



Prifysgol Cymru
Y Drindod Dewi Sant
University of Wales
Trinity Saint David

Application of Machine Learning

Image classification using Artificial Neural Networks on the
CIFAR-10 Dataset

Project Report

<i>Name:</i>	<i>Shahriar Khan</i>
<i>Student ID:</i>	<i>2507023</i>
<i>Programme:</i>	<i>Data Science and Analytics</i>
<i>Module:</i>	<i>Application of Machine Learning</i>
<i>Module Code:</i>	<i>ACCA7025</i>
<i>Lecturer:</i>	<i>Mark Huntly</i>
<i>Internal Verifier:</i>	<i>Dr Seena Joseph</i>
<i>Date of submission:</i>	<i>18th May 2026</i>

Contents

1	Abstract	2
2	Introduction	2
3	Related Work	3
3.1	Recurrent Neural Networks (RNNs) for Comparison	3
4	Workflow Methodology	4
4.1	CNN Architecture Pipeline	4
4.2	Dataset Exploration	5
4.3	Model Development Process	6
4.4	Hyperparameter Configuration	6
5	Results	7
5.1	Accuracy Analysis	7
5.2	Loss Analysis	8
5.3	Model Evaluation	9
5.4	Classification Report	10
6	Discussion and Critical Analysis	12
7	Conclusion	13

1. Abstract

Artificial Neural Networks (ANNs) have become central to modern machine learning (ML), particularly in solving complex pattern recognition problems. Convolutional Neural Networks (CNNs), a specialized form of ANN, have demonstrated exceptional performance in image classification due to their ability to automatically learn spatial hierarchies from raw input data. Recent developments in deep learning, including transformer-based models and improved CNN architectures, have further enhanced classification performance on benchmark datasets [1].

This study presents the design, implementation, and evaluation of a CNN model applied to the CIFAR_10 dataset. A baseline model is initially constructed and subsequently improved through architectural modifications, including additional convolutional layers, dropout regularization, and batch normalization. The enhanced model is evaluated using accuracy, loss curves, a confusion matrix, and classification metrics. The final model achieves approximately 79% accuracy on the test dataset. Although the results demonstrate effective classification performance, certain limitations remain, particularly in distinguishing visually similar classes. The findings highlight the importance of model design, optimization, and hyperparameter tuning in improving deep learning performance.

2. Introduction

ANNs have significantly advanced AI by enabling systems to learn directly from data rather than relying on manually engineered features [2]. Among these networks, CNNs have proven highly effective in image classification tasks due to their ability to exploit spatial structures within images.

Image classification involves assigning a label to an input image based on its content. Traditional ML approaches relied heavily on handcrafted features, which limited scalability and performance. CNNs overcome this by learning hierarchical representations automatically, from low-level features such as edges to high-level semantic concepts [3]. The CIFAR_10 dataset is widely used as a benchmark for evaluating image classification models [4]. It consists of 60,000 color images divided into ten classes, with each image having a resolution of 32×32 pixels [5]. Despite its relatively small size, the dataset presents challenges due to low resolution and similarity between certain categories [6]. The objective of this report is to develop a CNN model for image classification using CIFAR_10 and critically evaluate its performance. The report also explores current research trends to provide context for the results obtained.

3. Related Work

Recent research in image classification has shown significant progress due to improvements in CNN architectures and the introduction of alternative approaches such as vision transformers. CNN-based approaches remain highly relevant, as they provide efficient feature extraction and strong generalization performance. Modern CNN research focuses on improving network depth, reducing overfitting, and enhancing feature representation. Enhanced CNN architectures that integrate deeper layers and regularization techniques have been shown to improve classification accuracy on CIFAR-10 beyond baseline models [6]. One important development is EfficientNetV2, which introduces training-aware scaling and improved architecture design. This model achieves higher accuracy while maintaining efficient training performance by optimizing both network size and learning strategy [7].

In addition to CNN improvements, transformer-based models have recently emerged as strong competitors. Vision Transformers (ViT) process images as sequences of patches rather than using convolutional operations. These models have demonstrated high performance when trained on large datasets [8].

Hybrid models that combine convolution and transformer layers have also been proposed. For example, convolutional vision transformers integrate the strengths of both approaches, achieving improved efficiency and performance [9].

3.1 Recurrent Neural Networks (RNNs) for Comparison

RNNs are designed for sequential data processing, where the order of inputs is significant. Unlike CNNs, which capture spatial features, RNNs maintain temporal dependencies through recurrent connections [2]. RNNs are widely used in applications such as natural language processing, speech recognition, and time-series analysis. Advanced variants such as LSTM and GRU address issues like vanishing gradients, enabling the handling of long sequences [2]. However, in the context of CIFAR_10 image classification, RNNs are not the most suitable choice. The dataset consists of static images, where spatial feature extraction is more important than temporal relationships. CNNs are therefore preferred for this task due to their ability to capture spatial hierarchies efficiently [3]. Nevertheless, RNNs can complement CNNs in applications such as video classification, where sequential frames are analyzed over time. This comparison highlights the importance of selecting an appropriate model based on the structure of the data.

4. Workflow Methodology

4.1 CNN Architecture Pipeline

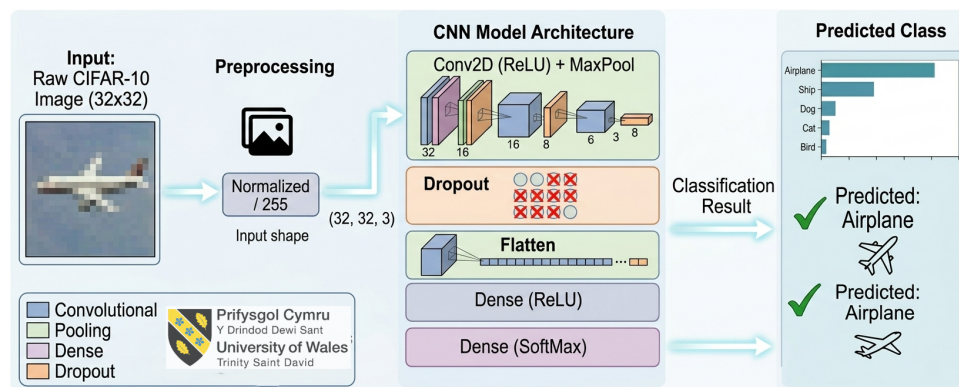


Figure 1: CNN Model Architecture Pipeline

The model begins with CIFAR_10 images, which are normalized to improve numerical stability. Convolutional layers are applied to extract spatial features, followed by pooling layers to reduce dimensionality.

Dropout is included to reduce overfitting by randomly deactivating neurons during training. This improves model generalization and prevents memorization of the training data. Finally, classification is performed using fully connected layers with a softmax activation function.

4.2 Dataset Exploration

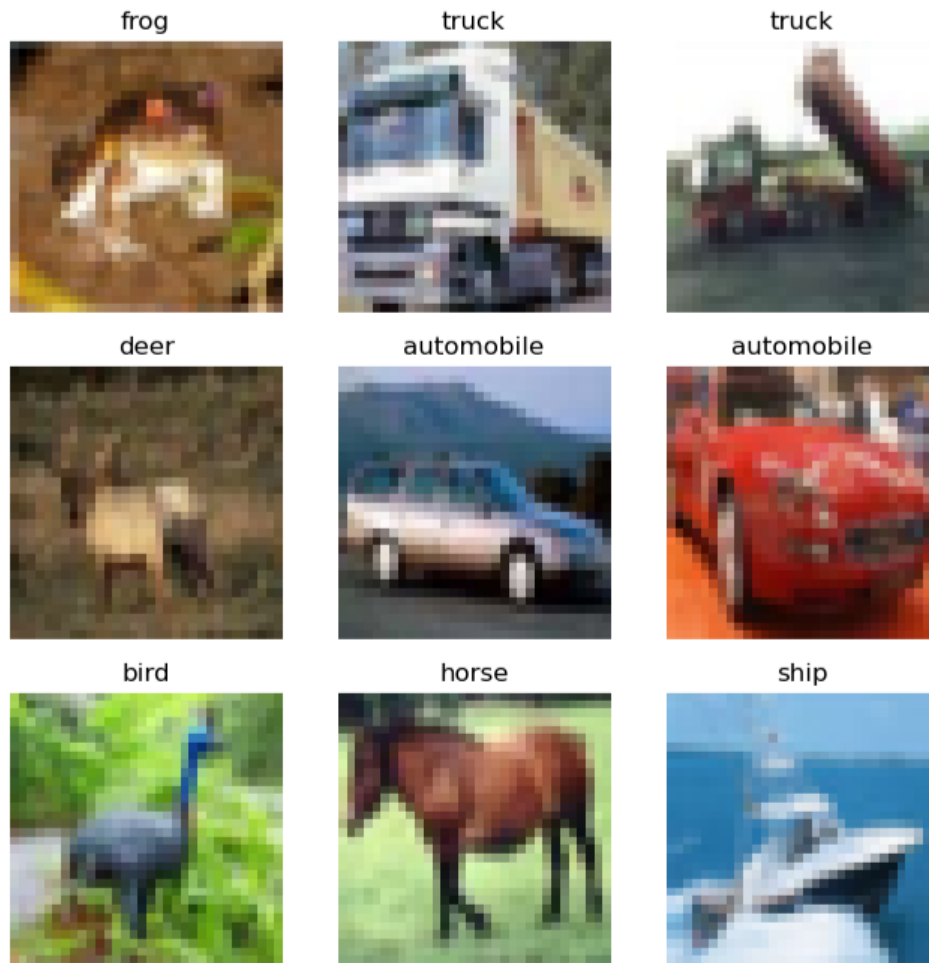


Figure 2: Dataset Classes

The CIFAR-10 dataset is balanced, ensuring equal representation across all classes. Each category contains 6,000 images, making it suitable for supervised learning [6].

However, several challenges exist. The low resolution limits the amount of information available, and similar categories (such as animals) can lead to classification errors. These issues affect model performance and increase the difficulty of feature extraction.

4.3 Model Development Process

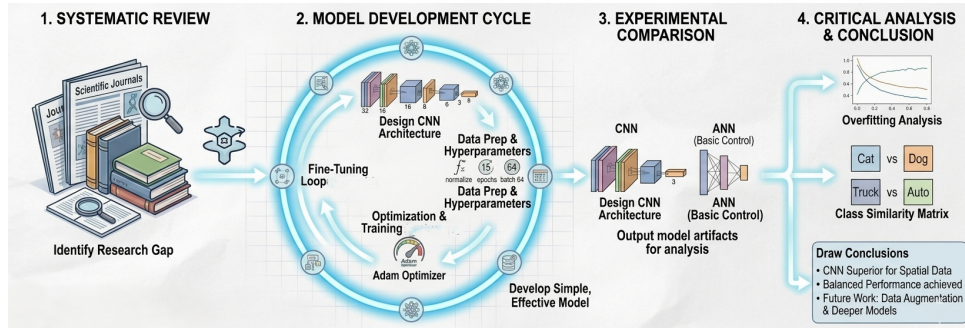


Figure 3: Model Development Lifecycle

The development process involves iterative stages, including model design, training, validation, and optimization. This cycle ensures continuous improvement in model performance.

4.4 Hyperparameter Configuration

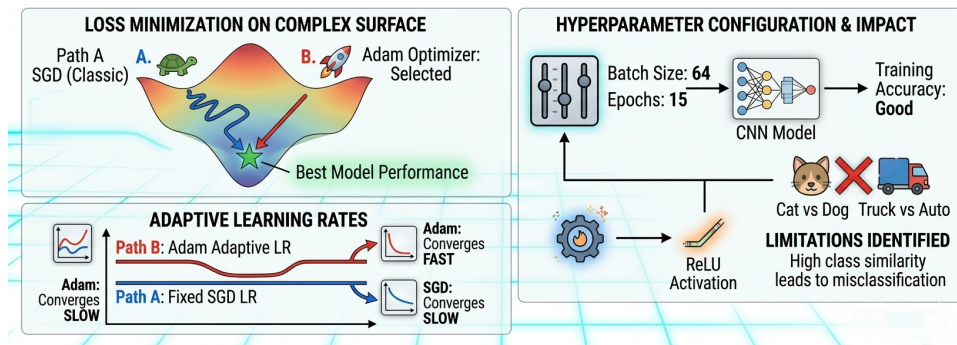


Figure 4: Optimisation Process (Adam Optimiser)

The model uses the Adam optimiser, which adapts the learning rate dynamically during training [10]. Key hyperparameters include batch size, number of epochs, and dropout rates, all tuned to achieve optimal performance.

5. Results

5.1 Accuracy Analysis

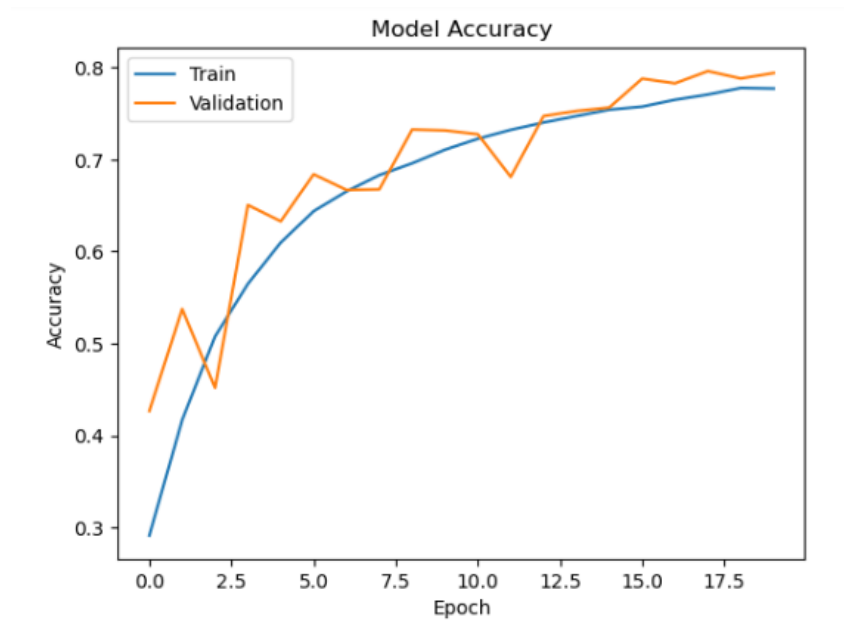


Figure 5: Training and Validation Accuracy

The training accuracy increases steadily from around 30% to approximately 78%, while validation accuracy stabilises 79%. This indicates that the model successfully learns meaningful patterns without severe overfitting.

```
[12]: test_loss, test_accuracy = model.evaluate(test_images, test_labels)
      print(test_accuracy)
```

313/313 ————— 6s 17ms/step - accuracy: 0.7926 - loss: 0.6299
0.7925999760627747

Figure 6: Final Model Accuracy

5.2 Loss Analysis

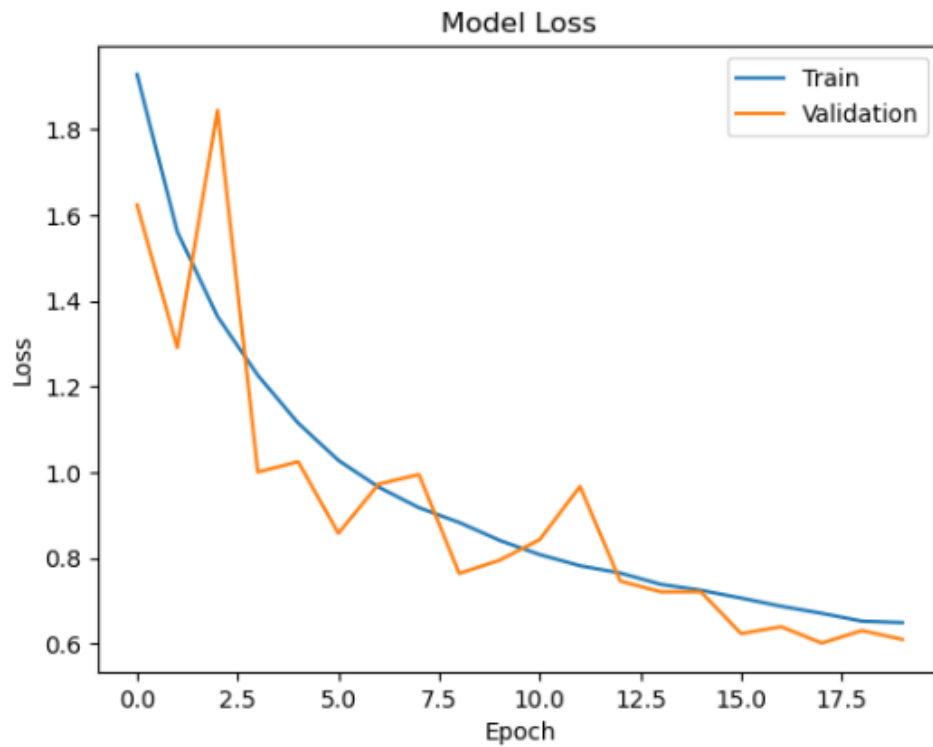


Figure 7: Training and Validation Loss

The loss curve shows a consistent decrease during training, indicating effective optimization. Slight fluctuations in validation loss suggest minor overfitting in later epochs.

5.3 Model Evaluation

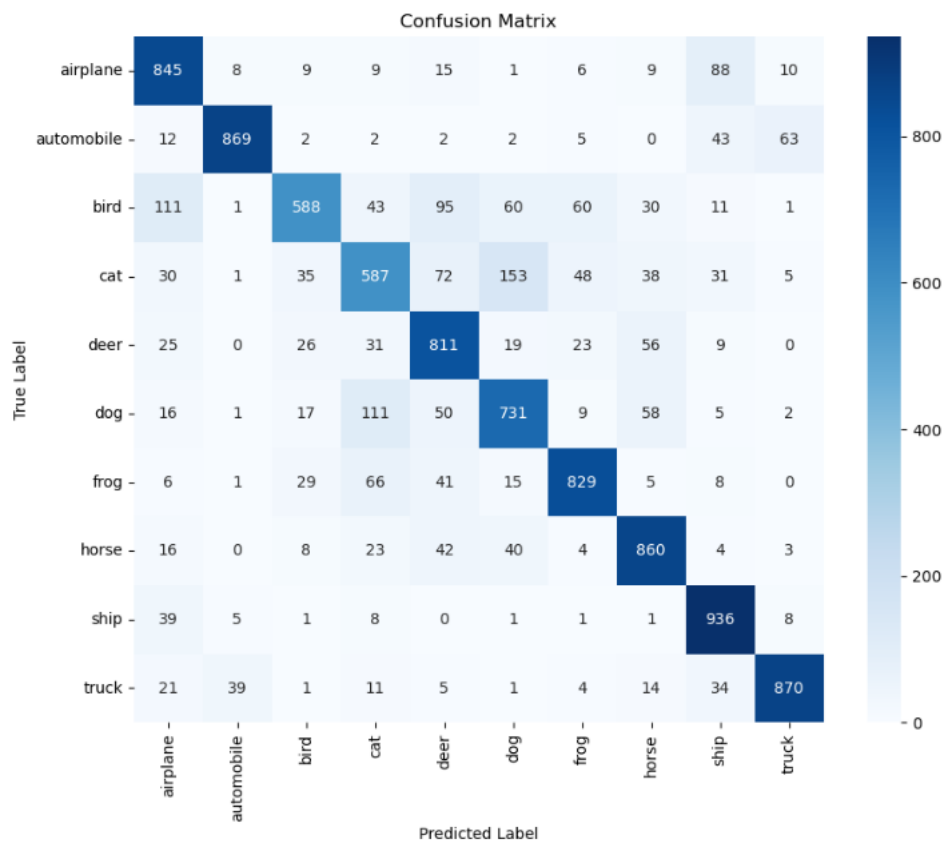


Figure 8: Confusion Matrix

The confusion matrix highlights classification performance across different classes. The model performs well on distinct categories such as ships and trucks but struggles with visually similar classes like cats and dogs.

5.4 Classification Report

	precision	recall	f1-score	support
airplane	0.75	0.84	0.80	1000
automobile	0.94	0.87	0.90	1000
bird	0.82	0.59	0.69	1000
cat	0.66	0.59	0.62	1000
deer	0.72	0.81	0.76	1000
dog	0.71	0.73	0.72	1000
frog	0.84	0.83	0.83	1000
horse	0.80	0.86	0.83	1000
ship	0.80	0.94	0.86	1000
truck	0.90	0.87	0.89	1000
accuracy			0.79	10000
macro avg	0.79	0.79	0.79	10000
weighted avg	0.79	0.79	0.79	10000

Figure 9: Precision, Recall and F1-score

Performance metrics such as precision and recall varied across classes. Higher values were observed for simple categories, while complex classes had lower scores. This imbalance highlights the challenges associated with feature similarity and dataset limitations.

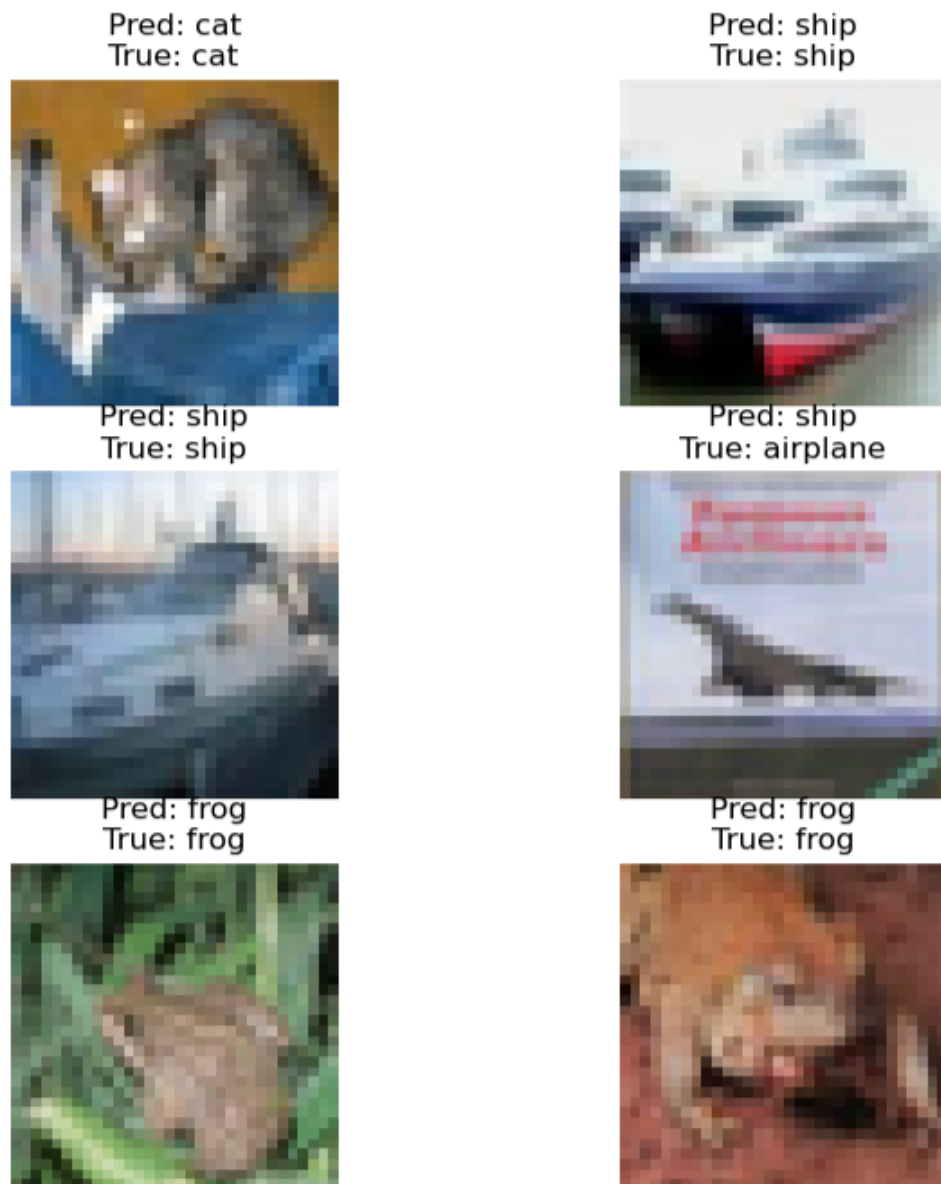


Figure 10: Comparison of predicted vs. true labels on sample images.

This picture shows several low-resolution images with their model predictions (Pred) and actual labels (True). Most images are classified correctly (e.g., cat, ship, frog), but one image is misclassified, where an airplane is predicted as a ship, highlighting a model error. This demonstrates both the model's accuracy and its limitations.

6. Discussion and Critical Analysis

The experimental results indicate that CNNs remain highly effective for image classification tasks, particularly when applied to structured datasets such as CIFAR_10. The achieved accuracy of approximately 79% suggests that the model is capable of learning meaningful hierarchical representations from the input data. This aligns with findings from recent studies, which highlight the strong feature extraction capability of CNN-based models in visual recognition tasks [1].

One of the key strengths of the implemented model is the use of regularisation techniques, such as dropout and batch normalisation. These techniques reduce overfitting by preventing the model from relying too heavily on specific neurons during training, leading to improved generalisation performance. Recent research demonstrates that enhanced CNN architectures incorporating these techniques achieve significantly better performance on classification benchmarks [6].

Despite these strengths, the model exhibits clear limitations. A major issue lies in its difficulty in distinguishing between visually similar classes, such as cats and dogs. This challenge is often associated with insufficient feature discrimination and the limited resolution of the CIFAR-10 dataset. Additionally, the architecture used in this study remains relatively shallow compared to modern CNN models. More advanced architectures, such as EfficientNetV2, have shown that deeper and more optimised networks can achieve higher accuracy through efficient scaling and improved training strategies [7]. Another important consideration is the emergence of transformer-based models. Vision Transformers (ViT) and related architectures have demonstrated strong performance by capturing global dependencies within images rather than focusing solely on local features [8].

Unlike CNNs, which rely on convolution operations, transformer-based methods use attention mechanisms to analyse the relationships between different regions of an image. Hybrid architectures, such as convolutional vision transformers, further combine the advantages of both approaches, leading to improved efficiency and performance [9], [11].

Furthermore, the absence of comprehensive data augmentation techniques in this study represents a limitation. Techniques such as rotation, flipping, and scaling are widely recognised for their ability to improve model robustness by increasing dataset variability. Studies show that augmentation plays a crucial role in enhancing generalisation, particularly when working with relatively small datasets like CIFAR-10 [6]. From an optimisation perspective, while the Adam optimiser provides efficient convergence, alternative training strategies could further improve performance. For example, adaptive learning rate scheduling and progressive training mechanisms have been successfully applied in modern architectures to achieve better results [7].

In summary, although the model developed in this study demonstrates solid performance,

it does not fully exploit recent advancements in deep learning. While CNNs continue to provide a strong baseline, the comparison with modern architectures highlights opportunities for improvement. Future work should explore deeper architectures, hybrid CNN-transformer models, and advanced data augmentation techniques to achieve higher classification accuracy and improved generalisation.

7. Conclusion

This report presented a CNN-based approach to image classification using the CIFAR-10 dataset. The model achieved approximately 79% accuracy, demonstrating that CNNs remain effective for feature extraction and classification tasks. However, several limitations were identified, including difficulty in distinguishing similar classes and the use of a relatively simple architecture. Future work could involve implementing deeper networks, using advanced architectures such as EfficientNetV2, and exploring transformer-based models. These approaches are likely to improve performance and provide better generalisation. Overall, the study highlights both the strengths and limitations of CNN-based image classification and emphasises the importance of continuous improvement in model design.

References

- [1] L. Chen, S. Li, Q. Bai, J. Yang, S. Jiang, and Y. Miao, “Review of image classification algorithms based on convolutional neural networks,” *Remote Sensing*, vol. 13, no. 22, 2021.
- [2] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*, vol. 1. MIT press Cambridge, 2016.
- [3] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [4] B. Kumar, K. Bavisetti, A. Vinay, S. Battula, K. Reddy, and A. K. Varma, “Performance investigation on image classification methods for cifar-10,” in *2024 Asian Conference on Intelligent Technologies (ACOIT)*, pp. 1–6, IEEE, 2024.
- [5] A. Krizhevsky, G. Hinton, *et al.*, “Learning multiple layers of features from tiny images,” 2009.
- [6] X. Yang, S. Yu, and W. Xu, “Enhanced convolutional neural networks for improved image classification,” *arXiv preprint arXiv:2502.00663*, 2025.
- [7] M. Tan and Q. Le, “Efficientnetv2: Smaller models and faster training,” in *International conference on machine learning*, pp. 10096–10106, PMLR, 2021.
- [8] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, *et al.*, “An image is worth 16x16 words: Transformers for image recognition at scale,” *arXiv preprint arXiv:2010.11929*, 2020.
- [9] H. Wu, B. Xiao, N. Codella, M. Liu, X. Dai, L. Yuan, and L. Zhang, “Cvt: Introducing convolutions to vision transformers,” in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 22–31, 2021.
- [10] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.
- [11] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, “Swin transformer: Hierarchical vision transformer using shifted windows,” in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 10012–10022, 2021.