A Neural Network Model for Fashion-MNIST Image Classification Executive Summary

1. Aim

The aim of this analysis is to design, implement and evaluate ANN (Artificial Neural Networks and CNN (Convolutional Neural Network) models for image classification, particularly using the Fashion-MNIST dataset. The target to achieve is to demonstrate understanding of network topology behaviour of learning and generalisation.

2. Methodology

The Framework and Tools: Implemented in Python using PyTorch with Torchvision, scikit learn and Matplotlib for training, visualisation and evaluation.

The Dataset: Fashion-MNIST data set, loaded through PyTorch (60,000 grayscale images training and 10,000 test grayscale images (10 classes).

Division of Data: Stratified 90 percent training and 10 percent validation split from original training data. The test set only used for final evaluation.

Preprocess: Normalisation $\mu = 0.2860$ and $\sigma = 0.3530$

Augmentation (only Training): Random Affine ±12°, translation ±6 %, horizontal flip p = 0.25 and random erasing p = 0.25.

Models: ANN: two connected hidden layesr 256, 128 neurons and ReLU and DropOut 0.2 . CNN: three convolutional block (32-64-128 filters and bathnorm and maxpool) is FC 192 and Dropout 0.35.

Training: Adamw lr = 3 * 10^{-4} , weight decay = 3 * 10^{-4}, Cross Entropy loss with label smoothing 0.05, ReduceLROnPlateau scheduler, early stopping patience = 5, mixed precision AMP, gradient clipping 1.0.

Evaluation: Accuracy, Loss, Confusion matrix, Classification repot, Pre class accuracy and Misclassified sample visualisation.

3. Results:

Model	Parameters	Test Accuracy	Test Loss
ANN	200 k	88.97 %	0.5609
CNN	1.2 M	91.85 %	0.5043

4. Analysis and Evaluation:

CNN achieved a + 3 percent accuracy improvement and a lower loss in comparison to ANN, showing its ability to capture spatial hierarchies through convolution filters and sharing of weigh. ANN converged faster with 15 epochs but resulted in overfitting in early layers while CNN stailised in 20 epochs and had better generalisation. The greatest confusion occurred between coat ans shirts due to similar visual while sneakers, sandals and bag were perfectly classified. Augmentation and label smoothing improved validation accuracy by 20 percent and dropout regularisation reduced the overfitting.

This experiment confirms CNNs outperformance in comparison to ANN for image classification, showing us an achievement of 91.85 percent test accuracy verses 88.97 percent for ANN. For visual task convolution architecture is more effective and a generalisable approach. For future experiments and improvents deeper residual architectures ReNet like, cosine learning rate scheduling and advance augmentation like Mixcup/ Cut Mix to test accuracy beyond 92 percent can be considered.