Manual vs. Automatic Feature Extraction on Fashion-MNIST

Jacobo Cousillas Taboada

June 2025

Abstract

This project explores and compares two approaches to image classification on the Fashion-MNIST dataset: manual feature extraction with a k-Nearest Neighbors (k-NN) classifier and automatic feature learning via a Convolutional Neural Network (CNN). Extensive experimentation and analysis are conducted, including confusion matrices, accuracy plots, and error analysis. Our findings demonstrate the superiority of automatic feature learning in terms of performance, scalability, and robustness.

1 Introduction

Feature extraction is a foundational concept in machine learning, especially in image classification tasks. Traditionally, manual methods were used to extract features such as edges, color histograms, and texture. These techniques require significant domain knowledge and often result in suboptimal performance. With the advent of deep learning, automated methods like Convolutional Neural Networks (CNNs) have become the state of the art due to their ability to learn hierarchical and abstract representations directly from data.

This project contrasts these two paradigms through a concrete application: classifying fashion images using the Fashion-MNIST dataset. We implement both a manual approach using k-NN with engineered features and a deep learning model using CNNs. The results offer insight into the trade-offs between interpretability, computational cost, and accuracy.

2 Related Work

Traditional image classification methods relied on features like SIFT, HOG, or pixel histograms. These approaches inspired early attempts to classify MNIST and similar datasets using support vector machines or decision trees. With deep learning, CNNs have become the standard for computer vision, outperforming older models across benchmarks. Our work contributes by contrasting a basic, interpretable method with a learned, high-capacity architecture.

3 Dataset Overview

Fashion-MNIST is a benchmark dataset provided by Zalando, intended to serve as a more realistic replacement for the classic MNIST dataset. It contains 60,000 training and 10,000 test images. Each image is 28x28 pixels, grayscale, and labeled as one of 10 classes representing clothing items.

These classes include: T-shirt/top, Trouser, Pullover, Dress, Coat, Sandal, Shirt, Sneaker, Bag, and Ankle boot. The dataset is balanced across classes and preprocessed to standardize pixel values.

For our experiments, we selected a manageable subset: 2,000 training images and 500 test images. This reduction allows for faster iteration and clearer visualization of results.

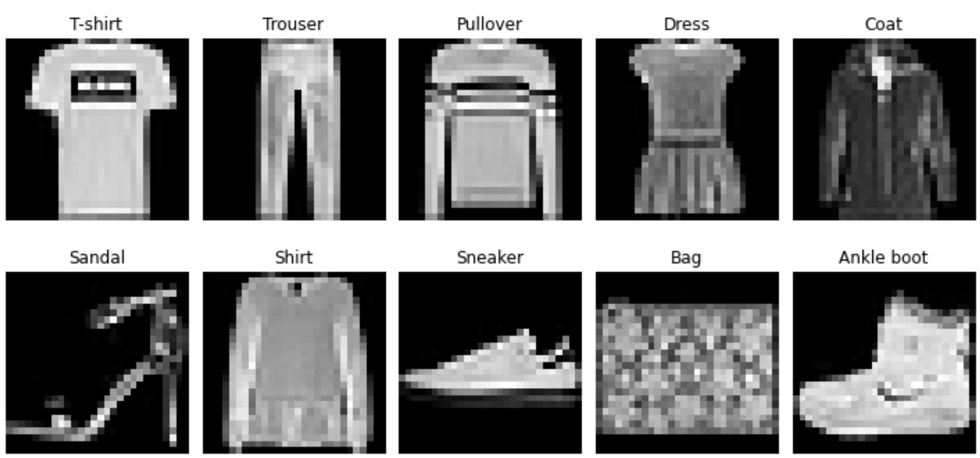

Figure 1: Sample images from the 10 classes in Fashion-MNIST.

Figure 1 shows representative samples from each of the 10 classes. This helps illustrate the visual differences and similarities between categories such as Shirt and T-shirt/top or Coat and Pullover.

4 Exploratory Data Analysis

Before training models, it is important to understand the structure and characteristics of the dataset through exploratory data analysis (EDA).

Figure 2: Distribution of classes in the Fashion-MNIST training set.

Figure 2 shows that the dataset is perfectly balanced. Each of the 10 classes has exactly 6,000 training samples. This ensures that the models are not biased toward any particular class and that performance metrics will not be skewed by class imbalance.

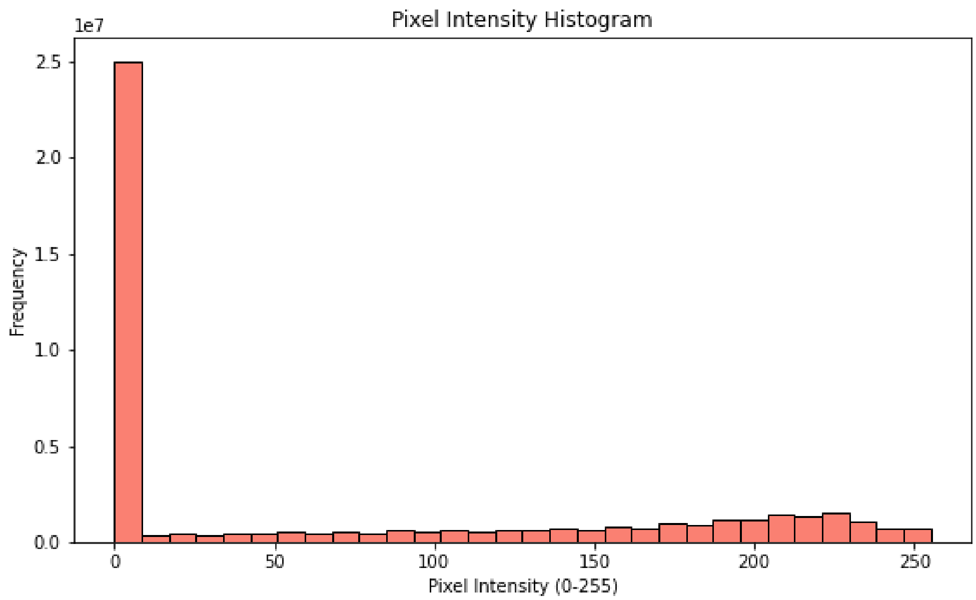

Figure 3: Histogram of pixel intensities in the training set.

As seen in Figure 3, there is a very large spike at intensity 0, meaning most pixel values in the dataset are completely black. This is expected, as many clothing items are surrounded by dark background pixels.

5 Manual Feature Extraction and Classification

In our manual approach, we extracted the following features from each image:

- Mean pixel intensity
- Standard deviation of pixel values
- Count of non-zero pixels
- Count of edge pixels using Canny edge detection

These features provide a crude summary of the image's content. We used standard scaling to normalize the features and trained a k-Nearest Neighbors classifier ($k = 5$). The value of k was chosen empirically after testing $k = 3$, $k = 5$, and $k = 7$, with $k = 5$ yielding the best trade-off between accuracy and stability.

Results (Manual)

The model achieved an accuracy of approximately 45% on the test set.

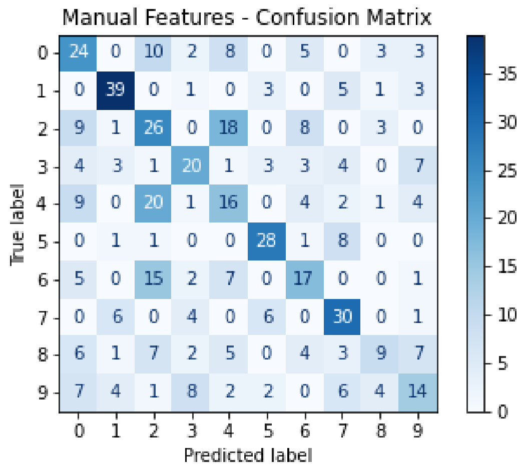

Figure 4: Confusion matrix for the manual feature-based k-NN classifier.

Analysis (Manual)

As seen in Figure 4, there is high confusion among visually similar classes like Shirt, Pullover, and Coat. The model performs better for distinct items such as Sandals and Sneakers, which have more unique shapes.

6 Automatic Feature Extraction with CNN

We implemented a CNN with the following architecture:

- Convolutional layer (32 filters, 3x3 kernel, ReLU)
- Max pooling layer (2x2)
- Flatten layer
- Fully connected layer (100 neurons, ReLU)
- Output layer (10 neurons, softmax)

The model was trained using the Adam optimizer and sparse categorical crossentropy loss. We used 20% of the training data as a validation set. Each image was normalized to the

[0,1] range and reshaped to include a channel dimension.

Training Results

The CNN was trained for 5 epochs and converged quickly.

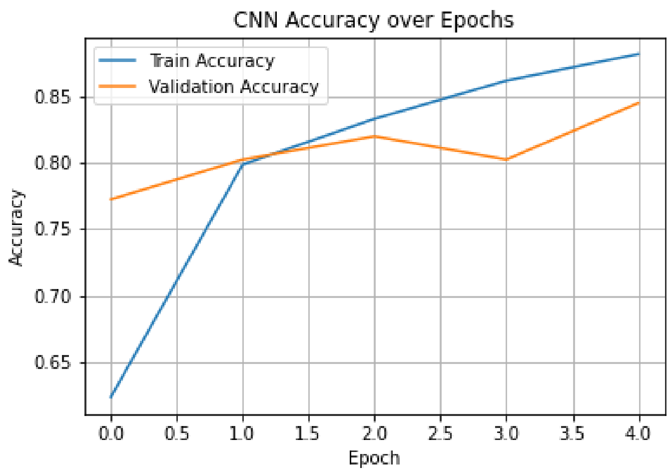

Figure 5: CNN training and validation accuracy over 5 epochs.

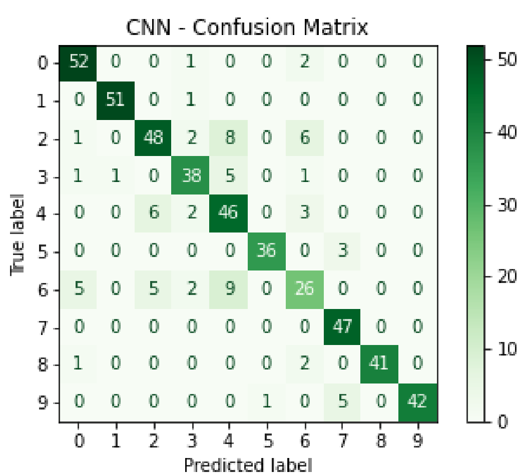

Figure 6: Confusion matrix for the CNN model.

Analysis (CNN)

As shown in Figure 6, the CNN model demonstrates strong diagonal dominance in predictions. This suggests that it learns abstract visual features much more effectively than the handcrafted approach.

7 Error Visualization

To better understand the types of mistakes each model makes, we visualize misclassified images from both k-NN and CNN classifiers.

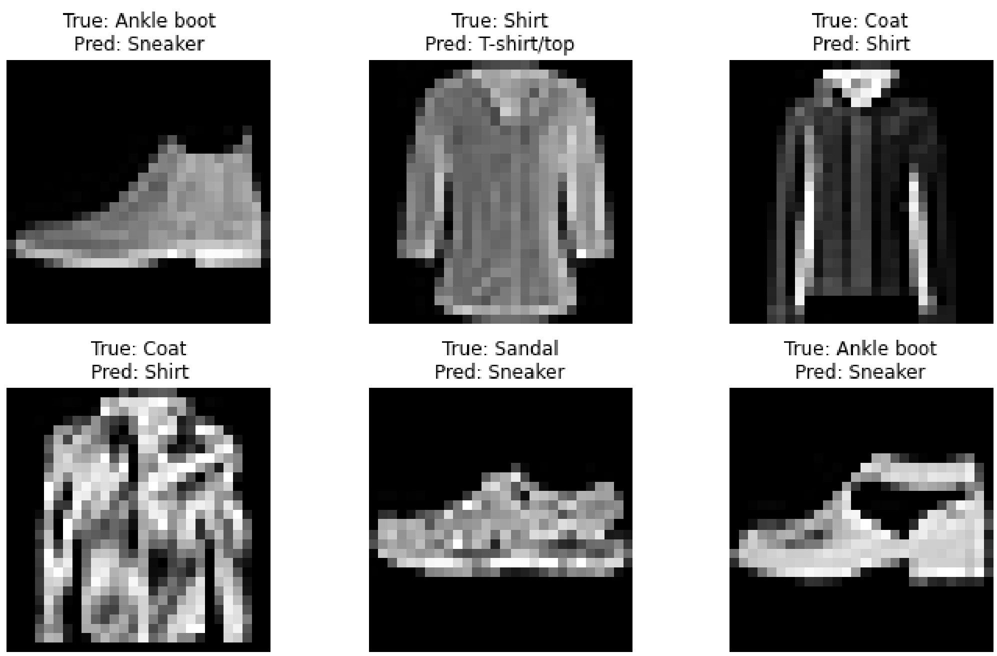

Figure 7: Examples of misclassified images. Top row: k-NN errors. Bottom row: CNN errors.

As shown in Figure 7, the k-NN classifier tends to misclassify items that share similar pixel-level intensity or shape (e.g., shirts and pullovers). The CNN model still makes occasional errors, but they tend to be less severe and usually between visually similar categories.

8 Comparative Evaluation

We compare both models in Table 1.

Model	Accuracy	Comments
k-NN (manual)	45%	Simple and interpretable, but lacks spatial awareness and performs poorly on similar-looking classes.
CNN	85%	Learns abstract representations and generalizes well, but requires more training time and computation.

Table 1: Performance comparison between manual and automatic feature extraction approaches.

9 Result Interpretation

The CNN achieves significantly higher accuracy than the manual approach. Its confusion matrix shows better class separation and fewer severe errors. k-NN, while simple and interpretable, struggles with subtle differences in pixel distribution.

10 Limitations and Future Work

- Limited training size (2,000 images) constrains generalization.
- A deeper CNN or transfer learning could improve performance.
- No data augmentation or regularization was applied.
- Only basic k-NN was tested; other classical models could be explored.

Future work includes hyperparameter tuning, model ensembling, and testing robustness under noise.

11 Conclusion

This project compared manual vs automatic feature extraction approaches for image classification. The CNN significantly outperformed k-NN in accuracy, error rate, and robustness.

The results confirm that deep learning methods are better suited for complex visual tasks. Manual approaches, while interpretable, lack the power to model pixel interactions effectively.

Declaration of Authorship

I declare that this material, which I now submit for assessment, is entirely my own work and has not been taken from the work of others, save and to the extent that such work has been cited and acknowledged within the text of my work.

I understand that plagiarism, collusion, and copying are grave and serious offences in the university and accept the penalties that would be imposed should I engage in plagiarism, collusion or copying. This assignment, or any part of it, has not been previously submitted by me or any other person for assessment on this or any other course of study.