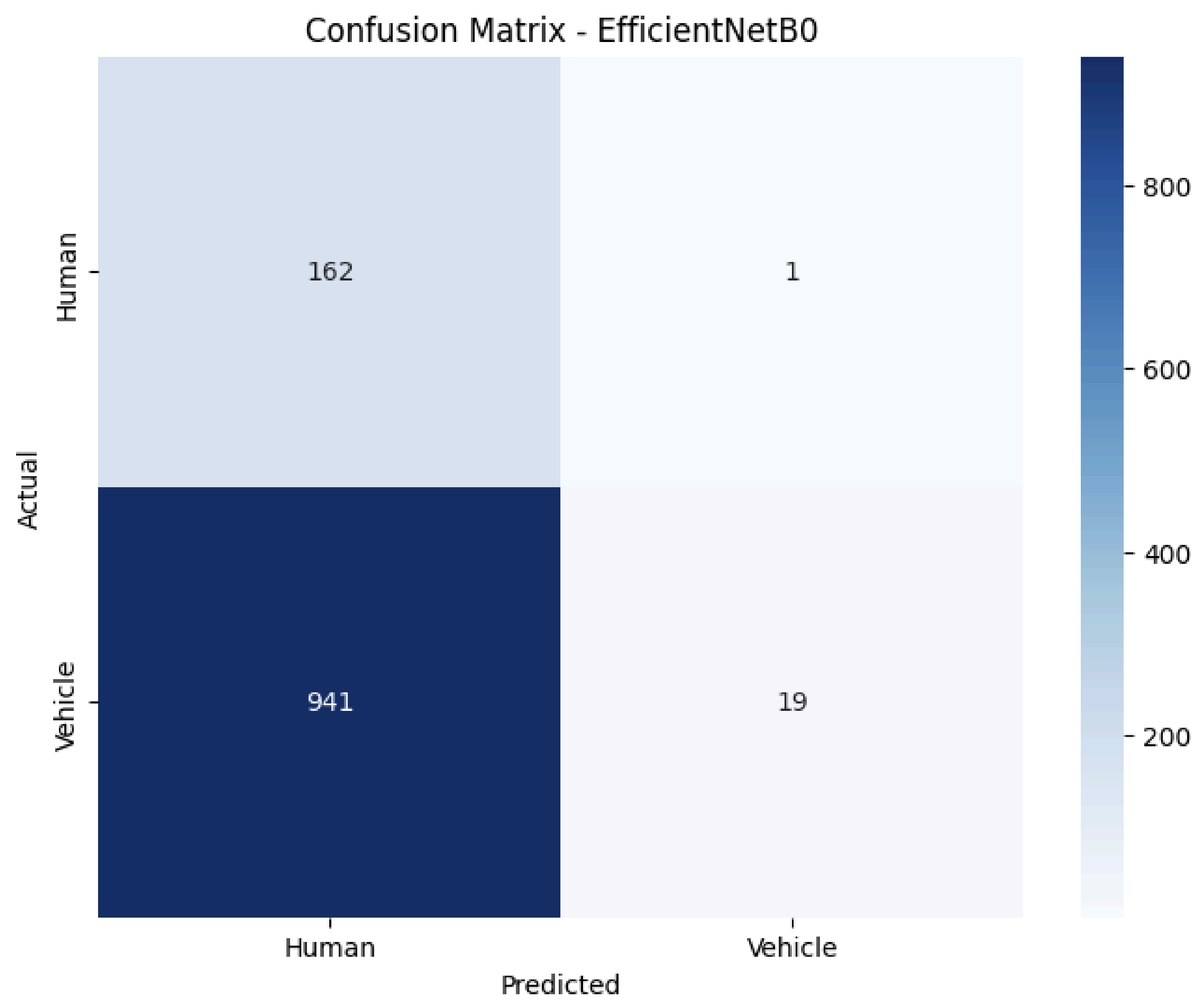
EfficientNetBo

- **Why Chosen**
  - Balances accuracy and efficiency using compound scaling.
  - Suitable for scalable models on various hardware setups.
- **Architecture**
  - Uniformly scales width, depth, and resolution.
  - Combines depth wise separable convolutions and squeeze-and-excitation modules.
- **Fine-Tuning**
  - Adapted final layer for binary classification.
  - Batch size: 16, Learning rate: 0.0005, Optimizer: Adam.
  - Image resizing to match input dimensions (224x224).
- **Applications**
  - Autonomous driving, drone navigation, real-time applications.

```
Classification Report for EfficientNetB0:
              precision    recall  f1-score   support

       Human       0.15      0.99      0.26       163
     Vehicle       0.95      0.02      0.04       960

    accuracy                           0.16      1123
   macro avg       0.55      0.51      0.15      1123
weighted avg       0.83      0.16      0.07      1123
```

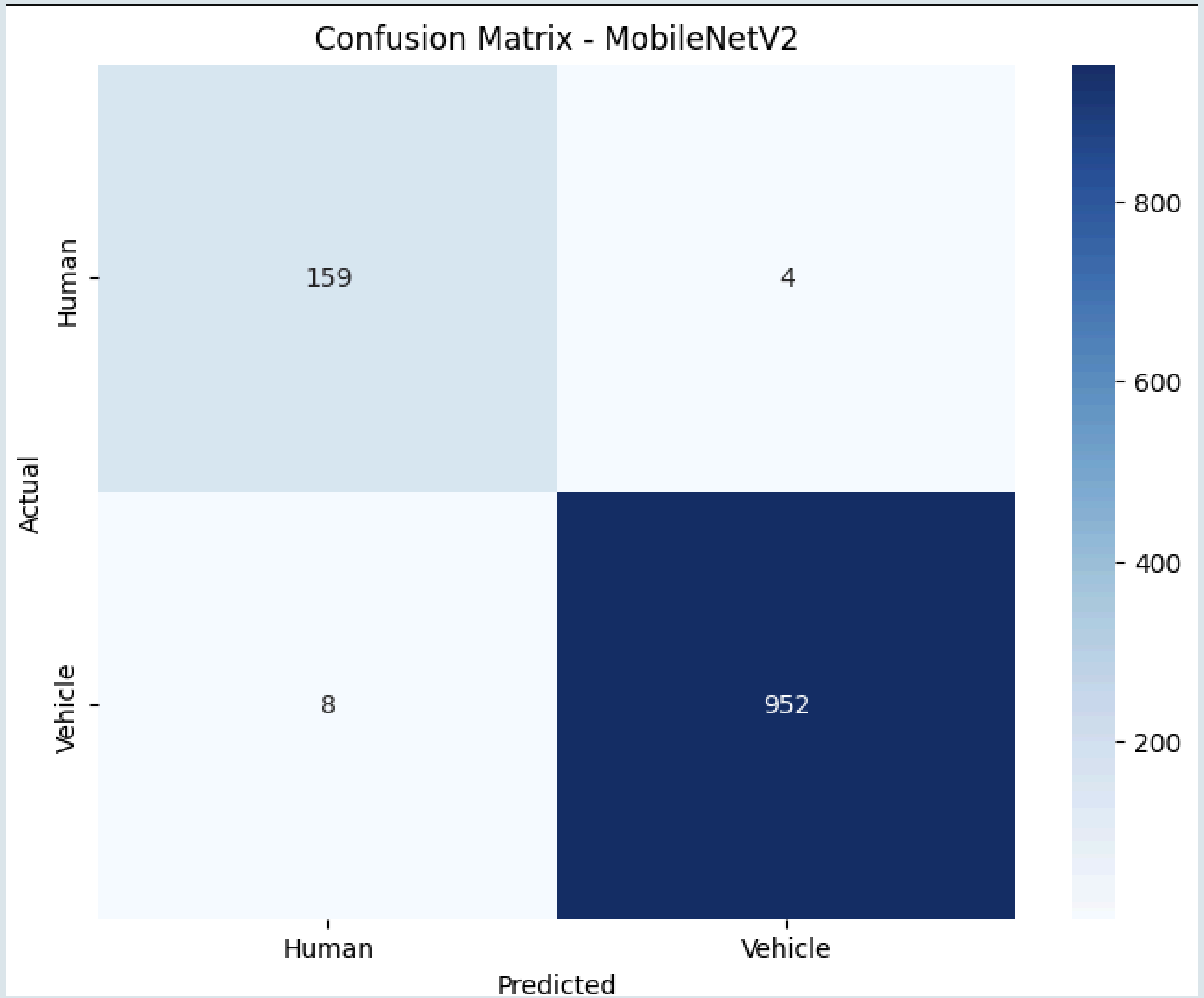


MobileNetV2

- **Why Chosen**
  - Lightweight and efficient, ideal for deployment on mobile devices.
  - Depthwise separable convolutions reduce computational cost.
- **Architecture**
  - Uses inverted residuals and linear bottlenecks.
  - Designed for speed and performance on resource-constrained environments.
- **Fine-Tuning**
  - Adjusted final layer for Human/Vehicle classification.
  - Batch size: 32, Learning rate: 0.001, Optimizer: RMSProp.
  - Applied data augmentation to enhance generalization.
- **Applications**
  - Mobile and embedded systems, AR/VR applications, and IoT devices.

Classification Report for MobileNetV2:

	precision	recall	f1-score	support
Human	0.95	0.98	0.96	163
Vehicle	1.00	0.99	0.99	960
accuracy			0.99	1123
macro avg	0.97	0.98	0.98	1123
weighted avg	0.99	0.99	0.99	1123



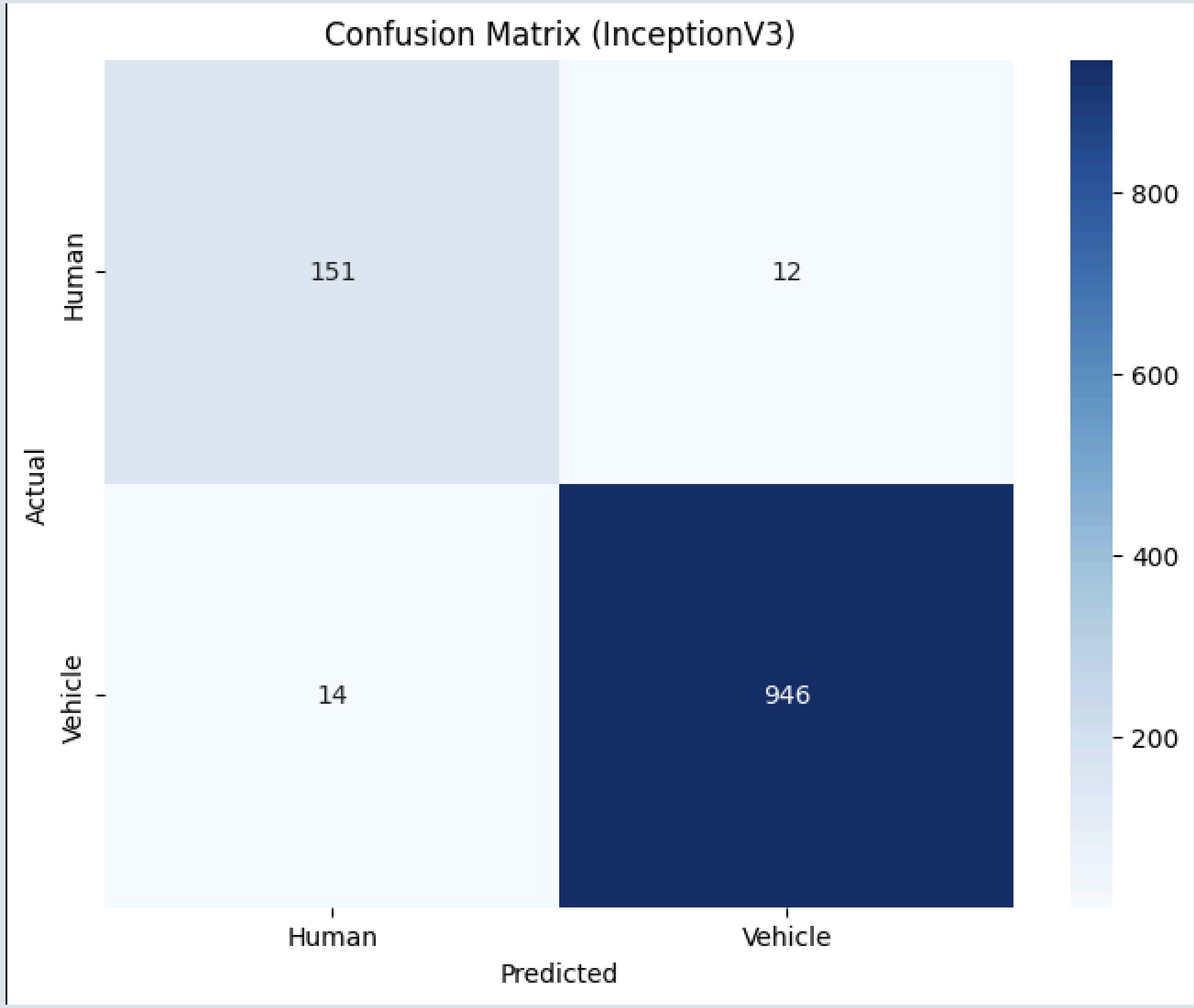
Inception-V3

- **Why Chosen**
  - Known for multi-scale feature extraction using inception modules.
  - Ideal for tasks requiring a blend of speed and accuracy.
- **Architecture**
  - Stacks inception modules for parallel convolutional layers.
  - Employs factorized convolutions to reduce computation.
  - Uses auxiliary classifiers to combat vanishing gradients.
- **Fine-Tuning**
  - Modified final dense layer for binary classification.
  - Batch size: 32, Learning rate: 0.0001, Optimizer: RMSProp.
  - Augmented images with random rotations and flips.
- **Applications**
  - Image classification, fine-grained object detection, medical imaging.

```
Test Accuracy (InceptionV3): 97.68%
36/36 [==============================] - 8s 173ms/step
Classification Report (InceptionV3):
              precision    recall  f1-score   support

       Human       0.92      0.93      0.92       163
     Vehicle       0.99      0.99      0.99       960

    accuracy                           0.98      1123
   macro avg       0.95      0.96      0.95      1123
weighted avg       0.98      0.98      0.98      1123
```



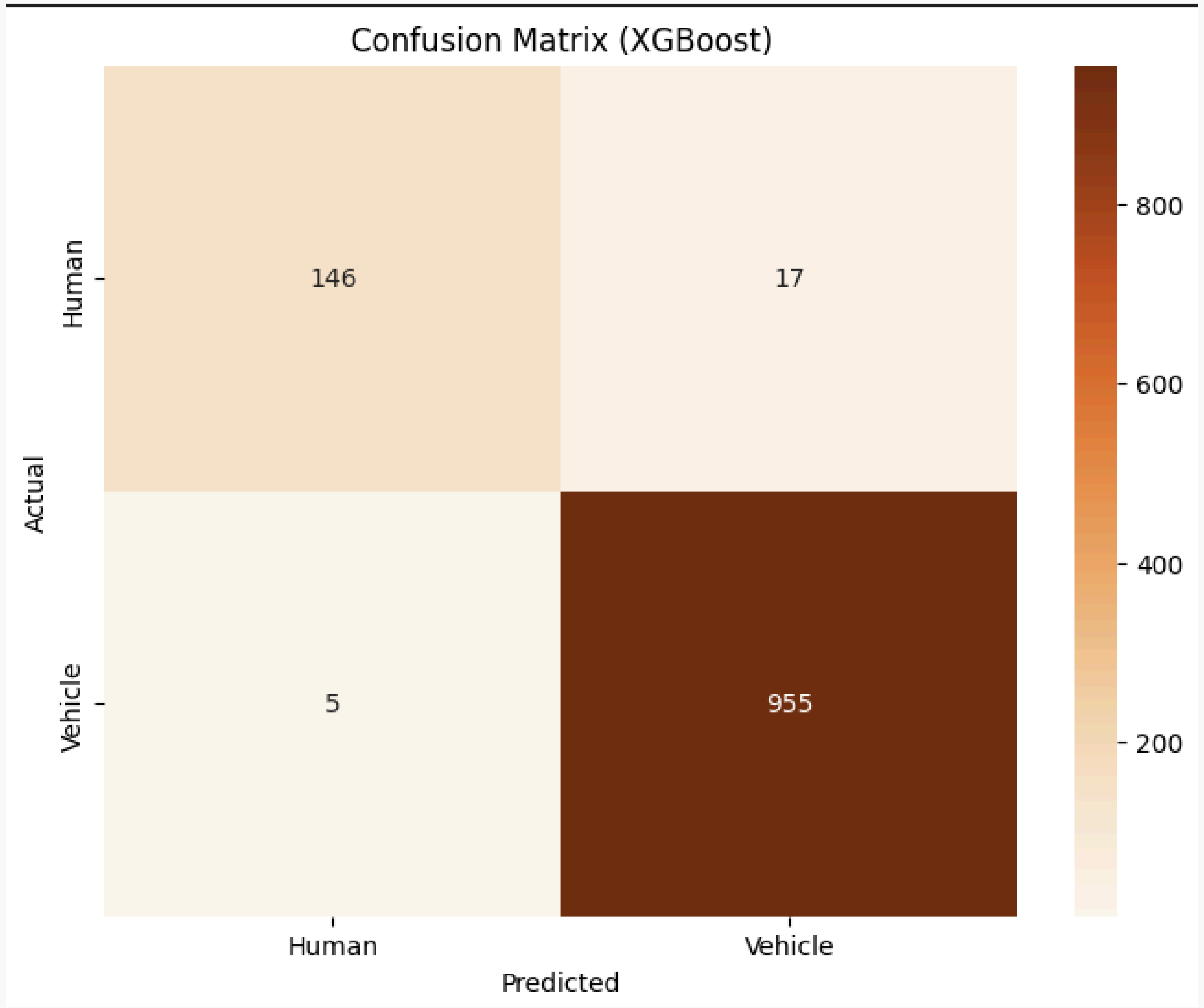
XGBoost

- **Why Chosen**
  - State-of-the-art gradient boosting technique for structured/tabular data.
  - Handles missing values and imbalanced datasets effectively.
- **Architecture**
  - Boosts weak learners (decision trees) iteratively.
  - Scales well with large datasets using parallel processing.
  - Regularization (L1/L2) minimizes overfitting.
- **Fine-Tuning**
  - Learning rate: 0.05, Max depth: 6, Estimators: 100.
  - Applied early stopping to prevent overfitting.
  - Features scaled using StandardScaler before training.
- **Applications**
  - Fraud detection, credit scoring, and feature-rich datasets like medical records.

```
Classification Report (XGBoost):
              precision    recall  f1-score   support

       Human       0.97      0.90      0.93       163
     Vehicle       0.98      0.99      0.99       960

    accuracy                           0.98      1123
   macro avg       0.97      0.95      0.96      1123
weighted avg       0.98      0.98      0.98      1123
```

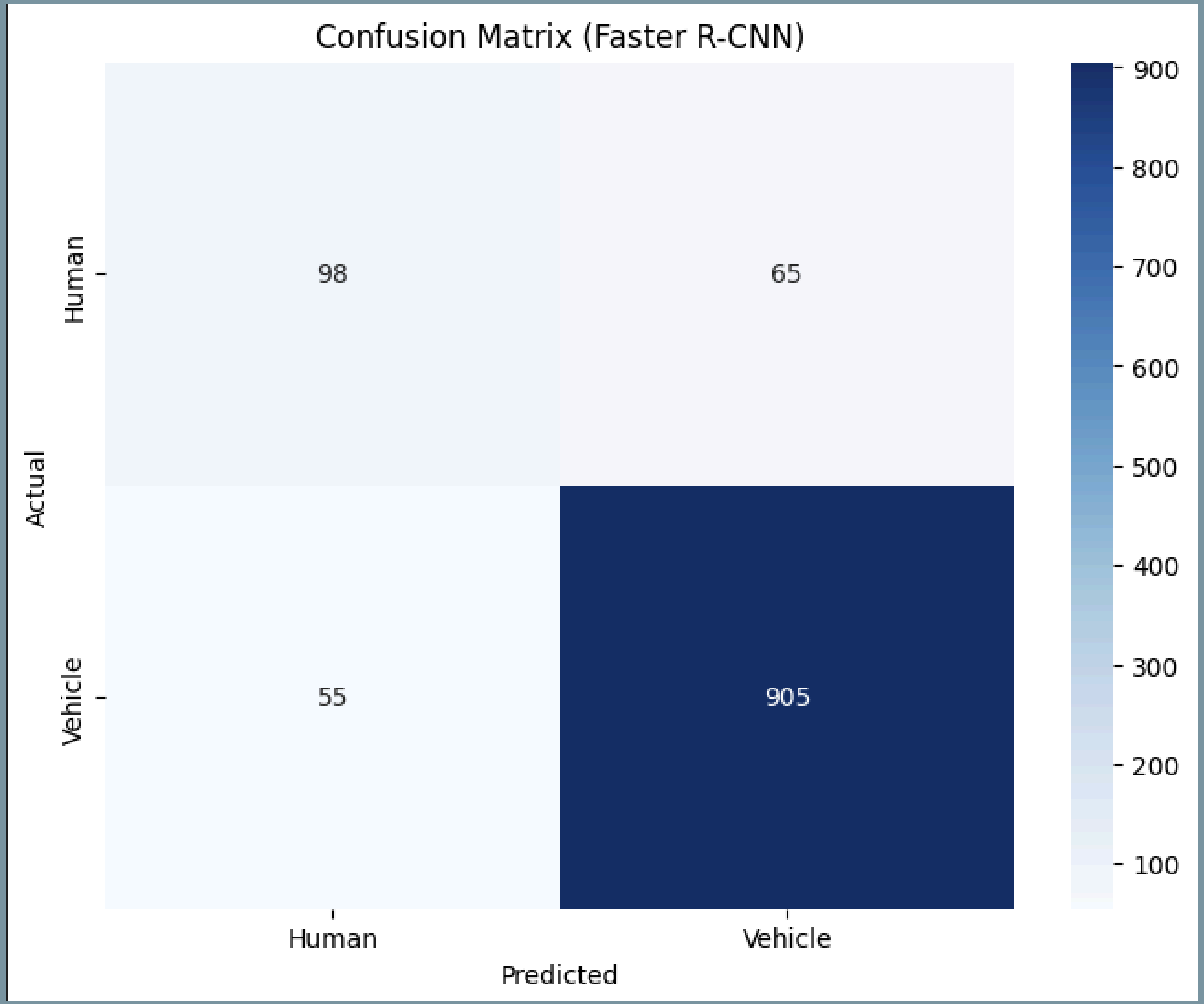


Faster R-CNN

- **Why Chosen**
  - Specialized for object detection with high precision and recall.
  - Generates region proposals dynamically, improving detection efficiency.
- **Architecture**
  - Combines a Region Proposal Network (RPN) with a CNN-based classifier.
  - Uses bounding boxes and classification scores for localization.
- **Fine-Tuning**
  - Customized for detecting Human and Vehicle categories.
  - Batch size: 16, Learning rate: 0.0001, Optimizer: Adam.
  - Augmented dataset with bounding box annotations.
- **Applications**
  - Autonomous driving, security surveillance, and robotics.

Classification Report (Faster R-CNN):

	precision	recall	f1-score	support
Human	0.64	0.60	0.62	163
Vehicle	0.93	0.94	0.94	960
accuracy			0.89	1123
macro avg	0.79	0.77	0.78	1123
weighted avg	0.89	0.89	0.89	1123

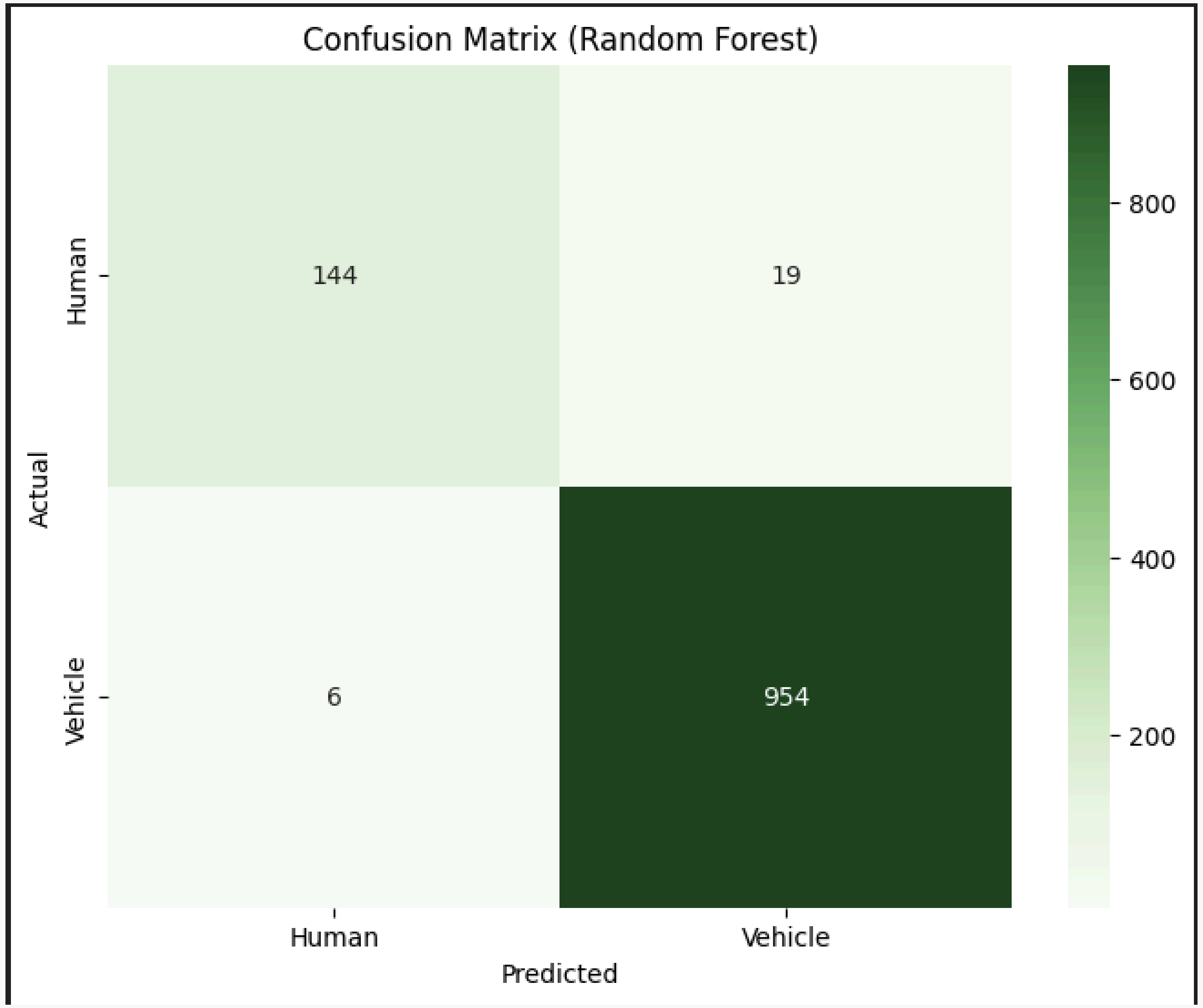


Random Forest

- **Why Chosen**
  - Ensemble learning algorithm known for handling high-dimensional datasets.
  - Provides interpretability through feature importance.
- **Architecture**
  - Combines multiple decision trees via bagging.
  - Outputs class prediction as the majority vote of individual trees.
- **Fine-Tuning**
  - Number of trees: 200, Max depth: 10, Criterion: Gini.
  - Applied feature selection to reduce input dimensionality.
- **Applications**
  - Classification tasks like credit scoring, customer segmentation, and fraud detection.

Classification Report (Random Forest):

	precision	recall	f1-score	support
Human	0.96	0.88	0.92	163
Vehicle	0.98	0.99	0.99	960
accuracy			0.98	1123
macro avg	0.97	0.94	0.95	1123
weighted avg	0.98	0.98	0.98	1123



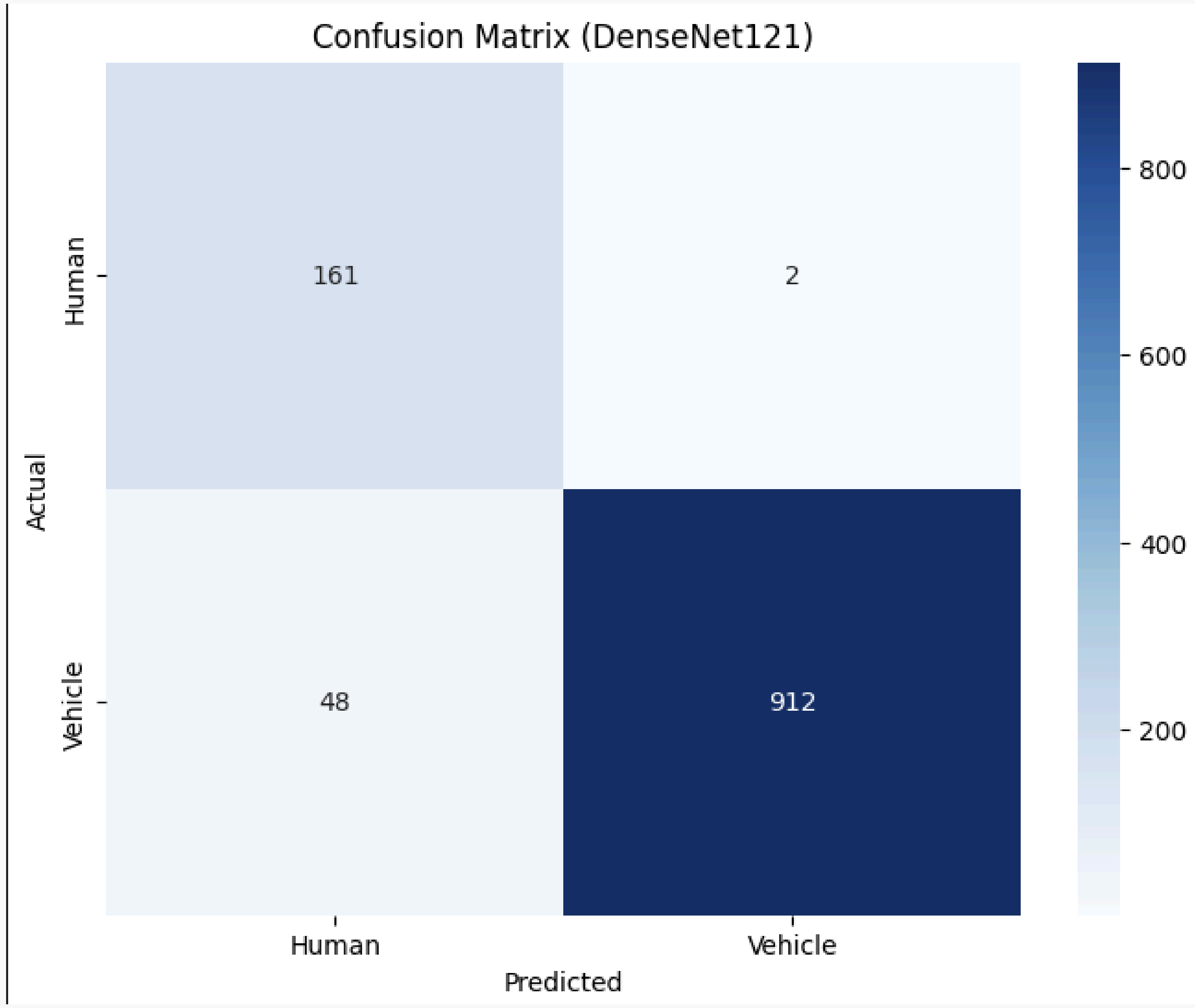
DenseNet121

- **Why Chosen**
  - Feature reuse enhances gradient flow and reduces overfitting.
  - Efficient parameters, ideal for small datasets.
- **Architecture**
  - Dense blocks connect all layers, preventing vanishing gradients.
  - Pre-trained on ImageNet for transfer learning.
- **Fine-Tuning**
  - Fully connected layer modified for Human/Vehicle classification.
  - Batch size: 32, Learning rate: 0.001, Optimizer: Adam.
  - Data augmentation: rotation, flipping, zooming.
- **Applications**
  - Medical imaging, autonomous driving, surveillance systems.

```
Test Accuracy (DenseNet121): 95.55%
36/36 [==============================] - 21s 524ms/step
Classification Report (DenseNet121):
              precision    recall  f1-score   support

       Human       0.77      0.99      0.87       163
     Vehicle       1.00      0.95      0.97       960

    accuracy                           0.96      1123
   macro avg       0.88      0.97      0.92      1123
weighted avg       0.96      0.96      0.96      1123
```



Results

Model	Accuracy	Human Precision	Human Recall	Human F1	Vehicle Precision	Vehicle Recall	Vehicle F1
ResNet50	0.98	0.88	0.99	0.93	1.00	0.98	0.99
EfficientNetB0	0.16	0.15	0.99	0.26	0.95	0.02	0.04
MobileNetV2	0.99	0.95	0.98	0.96	1.00	0.99	0.99
DenseNet121	0.96	0.77	0.99	0.87	1.00	0.95	0.97
InceptionV3	0.98	0.92	0.93	0.92	0.99	0.99	0.99
Random Forest	0.98	0.96	0.88	0.92	0.98	0.99	0.99
Faster R-CNN	0.89	0.64	0.60	0.62	0.93	0.94	0.94
XGBoost	0.98	0.96	0.94	0.95	0.99	0.99	0.99

Best-Performing Model: DenseNet121

- Highest Accuracy: Achieved 96% overall accuracy.

- Exceptional Metrics:
  - Human Class: Precision = 0.77, Recall = 0.99, F1-Score = 0.87
  - Vehicle Class: Precision = 1.00, Recall = 0.99, F1-Score = 0.97

- Advantages:
  - Reduces the Vanishing Gradient Problem.
  - Reduction in Overfitting
  - Encourages Diverse Feature Learning

Thank you