Executive Summary – Fashion-MNIST Classification

This coursework focused on developing, training, and evaluating two neural network architectures — an **Artificial Neural Network (ANN)** and a **Convolutional Neural Network (CNN)** — for the **Fashion-MNIST image classification** task. The dataset consists of 70,000 grayscale images (28×28 pixels) representing 10 categories of clothing items, such as T-shirt, Trouser, Dress, Coat, and Shirt. The primary objective was to design, train, and compare both architectures in terms of classification performance, generalization ability, and interpretability.

Model Overview

The **ANN (Multilayer Perceptron)** was implemented with three fully connected layers (512, 256, and 128 neurons) using **ReLu activation** and **dropout (0.3)** for regularization. The **CNN** architecture comprised three convolutional layers (32, 64 and 128 filters), each followed by **batch normalization** and **max pooling**. Used **GlobalAveragePooling2D** before a 128-unit dense head to reduce parameters and improve generalization. Both models used the **Adam optimizer**, **categorical cross-entropy loss**, and were trained for 15 epochs.

Major Findings

The CNN surpassed the ANN on the Fashion-MNIST dataset, securing a better test accuracy of **88.93%** as against **87.37%** of the ANN, which is an indicator of better **generalization** to new data. The ANN had a slightly lower training loss but still the CNN showed a lower validation loss, which is a sign of its stronger ability to extract transferable features. The two models were characterized by **stable** training and **no major overfitting**. Both networks had difficulties in dealing with categories that were very similar to each other like **Shirt**, **Coat**, and **Pullover**, but at the same time, the CNN was able to further **minimize** the misclassifications done in other categories, thus, proving that it is more advantageous to use convolutional architectures rather than dense ANN structures in image classification.

Innovations and Techniques

In this study, the main innovations were the use of **dropout** and **batch normalization** to enhance stability during training. Moreover, the incorporation of validation sets, confusion matrices, and classification reports allowed for a detailed performance assessment. Furthermore, the analytical visualization of confusions among closely related categories provided interpretative insights in addition to the numerical metrics.

Conclusions and Insights

The comparison showed that CNNs are the best for the image tasks in general but their superiority gets reduced without proper regularization and **data augmentation**. However, systematic errors on the **Shirt / T-shirt / Pullover / Coat** cluster remain the main shortcoming. This project was successful in meeting all objectives by the implementation, training, and critical analysis of two neural architectures for the Fashion-MNIST classification, thus demonstrating a good understanding of deep learning principles and evaluation practices.