IMAGE CLASSIFICATION

Group 8

Anirudh Pal(22040)
Vipul Sharma(22369)
Sundram Anand (22342)
Gargi Sujit Pande(22127)

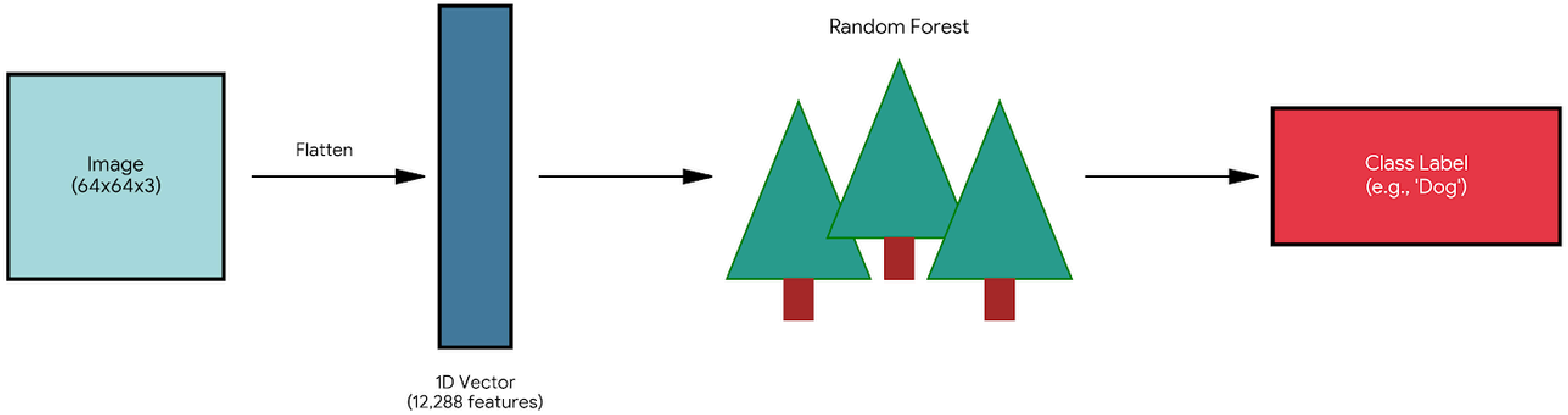
RANDOM FOREST

Anirudh Pal (22040)

CAN IT CLASSIFY

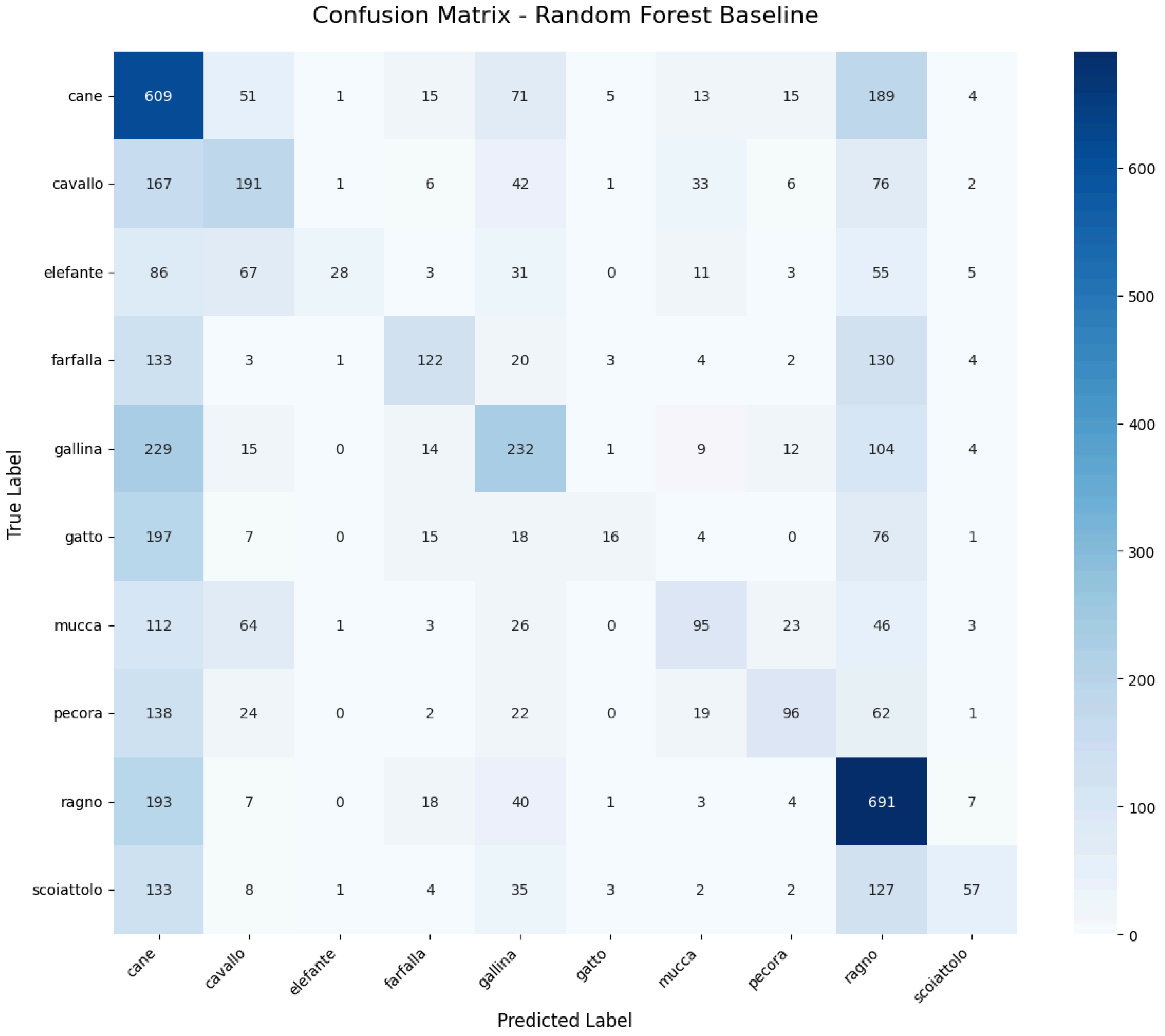
- Does it classify the images correctly , and how accurately ?

METHODOLOGY



- Input Data: Custom 10 class animal dataset.
- Preprocessing: Flattened 2D images into 1D vectors.
- Feature Space: 12,288 independent pixel features per image.
- Model: Random Forest Classifier (100 Estimators).
- Core Assumption: Can the model learn to classify objects based solely on pixel intensity patterns?

OVERALL PERFORMANCE

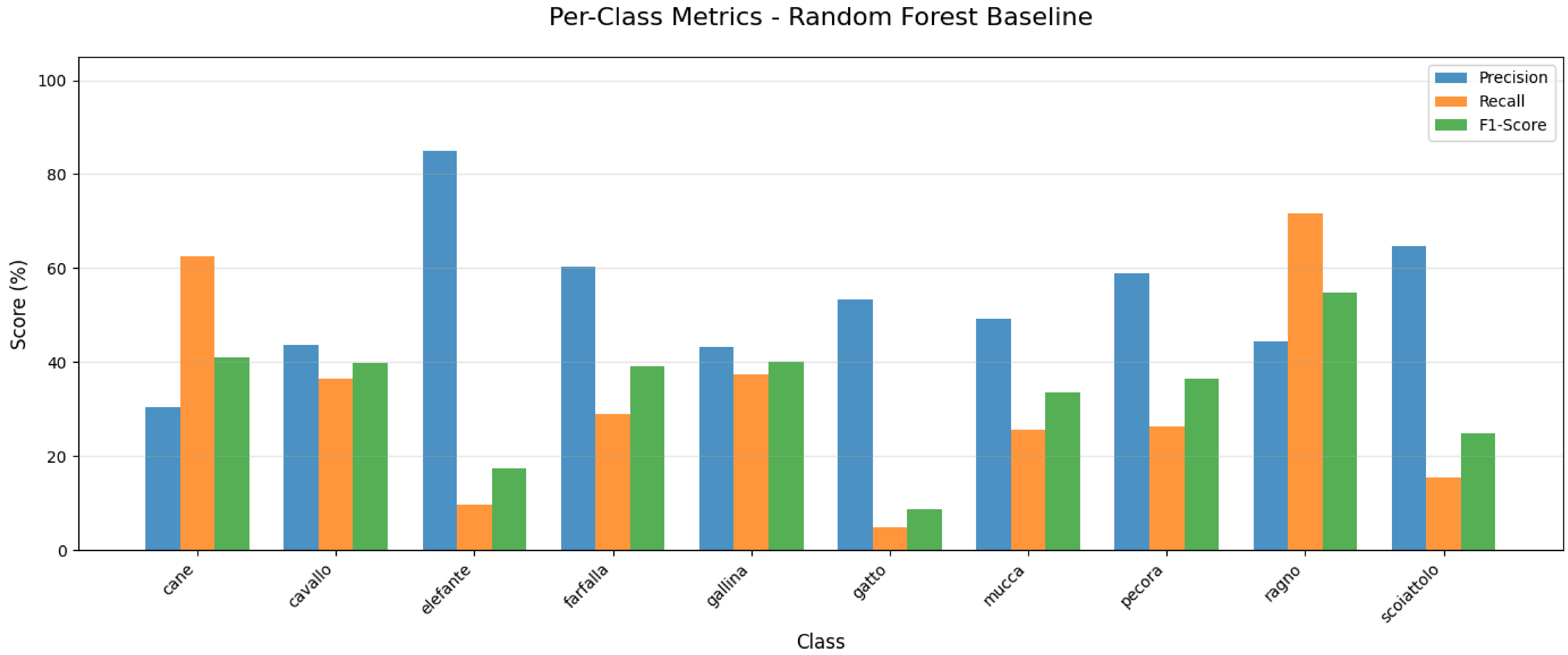


- The Confusion matrix shows that the accuracy ,precision, recall and F1-score wasn't good.

```
=========================================
OVERALL METRICS
=========================================
Accuracy:  40.81%
Precision: 53.33%
Recall:    31.86%
F1-Score:  33.57%
```

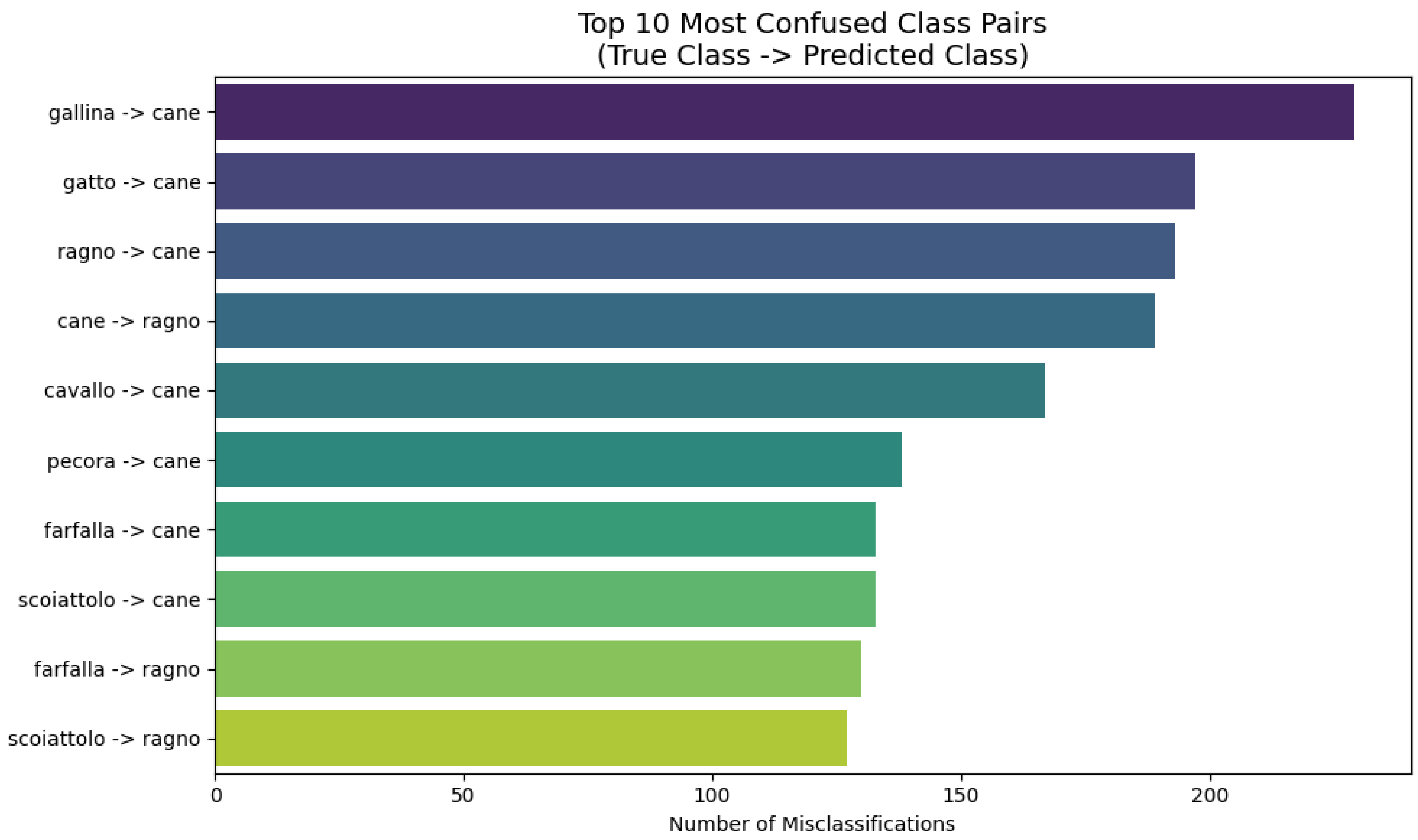
(The values of the evaluation metrics)

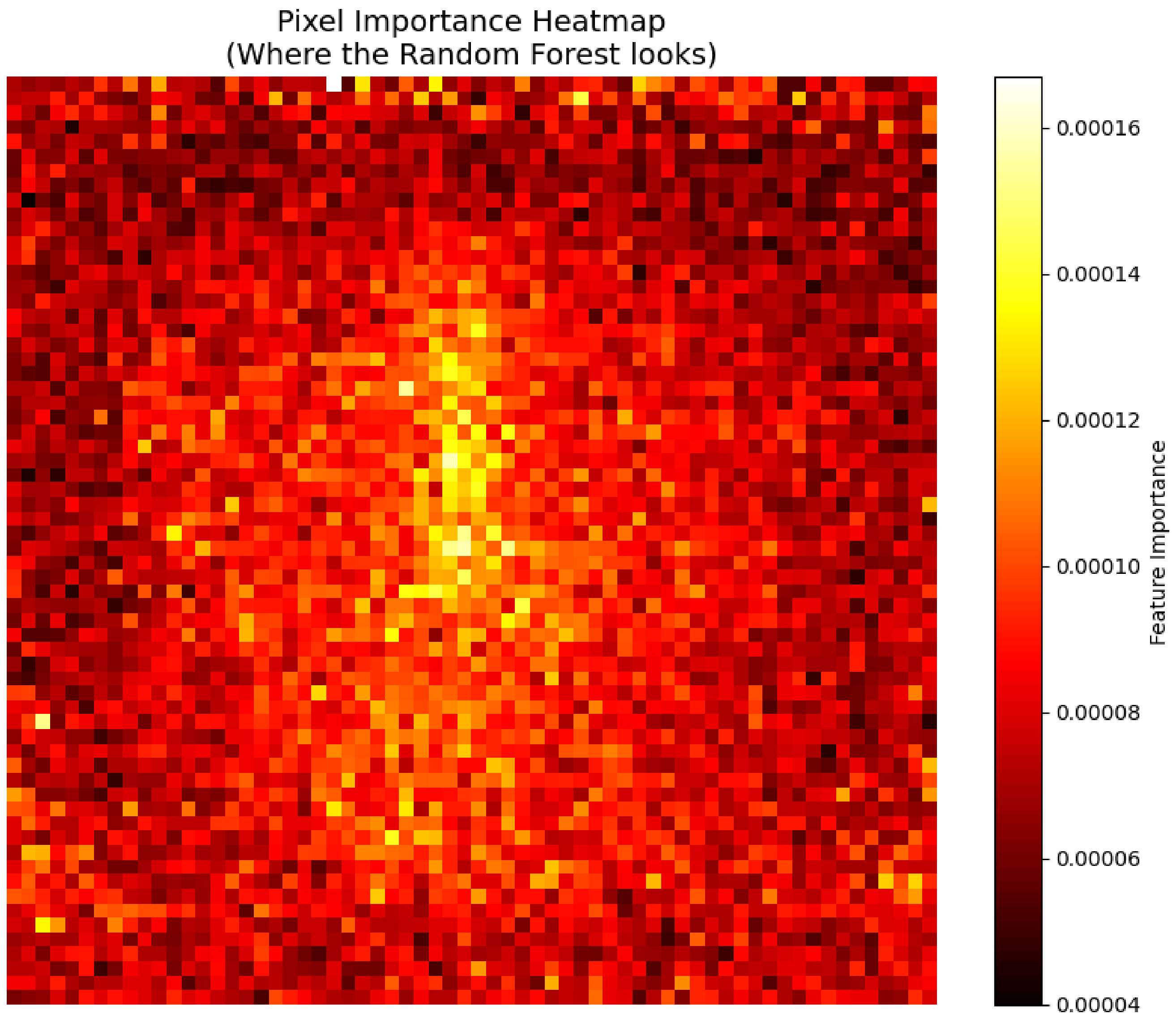
CLASS-SPECIFIC BREAKDOWN



- The "Easy" Classes: Ragno (Spider) and Cane (Dog) show the highest accuracy (Recall > 60%).
- The "Hard" Classes: Gatto (Cat) and Scoiattolo (Squirrel) are near total failure (< 15% Accuracy).
- The model struggles to differentiate between animals with similar textures (e.g., Cow vs. Horse) because it cannot see the body shape.

WHY IT FAILS

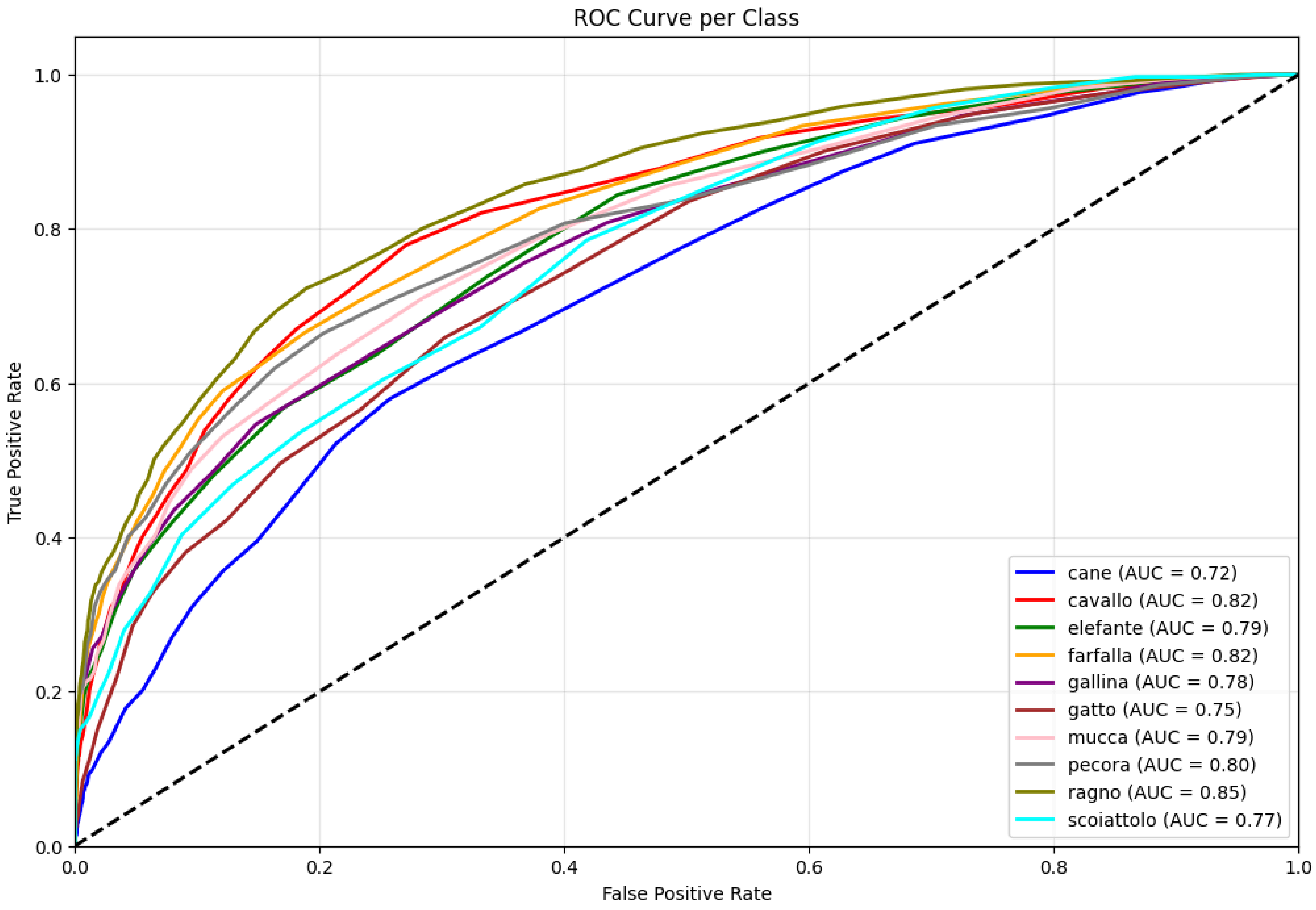




- The model most frequently confuses Cane (Dog) with Gallina (Hen) and on 2nd most Gatto (Cat) with Cane (Dog)

- The heatmap on the right proves the model only looks at the center of the image
- It completely misses critical features like legs, ears, or tails (the edges).

AUC ANALYSIS



- Low Area Under Curve (AUC): Most classes likely hover between 0.60 and 0.70 (where 0.50 is random guessing)
- High False Positive Cost: To achieve a high True Positive Rate (e.g., catching 80% of the dogs), the curve shows we would accept a massive False Positive Rate.
- The curves for difficult classes (like Gatto/Cat) are closer to the diagonal line.

CONCLUSION

Critical Findings:

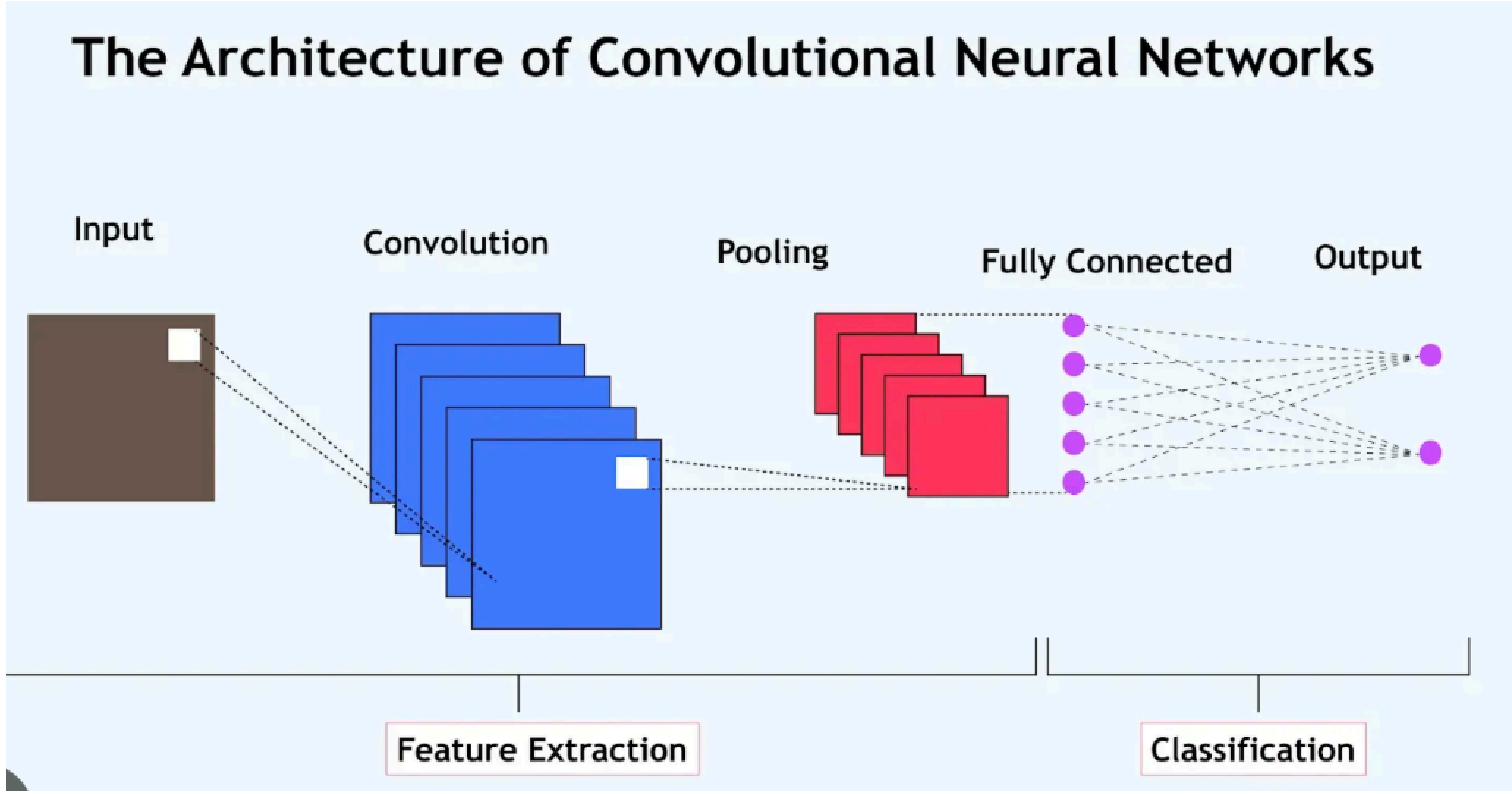
- Random Forests cannot learn complex image structures from raw pixels alone (Accuracy stuck at ~40%).
- The model lacks Spatial Invariance—it sees color intensity but misses shape, hierarchy, and texture.
- Better data representation (Features) is more important than a complex model.

Future Direction: To achieve human-level accuracy, we must move to Convolutional Neural Networks (CNNs)

IMAGE CLASSIFICATION USING CNN

GARGI PANDE (22127)

HOW DOES IT WORK?



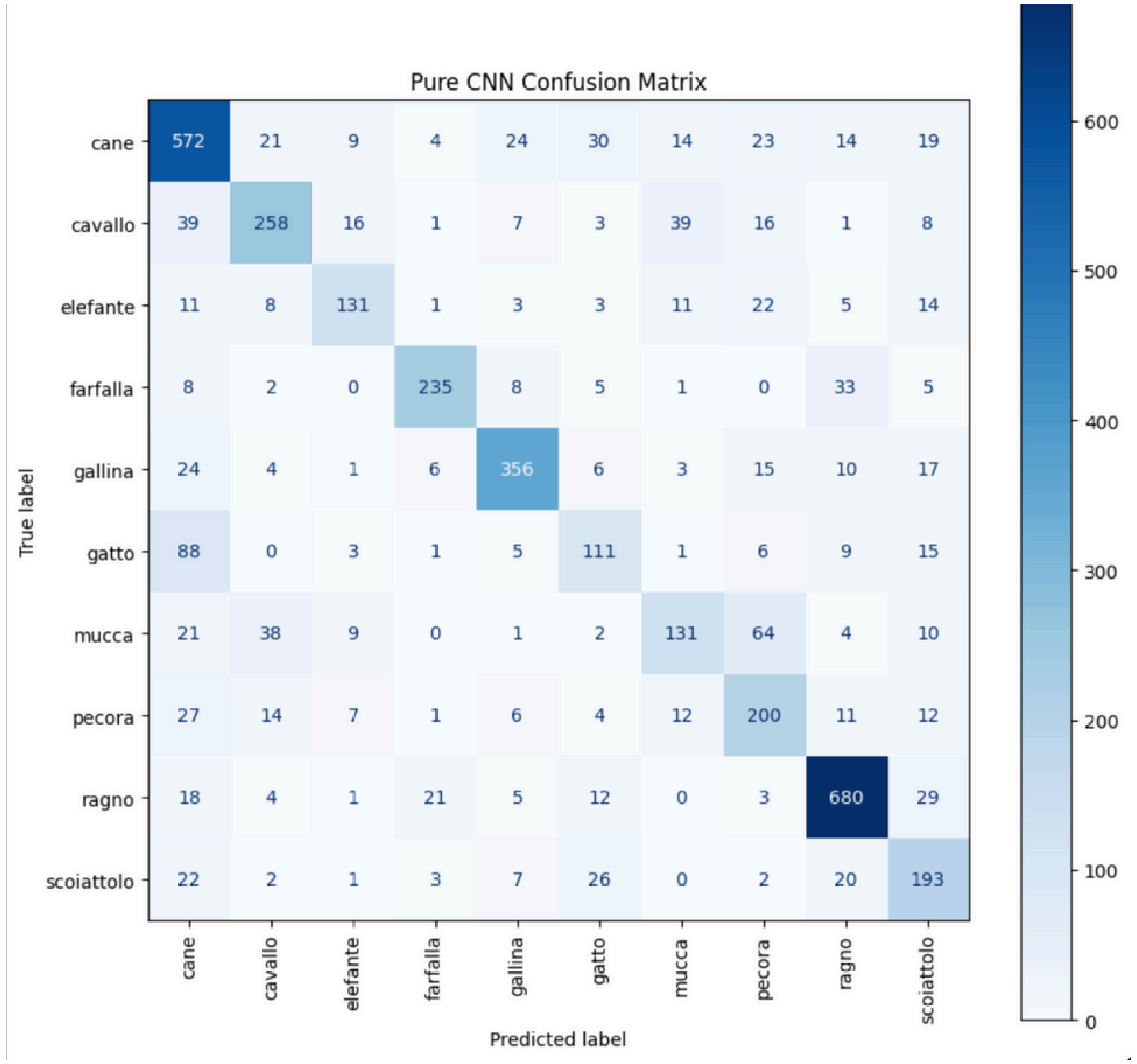
A Convolutional Neural Network (CNN) works by applying a series of convolutional layers to an input image, automatically learning hierarchical features like edges and patterns, followed by pooling and fully connected layers to classify the image based on learned features.

OVERALL PERFORMANCE OF THE CNN MODEL

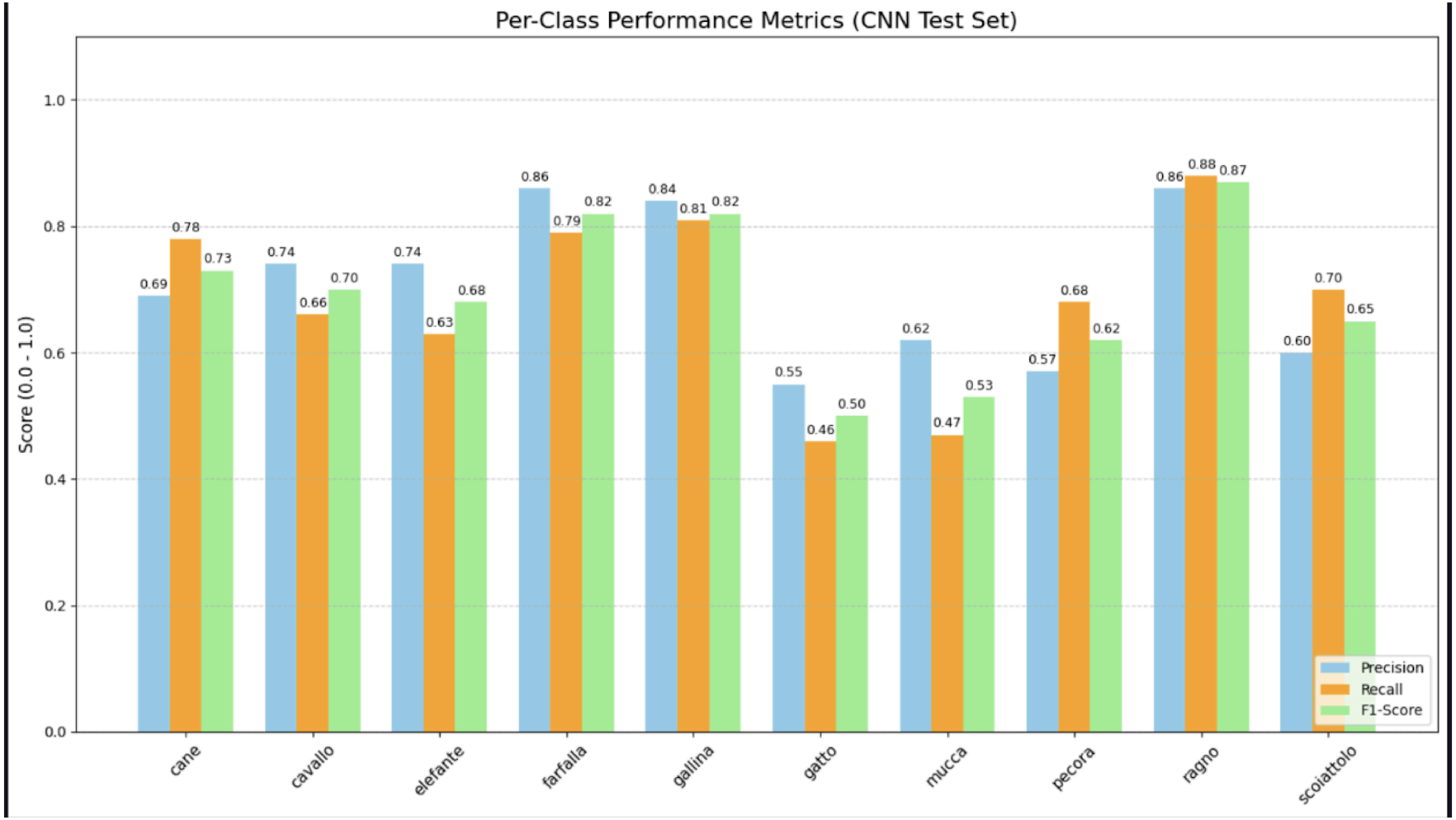
```
--- Evaluating Best CNN on Test Set ---
              precision    recall  f1-score   support

        cane       0.69      0.78      0.73       730
     cavallo       0.74      0.66      0.70       388
    elefante       0.74      0.63      0.68       209
    farfalla       0.86      0.79      0.82       297
     gallina       0.84      0.81      0.82       442
       gatto       0.55      0.46      0.50       239
       mucca       0.62      0.47      0.53       280
      pecora       0.57      0.68      0.62       294
       ragno       0.86      0.88      0.87       773
  scoiattolo       0.60      0.70      0.65       276

    accuracy                           0.73      3928
   macro avg       0.71      0.69      0.69      3928
weighted avg       0.73      0.73      0.73      3928
```

- Farfalla (Butterfly) and Ragno (Spider) are the best performers, with high precision, recall, and f1-scores.
- Gatto (Cat) has the lowest performance across precision, recall, and f1-score, indicating that the model struggles to classify cats correctly.
- Class Imbalance might be affecting performance, especially for classes like Gatto, which have lower scores.
- The weighted average is slightly higher than the macro average, suggesting the model is performing better on classes with higher support.

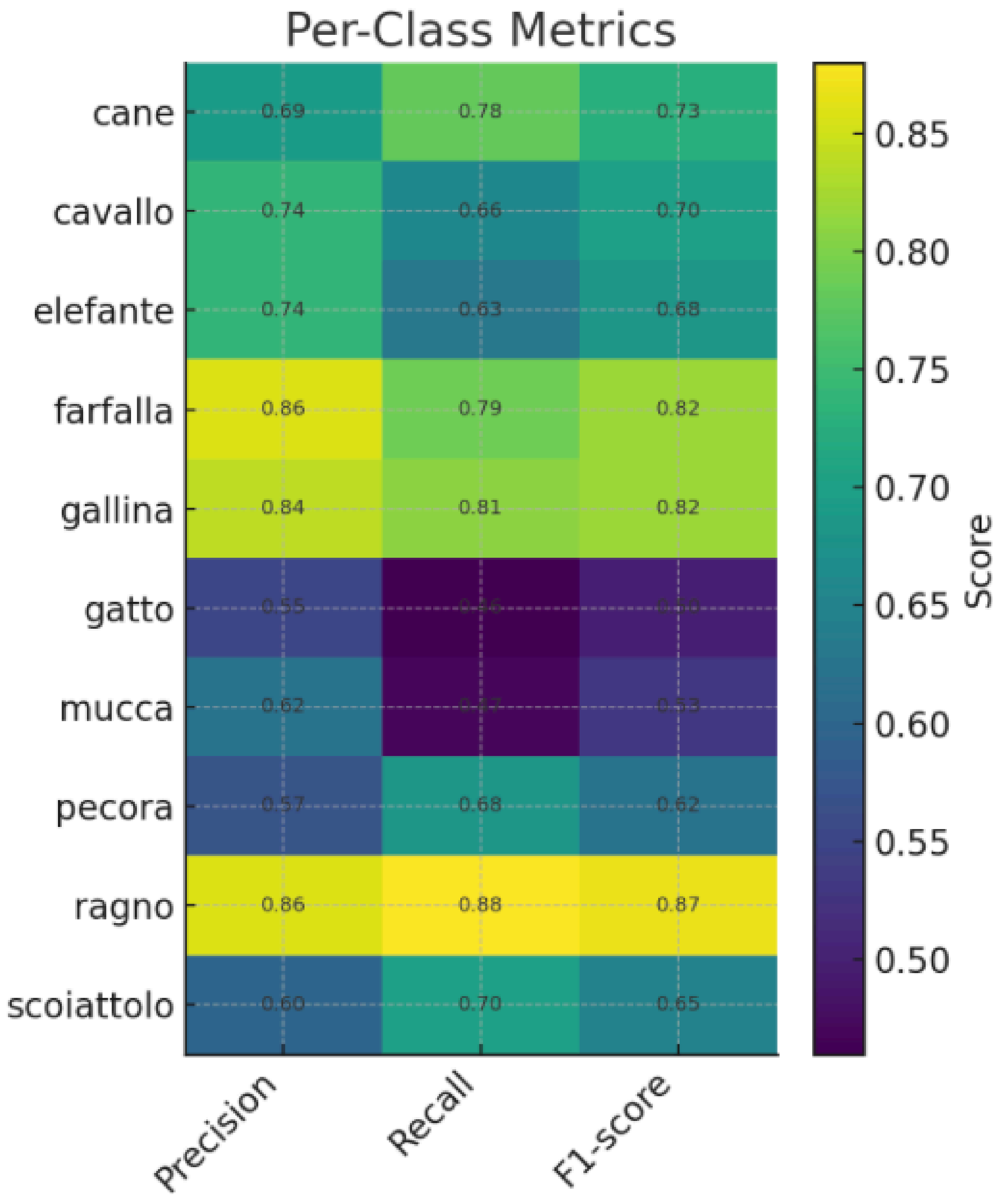
CONFUSION MATRIX OF THE CNN MODEL

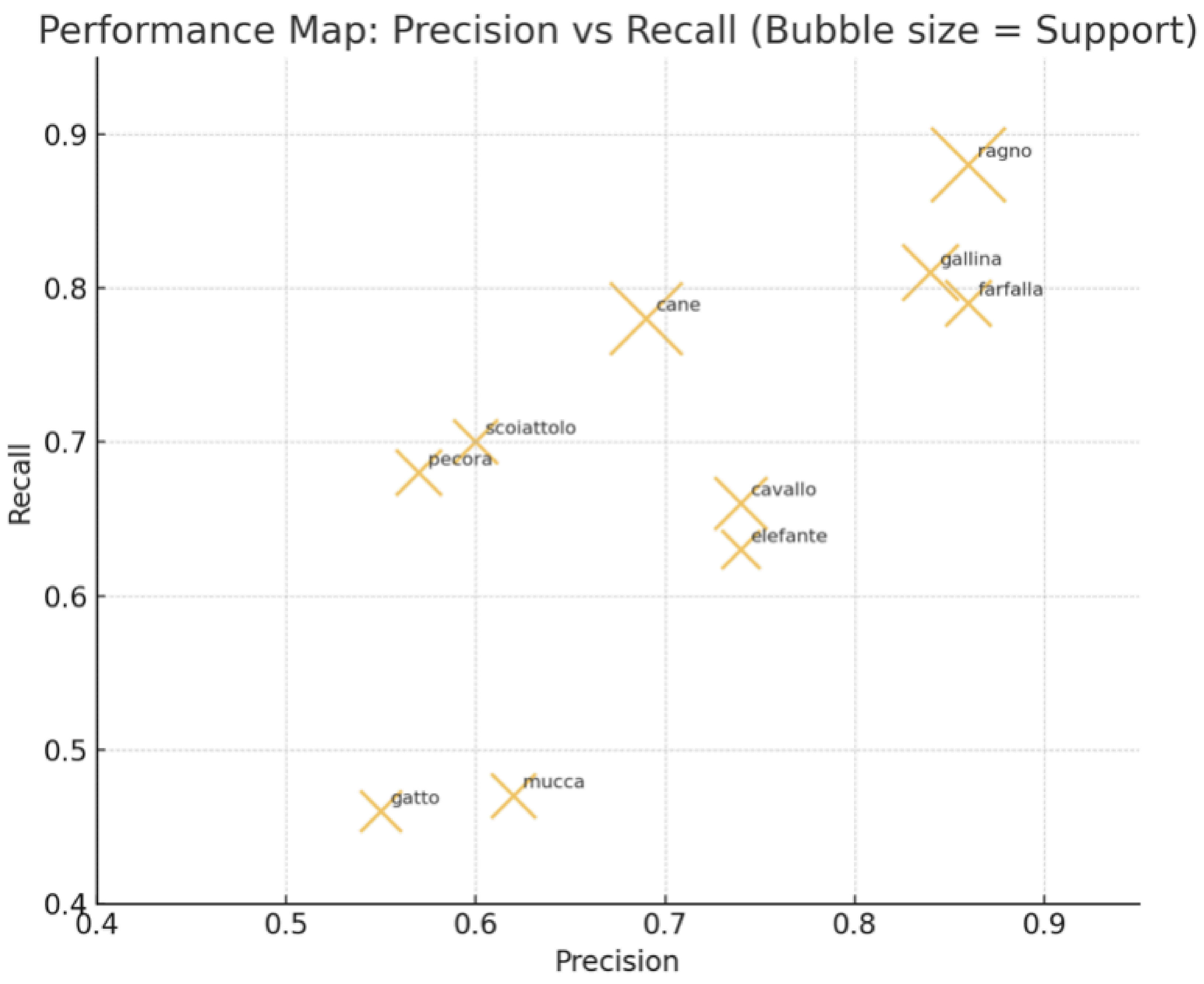


CLASS BASED PERFORMANCE

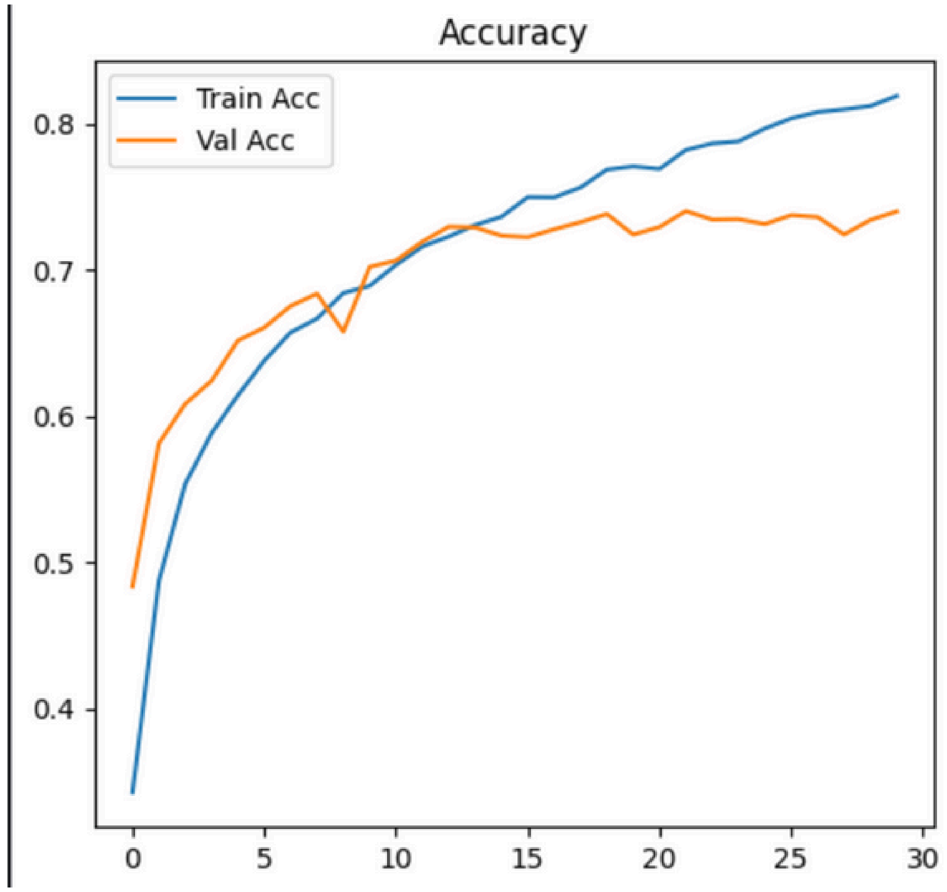


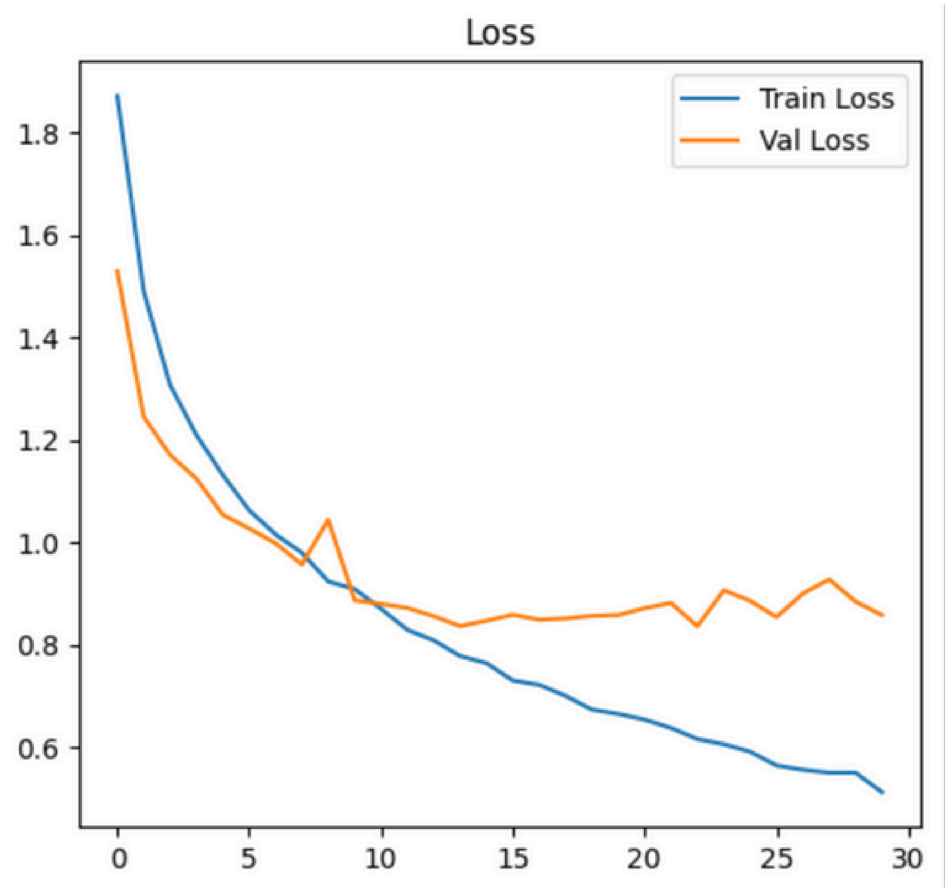
TO UNDERSTAND PERFORMANCE BY CLASS





TRAINING AND VALIDATION METRICS





Validation Accuracy:

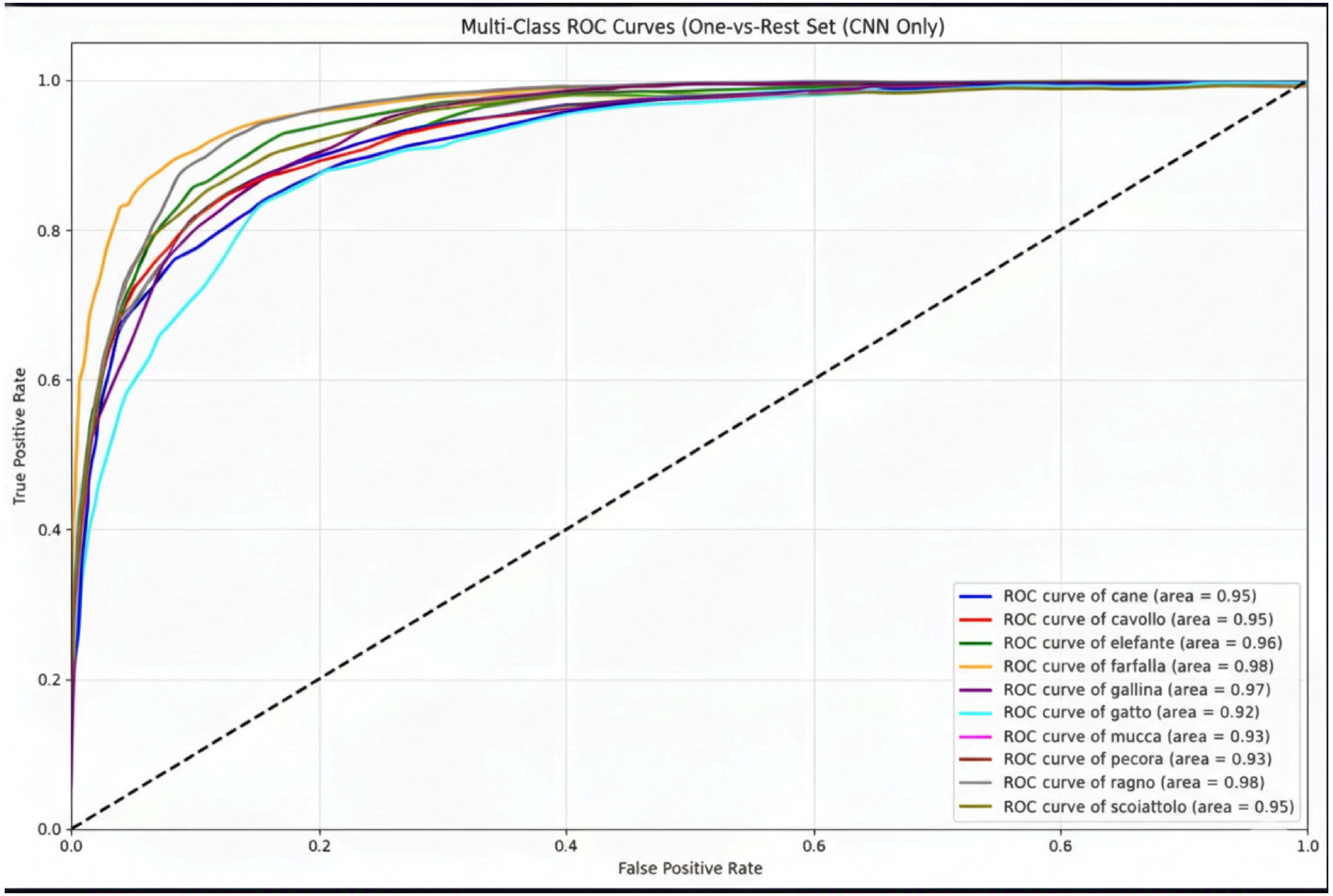
- The validation accuracy also improves initially but starts to level off around epoch 15.
- Indicates that the model may be starting to overfit and struggle to generalize to unseen data.

Validation Loss:

- The validation loss starts high but decreases initially, then begins to rise slightly after epoch 10.
- The increasing validation loss after epoch 10 further supports the overfitting issue, as the model is memorizing the training data but not generalizing well on the validation set.

- While the model performs well on training data (high accuracy, low loss), it shows signs of overfitting by failing to generalize to new data.

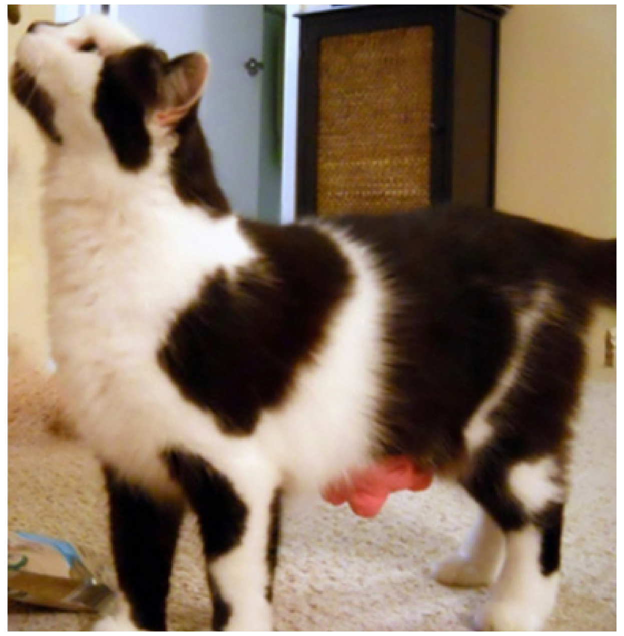
ANALYSIS OF ROC CURVE



- The model is generally performing well for most classes with AUCs above 0.90.
- There is a clear discrepancy in performance, where certain classes (like Ragno, Farfalla) are easily distinguishable, while others (like Gatto, Mucca) are more challenging for the model.
- The model is stronger in differentiating between visually or contextually distinct classes (like Ragno and Farfalla) than those that might have more similarities in features (e.g., Gatto vs Cane).

LIMITATIONS WITH CNN

- While CNNs are great at learning hierarchical features, they may still struggle if the visual features in the dataset are not distinct enough or the image quality is not optimal. The netCNNs generally perform well with distinct objects, but subtle differences between classes (like cats and cows) can challenge the model, especially if it has a simpler architecture or hasn’t been tuned well for these types of subtleties.
- The model might not be complex enough to capture the more nuanced differences between these two classes.



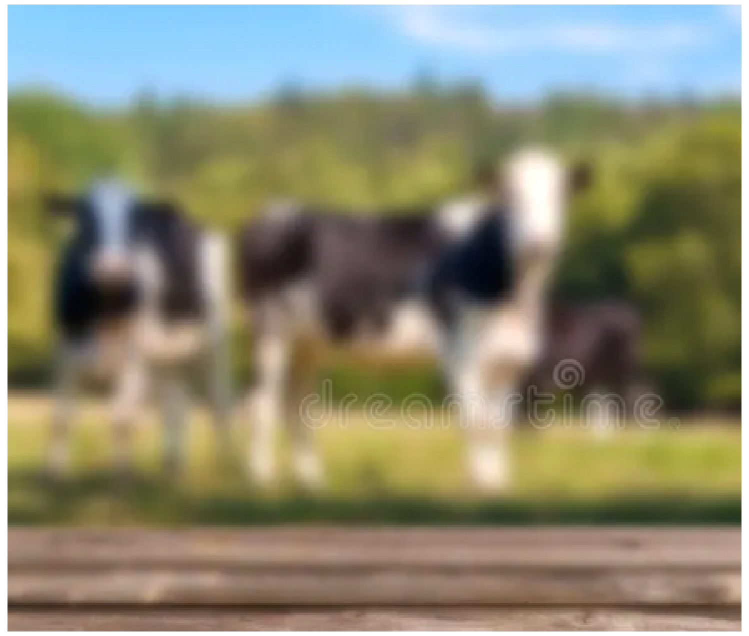
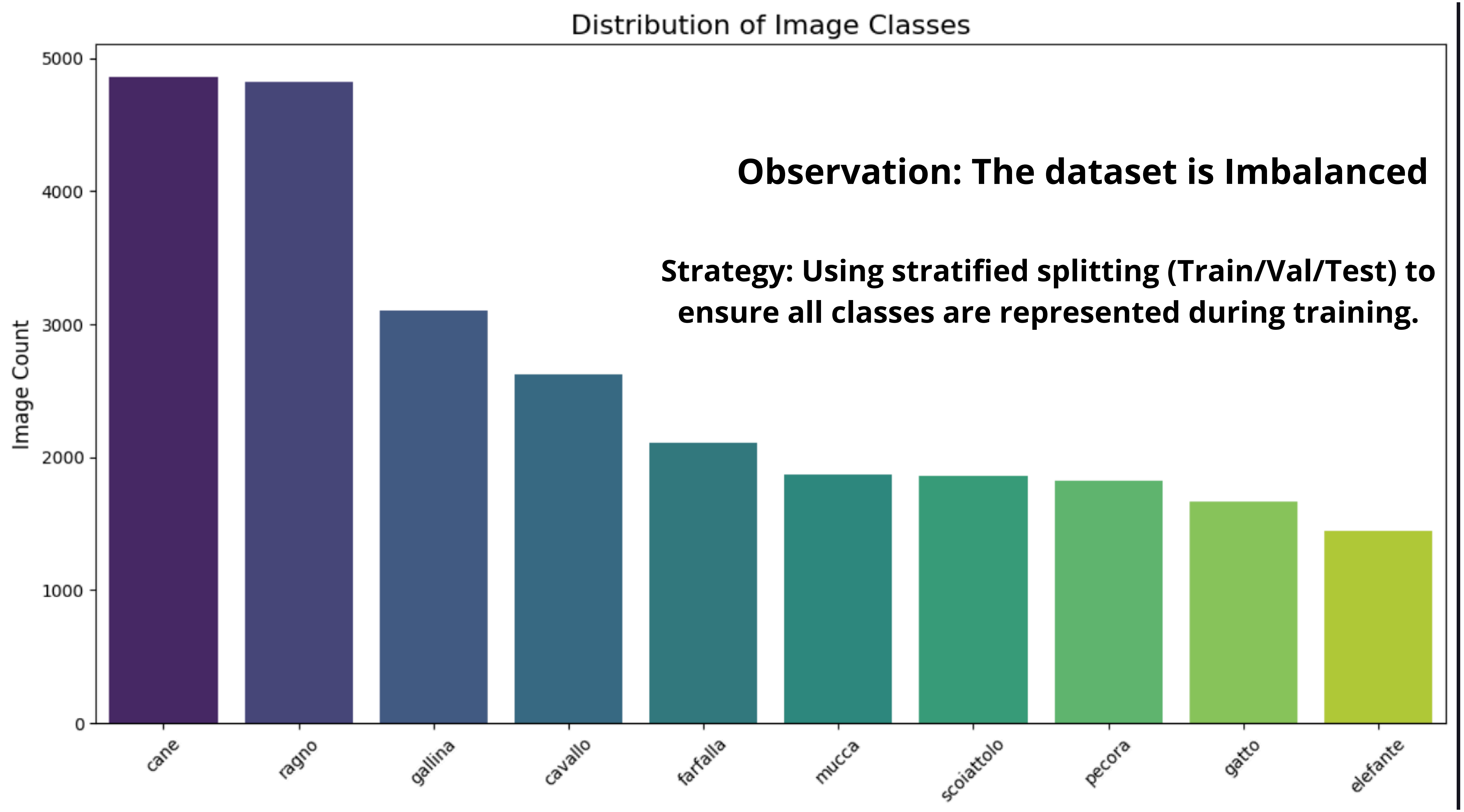


IMAGE CLASSIFICATION USING HYBRID CNN + RANDOM FOREST

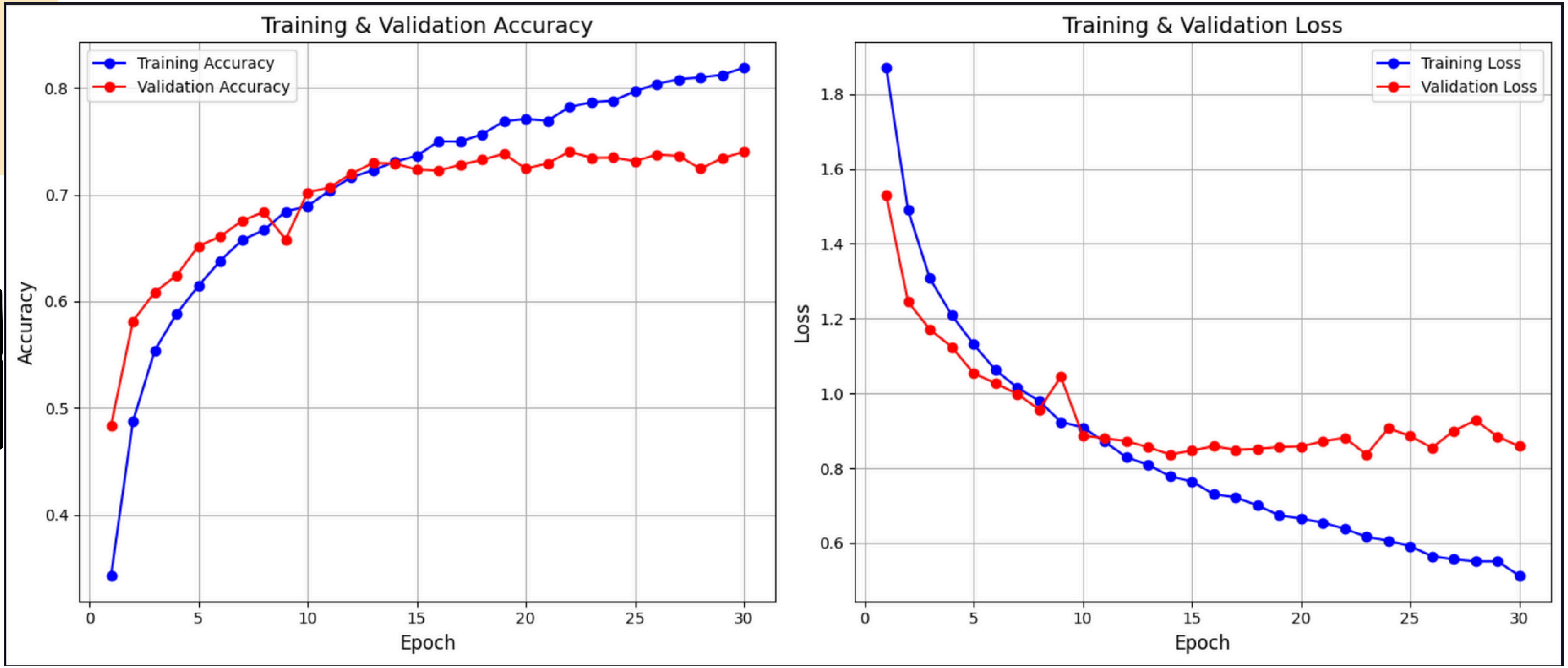
VIPUL SHARMA(22369)

Objective: Classify images into 10 distinct animal categories (Cane, Cavallo, Elefante, Farfalla, Gallina, Gatto, Mucca, Pecora, Ragno, Scoiattolo).



Observation: The dataset is Imbalanced

Strategy: Using stratified splitting (Train/Val/Test) to ensure all classes are represented during training.



Accuracy: The Training Accuracy keeps climbing to over 80%, but the Validation Accuracy plateaus around 74%.
This gap indicates Overfitting. The model is memorizing the training data but struggling to generalize to new data.
Loss: Training error dropped, while validation error stabilized after Epoch 15.

Outcome: The best model was saved for feature extraction.

Stage 1: Feature Extraction (CNN)

Input: Receives raw 150×150 RGB images.

Processing: Passes through 3 Convolutional Blocks (Conv2D + ReLU + MaxPool) to learn spatial patterns like edges and textures.

Vectorization: A **Flatten Layer converts the 3D maps into a 1D vector** (41,472 units).

Output: A fully connected Dense Layer compresses this into a rich 512-dimensional feature vector.

Stage 2: Classification (Random Forest)

Input: Takes the 512 numeric features extracted by the CNN.

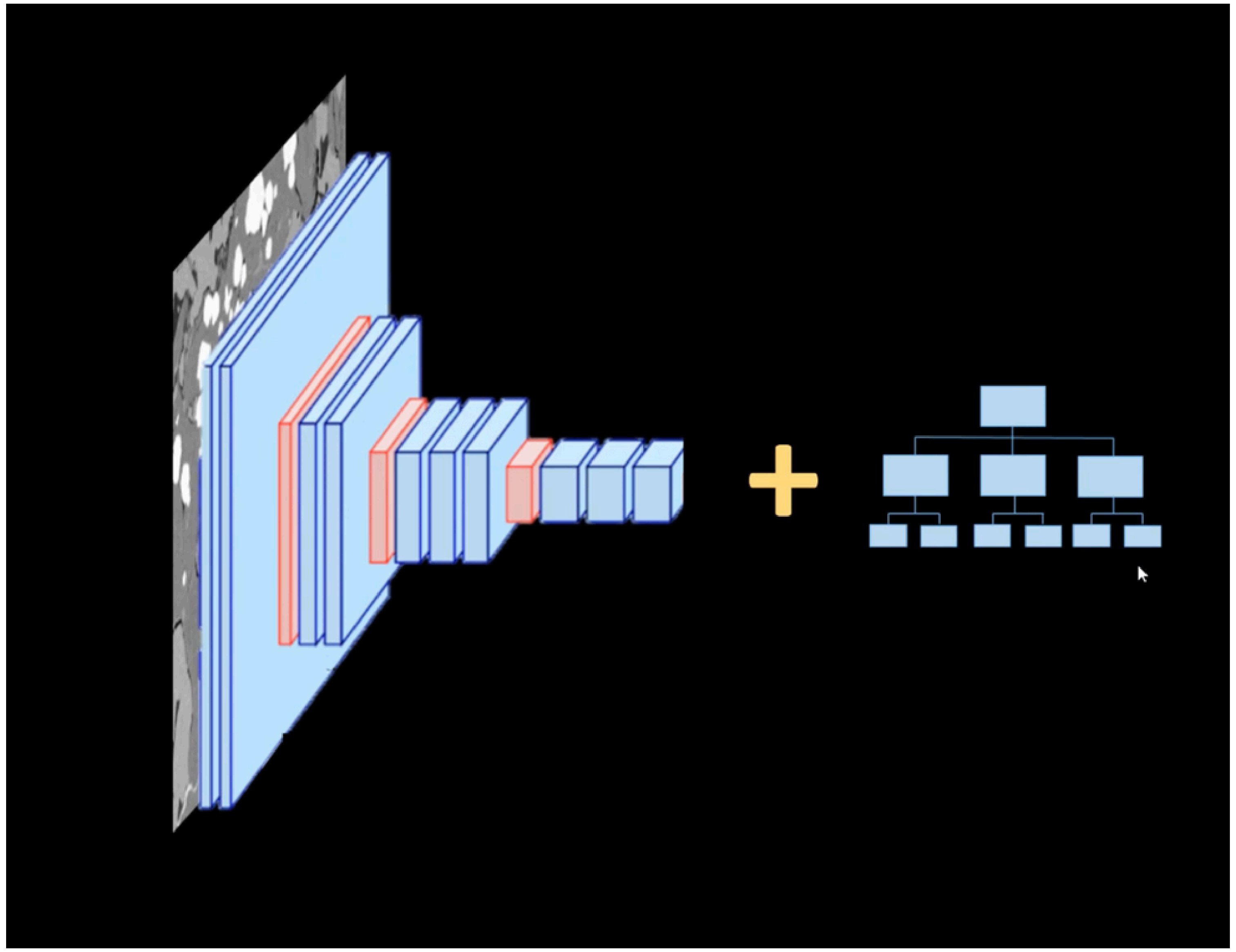
Logic: An ensemble of 100 Decision Trees analyzes these features.

Result: The final class label is decided by a majority vote among the trees.

The Architecture:

Feature Extractor (CNN): A Convolutional Neural Network processes raw images to detect edges, textures, and patterns. We remove the final "Dense" layers.

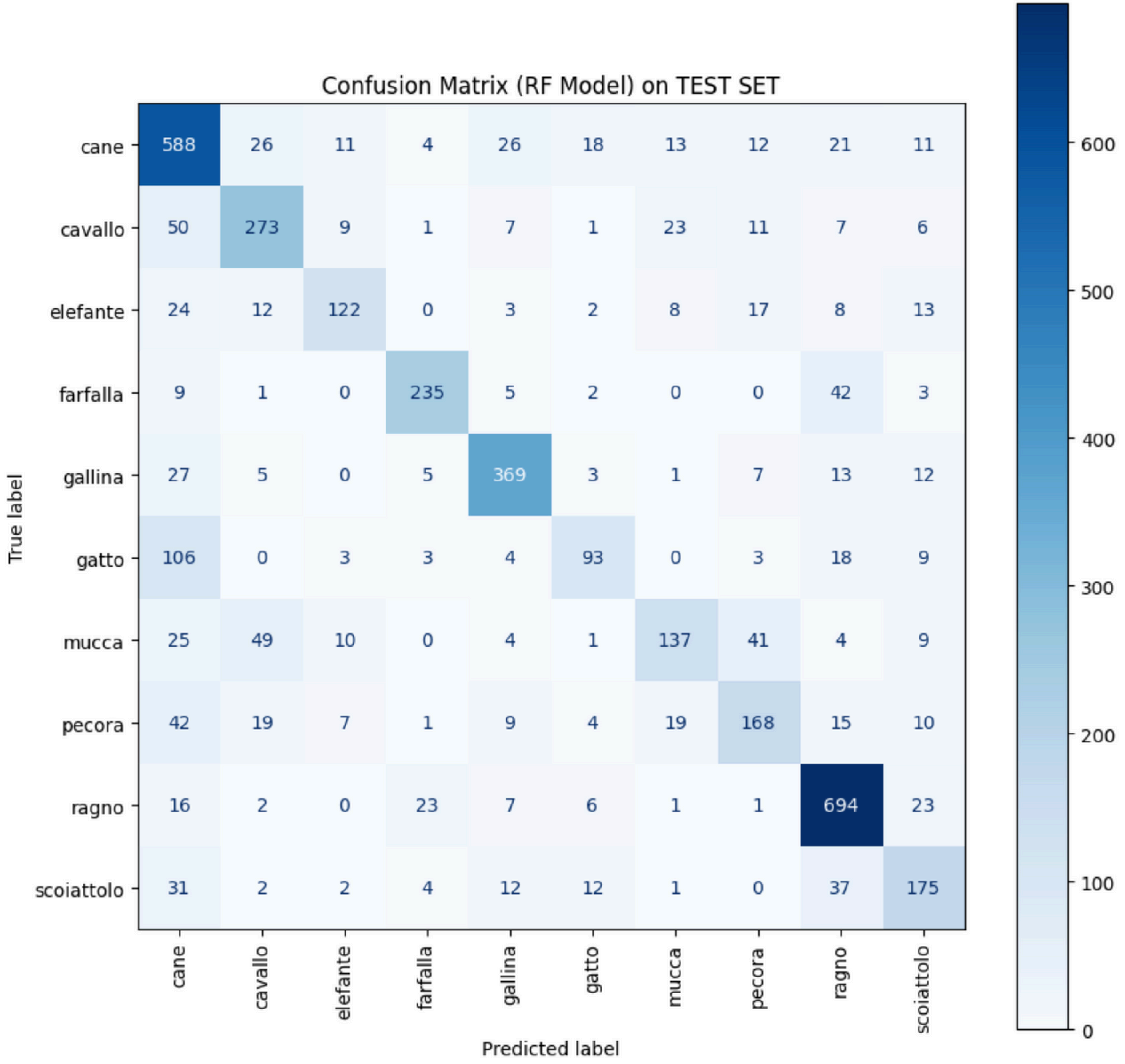
Classifier (Random Forest): The extracted features (vectors of 512 numbers) are fed into a Random Forest for the final decision.



Why this approach?

Prevents Overfitting: Avoids standard CNN dense layers, which often memorize small datasets.

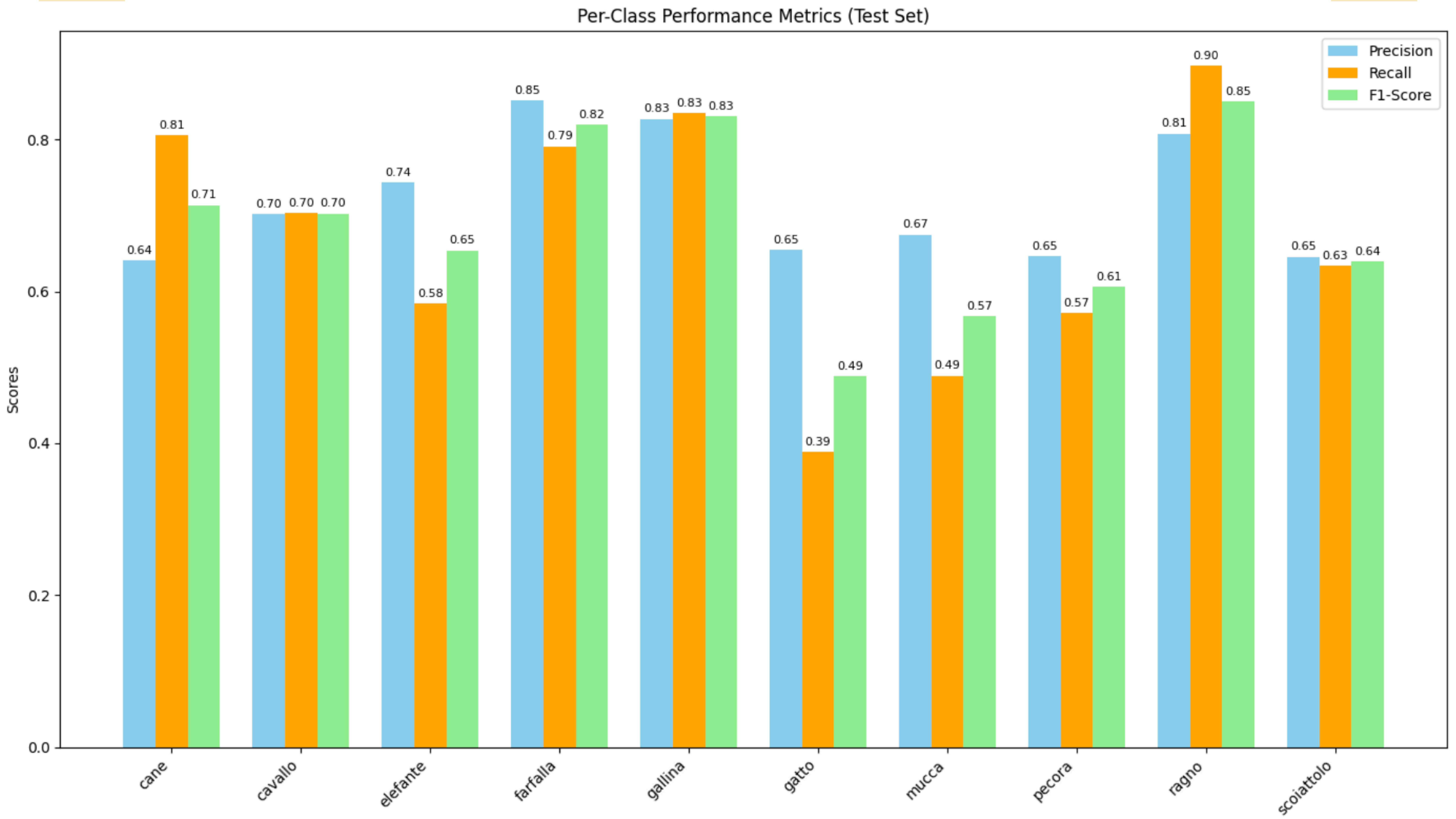
Increases Robustness: Random Forest ensembles generalize better by averaging predictions from multiple trees.



Confusions:

Gatto (Cat) vs. Cane (Dog): There is noticeable confusion between these two (e.g., 106 cats predicted as dogs), likely due to **similar body shapes and fur textures**.

Mucca (Cow) vs. Pecora (Sheep): Some overlap exists here, likely due to **grazing backgrounds**



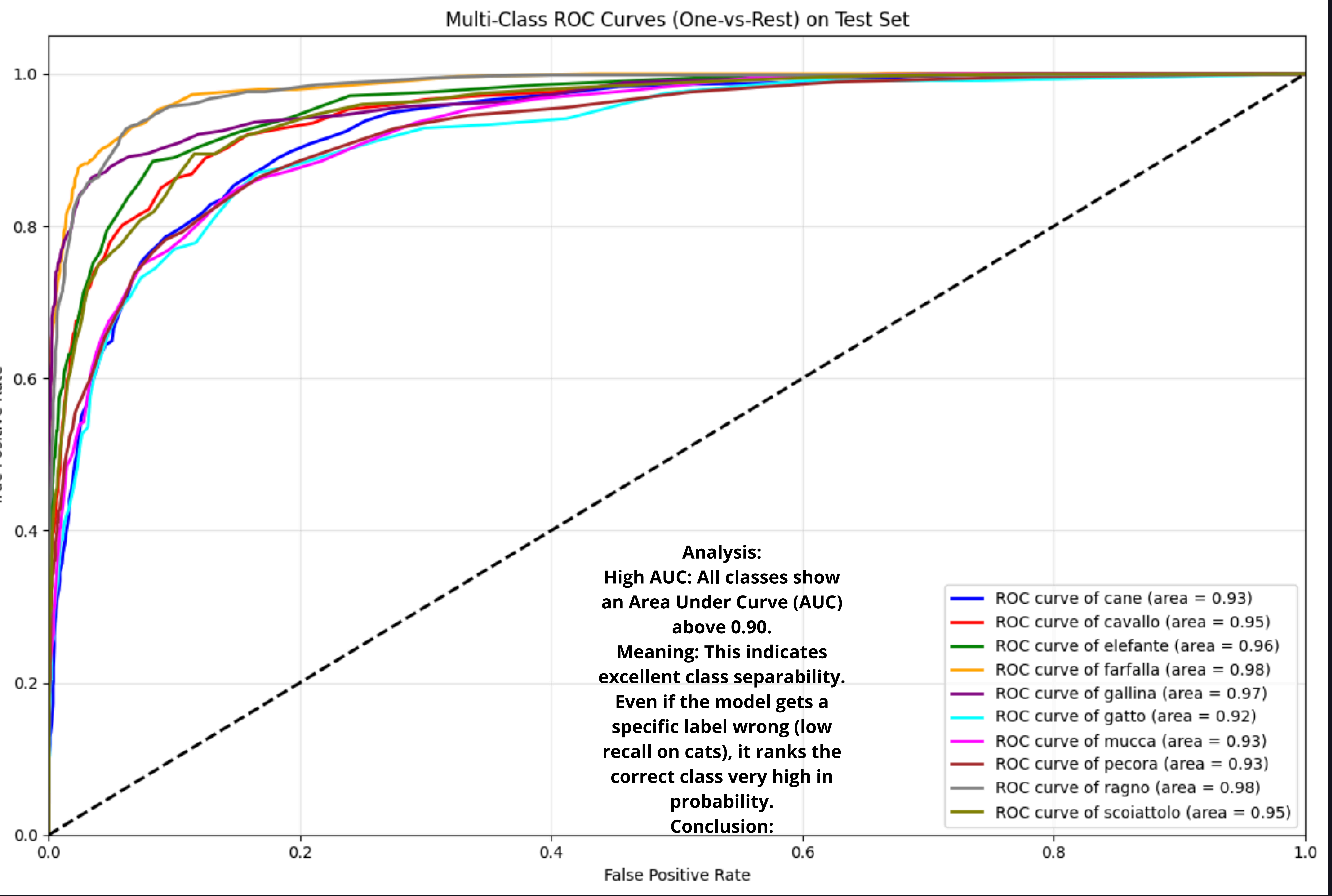
Per-Class Performance Metrics (Test Set)
Precision
Recall
F1-Score
Scores
0.0
0.2
0.4
0.6
0.8
0.64
0.81
0.71
0.70
0.70
0.70
0.74
0.58
0.65
0.85
0.79
0.82
0.83
0.83
0.83
0.65
0.39
0.49
0.67
0.49
0.57
0.65
0.57
0.61
0.81
0.90
0.85
0.65
0.63
0.64
cane
cavallo
elefante
farfalla
gallina
gatto
mucca
pecora
ragno
scoiattolo

```
--- FINAL RESULTS (Custom CNN + Random Forest on TEST SET) ---
Overall Accuracy: 72.66%
Area Under Curve (AUC, Weighted): 0.9536

--- Full Classification Report (Precision, Recall, F1) ---
              precision    recall  f1-score      support
cane           0.640523  0.805479  0.713592   730.000000
cavallo        0.701799  0.703608  0.702703   388.000000
elefante       0.743902  0.583732  0.654155   209.000000
farfalla       0.851449  0.791246  0.820244   297.000000
gallina        0.827354  0.834842  0.831081   442.000000
gatto          0.654930  0.389121  0.488189   239.000000
mucca          0.674877  0.489286  0.567288   280.000000
pecora         0.646154  0.571429  0.606498   294.000000
ragno          0.807916  0.897801  0.850490   773.000000
scoiattolo     0.645756  0.634058  0.639854   276.000000
accuracy       0.726578  0.726578  0.726578     0.726578
macro avg      0.719466  0.670060  0.687409  3928.000000
weighted avg   0.726104  0.726578  0.720238  3928.000000
```

Best Performer: 'Farfalla' (**Butterfly**) and 'Ragno' (**Spider**) have the highest scores (F1 > 0.80). Their visual f**eatures (wings, many legs)** are distinct from mammals.

Weakest Performer: 'Gatto' (**Cat**) has low Recall (0.39), meaning the model misses identifying many actual cats.

Overall: The Random Forest achieved a 72.66% accuracy on the test set, which is a strong result given the class imbalance.



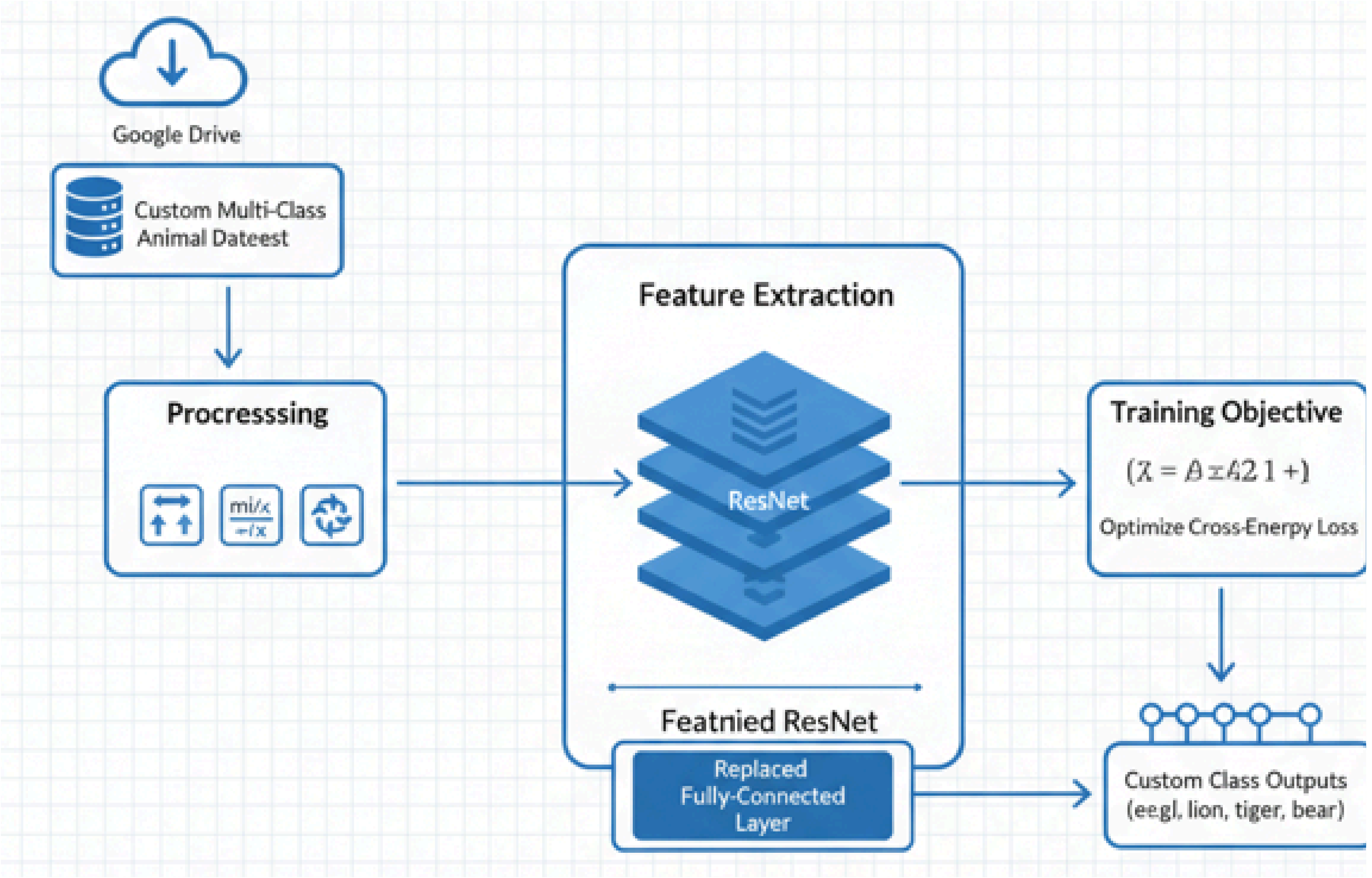
Multi-Class ROC Curves (One-vs-Rest) on Test Set
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
False Positive Rate
Analysis:
High AUC: All classes show an Area Under Curve (AUC) above 0.90.
Meaning: This indicates excellent class separability. Even if the model gets a specific label wrong (low recall on cats), it ranks the correct class very high in probability.
Conclusion:
ROC curve of cane (area = 0.93)
ROC curve of cavallo (area = 0.95)
ROC curve of elefante (area = 0.96)
ROC curve of farfalla (area = 0.98)
ROC curve of gallina (area = 0.97)
ROC curve of gatto (area = 0.92)
ROC curve of mucca (area = 0.93)
ROC curve of pecora (area = 0.93)
ROC curve of ragno (area = 0.98)
ROC curve of scoiattolo (area = 0.95)

Robust Performance: Accurately identifies animals even with obstructions or tricky angles.
Strong Precision: Confidently classifies visually distinct creatures like spiders and elephants.

RESNET

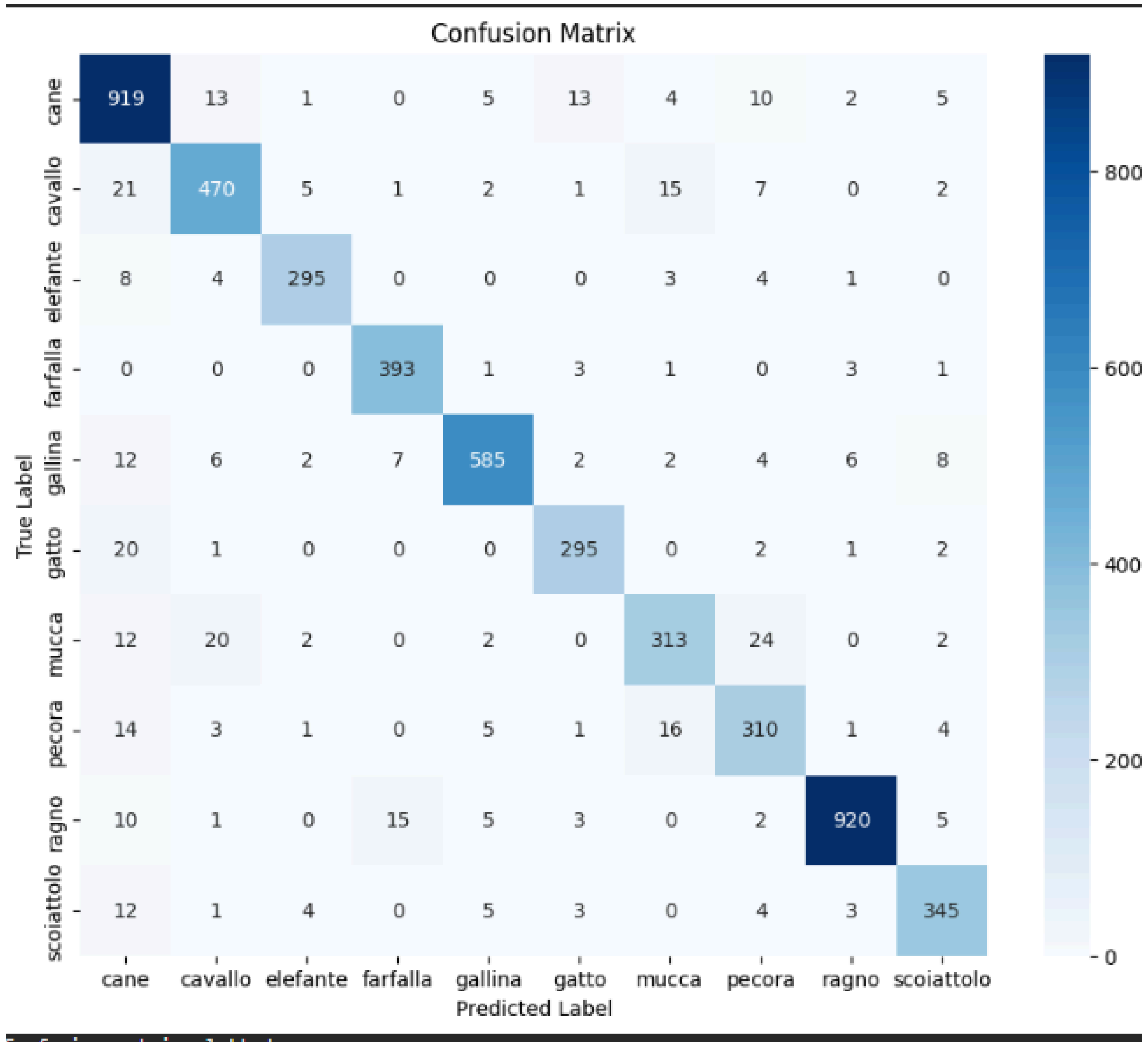
Sundram Anand(22342)

METHODOLOGY



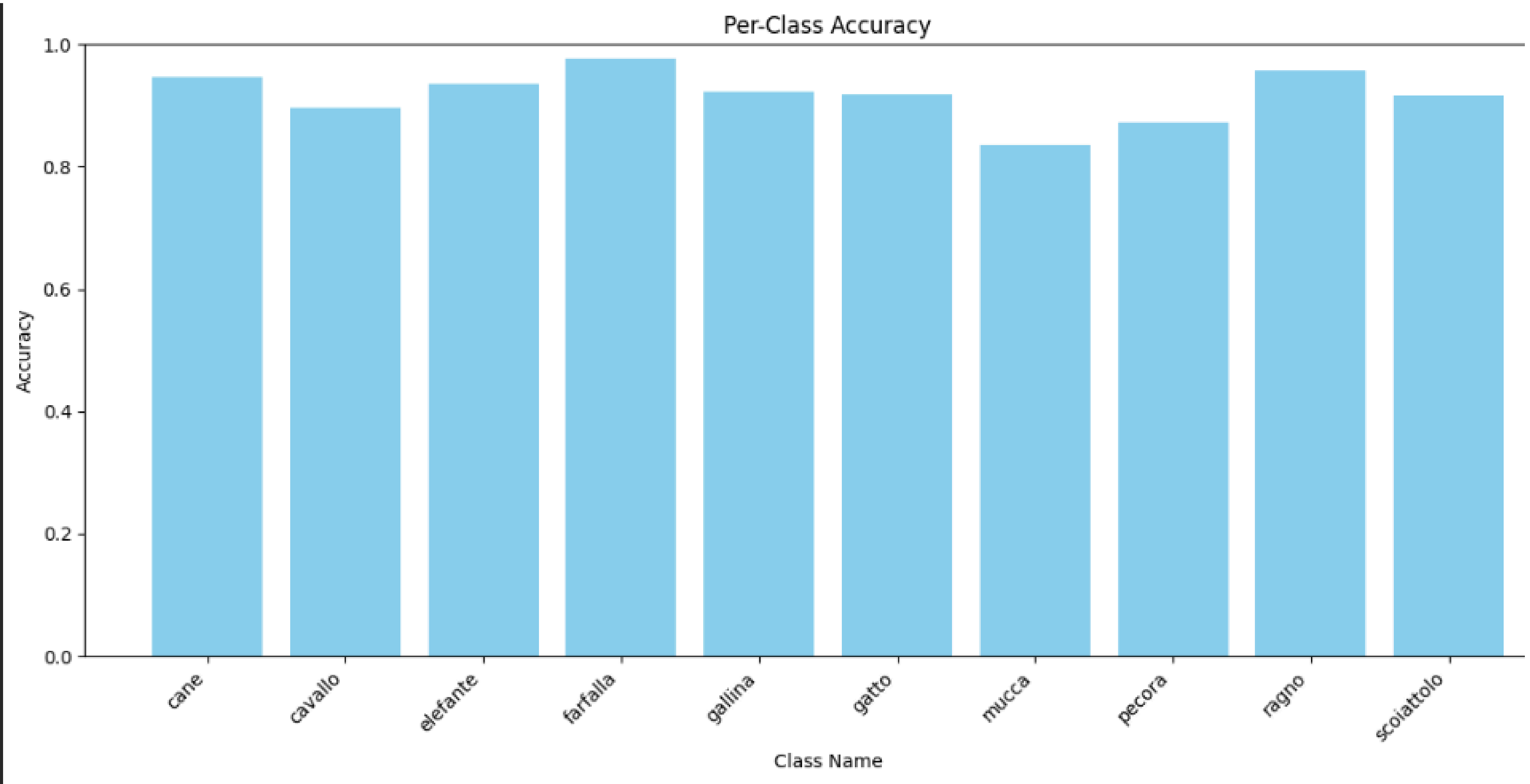
- Input Data: Custom multi-class animal dataset loaded from Google Drive.
- Preprocessing: Applied resizing, normalization, and data augmentation using torchvision transforms.
- Feature Extraction: Deep hierarchical features learned automatically by the pretrained ResNet backbone.
- Model: Modified ResNet architecture with a replaced fully-connected layer for custom class outputs.
- Training Objective: Optimize cross-entropy loss to learn discriminative visual patterns from images.
- Core Question: Can the model accurately classify animals using learned feature representations instead of raw pixels?

OVERALL PERFORMANCE



The performance is excellent. The model achieved a 92.5% validation accuracy, which is significantly higher than the baseline. It shows strong generalization across all 10 classes, with even the most difficult class (Cow) achieving over 80% accuracy.

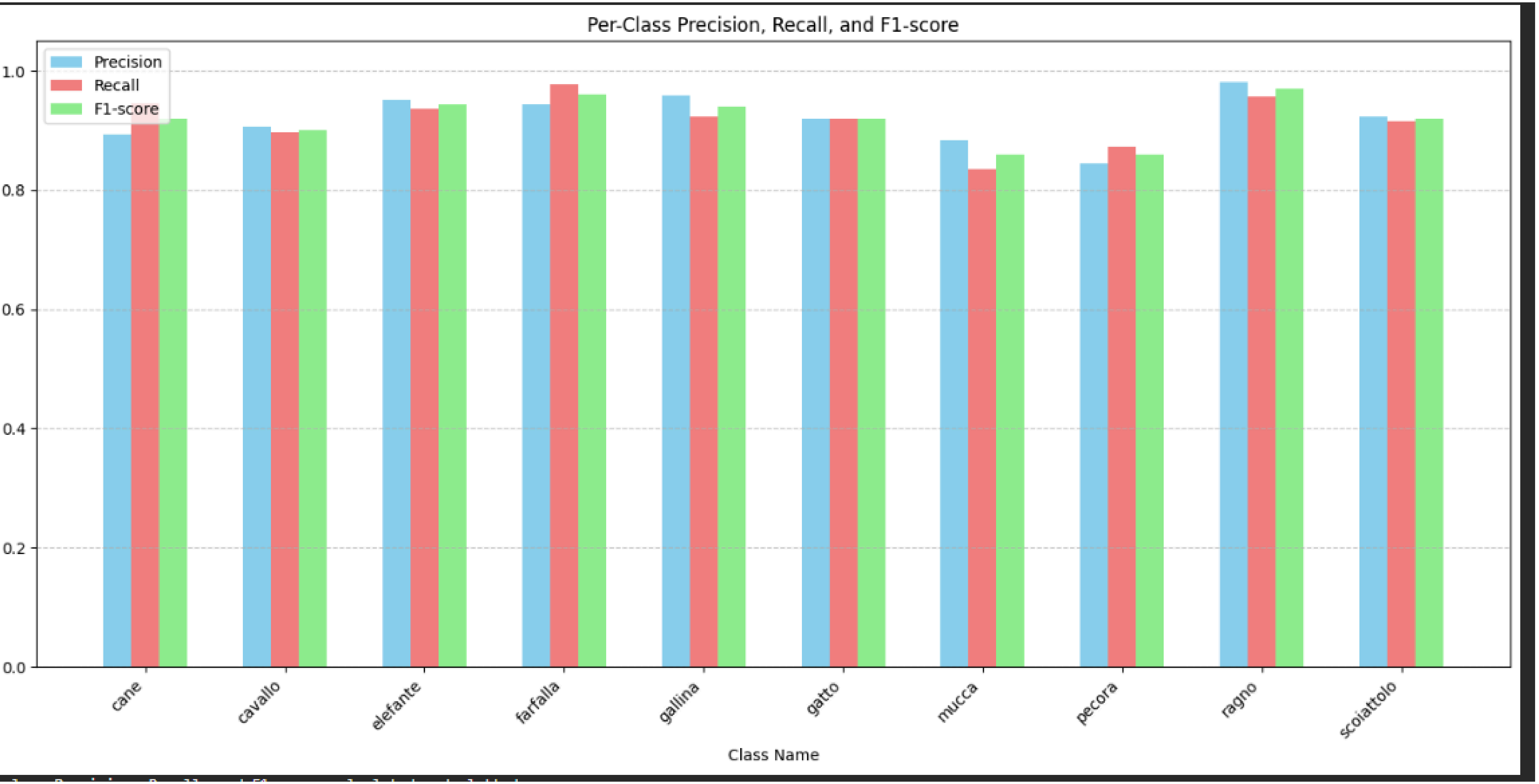
PER-CLASS ACCURACY




```
Accuracy for class 'cane': 0.95
Accuracy for class 'cavallo': 0.90
Accuracy for class 'elefante': 0.94
Accuracy for class 'farfalla': 0.98
Accuracy for class 'gallina': 0.92
Accuracy for class 'gatto': 0.92
Accuracy for class 'mucca': 0.83
Accuracy for class 'pecora': 0.87
Accuracy for class 'ragno': 0.96
Accuracy for class 'scoiattolo': 0.92
Per-class accuracy plotted and displayed.
```

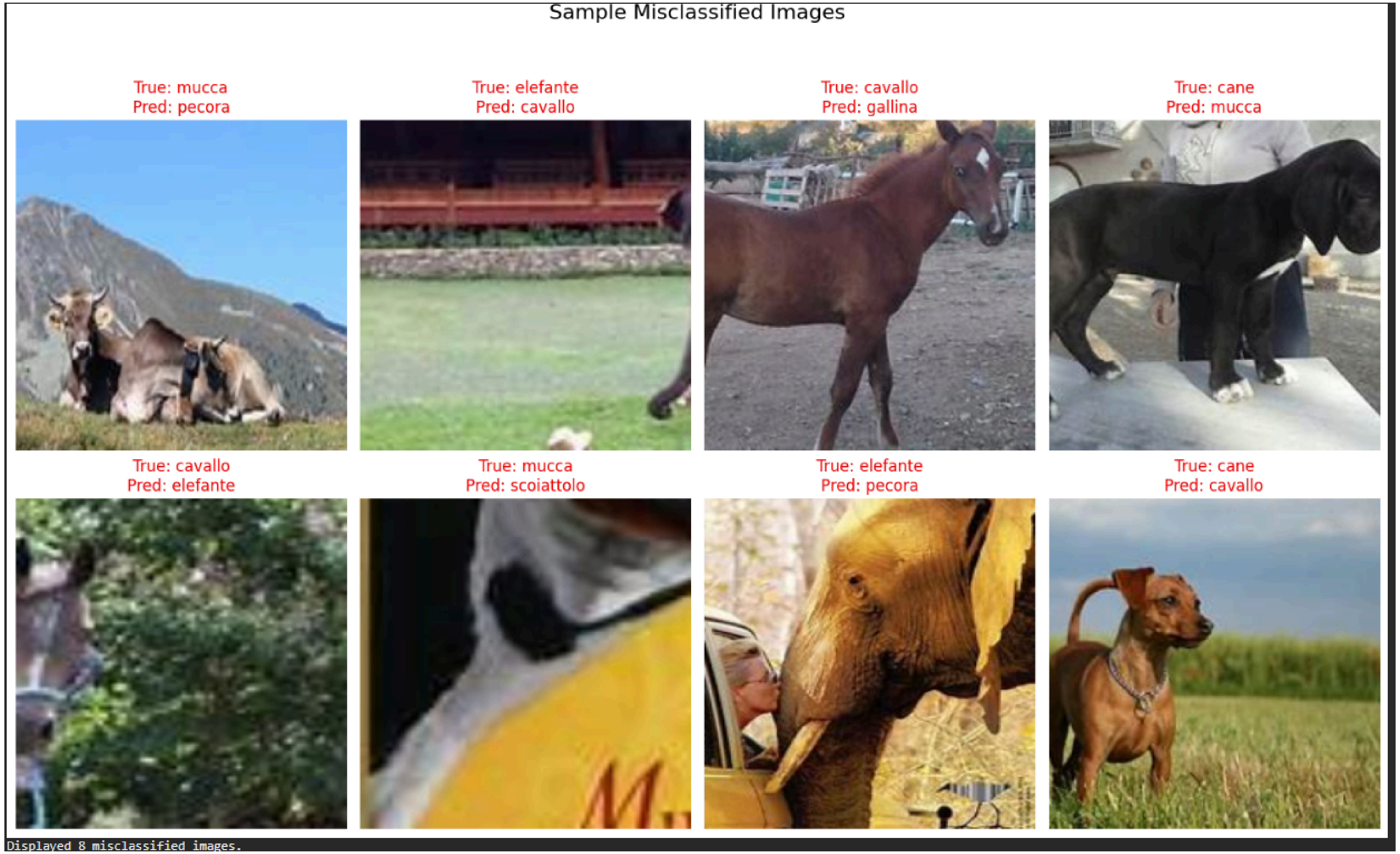
- Most classes achieve high accuracy (above 90%), showing that the model is performing consistently well across the dataset.
- ‘Farfalla’ (0.98) and ‘Ragno’ (0.96) are the best-performing classes, indicating strong feature separability for these categories.
- ‘Mucca’ (0.83) and ‘Pecora’ (0.87) show relatively lower accuracy, suggesting that these classes may have more visual similarity or need additional training data.

PER-CLASS PRECISION,RECALL, AND F1 SCORE



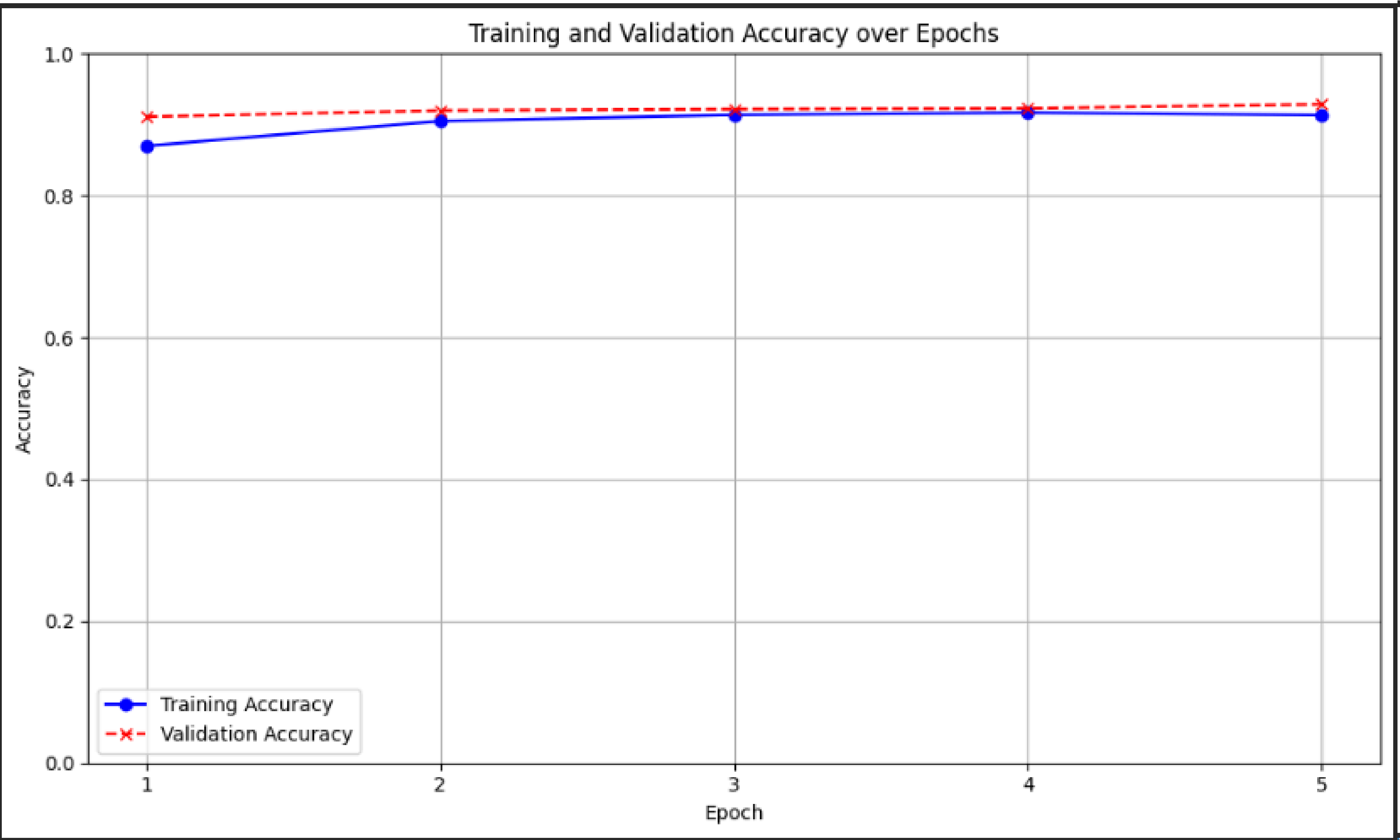
- Most classes show high precision, recall, and F1-scores (mostly above 0.90), indicating strong and consistent performance.
- Farfalla, ragno, and gallina perform the best across all metrics.
- Mucca and pecora have slightly lower scores, suggesting these classes are more challenging to classify.
- Overall, the model shows balanced behavior with minimal difference between precision, recall, and F1.

MISCLASSIFIED IMAGES

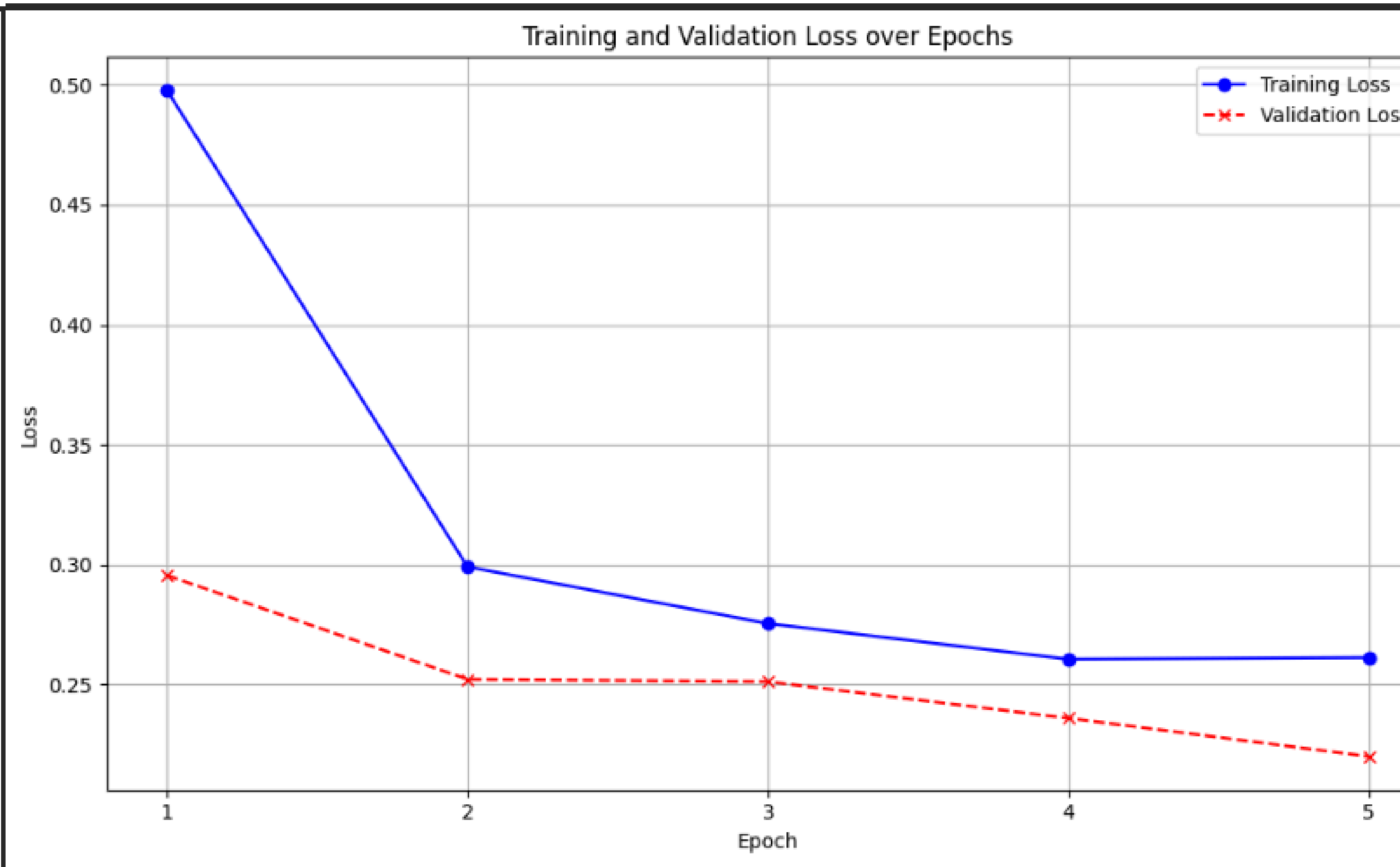


- Many mistakes occur between visually similar classes (e.g., mucca–pecora, cavallo–elefante).
- Cluttered backgrounds and complex scenes make classification harder.
- Some errors are due to cropped, blurred, or partially visible animals.
- The model sometimes relies too much on color/texture, causing confusion between look-alike classes.
- Around 372 images were not classified, indicating possible data quality issues or model limitations.

Training and Validation Accuracy/Loss over Epochs

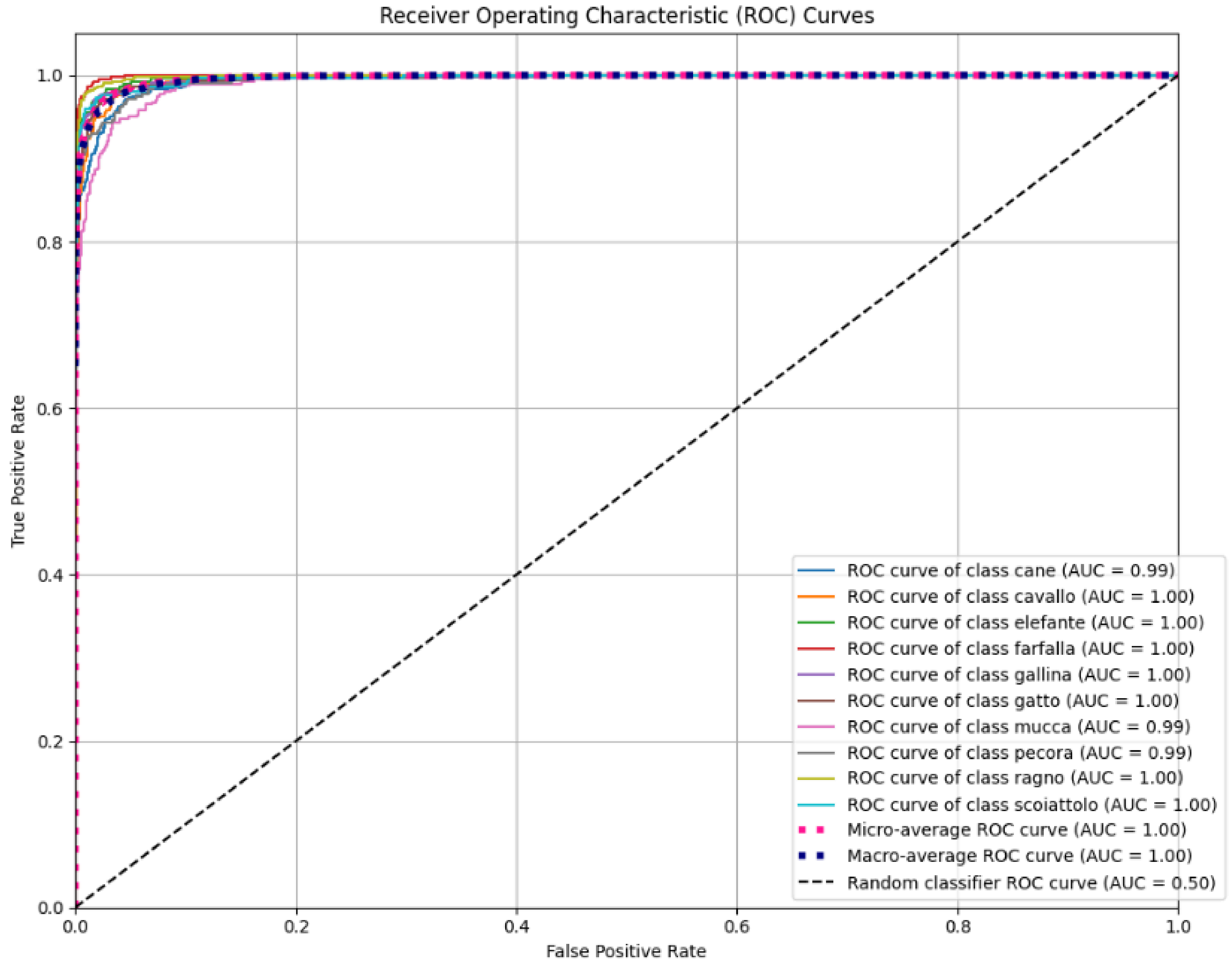


Training and Validation accuracy plot displayed.



Training and Validation loss plot displayed.

ROC CURVE




```
ROC curves for all classes, micro-average, and macro-average plotted.
```

- Most classes achieved AUC = 1.00, showing near-perfect classification.
- A few classes have AUC = 0.99, still indicating very strong performance.
- Micro and macro AUC = 1.00, confirming consistent results across all classes.
- All ROC curves stay close to the top-left corner, far above the random baseline

CONCLUSION

- The ResNet-based model achieved strong overall performance, with high precision, recall, and F1-scores across most classes.
- Misclassifications primarily occurred between visually similar animals or images with poor quality or cluttered backgrounds.
- ROC curves and AUC values close to 1.0 confirm that the model is highly effective at distinguishing between classes.
- Further improvements can be made by enhancing data quality, increasing training data, and refining hyperparameters to reduce remaining errors.

THANK YOU