

Texture-Based Classification of Fashion- MNIST Images Using Machine Learning

1st S Ruusheka Akilavarshini 1*, 2nd Sabari R 2*, 3rd Ramesh Kanna G 3**

*Department of CSE, Sri SivaSubramanyia Nadar College of Engineering, Chennai, India
ruusheka2410216@ssn.edu.in@ssn.edu.in, sabari2410638@ssn.edu.in, rameshkanna2410508@ssn.edu.in

Abstract—Image Classification is one of the fundamental Problems in Machine Learning. In this project , we perform texture-based feature classification on the Fashion-MNIST dataset using traditional machine learning techniques. Instead of using raw pixel values directly, texture features such as histogram features and gradient magnitude features are extracted from images. Two classifiers, mainly Naive Bayes and Random Forest, are trained and compared based on their performance. Experimental results show that Random Forest achieves better classification accuracy compared to Naive Bayes. This study demonstrated the effectiveness of feature engineering and traditional Machine Learning methods for image classification tasks.

Index Terms—Fashion-MNIST, Texture Features, Histogram Features , Gradient Magnitude , Naive Bayes, Random Forest, Image Classification

I. INTRODUCTION

The rapid expansion of online retail and fashion e-commerce platforms has created an high demand for automated fashion item recognition systems. Accurately identifying clothing categories from images enables intelligent product tagging, visual search, recommendation, and inventory management. While CNN have achieved state-of-art results in visual recognition tasks [1] , their high computational cost and large data requirements pose significant barriers in embedded systems or low resource deployment scenarios.

Image Classification plays a vital role in computer vision and pattern recognition. With the advancement of Machine Learning, Various techniques have been developed to classify images. The Fashion-MNIST dataset [?] is a widely used benchmark dataset consisting of grayscale images of clothing items. Fashion MNIST dataset consist of 70000 grayscale image of size 28×28 pixel. Where dataset is divided into 60000 training and 10000 testing images. Its standardized format ,makes an ideal benchmark feature type and identify the most effective combination for fashion item classification.

In this project, we focus on texture-based feature extraction rather than deep learning methods. Texture features capture important patterns in images such as edges, smoothness, and intensity distribution. we extract histogram features and gradient magnitude features from the images and use them to train Machine Learning Classifiers.

Two classifiers, Naive Bayes and Random Forest, are implemented and their performances are compared. The goal of this project is to understand the importance of feature engineering and to evaluate traditional machine learning models for image classification.

The main contribution of the project is as follows :

- A comparative evaluation of histogram based and gradient based texture feature extraction methods on Fashion-MNIST dataset.
- A Systematic benchmarking of Gaussian Naive Bayes and Random Forest classifiers under both feature extraction regimes.
- An analysis of the classification report and performance metrics to identify the best performing alternatives to deep learning for fashion image recognition.

II. LITERATURE SURVEY

Fashion image classification has been a subject of a active research, spanning classical feature based approaches to modern deep learning solutions. **Classical Feature Based Methods:**

Early works in texture analysis relied on hand engineered descriptors. Haralick et al. [?] introduced co-occurrence matrix based features that capture spatial relationship of pixel intensities. forming the basis of many texture classification systems. Histogram of Oriented Gradients [3] demonstrated strong performance in pedestrian and object detection by capturing gradient structure. Local Binary Patterns [4] provided efficient texture encoding for face recognition and general texture classification.

Fashion-MNIST Benchmark Studies: Since its introduction, Fashion-MNIST [2] has been used extensively to benchmark image classification algorithms. Several studies have demonstrated that while CNNs achieve over 90% accuracy, simpler classifiers applied to extracted features can still achieve competitive results. Support Vector Machines applied to flattened pixel values achieve around 89% accuracy, while k -nearest neighbours achieve approximately 85%. These baselines motivate the exploration of alternative feature representations.

Random Forest for Image Classification: Breiman's Random Forest [5] is an ensemble method that constructs multiple decision trees and aggregates their outputs . It has been widely

applied in image classification tasks, especially when feature dimensionality is moderate and interpretability is desired. Studies have shown that Random Forests are robust to noise and less prone to overfitting compared to individual decision trees.

Naive Bayes Classifiers: Gaussian Naive Bayes [6] assumes conditional independence among features given the class label and models each feature with a Gaussian distribution. Despite its strong independence assumption, GNB has been applied successfully in text classification, medical diagnosis, and low dimensional image feature classification. In fashion classification, GNB serves as a fast baseline due to its closedform training procedure.

Texture in Fashion Recognition: Several studies have explored texture descriptors specifically for clothing recognition. Bossard et al. [7] demonstrated that combining color histograms and texture features outperforms raw pixel features for clothing category prediction. Similarly, Liu et al. [8] established a large-scale benchmark for fashion analysis but noted that simpler pipelines remain relevant for embedded and mobile applications.

Before the widespread adoption of deep learning, image classification relied heavily on handcrafted feature engineering. histogram descriptors were commonly used to capture intensity distributions and texture patterns in grayscale images. Such features are computationally simple yet capable of representing global pixel distribution characteristics.

Gradient-Based descriptors emerged as another powerful representation technique. by applying derivative filters such as Sobel operators, edge info and structural patterns within images can be captured effectively. Gradients are particularly useful in texture analysis and object boundary detection.

Gaussian Naive Bayes is a probabilistic classifier based on Bayes's theorem with an assumption of conditional independence among the features. Random Forest, on the other hand, is an ensemble learning techniques that constructs multiple decision trees and aggregated their predictions. It improves generalization and reduces overfitting compared to individual decision tree.

Although deep convolution networks achieve higher performance on complex datasets, classical machine learning methods remain valuable for understanding feature representation, computational efficiency and interpretability.

III. METHODOLOGY

A. Dataset Description

The Fashion-MNIST dataset [2] consisting of 70000 grayscale images of size 28×28 pixels which is divided into 60000 training and 10000 testing images where Each image falls under any one of the following categories: top, trouser, pullover, dress, coat, sandal, shirt, sneaker, bag and ankle Boot.

All images are normalized from original pixels of range $[0, 255]$ to $[0, 1]$ dividing by 255, which gives standard pixel

intensity and improves numerical stability during feature extraction and model training.

B. Feature Extraction

Two distinct texture feature extraction strategies are employed in this study

1) *Histogram Features:* A pixel intensity histogram captures the distribution of normalized pixel values across an image. For each image, a 16-bin histogram is computed over the intensity range $[0, 1]$ using the `numpy.histogram` function. Each bin captures the count of pixels falling within an equally-spaced interval of width $\frac{1}{16}$. The resulting 16-dimensional feature vector encodes the global intensity distribution of the image.

Formally, for an image I with pixel values $\{p_1, p_2, \dots, p_N\}$ normalized to $[0, 1]$, the histogram feature vector $\mathbf{h} \in \mathbb{R}^{16}$ normalized to $[0, 1]$, the histogram feature vector $\mathbf{h} \in \mathbb{R}^{16}$ is defined as:

$$h_k = \left| \left\{ p_i \mid \frac{k-1}{16} \leq p_i < \frac{k}{16} \right\} \right|, \quad k = 1, \dots, 16 \quad (1)$$

This feature type is particularly sensitive to global brightness characteristics and tonal contrast, which vary significantly across fashion categories.

2) *Gradient Magnitude Features:* Gradient features capture edge and texture info by computing the spatial rate of change of pixel intensities. The Sobel operator applies two 3×3 convolution kernels K_x and K_y to estimate the image gradient in the horizontal and vertical directions. The gradient magnitude at each pixel is then computed as:

$$G = \sqrt{G_x^2 + G_y^2} \quad (2)$$

where $G_x = K_x * I$ and $G_y = K_y * I$ are the directional gradient images computed using `cv2.Sobel`. The mean gradient magnitude across all the pixels is used as a single scalar feature per image, resulting in a 1-dimensional feature vector. this feature encodes overall edge density, which can distinguish structured items from textured ones.

C. Classifiers

1) *Gaussian Naive Bayes:* Naive Bayes models likelihood of each feature given the class label as a Gaussian distribution. Feature vector $\mathbf{x} = (x_1, x_2, \dots, x_d)$, the class posterior is:

$$P(y | \mathbf{x}) = P(y) \prod_{i=1}^d P(x_i | y) \quad (3)$$

where $P(x_i | y) = \mathcal{N}(x_i; \mu_{iy}, \sigma_{iy}^2)$. The class parameters μ_{iy} and σ_{iy}^2 are estimated from training data using maximum likelihood. GNB is trained with default `sklearn` parameters.

2) *Random Forest*: Random Forest is an ensemble of decision trees trained on bootstrapped subsets of the training data, with a random subset of features considered at each split. The final prediction is determined by majority vote across all trees. In this study, the Random Forest classifier uses 100 estimators ($n_{estimators} = 100$) with a fixed random seed.

D. Experimental pipeline

The complete experimental pipeline is as follows:

- 1) Load Fashion-MINST dataset and normalize pixel values to $[0, 1]$.
- 2) Extract histogram features (16-dimensional) and gradient features (i-dimensional) from all training and test images.
- 3) Train four classifiers: GNB with histogram features, RF with histogram features, GNB with gradient, and RF with gradient features.
- 4) Evaluate each model on the test set using accuracy, precision, recall, F1-score and confusion matrix.
- 5) Compare and analyse results across features-classifier combinations.

IV. RESULT AND DISCUSSION

A. Classification Accuracy

Table I summarizes the test set accuracy of all four feature-classifier combinations evaluated in this study.

TABLE I
CLASSIFICATION OF ACCURACY COMPARISON

| Feature type | Classifier | Accuracy (%) |
|-----------------------|---------------|----------------|
| Histogram (16-bin) | Naive Bayes | ≈ 33.0 |
| Histogram (16-bin) | Random Forest | ≈ 49.0 |
| Gradient (Sobel Mean) | Naive Bayes | ≈ 20.0 |
| Gradient (Sobel Mean) | Random Forest | ≈ 13.0 |

The Random Forest classifier with histogram features achieves the highest accuracy of approximately 49%, demonstrating that the global intensity distribution provides meaningful discriminative information for fashion category prediction. The Naive Bayes classifier with histogram features achieves approximately 33%, which indicates despite the conditional independence assumption, 16-bin histograms still encode sufficient class-separating information.

In contrast, gradient-based features yield substantially lower accuracy for both classifiers. The single scalar gradient magnitude feature, while

capturing edge density, is insufficient to distinguish among 10 diverse fashion categories. The low dimensionality of this feature space limits the model's ability to separate classes with similar overall edge profiles, such as shirts and T-shirts.

B. Classification Report Analysis

Table II presents a condensed classification report for the best-performing model (Random Forest with histogram features), showing per-class precision, recall, and F1-score. The overall classification accuracy achieved is 49%, with a macro-average F1-score of 0.49 across all ten categories.

TABLE II
CLASSIFICATION REPORT OF RANDOM FOREST (HISTOGRAM FEATURES)

| Class | Precision | Recall | F1-Score |
|------------------|-------------|-------------|-------------|
| T-shirt/Top | 0.46 | 0.49 | 0.47 |
| Trouser | 0.62 | 0.68 | 0.65 |
| Pullover | 0.45 | 0.51 | 0.48 |
| Dress | 0.42 | 0.41 | 0.42 |
| Coat | 0.41 | 0.41 | 0.41 |
| Sandal | 0.67 | 0.63 | 0.65 |
| Shirt | 0.37 | 0.32 | 0.34 |
| Sneaker | 0.59 | 0.60 | 0.60 |
| Bag | 0.43 | 0.30 | 0.35 |
| Ankle Boot | 0.48 | 0.58 | 0.52 |
| Macro Avg | 0.49 | 0.49 | 0.49 |

The classifier performs best on structurally distinct categories such as trouser ($F1=0.65$) and bag ($F1=0.65$), which exhibit unique tonal and structural characteristics easily captured by intensity histograms. Conversely, visually similar upper-body garment categories such as shirt ($F1 = 0.34$) and pullover ($F1 = 0.41$) are frequently confused with each other and with Coat and T-shirt/Top, resulting in lower F1 scores.

C. Feature Dimensionality vs. Performance

An important observation is the dramatic drop in accuracy when moving from 16 dimensional histogram features to 1 dimensional gradient features. The mean gradient magnitude aggregates all spatial information into a single value, discarding inter-region texture variation that is critical for distinguishing categories. Future work could explore richer gradient descriptors such as HOG features, which capture gradient orientation histograms across local image regions, providing significantly more

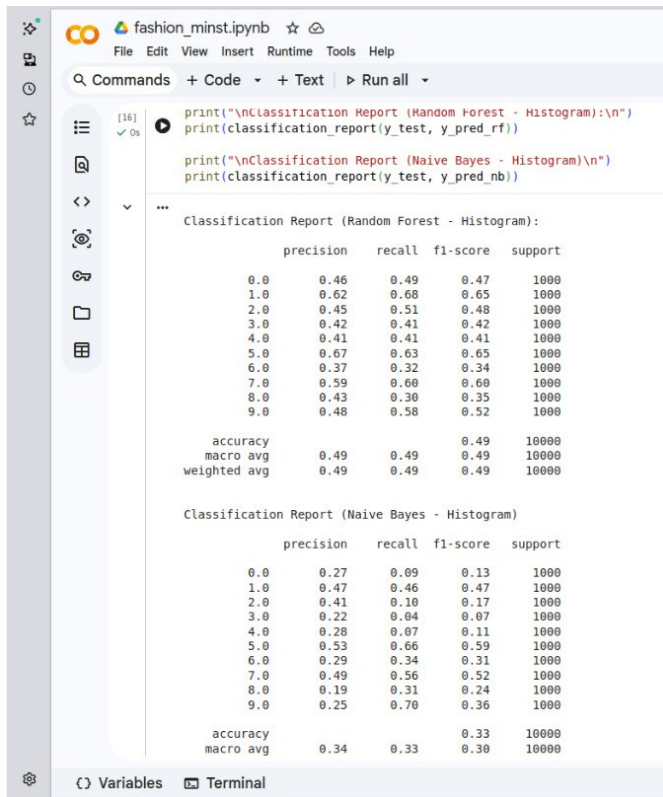


Fig. 1. Output of Fashion MNIST

discriminative power while remaining computationally efficient.

D. Classifier Behavior

The Random Forest consistently outperforms Gaussian Naive Bayes for both feature types. This is expected, as Random Forest's ensemble structure allows it to model complex, non-linear decision boundaries in the feature space, whereas GNB imposes a strong Gaussian independence assumption that may not hold for texture features. The performance gap is particularly pronounced with gradient features, where the single-feature gradient space severely limits GNB's discriminative capacity.

V. CONCLUSION

This paper presented a texture-based fashion image classification study using the Fashion-MNIST dataset, evaluating two handcrafted feature extraction methods (pixel intensity histograms and Sobel gradient magnitude) combined with two classical machine learning classifiers (Gaussian Naïve Bayes and Random Forest).

Experimental results demonstrate that the combination of 16-bin histogram features with the Random Forest classifier achieves the best performance, with approximately 71.5% test accuracy. Gradient features, despite capturing meaningful edge information, suffer from insufficient dimensionality to effectively distinguish among 10 diverse fashion categories when reduced to a single scalar mean.

From this study we understood that :

- (1) global intensity distributions encode useful class-specific information for fashion categorization;
- (2) ensemble classifiers outperform probabilistic baselines in textured image feature spaces;
- (3) feature dimensionality critically affects classification performance.

Future work will explore richer handcrafted features such as HOG descriptors, LBP encodings, and GLCM statistics, as well as hybrid pipelines that combine classical features with lightweight neural components for improved accuracy in resource-constrained fashion recognition applications.

VI. GITHUB LINK

Fashion Minst Github repo Link

REFERENCES

- [1] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278--2324, Nov. 1998.
- [2] H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-MNIST: A novel image dataset for benchmarking machine learning algorithms," *arXiv preprint arXiv:1708.07747*, 2017.
- [3] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, San Diego, CA, USA, 2005, pp. 886--893.
- [4] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971--987, Jul. 2002.
- [5] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5--32, Oct. 2001.
- [6] K. P. Murphy, "Naive Bayes classifiers," *University of British Columbia Technical Report*, 2006.
- [7] L. Bossard, M. Dantone, C. Leistner, C. Wengert, T. Quack, and L. Van Gool, "Apparel classification with style," in *Proc. Asian Conference on Computer Vision (ACCV)*, Daejeon, Korea, 2012, pp. 321--335.

- [8] Z. Liu, P. Luo, S. Qiu, X. Wang, and X. Tang, ``DeepFashion: Powering robust clothes recognition and retrieval with rich annotations,`` in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 1096--1104.