

Comparative Analysis of Neural Network Architectures on Fashion-MNIST

Name: Chandra Sekhar Reddy Puli

P no: P2906314

SUMMARY: The present document presents the execution and assessment of two different neural network architectures, namely an ANN and a CNN, on the Fashion-MNIST dataset for image classification. The discussion also includes the basic models, the improved versions of the models, and the main conclusions that were reached through the analysis of the performance.

OVERVIEW OF NEURAL NETWORK MODELS:

Two distinct models were designed and trained to compare fundamental architectural differences.

Artificial Neural Network (ANN / MLP): A Multi-Layer Perceptron was implemented as a baseline for dense, fully connected architectures. The enhanced version consists of:

- An input layer that **flattens** the 28x28 pixel images into a 784-element vector.
- Two hidden dense layers with **128 and 64 neurons**, respectively.
- **Leaky ReLU** activation functions in the hidden layers to prevent the "dying ReLU" problem and allow for a small, non-zero gradient when the unit is not active.
- An output layer with **10 neurons** (one for each class) using a **Softmax** activation function to produce a probability distribution.

Convolutional Neural Network (CNN): A CNN was designed to leverage the spatial hierarchy of image data. The enhanced version features a deeper architecture:

- **Three convolutional blocks**, each using a **Leaky ReLU** activation function. The blocks progressively increase feature map depth ($32 \rightarrow 64 \rightarrow 128$ filters) to learn patterns of increasing complexity.
- **Max Pooling layers** after the first two blocks to downsample the feature maps, reducing computational load and providing a degree of translational invariance.
- A **Flatten** layer to convert the final 2D feature maps into a 1D vector.
- A final classifier consisting of a **Dense layer (128 neurons)** and a **Softmax output layer (10 neurons)**.

MAJOR FINDINGS: The primary goal was to quantify the performance difference between the ANN and CNN and to measure the impact of architectural enhancements.

Overall Accuracy and Performance Comparison: In all the experiments, CNN showed vibrant superiority in performance. The main reason for the advantage of CNN over ANN is its capability to convolve the filters and to detect the spatial features of the images like edges, patterns, and textures. The input image in ANN is treated as an unstructured vector of pixels which makes it less capable than CNN. The table below shows the final test accuracies for both the original and enhanced models:

Model Type	Baseline Accuracy	Enhanced Accuracy	Improvement
ANN (MLP)	88.51%	89.27%	+0.76%
CNN	91.39%	92.15%	+0.76%

CLASS-SPECIFIC CHALLENGES: Analysis of the confusion matrices revealed that both models, especially the ANN, struggled to differentiate between visually similar classes. The most frequent misclassifications occurred between:

- Shirt
- T-shirt/top
- Pullover
- Coat

The enhanced CNN, while still showing some confusion in these categories, made significantly fewer errors than the ANN, highlighting its superior feature extraction capabilities.

HIGHLIGHTS: Several techniques, which were also used to improve the model's performance, were also used to provide a thorough analysis:

Higher Model Complexity: The depth of the ANN and CNN was increased by an extra hidden layer and an extra convolutional block, respectively. This enabled the models to better and more complexly represent the data thus, directly improving test accuracy.

Activation Function Testing: The standard ReLU activation was exchanged with Leaky ReLU. This slight alteration guarantees that the neurons will not be completely inactive during the training process, thereby facilitating a less turbulent learning process and preventing the issue of vanishing gradients from occurring in the first place.

Large Visualization Techniques: Confusion matrices based on Seaborn were one of the most important tools that allowed us to move beyond the use of just one accuracy metric for our evaluation. The confusion matrices provided with this visualization were so easy to interpret that one could immediately see the model's performance for each class and know in which cases the items of clothing were mistaken for each other, along with the estimated quantity of confusion in numerical terms.

CONCLUSION: The findings of the experiments resulted in clear-cut conclusions. Convolutional Neural Networks (CNNs) are invariably the best option for classifying images, as they are always approximately 3% better than Artificial Neural Networks (ANNs) due to their capacity to deal with spatial data intrinsically. The increase in accuracy that resulted when making models deeper supports the notion that architectural depth is a decisive factor in the performance of complex tasks. On the other hand, accuracy analysis must include the confusion matrix as a primary tool for uncovering the model's predictable weaknesses, for instance, a tendency to confuse 'Shirts' with 'Coats'. Such deep understanding is critical because it helps direct future work to specific solutions like using data augmentation to assist the model in learning the subtle differences that distinguish similar classes.