Neural Network Models for Object Recognition Using Multi-Track ML Approaches

A study of Convolutional Neural Networks with CIFAR-10 Dataset

INTRODUCTION

- Object recognition enables computers to identify and classify visual elements within images.
- Machine learning and deep learning have significantly advanced recognition accuracy in recent years (Sarkar *et al.,* 2024).
- This presentation explores two neural network approaches using the CIFAR-10 dataset.

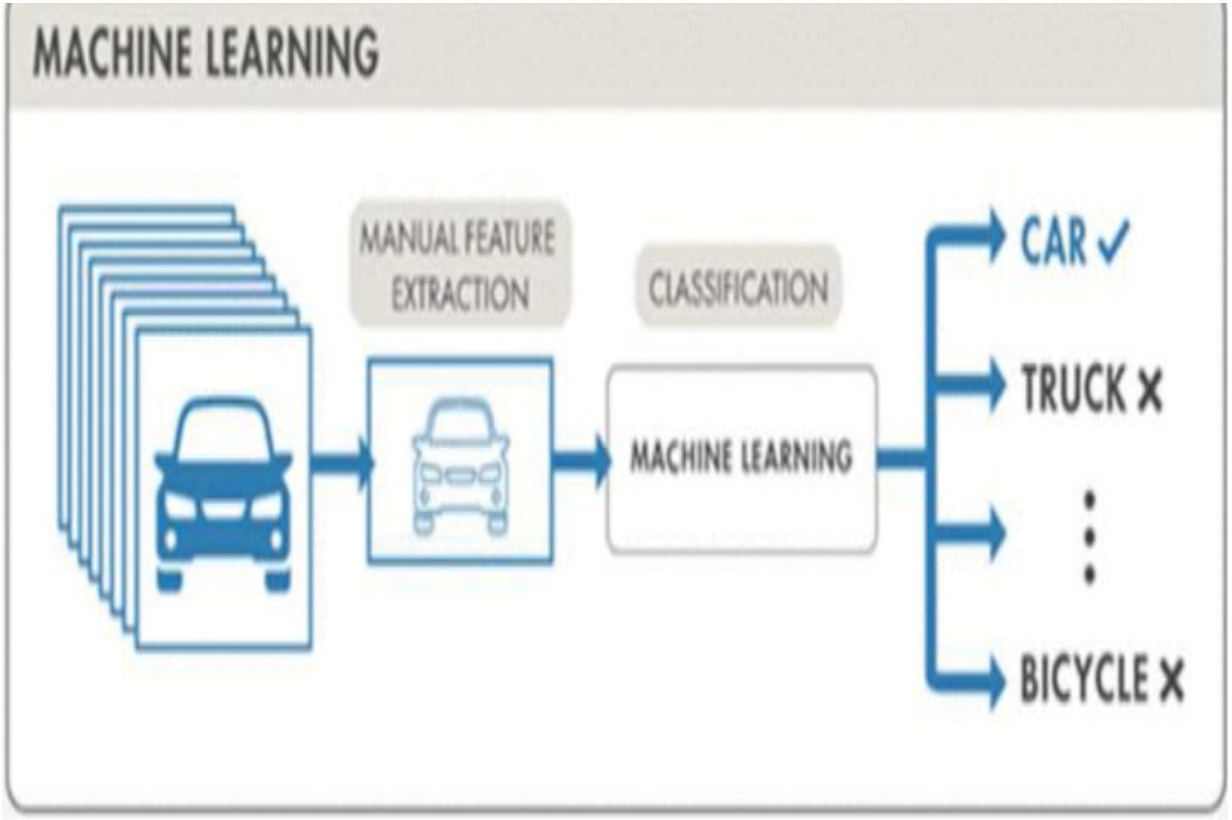


Figure 1: Object recognition using machine learning (Sarkar *et al.,* 2024)

DATASET: CIFAR-10 OVERVIEW

- 60,000 colour images (32 × 32 × 3).
- 10 classes: Airplane, automobile, bird, cat, deer, dog, frog, horse, ship, truck
- **Training**: 50,000 images Validation/Test: 10,000 images
- **Preprocessing**: Normalisation, one-hot encoding.
- **Data Augmentation**: Horizontal flip, rotation, zoom, shift

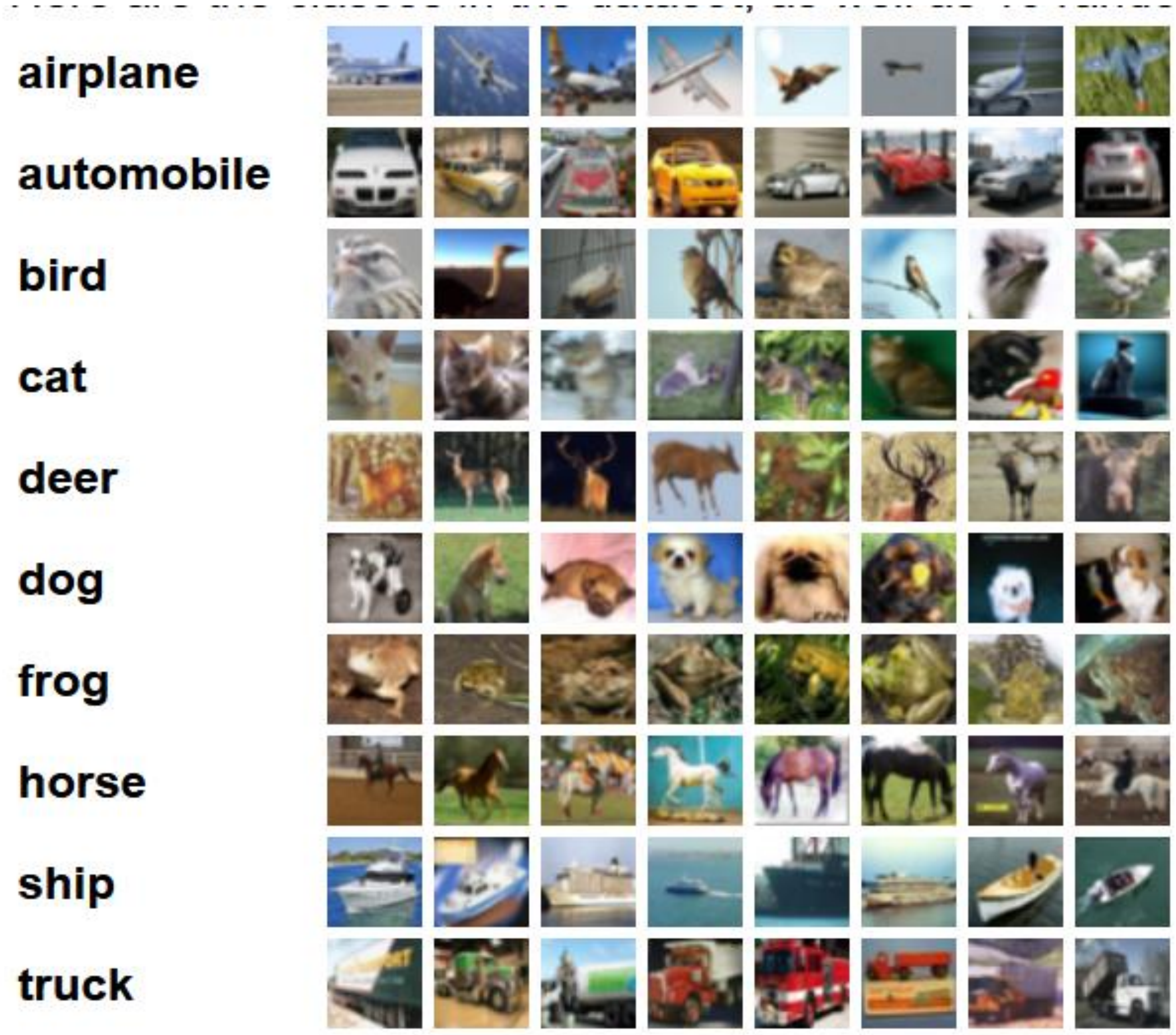


Figure 2: CIFAR-10 Dataset

Krizhevsky *et al.*, 2009

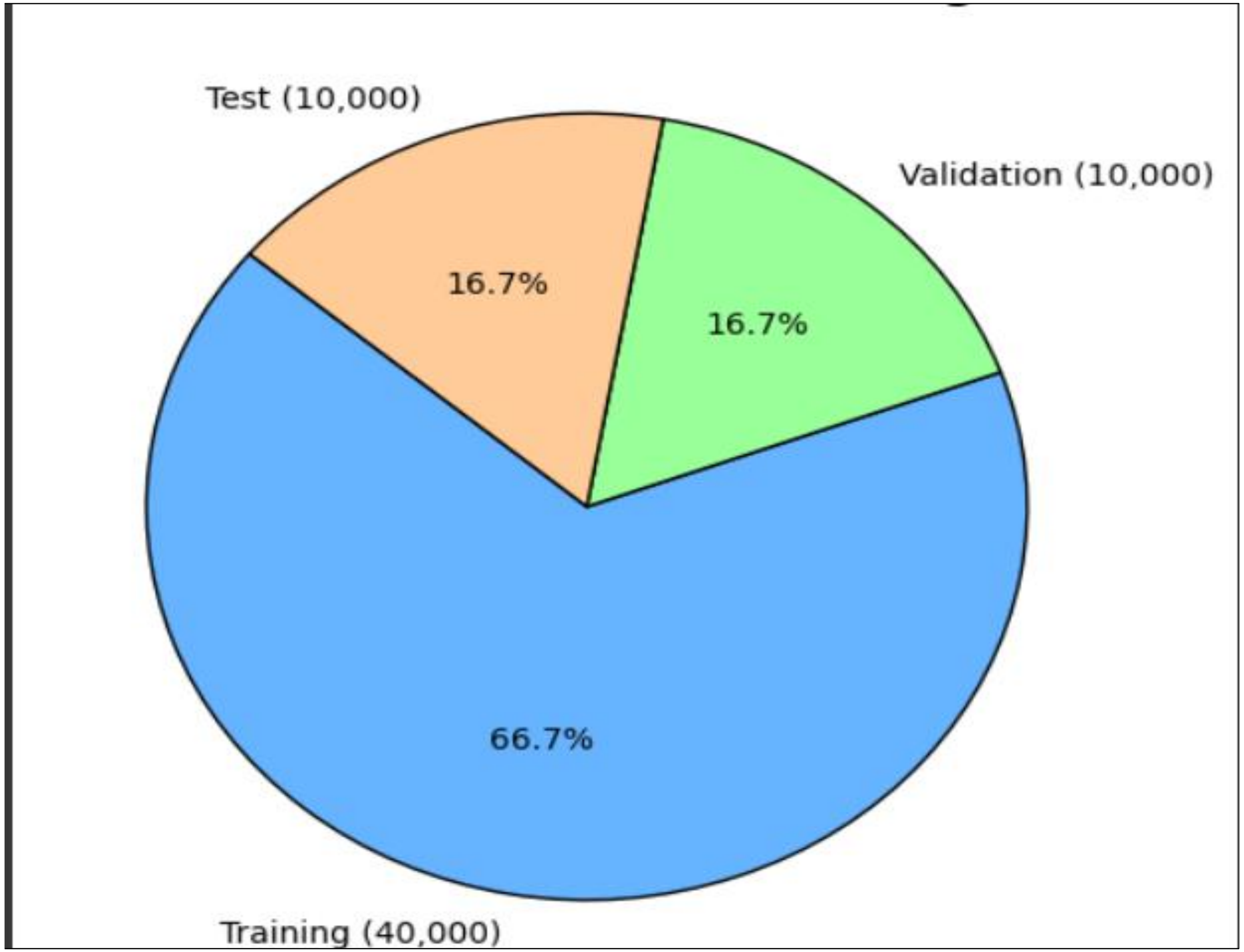
DATA PREPARATION AND PREPROCESSING

DATASET PARTIONINIG

- Dataset divided into:
  - Training: 40,000 (66.7%)
  - Validation: 10,000 (16.7%)
  - Test: 10,000 (16.7%)

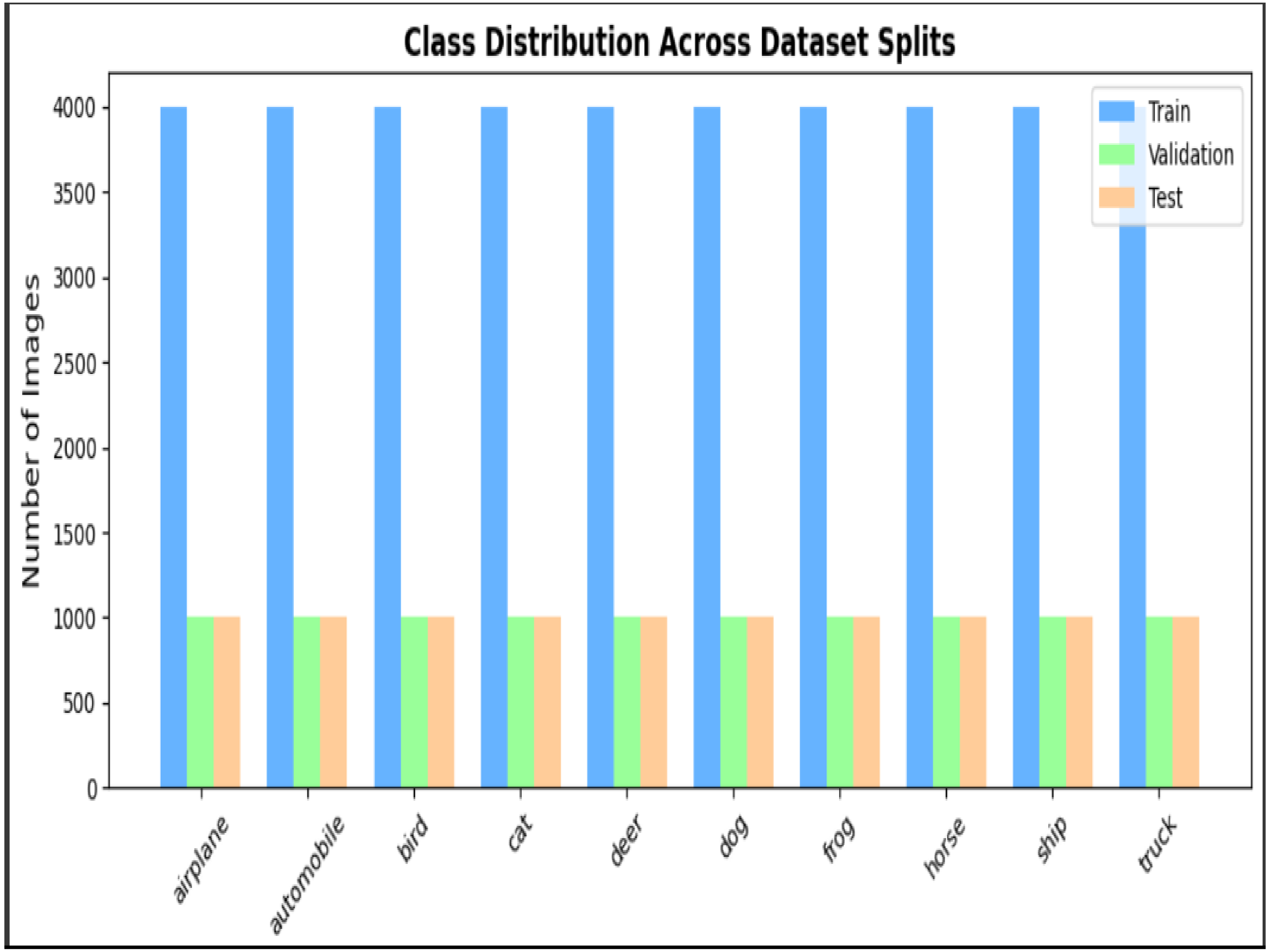
- Used **stratified sampling** to preserve class balance across subsets.
- Validation set used to monitor model generalisation and prevent overfitting.

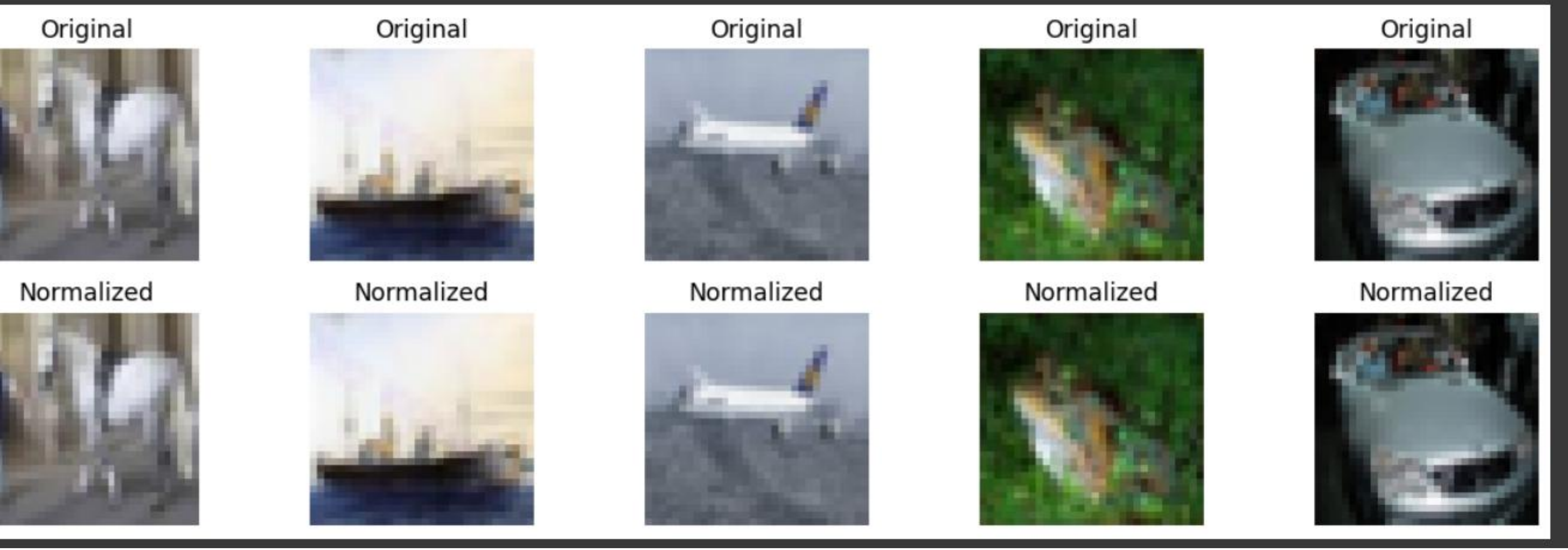
(Géron, 2022)



METADATA AND CLASSS DISTRIBUTION

- Image dimensions: **32 × 32 × 3 (RGB)**
- Data type: **float32**
- Pixel range: **0.0 – 1.0** *(post-normalisation)*
- Mean RGB intensity: **(0.49, 0.48, 0.45)**
- Standard deviation: **≈ 0.24 across channels**
- Dataset split: **Train 40,000 | Validation 10,000 | Test 10,000.**
- Class balance preserved across splits (as shown in chart).

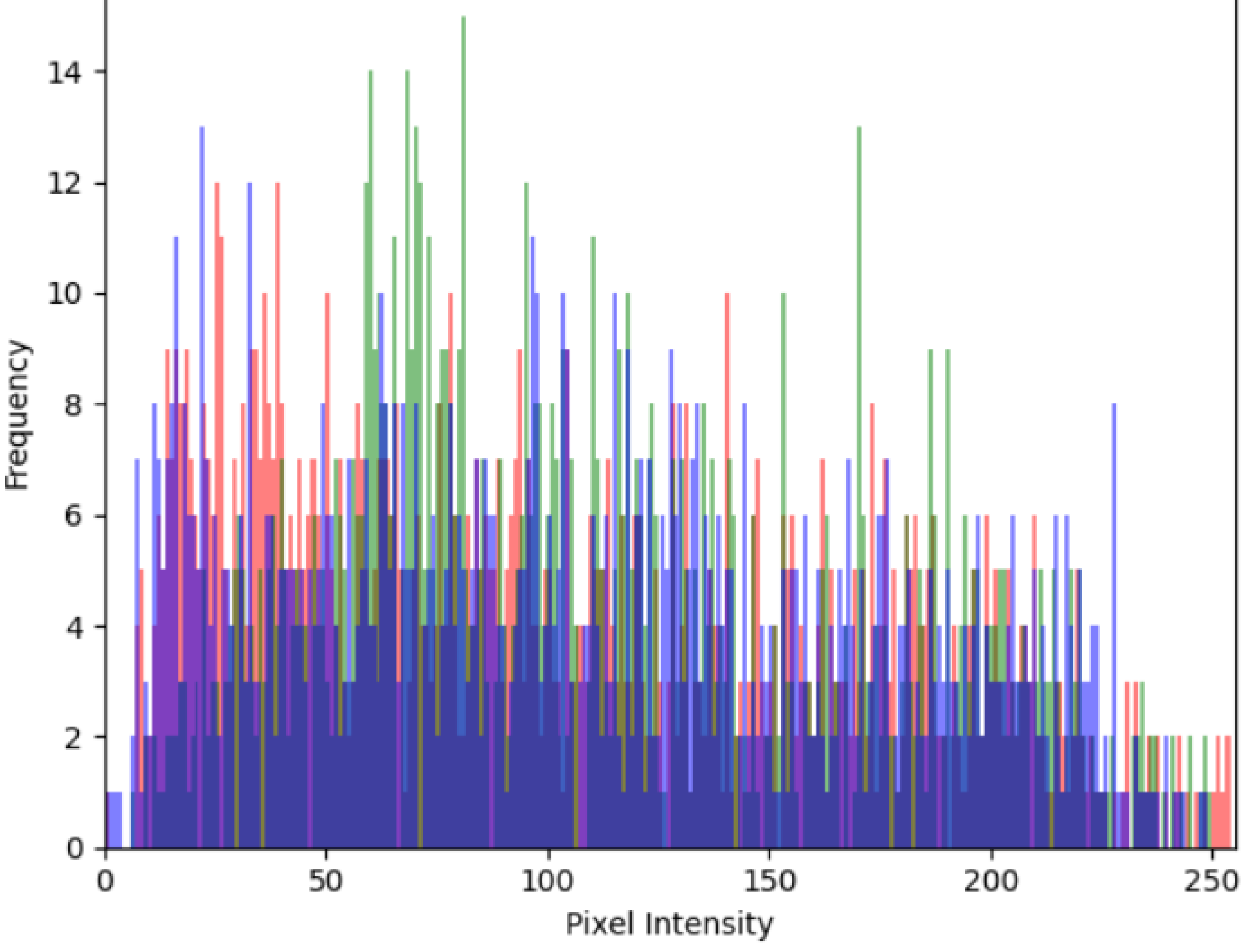


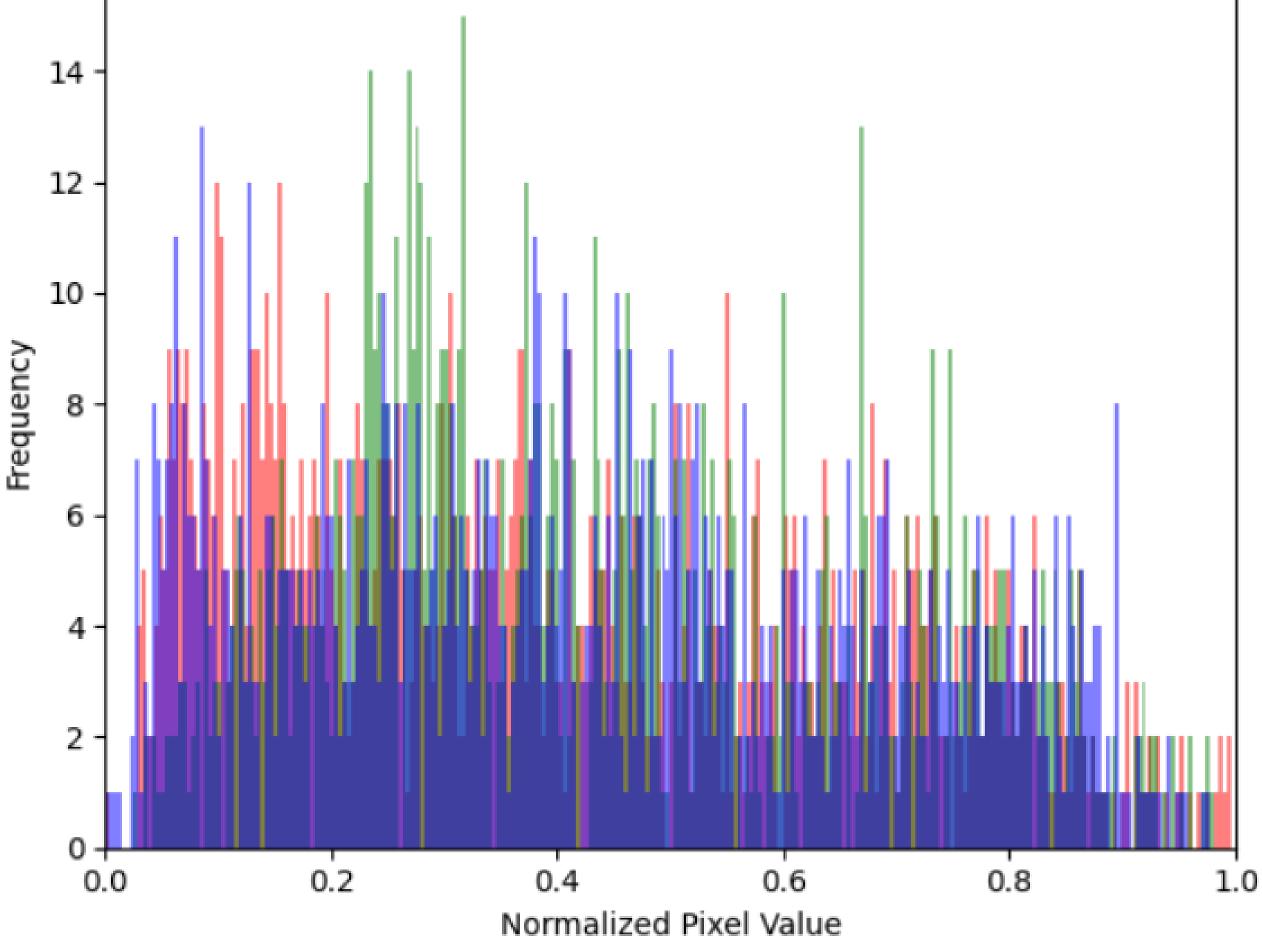


Normalisation rescales pixel intensity values from [0–255] to [0–1] without distorting image quality.

NORMALISATION

- Pixel values in the CIFAR-10 dataset range from 0–255.
- To improve model performance and stability, these values were scaled to the [0,1] range.
- Normalisation ensures faster convergence and reduces gradient instability during training (Rawat and Wang, 2017).





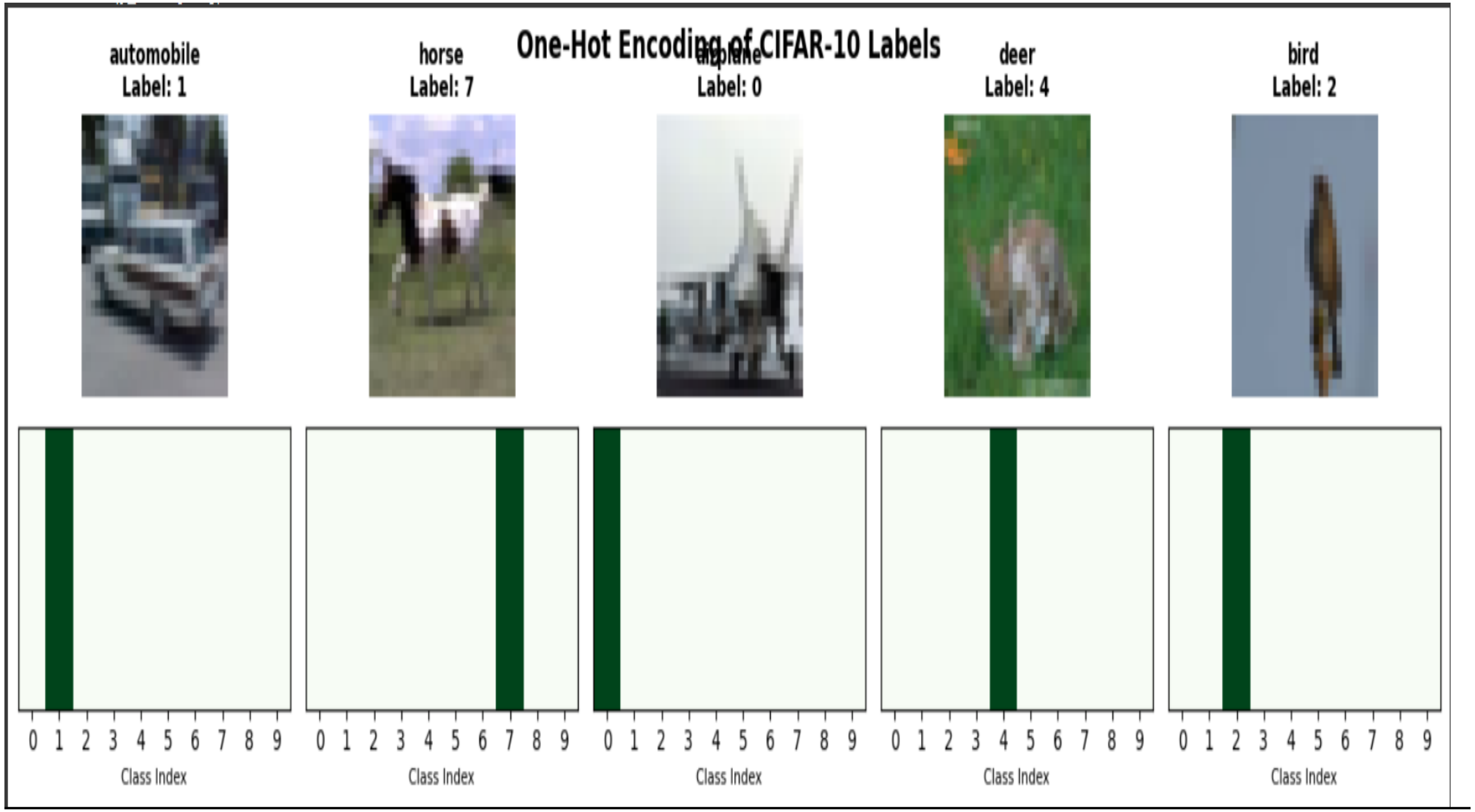
NORMALISATION HISTOGARM COMPARISON

- Before normalisation: Pixel values spread between 0–255.
- After normalisation: All values lie within 0–1 range.
- Distribution shape preserved and only the scale was adjusted.
- Normalisation reduces computation cost and enhances gradient flow.

ONE - HOT ENCODING

- CIFAR-10 has 10 categorical classes.
- Labels converted into binary vectors (length = 10).
- Each vector has a single “hot” value representing the class.
- Prevents ordinal bias between categories.
- Supports soft-max activation and cross-entropy loss.

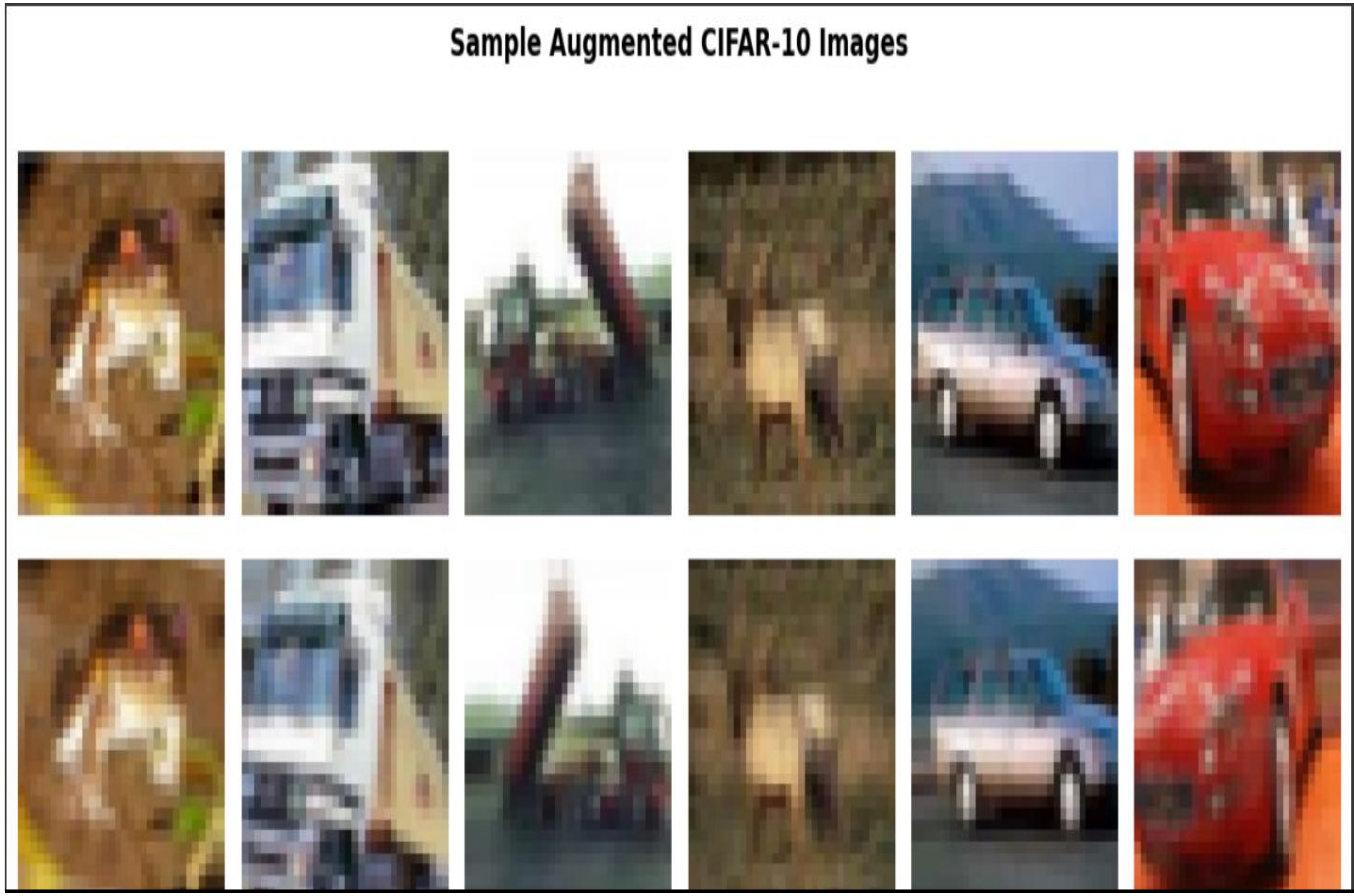
(Zhu, Qiu and Fu, 2024).



DATA AUGUMENTATION

- Applied random transformations such as rotation, horizontal flips, and width/height shifts to training images only.
- Augmentation increases dataset diversity without collecting new images.
- Helps prevent overfitting by teaching the model to recognise objects from different angles and positions

(Shorten and Khoshgoftaar, 2019).



MODEL ARCHITECTURE

OVERVIEW OF CONVOLUTIONAL NEURAL NETWORK WORKFLOW

- CNN extracts hierarchical features from input images using convolution and pooling layers.
- Enables automated feature learning for accurate image recognition in CIFAR-10.
- Minimises loss by adjusting kernels and weights during each iteration.

(Yamashita *et al.*, 2018).

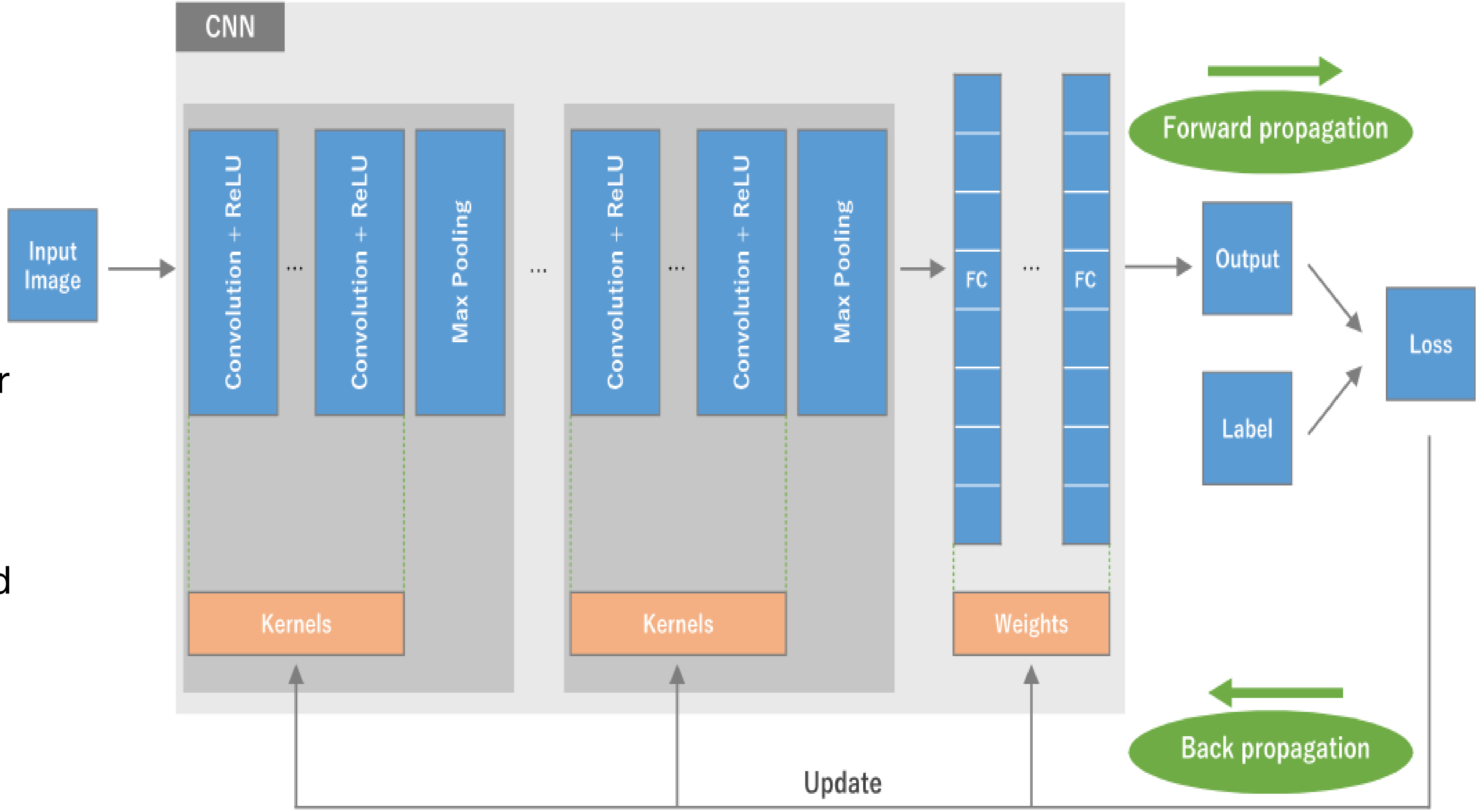


Figure 3: An overview of CNN training process

(Yamashita *et al.*, 2018).

CNN ARCHITECTURE

- **Convolution layers** detect local patterns such as edges or textures.
- **ReLU layers** introduce non-linearity for better feature learning.
- **Pooling layers** reduce spatial size and control overfitting.
- **Fully connected layers** combine extracted features for final classification.
- Output layer uses **Softmax** to predict class probabilities (e.g., airplane, car, dog).

(Rawat and Wang (2017).

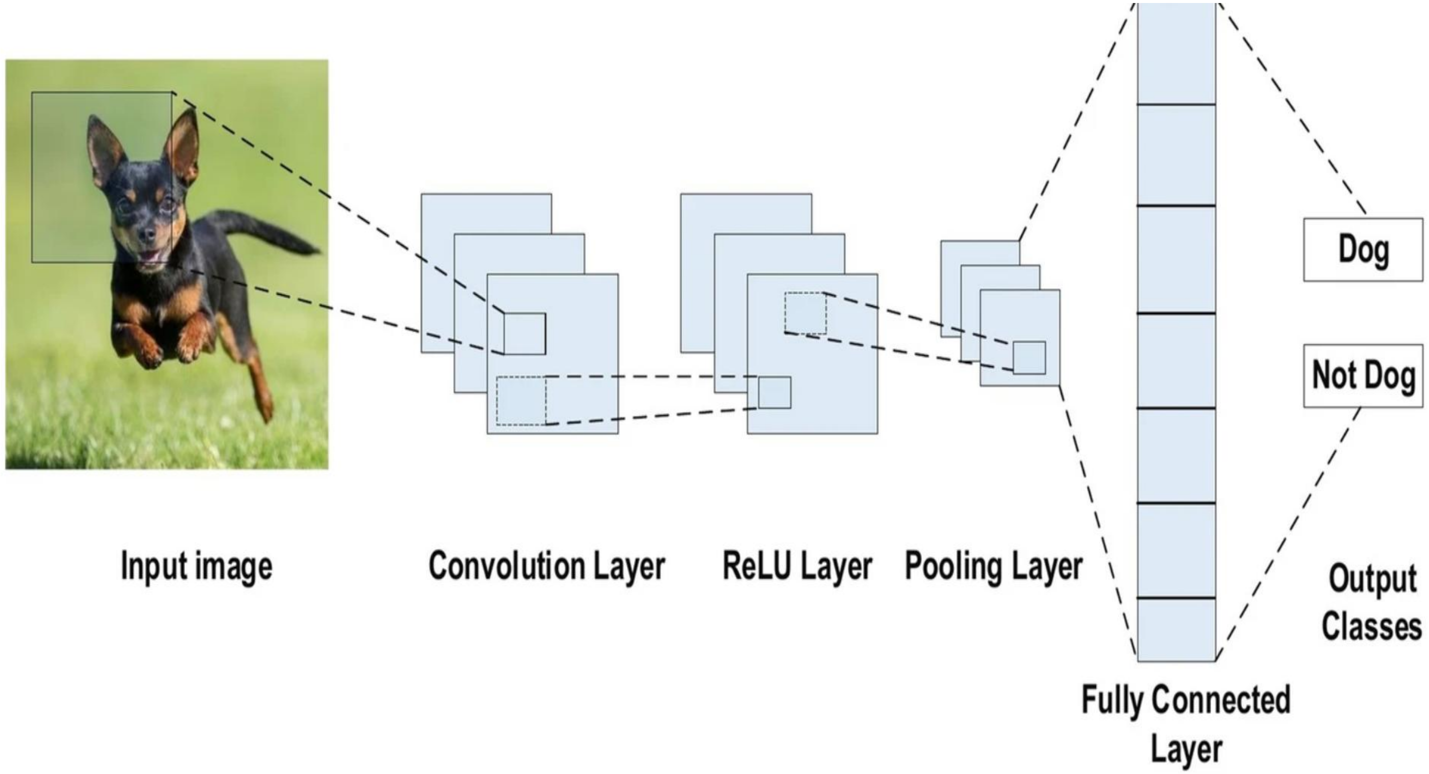


Figure 4: An example of a CNN Architecture

(Alzubaidi *et al.,* 2024)

REASON FOR CHOSEN MODEL

- Deep learning automates feature extraction.
- CNNs learn complex visual patterns directly from raw pixel data.
- Outperforms traditional ML for large-scale image datasets like CIFAR-10.
- Simplifies workflow while improving accuracy and generalisation.

(Bhatt *et al*., 2021).

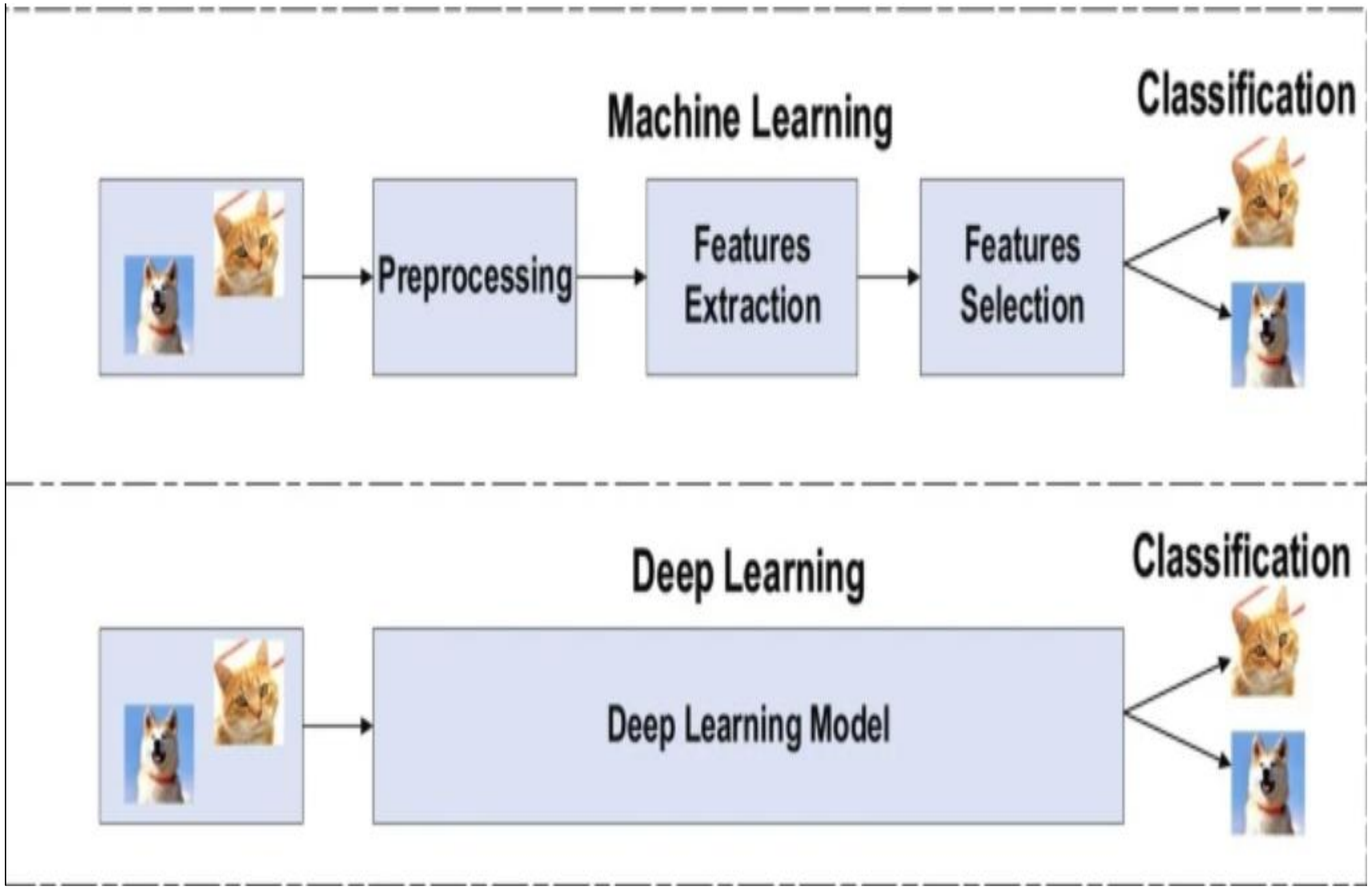


Figure 5: Difference between Machine Learning and Deep Learning (Alzubaidi *et al.,* 2024)

MODEL 1

BASELINE CNN ARCHITECTURE

- **3 convolutional–pooling blocks** progressively extract spatial features and reduce image dimensions (from 32×32 → 4×4).

- Model 1 is lightweight with **≈356k trainable parameters**, suitable for efficient training on CIFAR-10.

Model: "sequential"

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 32, 32, 32)	896
max_pooling2d (MaxPooling2D)	(None, 16, 16, 32)	0
conv2d_1 (Conv2D)	(None, 16, 16, 64)	18,496
max_pooling2d_1 (MaxPooling2D)	(None, 8, 8, 64)	0
conv2d_2 (Conv2D)	(None, 8, 8, 128)	73,856
max_pooling2d_2 (MaxPooling2D)	(None, 4, 4, 128)	0
flatten (Flatten)	(None, 2048)	0
dense (Dense)	(None, 128)	262,272
dense_1 (Dense)	(None, 10)	1,290

Total params: 356,810 (1.36 MB)
Trainable params: 356,810 (1.36 MB)
Non-trainable params: 0 (0.00 B)

EPOCHS AND TRAINING STRATEGY

Purpose: To balance efficient learning with prevention of overfitting and optimise model generalisation.

- Optimiser: Adam (adaptive learning for faster convergence).
- Learning Rate: 0.001, reduced automatically on plateau.

Epochs: 15 with batch size of 64 but stopped early at epoch 9 (using Early Stopping).

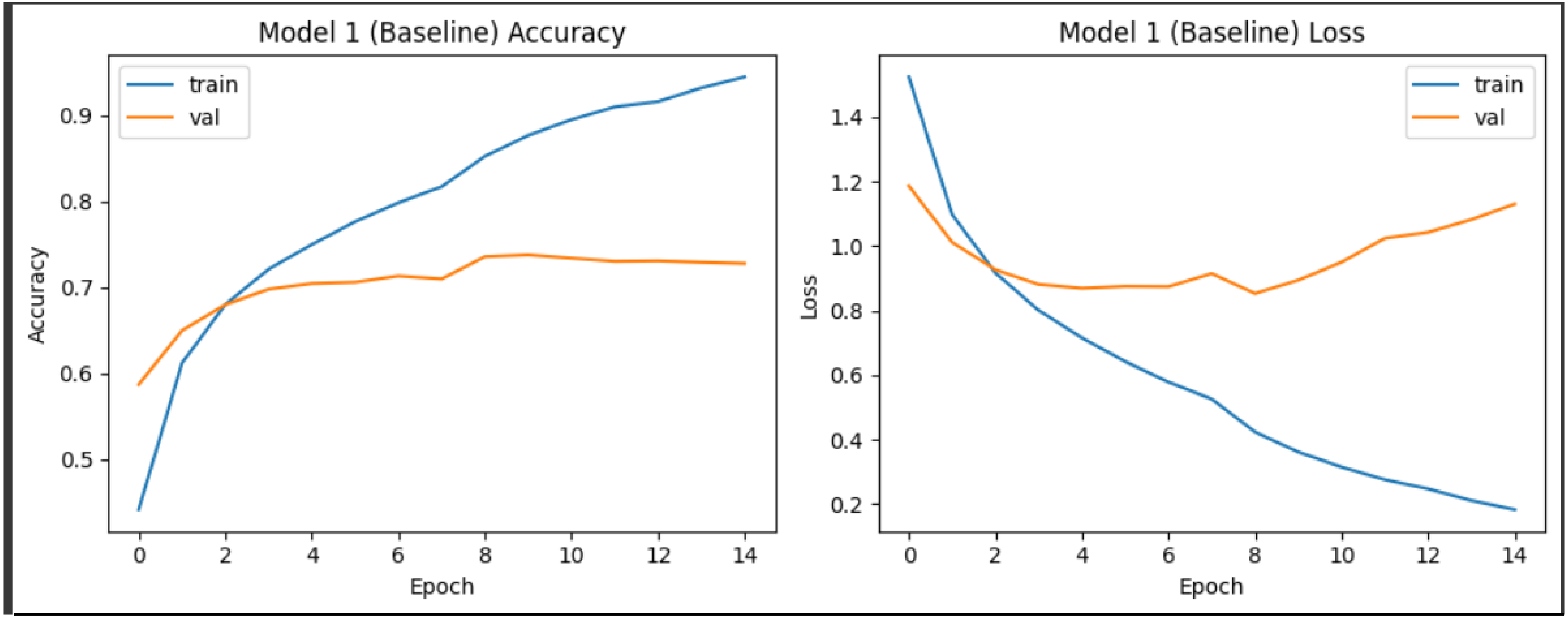
- Validation Split: 20% of training data reserved for validation.

Callbacks:

- *ReduceLROnPlateau* – lowers learning rate when validation loss stops improving.

TRAINING PERFORMANCE

- Training accuracy improved consistently, reaching 95% by the final epoch.
- Validation accuracy plateaued around 73%, suggesting mild overfitting as training continued.
- Training loss declined consistently, while validation loss began to rise after epoch 8, confirming the same overfitting trend.

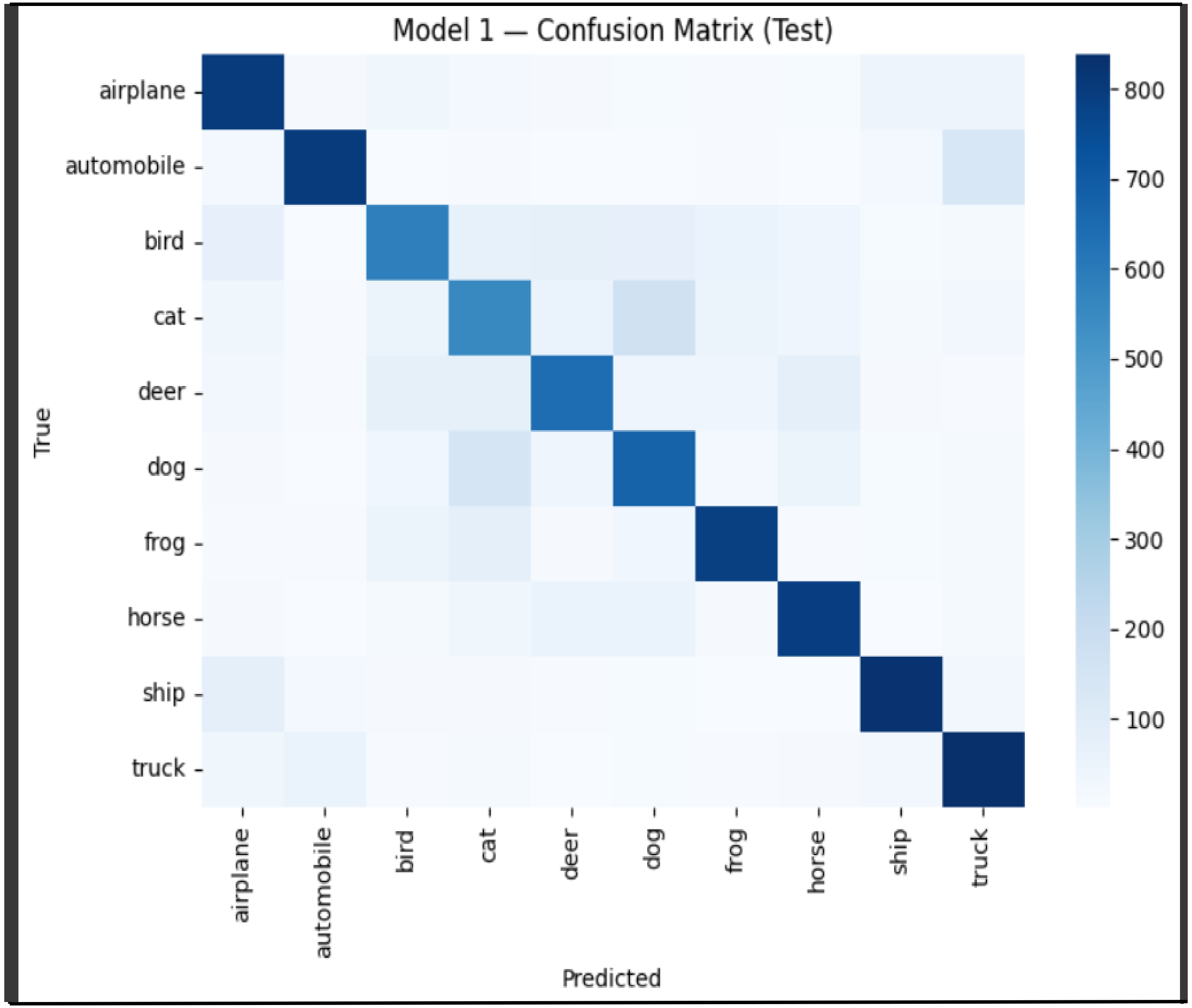


EVALUATION METRICS

Metric	Training Set	Validation Set	Test Set	Interpretation
Accuracy	0.8753	0.7358	0.7289	Good learning on training data; validation and test results are consistent, showing generalisation.
Loss	0.366	0.8529	0.8782	Slight overfitting as validation/test loss remain higher than training loss.
Epoch at Best Model	8	-	-	Early stopping prevented further overfitting, saving the best-performing weights.

CONFUSION MATRIX

- Strong diagonal pattern shows correct class predictions across all classes
- Misclassifications occur mainly between similar classes (e.g., cats vs. dogs, airplanes vs. ships).
- Highest performance was achieved for automobiles, ships, and trucks due to their distinctive visual features.
- Overall test accuracy was approximately **73%**, reflecting solid generalisation for a baseline CNN.



MODEL 2

ENHANCED CNN ARCHITECTURE

- Built on the baseline CNN with additional convolutional layers to deepen feature extraction.
- Integrated Batch Normalisation after each convolutional layer to stabilise learning and accelerate convergence.
- Added Dropout layers to reduce overfitting and improve generalisation.
- Model includes approximately 530k trainable parameters, balancing performance and efficiency.

Layer (type)	Output Shape	Param #
input_layer_1 (InputLayer)	(None, 32, 32, 3)	0
augmentation (Sequential)	(None, 32, 32, 3)	0
conv2d_3 (Conv2D)	(None, 32, 32, 32)	896
batch_normalization (BatchNormalization)	(None, 32, 32, 32)	128
activation (Activation)	(None, 32, 32, 32)	0
conv2d_4 (Conv2D)	(None, 32, 32, 32)	9,248
batch_normalization_1 (BatchNormalization)	(None, 32, 32, 32)	128
activation_1 (Activation)	(None, 32, 32, 32)	0
max_pooling2d_3 (MaxPooling2D)	(None, 16, 16, 32)	0
dropout (Dropout)	(None, 16, 16, 32)	0
conv2d_5 (Conv2D)	(None, 16, 16, 64)	18,496
batch_normalization_2 (BatchNormalization)	(None, 16, 16, 64)	256
activation_2 (Activation)	(None, 16, 16, 64)	0
conv2d_6 (Conv2D)	(None, 16, 16, 64)	36,928
batch_normalization_3 (BatchNormalization)	(None, 16, 16, 64)	256
activation_3 (Activation)	(None, 16, 16, 64)	0
max_pooling2d_4 (MaxPooling2D)	(None, 8, 8, 64)	0
dropout_1 (Dropout)	(None, 8, 8, 64)	0
conv2d_7 (Conv2D)	(None, 8, 8, 128)	73,856
batch_normalization_4 (BatchNormalization)	(None, 8, 8, 128)	512
activation_4 (Activation)	(None, 8, 8, 128)	0
max_pooling2d_5 (MaxPooling2D)	(None, 4, 4, 128)	0
dropout_2 (Dropout)	(None, 4, 4, 128)	0
flatten_1 (Flatten)	(None, 2048)	0

EPOCHS AND TRAINING STRATEGY

- **Purpose:** To enhance the baseline model's learning efficiency and generalisation by extending training duration and integrating adaptive callbacks.
- **Epochs:** Trained up to *25 epochs* (stabilised around *epoch 20*).

    Batch Size: 64 (kept consistent for comparability).

    Learning Rate: Initial value 0.001, reduced adaptively when validation loss plateaued.

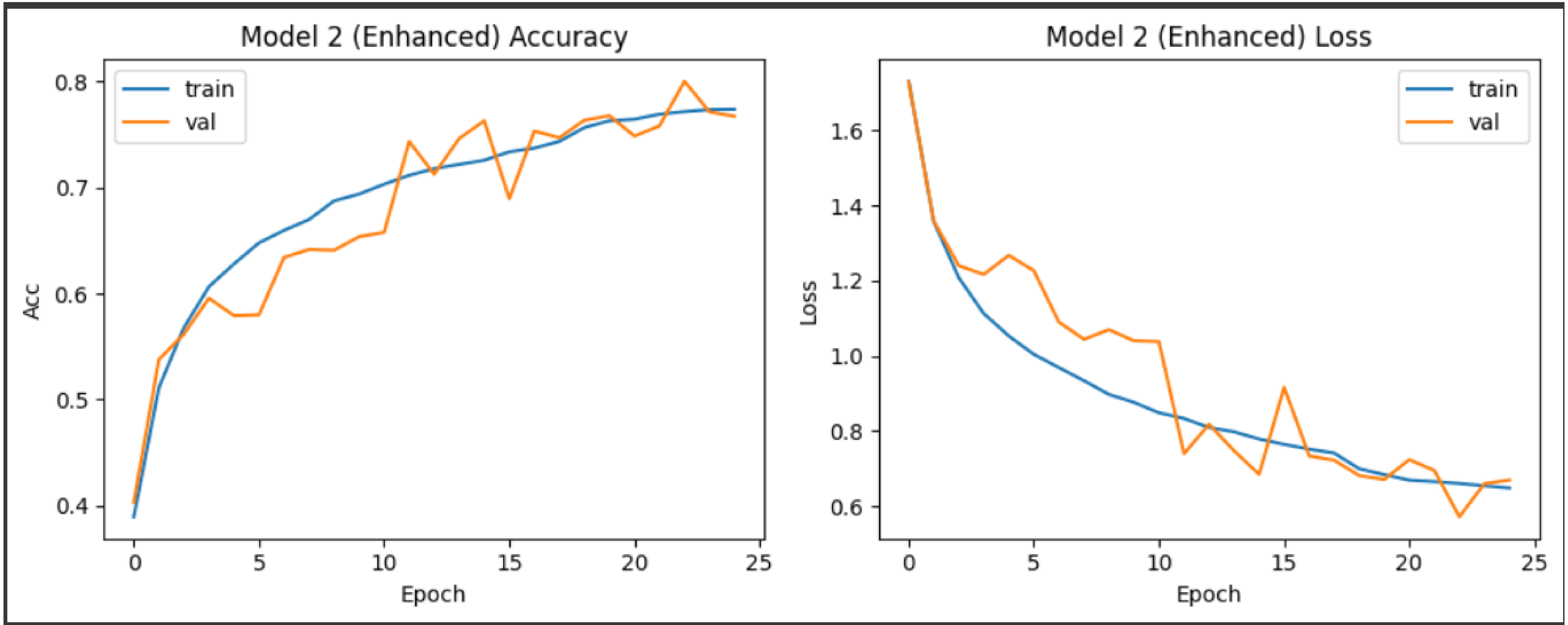
    Validation Split: 20% of training data reserved for model tuning.

- **Callbacks:** Adam Optimiser

    ReduceLROnPlateau: Automatically lowered learning rate to sustain improvement.

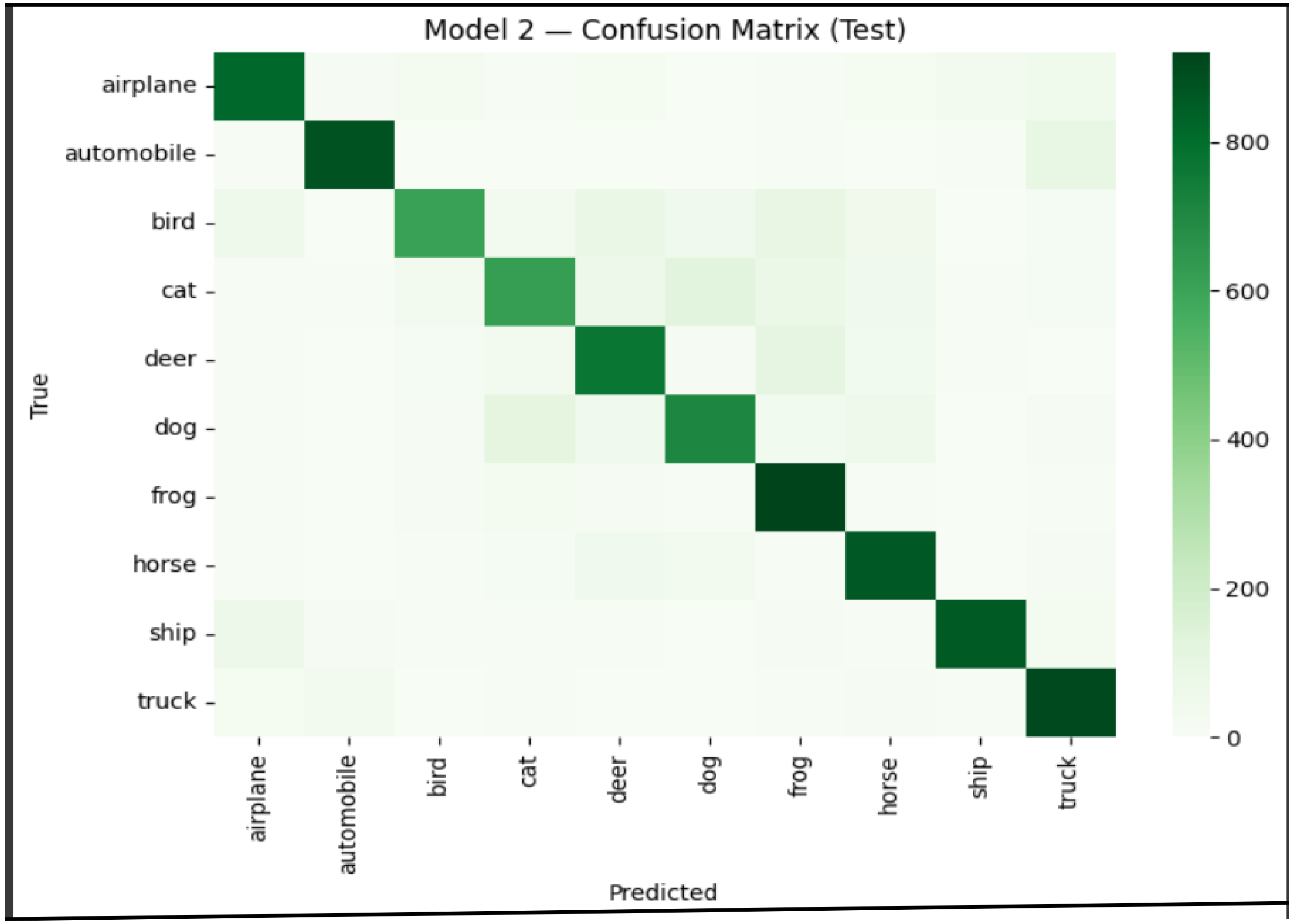
TRAINING PERFORMANCE

- Training and validation accuracy both increased steadily, converging around 77–80% by the final epoch.
- Validation performance remained close to training accuracy, indicating reduced overfitting.
- Both training and validation loss decreased consistently, confirming stable optimisation.



CONFUSION MATRIX AND INSIGHTS

- Strong diagonal dominance indicates improved class prediction accuracy across all categories.
- Misclassifications were reduced, especially between similar classes such as cats and dogs.
- Distinct classes like automobiles, ships, and trucks maintain high recognition accuracy.



EVALUATION METRICS

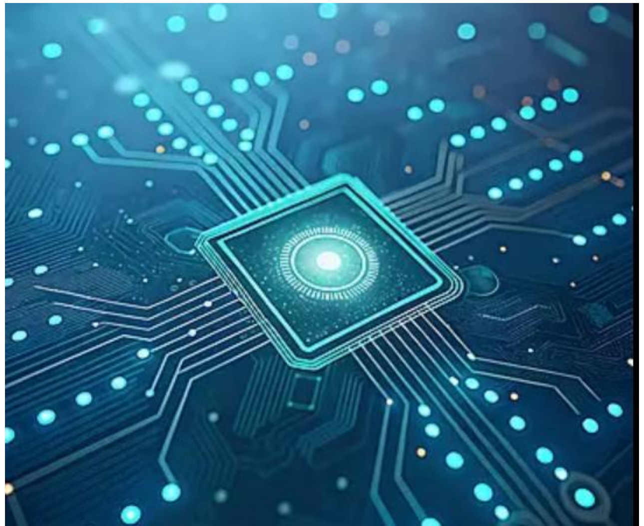
Metric	Training Set	Validation Set	Test Set	Interpretation
Accuracy	0.8270	0.8001	0.7950	Higher accuracy across all sets compared to Model 1. Validation and test scores are closely aligned, showing improved generalisation.
Loss	0.4918	0.5717	0.5996	Lower loss across datasets, confirming better optimisation and stability. Minor gap suggests mild overfitting but within acceptable limits.
Epoch at Best Model	20	-	-	Model stabilised around epoch 20 before early stopping prevented overfitting.

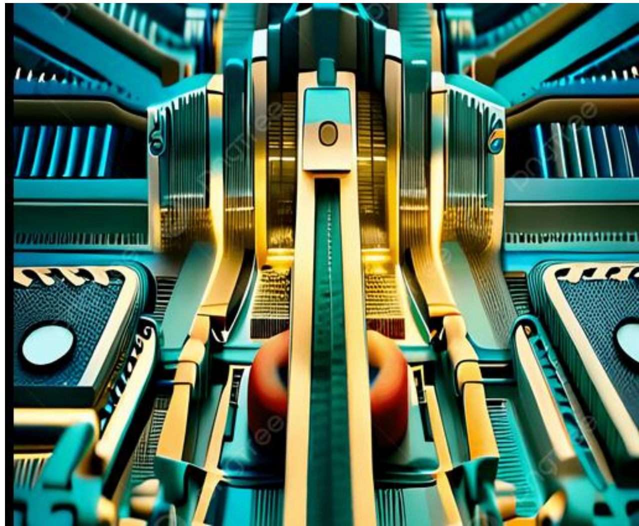
MODEL 1 VS MODEL 2 COMPARSION

Metric	Model 1 (Baseline CNN)	Model 2 (Enhanced CNN)	Observation
Train Accuracy	0.8753	0.8270	Slightly lower due to regularisation but indicates reduced overfitting.
Validation Accuracy	0.7358	0.8001	Significant improvement, showing stronger generalisation.
Test Accuracy	0.7289	0.7596	Higher test accuracy confirms better real-world performance.
Train Loss	0.366	0.4918	Increased slightly due to dropout, helping generalisation.
Validation Loss	0.8529	0.5717	Reduced drastically, showing improved stability and lower overfitting.
Overfitting Trend	High (validation loss increased)	Reduced (training and validation closely aligned)	Enhanced architecture improved balance.

REFLECTION AND CONCLUSION

- Demonstrated the impact of architectural refinement on model performance and stability.
- Future improvements could include data augmentation, deeper architectures, or transfer learning.





References

- Alzubaidi, L., Zhang, J., Humaidi, A.J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M.A., Al-Amidie, M. and Farhan, L. (2021) ‘Review of deep learning: concepts, CNN architectures, challenges, applications, future directions’, *Journal of Big Data*, 8(1), p. 53.

- Bhatt, D., Patel, C., Talsania, H., Patel, J., Vaghela, R., Pandya, S., Modi, K. and Ghayvat, H. (2021) ‘CNN variants for computer vision: History, architecture, application, challenges and future scope’ *Electronics*, 10(20), p. 2470.

- Géron, A. (2022) *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly Media, Inc.

- Rawat, W. and Wang, Z. (2017) ‘Deep convolutional neural networks for image classification: A comprehensive review’, *Neural computation 29*(9), pp. 2352-2449.

- Sarkar, T., Rakhra, M., Sharma, V., Takkar, S. and Jairath, K. (2024) ‘Comparative Study of Object Recognition Utilizing Machine Learning Techniques’, *International Conference on Communication, Computer Sciences and Engineering* pp. 726-731.

- Shorten, C. and Khoshgoftaar, T.M.|(2019) ‘A survey on image data augmentation for deep learning’, *Journal of big data* 6(1), pp.1-48.

- Yamashita, R., Nishio, M., Do, R.K.G. and Togashi, K. (2018) ‘Convolutional neural networks: An overview and application in radiology’ *Insights into imaging*, 9(4), pp. 611-629.

- Zhu, W., Qiu, R. and Fu, Y. (2024) ‘Comparative study on the performance of categorical variable encoders in classification and regression tasks’ *arXiv preprint arXiv:2401.09682*.