Executive Summary

This project presents an end-to-end implementation and comparative evaluation of two neural network architectures. The Artificial Neural Network (ANN) and the Convolutional Neural Network (CNN) for image classification using the Fashion-MNIST dataset. The study aims to investigate how architectural design choices affect predictive accuracy, training efficiency, and generalisation performance on low-resolution visual data. The Fashion-MNIST dataset, developed by Xiao, Rasul, and Vollgraf (2017), contains 70,000 grayscale images (28 × 28 pixels) across ten apparel categories such as T-shirt/top, trouser, pullover, dress, coat, sandal, shirt, sneaker, bag, and ankle boot. It serves as a modern benchmark for testing deep learning models under controlled experimental settings without the complexity of large-scale datasets such as ImageNet. The experiment was implemented using TensorFlow and Keras frameworks, guided by open-source methodologies such as those demonstrated by Preda (2021) on Kaggle, which provide reproducible baselines for evaluating neural network performance.

Methodology Overview

The analysis followed a rigorous preprocessing and model-development pipeline. All images were normalised to a pixel intensity range of [0, 1] to stabilise gradient computations during backpropagation. From the original dataset, 50,000 samples were used for training, 10,000 were set aside for validation, and 10,000 for final testing to ensure unbiased performance assessment. The ANN treated each image as a flat vector of 784 features. Its architecture consisted of three dense layers with 512, 256, and 128 neurons, each activated by the ReLU function to enhance non-linearity. Dropout layers with a rate of 0.3 were inserted to reduce overfitting by randomly deactivating neurons during training.

The CNN, on the other hand, preserved the two-dimensional spatial structure of the images. It comprised three convolutional blocks with 32, 64, and 128 filters (3×3 kernels), each followed by ReLU activation and max-pooling layers to progressively reduce spatial dimensions while retaining feature salience. Batch normalisation was applied to stabilise learning and improve convergence. The extracted feature maps were flattened and passed to a fully connected dense layer of 128 neurons with dropout (0.4), before final classification through a ten-node softmax output layer. Both models were optimised using the Adam algorithm with a learning rate of 0.001 and trained for up to 30 epochs using a batch size of 128. Early stopping with a patience of five epochs was employed to prevent overtraining and preserve the best model weights based on validation loss.

Results and Key Findings

The CNN outperformed the ANN across all performance indicators. The CNN achieved 91.0% test accuracy, while the ANN achieved 88.9%. Training history revealed faster convergence for the CNN, which reached optimal performance after approximately nine epochs, compared to the ANN's twenty-two epochs. The CNN's validation loss curve also demonstrated smoother descent and reduced variance, suggesting stronger generalisation and resilience to overfitting.

From the classification reports, the CNN recorded a weighted F1-score of 0.91, compared to 0.89 for the ANN. Both models performed exceptionally on simple, high-contrast classes such as trouser, sandal, bag, and ankle boot (precision and recall ≥ 0.97). However, the CNN provided notable improvements in distinguishing similar garments such as shirt, coat, and pullover, where the ANN often misclassified instances due to texture and shading similarities. The confusion matrices further revealed that misclassifications were concentrated in visually similar

upper-body categories, while classes with distinct outlines or textures (e.g., sandal and trouser) achieved near-perfect separation.

Interpretation and Recommendations

The superior performance of the CNN can be attributed to its ability to learn hierarchical spatial features through convolutional operations. Whereas the ANN relies on fully connected layers that treat each pixel independently, the CNN leverages local receptive fields and shared weights, enabling translation invariance and efficient feature extraction. This architectural inductive bias allows the CNN to recognise edges, patterns, and contours regardless of their position in the image. The addition of batch normalisation in the CNN further stabilised gradient flow, while dropout mitigated overfitting, thus achieving a balance between bias and variance.

Nonetheless, both models encountered persistent difficulties in discriminating between shirts, coats, and pullovers, which share similar silhouettes and grayscale intensity patterns. These confusions suggest that the limited 28 × 28 resolution constrains the models' ability to extract fine-grained texture information. This observation aligns with findings by LeCun, Bottou, Bengio, and Haffner (1998), who emphasised that spatial context is vital for robust image recognition. The results reinforce the principle that model architecture, not merely dataset size, is the dominant determinant of performance efficiency.. Future improvements should include image augmentation, learning-rate scheduling, and deeper networks to reach state-of-the-art accuracy.

Conclusion

This study concludes that convolutional neural networks provide a structurally superior approach to image classification compared to traditional fully connected neural networks. The CNN achieved higher accuracy, faster convergence, and more stable validation metrics, demonstrating its suitability for real-world applications such as automated quality inspection, e-commerce tagging, and visual analytics in resource-constrained environments. The ANN remains a useful baseline but lacks the spatial inductive bias required for high-dimensional image tasks. Future improvements could include stronger data augmentation (rotation, zoom, and horizontal flips), adaptive learning-rate scheduling, and experimenting with deeper CNN variants such as ResNet or MobileNet to further improve performance.

In summary, the project achieves its intended objectives by demonstrating, through empirical evidence, that CNNs outperform ANNs in accuracy, efficiency, and generalisation on Fashion-MNIST. The findings contribute to understanding how architectural design and regularisation choices influence model behaviour, reinforcing the importance of structure-aware learning in modern computer vision.

References

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278–2324. https://doi.org/10.1109/5.726791

Preda, G. (2021). Fashion-MNIST classification using Keras CNN. Kaggle. https://www.kaggle.com/code/gpreda/fashion-mnist-with-keras-cnn

Xiao, H., Rasul, K., & Vollgraf, R. (2017). Fashion-MNIST: A novel image dataset for benchmarking machine learning algorithms. arXiv preprint arXiv:1708.07747.