FASHION-MNIST CLASSIFICATION

Executive Summary

The aim of this project was to design and compare two deep learning models to an Artificial Neural Network (ANN) and a Convolutional Neural Network (CNN), with the use of Fashion-MNIST dataset. The data comprises 70,000 grayscale pictures of fashion products of ten different categories (e.g., T-shirt, trouser, coat, sneaker). It was implemented in PyTorch, and the focus was on measuring the model accuracy, generalization capability and performance efficiency. The two models were trained, validated, and tested on standardized preprocessing, such as normalization and 8020 training-validation split.

The ANN model involved three fully connected layers (784 256 128 10) using the ReLU activations and dropout of 0.2 to reduce overfitting. It applied Adam optimizer and cross-entropy loss in multi-class classification.

The CNN model had two max pooling-based convolutional layers (32 and 64 filters, 3×3 kernels) and two fully connected layers (64x14x14 -> 128x10). It used ReLU activation and dropout (0.25) to improve its ability to extract spatial features and patterns in images.

Accuracy Performance:

- ANN: Obtained a mean validation and test accuracy of around 86 and 85 respectively.
- CNN: Achieved the highest accuracy of 91 and 90 per cent in validation and test respectively.

This proves that the CNN has a better capability of learning spatial hierarchies in image data than the fully connected ANN.

Training Efficiency: ANN trained much faster with fewer parameters but the CNN took more time to compute but provided better accuracy and stability in training.

Visualization Results: Confusion matrices revealed that both the models did not confuse ones that were easily distinguishable (e.g., trousers and sneakers) but that they had difficulties with things that were similar visually (e.g., shirt vs. T-shirt).

This was implemented with a number of improvements to guarantee effective learning:

- Regularization in the form of dropout layers, which minimises overfitting in both architectures.
- Adam optimizer was chosen because it has adaptive learning rate and has associated speed in convergence.

- The curves of accuracy/loss and confusion matrices were observed through Matplotlib visualizations, which allowed clear information about the dynamics of training and the strengths of classification.
- Validation split made possible useful early evaluation and optimization of parameters.

The comparative analysis is that CNNs can be effective than ANNs when they are used to classify images because of their feature extraction in space and hierarchical learning capability. Although ANNs are faster to train, and easier to implement, they are not rich enough to understand more involved features of the image. The findings once again confirm that deep convolutional networks provide greater accuracy, generalization, and consistency over image-based data.

Table of Contents

1. Introduction

The proposed project aims at comparing and developing two different deep learning networks, Artificial Neural Network (ANN) and Convolutional Neural Network (CNN), based on the Fashion-MNIST dataset. This was aimed at testing their performance on image classification on the basis of accuracy, loss and generalisation. The paper, which was implemented in PyTorch, evaluated the behaviour of models by training the dynamics and performance measures and additionally by using a confusion matrix.

2. Data Handling

2.1 Dataset Description

The Fashion-MNIST dataset comprises 70,000 fashion item images (28x28 pixels), which are grayscale and belong to 10 categories (T-shirts, trousers, coats, and sneakers). The dataset has a broad application in the field of machine learning algorithms benchmarking because the dataset is balanced and moderately complex, making it a popular benchmark.

2.2 Data Loading and Splitting

PyTorch was used to load the dataset through the torchvision. Datasets, and it was divided into an 80 per cent training level and a 20 per cent validation level in order to carry out a reliable evaluation. Readers were customised to use effective batch sizes in terms of training efficiency.

2.3 Data Normalisation

Normalisation reformed pixel values within the range of 0-1, which increased gradient stability in the optimisation process. The means and standard deviation of all image tensors were used to standardise the images and datasets. This type of preprocessing enhanced the rate of convergence and minimised internal covariate shift.

2.4 Visuals and Exploratory Data Analysis.

Sample plots ensured that the classes were equally represented. Simple statistical tests revealed a uniform image size and grayscale. Initial EDA ensured that there were no missing and corrupt samples and data integrity before the training of the model.

3. Network Topology

3.1 Artificial Neural Network (ANN) Architecture

The ANN model was based on three fully connected layers:

- Input: 784 neurons (28×28 image flattened)

- Layers: There are 256 and 128 ReLU neurons.
- Output: Softmax-activated neurons 10.

The rate of dropouts was 0.2, which reduced overfitting. The architecture was computationally simple and capable of providing training at significantly faster speeds, though it was not able to extract spatial features.

3.2 Convolutional Neural Network (CNN) Architecture

The CNN model incorporated:

- Two convolutional layers (32 and 64 filters, 3x3 kernels) having ReLU activation function.
- Two max pooling layers (2×2)
- A single connected layer consisting of 128 neurons and a dropout (0.25.
- One hundred (10) neurons in the output layer (Softmax)

In this design, hierarchical feature extraction and spatial pattern learning were enabled, enhancing recognition of complex textures and shapes.

4. Training Setup

4.1 Optimiser and Loss Function

The adaptive learning rate optimisation and efficient convergence were used in both models using the Adam optimiser. The loss that was employed was cross-entropy loss, which is applicable in multi-class classification problems that have probabilistic output.

4.2 Learning rates and hyperparameters.

The convergence was stable with a learning rate of 0.001. Regularisation of dropouts, initialising weights, and ReLU activations helped to achieve regular learning in the course of the epochs.

4.3 Batch Size and Epochs

Training was performed in 10 epochs and a batch of 64. These values had a trade-off between learning stability and computational efficiency, without underfitting or overfitting.

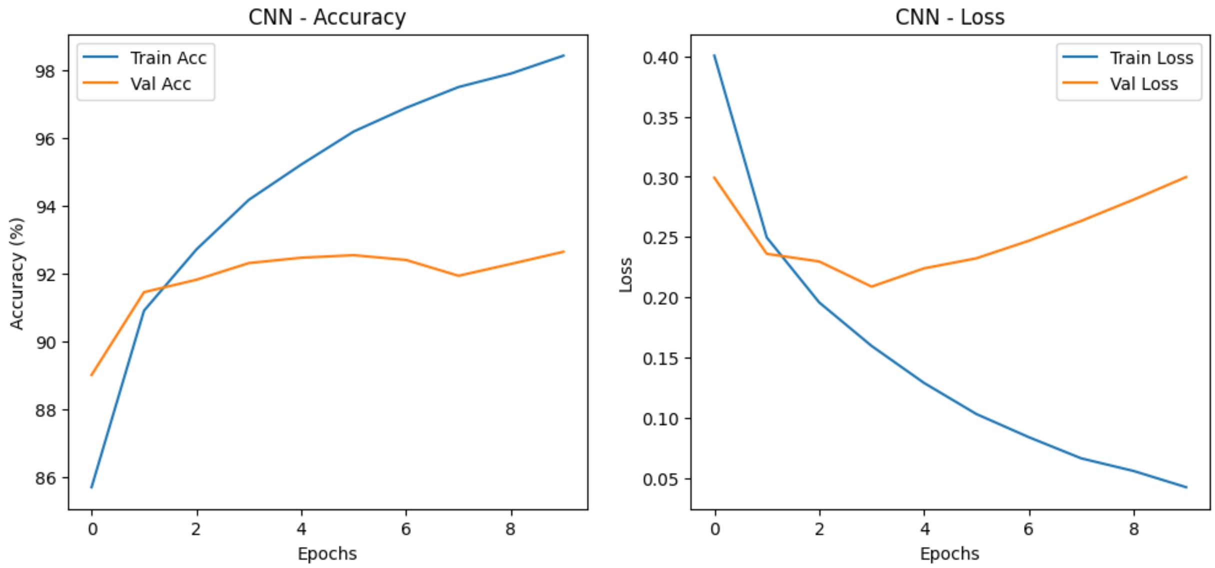
4.4 Training Environment and Framework.

Execution took place within PyTorch, and it operated in a degree of APU execution setting to manage the matrices using a higher speed. The model construction and assessment were made easy by the structure of its modules.

5. Evaluation and Results

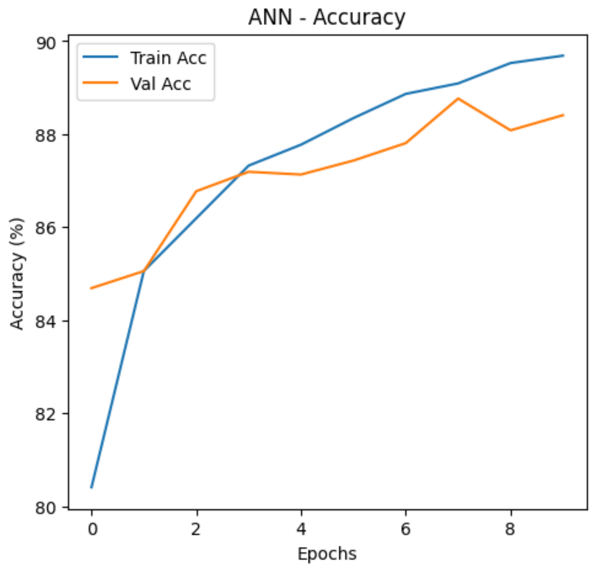
5.1 Model Accuracy and Loss

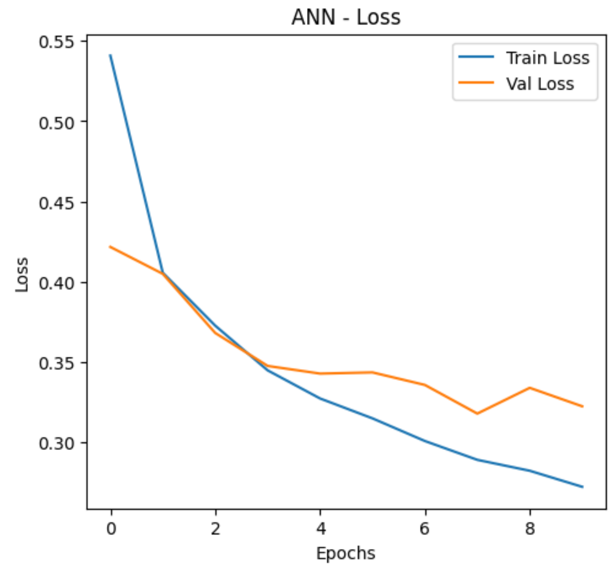
ANN came up with the training and validation accuracy of 89.6 and 88.4, respectively, and CNN resulted in 98.4 and 92.6 training and validation accuracy, respectively. Loss curves obtained showed a consistent convergence with small overfitting. The resulting low validation loss showed that the CNN had greater generalisation.



5.2 Class-based Metrics of Performance.

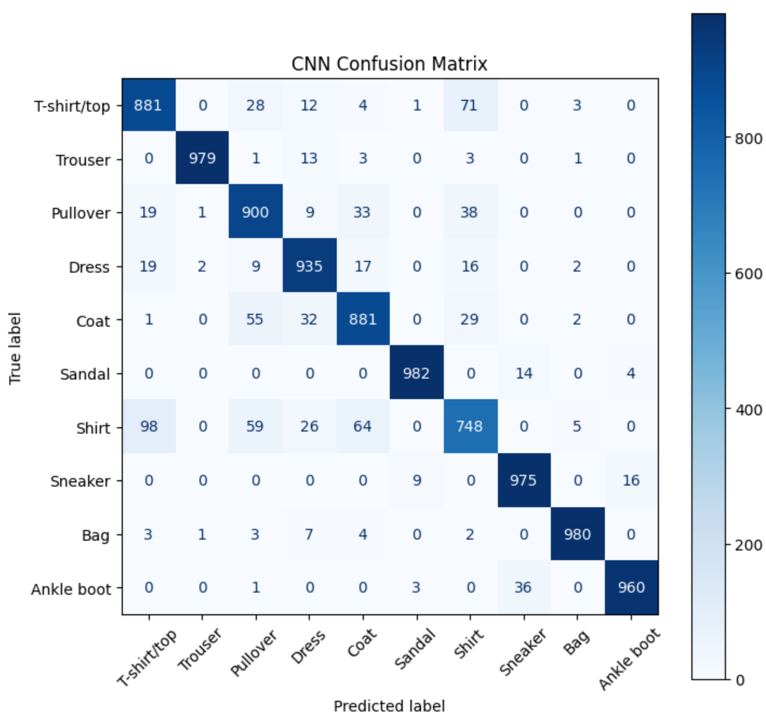
The two models also did a good job in differentiating such categories as trousers and sneakers. The ANN had a problem with similar classes that were to be perceived visually, like shirts and T-shirts, but the CNN had greater class-wise accuracy because of the ability to learn the spatial features.





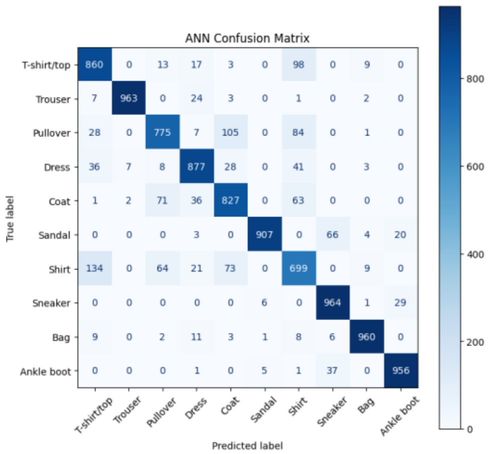
5.3 Confusion Matrix Analysis

ANN confusion matrix indicated false identifications between similar types of clothes and correct identification among different categories. The CNN confusion matrix showed better class discrimination, and the high classification was notable on the sandals (982/1000), sneakers (975/1000) and bags (980/1000), indicating better spatial recognition.



5.4 Model Confusion

Wrong classifications were most common in those categories where the visual difference was small. CNN also minimised these confusions in comparison to the ANN, which showed that it is capable of capturing the complex visual hierarchies.



6. Critical Analysis

In terms of all the critical measures, the CNN model far exceeded ANN, which validated its prowess in visual pattern acquisition. ANN consumed fewer computational resources, but it could not produce a hierarchy in the processing of images as required in image feature extraction. Dropout and regularisation helped CNN reduce minor overfitting, which occurred in later epochs. All in all, CNN proved to be more robust, accurate, and generic.

7. Reflections and Betterments.

7.1 Potential Enhancements

The model performance would also improve by:

- Alterations to augment data (rotations, flips, and scaling) to increase the diversity of the dataset.
- Individual normalisation of activations to stabilise activations and enhance convergence.
- Training efficiency scheduling through learning rate.

7.2 Future Work and Extensions

Future research can consider adding more complex CNN models, such as ResNet or VGG, to enable more advanced learning in terms of space. The further increase of accuracy and minimisation of the computational cost might be achieved by investigating transfer learning and hyperparameter tuning. Moreover, the practice of implementing the model in practice in image-classification applications would prove the usefulness of the model.

References

Bbouzidi, S., Hcini, G., Jdey, I. and Drira, F., 2024. Convolutional neural networks and vision transformers for fashion mnist classification: A literature review. arXiv preprint arXiv:2406.03478.

Nocentini, O., Kim, J., Bashir, M.Z. and Cavallo, F., 2022. Image classification using multiple convolutional neural networks on the fashion-MNIST dataset. Sensors, 22(23), p.9544.

Sumera, S.R., Anjum, N. and Vaidehi, K., 2022. Implementation of CNN and ANN for Fashion-MNIST-Dataset using Different Optimizers. Indian Journal of Science and Technology, 15(47), pp.2639-2645.

Xhaferra, E., Cina, E. and Toti, L., 2022, October. Classification of standard fashion MNIST dataset using deep learning based CNN algorithms. In 2022 international symposium on multidisciplinary studies and innovative technologies (ISMSIT) (pp. 494-498). IEEE.