Fashion MNIST Classification : NN Models

Karan Saraswat   
*M.S. in Industrial Engineering  
Texas A&M University*  
College Station, U.S.  
ksaraswat@tamu.eduGeetesh Challur  
*M.S. in Computer Science*  *Texas A&M University*  
College Station, U.S.  
geeteshch@tamu.edu  
Venkata Pitta  
*M.S. in Computer Science*  *Texas A&M University*  
College Station, U.S.  
pitta.venkata@tamu.edu

*Abstract*— In this article, both Convolutional Neural Network (CNN) and Dense Neural Network (DNN) models are developed and assessed for their effectiveness in image recognition task on Fashion MNIST dataset for classifying articles of clothing based on the performance of these models. The algorithmic frameworks are applied to the Fashion MNIST dataset, and a comprehensive evaluation of their performance is conducted. The outcomes of this study reveal that the DNN model achieves an accuracy rate of 86% on the MNIST dataset, while the CNN model demonstrates a commendable accuracy of 85%. The utilization of these models on the Fashion MNIST dataset showcases their capabilities in handling complex image recognition tasks.

Keywords—Convolutional Neural Network, Accuracy, Dense Neural Network, PCA, t-SNE, Cross Validation

# Introduction

Image classification is a fundamental aspect of computer vision that involves the automated categorization of images into predefined classes or labels. Leveraging advanced machine learning techniques, particularly Neural Networks, image classification algorithms learn to discern and extract distinctive features within images. A Dense neural network, or dense neural network, consists of layers where each neuron is connected to every neuron in the adjacent layers, enabling information flow throughout the network. Notable for their hierarchical architecture, CNN employ convolutional layers to automatically extract features from input data, capturing spatial hierarchies and enabling robust pattern recognition. This model consist of convolutional, pooling, and Dense layers, allowing it to efficiently learn and generalize complex visual patterns. Both DNN and CNN have revolutionized image-related tasks, exhibiting exceptional performance in image classification, object detection, and facial recognition.

Top of Form

This study offers a concise exploration of image classification models. Two neural network models are employed: (a) DNN, and (b) CNN specifically designed to tackle the issue of model overfitting. The subsequent sections of the paper include a summary of related work in [Section 2](#_LITERATURE_REVIEW), an overview of methods and the proposed algorithm in [Section 3](#_METHOD), classification experiments and discussions in [Section 4](#_Experiments), and concluding remarks with future directions in [Section 5](#_Conclusions).

# LITERATURE REVIEW

In the past few years, significant advancements have been made in constructing image detection and recognition classifiers across diverse datasets, employing a range of machine learning algorithms. Notably, deep learning has demonstrated substantial enhancements in accuracy across various datasets. Specifically, the convolutional neural networks revealed special effects on computer vision and image classification. Several noteworthy endeavors in this realm are outlined below:

Shobhit et al. [[1]](https://ieeexplore.ieee.org/document/8313740) introduced an innovative approach to classifying images of fashion articles. The research focused on training convolutional neural network (CNN) architectures within a deep learning model to effectively categorize the Fashion-MNIST dataset. To enhance the learning rate, three distinct CNN methods were implemented, featuring residual skip networks and batch normalization. The findings of the study revealed a notable 2% increase in accuracy rates for existing systems, showcasing the effectiveness of the proposed model.

Wang Di [[2]](https://ieeexplore.ieee.org/document/9368530) emphasizes the importance of deploying computationally efficient neural networks in real-world applications. Comparing Fully Connected Neural Network, CNN, MobileNetV1, and MobileNetV2 on Fashion-MNIST, MobileNetV2 emerges as the most effective, achieving a remarkable 93% accuracy. The findings underscore the practical significance of MobileNet's simplicity and heightened accuracy.

Yehya Abouelnaga et al. [[3]](https://ieeexplore.ieee.org/document/7881518) developed an ensemble of classifiers utilizing the K-Nearest Neighbors (KNN) algorithm. Integrated KNN with Convolutional Neural Networks (CNN) to address Overfitting concerns, implementing Principal Component Analysis (PCA) for dimensionality reduction. The amalgamation of these two classifiers resulted in a notable enhancement, yielding an improved accuracy by approximately 0.7%.

Edmira Xhaferra et al. [[4]](https://ieeexplore.ieee.org/document/9932737) investigates the efficacy of Convolutional Neural Networks (CNNs) in fashion industry applications, targeting challenges such as model overfitting and intricate garment classification. The study employs CNN-C1 and CNN-C2 architectures for image classification in the FASHION MNIST database. Results demonstrate that CNN-C2 outperforms CNN-C1, achieving an enhanced accuracy of 93.11%.

Rahul Chauhan et al. [[5]](https://ieeexplore.ieee.org/document/8703316) investigates Deep Learning algorithms, focusing on Convolutional Neural Networks (CNNs) that mirror the human cerebral cortex. His study evaluates CNN models' performance on MNIST and CIFAR-10 datasets, achieving an impressive 99.6% accuracy on MNIST and optimizing CIFAR-10 results through real-time data augmentation and dropout on a CPU unit.

# METHOD

## Data Preparation

In this study, we tackled the challenge of fashion image classification within the extensive Fashion MNIST dataset, which consists of 70,000 images. Employing two different CNN Networks, we aimed to improve efficiency in categorizing clothing types. The dataset contains images, each represented in grey scale with dimensions of 28x28.

The complexity of classification lies in the nuanced fashion divisions, where labels reflect the diverse types of clothing depicted in the images. Our study aims to enhance the accuracy and effectiveness of fashion image classification, ensuring a more refined understanding of the dataset's 10 distinct classes, as illustrated in Fig. 1.

A screenshot of a computer screen

Description automatically generated

Fig 1. Fashion MNIST [4]

1. *Data cleansing and transformation*

Dataset doesn’t have any missing values or corrupted images which makes it easier to use for our analysis. Data normalization is crucial for image classification as it ensures consistent feature scales, preventing dominance by certain features. It is done by scaling pixel values to a range between 0 and 1. Min-max normalization is applied to the data.

1. *Data Splitting*

The dataset is sourced from torch vison's fashion clothing image database, includes 60,000 training images and 10,000 test images. Training data is further split into validation sets keeping equal number of data points in each class.

## Exploratory Data Analysis

1. *Descriptive statistics*

Ensuring uniform data length across all classes is crucial to prevent analysis errors stemming from class imbalance. This is verified by visualizing class distribution with a count plot showing an equally distributed data in each class for split training and validation datasets given below:

A close-up of a graph

Description automatically generated

Fig 2. Data distribution among ten classes for training and validation

1. *Data Visualization*

In this study, Principal Component Analysis (PCA) is employed to reduce dataset dimensions while preserving key information. By assessing correlations between dimensions, it retains the most meaningful variables, discarding those with lower explained variance. The following visualization illustrates images before PCA and post-PCA application, showcasing the preservation of key information in the reduced dimensionality.

A screenshot of a computer screen

Description automatically generated

Fig 3. Fashion accessories picture before (top) and after PCA (bottom)

In the graph below, 87 components cover about 90% variance, while 194 components explain around 95%. As components increase, variance reaches 100%, aligning with the original 784 dimensions.

A graph with a blue line

Description automatically generated

Fig 4. Cumulative variance captured with respect to number of features

Following are plots of first two and first three components of PCA data, while each color a label: A colorful dots in a circle

Description automatically generated with medium confidence A diagram of a graph

Description automatically generated with medium confidence

Fig 5. PCA components plots: 2 components (left) and 3 components(right)

From the above plots, different categories are separated apart to some extent, but it's unclear how distinct the various categories are from one another. Even a plot with three components does a poor job of separating classes. Thus, t- Distributed Stochastic Neighboring Entities is applied on first five thousand rows to reduce the number of dimensions. The following plot shows significant improvement in classification wherein each class is properly clustered and separated from one another.

A colorful dots on a white background

Description automatically generated

Fig 6. t-SNE first two components plot

## Model Selection

1. *Algorithm selection*

Drawing insights from the visualizations in Zalando’s fashion-mnist repository, it becomes evident that the data is characterized by non-linearity and complex pattern relationships [6]. In addressing this complexity, neural networks stand out from k-nearest neighbors and Decision Tree algorithms by autonomously learning hierarchical feature representations. This inherent capability contributes to their superior performance on the dataset, enabling effective adaptation to intricate image patterns unlike k-nearest neighbors and decision tree algorithms.

Hence, in this study, the initial choice of a Dense Neural Network (DNN) is motivated by its capability to capture intricate patterns within the data. As a comparative analysis, the selection of a Convolutional Neural Network (CNN) is justified, given its specialization in image-related tasks and its efficient handling of spatial hierarchies of features through shared weight parameters.

1. *Model building*

more compact, you may use the solidus ( / ), the exp function, or appropriate exponents

1. *Model evaluation*

more compact, you may use the solidus ( / ), the exp function, or appropriate exponents. Italicize Roman symbols:

*a**b* 

# Experiments

## Authors and Affiliations

## Identify the Headings

because all subsequent material relates and elaborates on this one topic. If there are two or more sub-topics, the next

## Figures and Tables

#### Positioning Figures and Tables: Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation “Fig. 1”, even at the beginning of a sentence.

1. Table Type Styles

| Table Head | Table Column Head | | |
| --- | --- | --- | --- |
| Table column subhead | Subhead | Subhead |
| copy | More table copya |  |  |

1. Sample of a Table footnote. (*Table footnote*)
2. Example of a figure caption. (*figure caption*)

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an example, write the quantity “Magnetization”, or “Magnetization, M”, not just “M”. If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write “Magnetization (A/m)” or “Magnetization {A[m(1)]}”, not just “A/m”. Do not label axes with a ratio of quantities and units. For example, write “Temperature (K)”, not “Temperature/K”.

# Conclusions

##### Acknowledgment *(Heading 5)*

The preferred spelling of the word “acknowledgment” in America is without an “e” after the “g”. Avoid the stilted expression “one of us (R. B. G.) thanks ...”. Instead, try “R. B. G. thanks...”. Put sponsor acknowledgments in the unnumbered footnote on the first page.

##### References

[[1]](https://ieeexplore.ieee.org/document/8313740) S. Bhatnagar, D. Ghosal, and M. H. Kolekar, “Classification of fashion article images using convolutional neural networks,” in 2017 Fourth International Conference on Image Information Processing (ICIIP), 2017: IEEE, pp. 1-6.

[[2]](https://ieeexplore.ieee.org/document/9368530) Wang Di, “A comparative research on clothing images classification based on neural network models” in 2020 IEEE 2nd International Conference on Civil Aviation Safety and Information Technology, 2020, pp. 1-5.

[[3]](https://ieeexplore.ieee.org/document/7881518) Yehya Abouelnaga , Ola S. Ali , Hager Rady , Mohamed Moustafa, “ CIFAR-10: KNN-based ensemble of classifiers”, IEEE, March 2017.

[[4]](https://ieeexplore.ieee.org/document/9932737) Edmira Xhaferra, Elda Cina, and Luçiana Toti, “Classification of Standard FASHION MNIST Dataset Using Deep Learning Based CNN Algorithms” in International Symposium on Multidisciplinary Studies and Innovative Technologies, IEEE, October 2022.

[[5]](https://ieeexplore.ieee.org/document/8703316) Rahul Chauhan, Kamal Kumar Ghanshala, and R.C. Joshi, “Convolutional Neural Network (CNN) for Image Detection and Recognition” in 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC), IEEE, 2018.

[[6]](https://github.com/zalandoresearch/fashion-mnist) Github repository : <https://github.com/zalandoresearch/fashion-mnist>