Report on Fashion MNIST Classification with Machine Learning

Karan Athrey, Abhijit Sinha, Anusha Chatterjee
Emails: kathrey@asu.edu , asinh117@asu.edu , and achatt53@asu.edu

***Abstract*— This report explores the application of machine learning techniques to the Fashion MNIST dataset, focusing on classifying grayscale images of fashion items into ten categories. The project implements a comprehensive machine learning pipeline that compares the performance of deep learning models, particularly Convolutional Neural Networks (CNNs) and Residual Networks (ResNet), with traditional methods such as Logistic Regression, Support Vector Machines (SVM), and Random Forest. Preprocessing steps, including normalization and data reshaping, were employed to optimize model performance.Deep learning models achieved superior accuracy (~91%) and demonstrated robust generalization capabilities, while traditional models offered simplicity and interpretability but struggled with complex image patterns. Grad-CAM visualizations were utilized to enhance the interpretability of CNN predictions, building trust in model decisions.The results highlight the potential of machine learning in real-world applications, such as e-commerce, fashion retail, and manufacturing, where automated image classification can drive operational efficiency, improve customer experiences, and enable strategic decision-making. Future directions include addressing class-specific misclassifications, extending models to handle more complex datasets, and exploring hybrid techniques for improved accuracy. This study underscores the transformative role of machine learning in advancing image classification tasks across industries.**

***Index Terms*— Machine Learning, Deep Learning, CNN, ResNet, MLP, MNIST**

INTRODUCTION

The project demonstrates the potential of machine learning in enhancing operational efficiency, innovation, and competitive advantage. Using the Fashion MNIST dataset, the study implements a comprehensive machine learning pipeline to classify images. The primary objective is to evaluate the performance of a Convolutional Neural Network (CNN) and compare it with traditional machine learning models. The findings highlight the significance of machine learning in real-world applications, including e-commerce, fashion retail, and manufacturing.

Objective

The primary objective of this project is to design and implement a comprehensive machine learning pipeline capable of classifying fashion items with high accuracy. By comparing the performance of deep learning models with traditional approaches, the study identifies the strengths and trade-offs of each method. Advanced visualizations are created using Plotly to ensure an interactive and in-depth understanding of the data and results.

Significance

The outcomes of this project extend beyond academic interest, demonstrating real-world applications in various industries:

- **E-commerce platforms**: Enhancing product recommendations, optimizing search functionality, and improving inventory management.
- **Fashion retail**: Using AI to analyze trends, customer preferences, and streamline stocking and design decisions.
- **Social media platforms**: Employing image recognition for features like shoppable posts and user engagement.
- **Manufacturing**: Ensuring quality control on production lines by classifying products and detecting defects.

ADDRESSING THE HOW ASPECT

The project systematically tackled the **"how"** of classifying Fashion MNIST images through the following steps:

- **Dataset Preparation:** Normalized pixel values to [0,1], reshaped images for CNNs, and split the dataset into training, validation, and testing subsets. Traditional models used flattened images for compatibility.
- **Model Implementation:**

1. Deep Learning: Designed CNNs and ResNet architectures with layers for feature extraction, dimensionality reduction, and classification.

2. Used data augmentation and Adam optimizer for training.Traditional Models: Implemented Logistic Regression, SVM, Random Forest, and MLP for performance benchmarking.

- **Performance Measurement:** Used accuracy, loss trends, confusion matrices, and Grad-CAM visualizations to evaluate model effectiveness and interpretability.
- **Comparison:** Deep learning models outperformed traditional ones in accuracy (~91%) and scalability, while traditional models were computationally simpler.
- **Real-World Applications:** Addressed use cases in e-commerce (recommendations), fashion retail (trend analysis), and manufacturing (quality control).
- **Challenges and Improvements:** Tackled overfitting with data augmentation, normalized sparse pixel data, and proposed future work on colored datasets and hybrid models.

ADDRESSING THE WHY ASPECT

-Preprocessing Steps:

- **Normalization:** Scaled pixel values to [0,1] to enhance model convergence.
- **Reshaping:** Adjusted images to (28x28x1) for CNNs and flattened them for traditional models.
- **Splitting:** Created training, validation, and testing subsets.

-Models Implemented:

Deep Learning Models:

- CNN: Extracted spatial features for classification.
- ResNet: Used residual connections to handle deeper architectures.

Traditional Models:

- Logistic Regression, SVM, Random Forest, and MLP provided benchmarks.

ML Models Used in This project

Residual Network (ResNet):
A deep learning model utilizing residual connections to address vanishing gradient issues, enabling deeper architectures and achieving ~91% accuracy with robust generalization.

Convolutional Neural Network (CNN):
Extracts spatial features from images using convolutional and pooling layers, achieving high accuracy (~91%) for image classification tasks.

Logistic Regression:
A simple linear classifier used as a baseline, achieving ~83% accuracy but limited in handling complex image patterns.

Support Vector Machine (SVM):
Employs kernel methods for non-linear classification, achieving ~87% accuracy but computationally expensive for large datasets.

Random Forest:
An ensemble method aggregating decision trees, providing ~89% accuracy with reasonable performance on flattened image data.

K-Nearest Neighbors (KNN):
A lazy learner classifying images based on nearest neighbors, but less efficient for large datasets due to high memory usage.

Multi-Layer Perceptron (MLP):
A fully connected neural network achieving ~85% accuracy, limited in performance due to lack of spatial awareness in images.

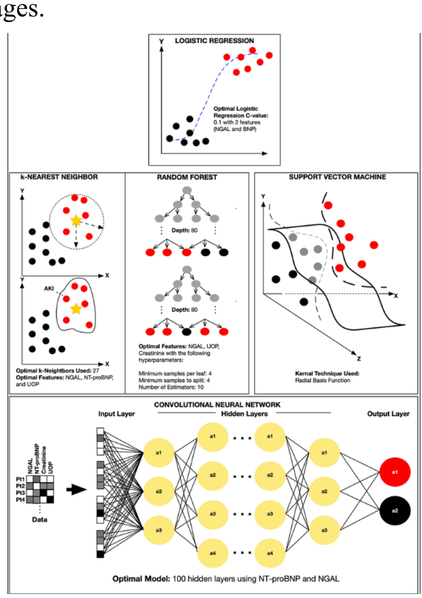
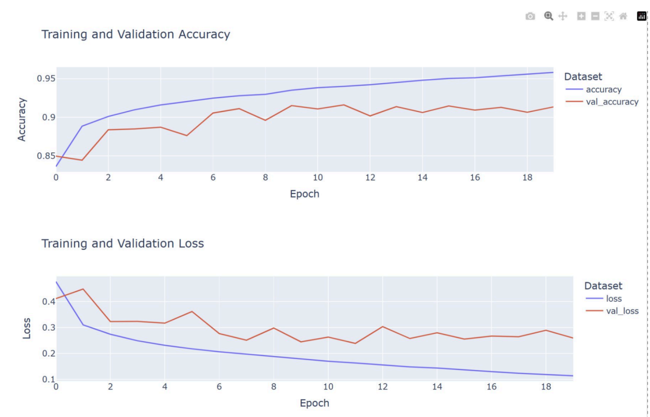


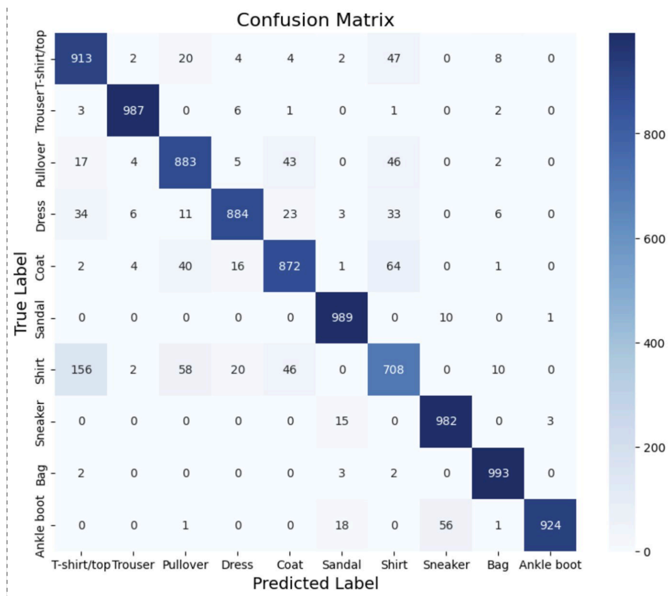
Fig: various types of algorithms

PERFORMANCE ANALYSIS

ResNet model analysis

Metric	Training	Validation
Final	95%	91%
Loss	0.1	0.3
Convergen	Steady	Stable






```
Classification Report
              precision    recall  f1-score   support

 T-shirt/top       0.81      0.91      0.86      1000
     Trouser       0.98      0.99      0.98      1000
    Pullover       0.87      0.88      0.88      1000
       Dress       0.95      0.88      0.91      1000
        Coat       0.88      0.87      0.88      1000
      Sandal       0.96      0.99      0.97      1000
       Shirt       0.79      0.71      0.74      1000
     Sneaker       0.94      0.98      0.96      1000
         Bag       0.97      0.99      0.98      1000
  Ankle boot       1.00      0.92      0.96      1000

    accuracy                           0.91     10000
   macro avg       0.91      0.91      0.91     10000
weighted avg       0.91      0.91      0.91     10000
```

Here is the Grad-cam mapping from the dataset

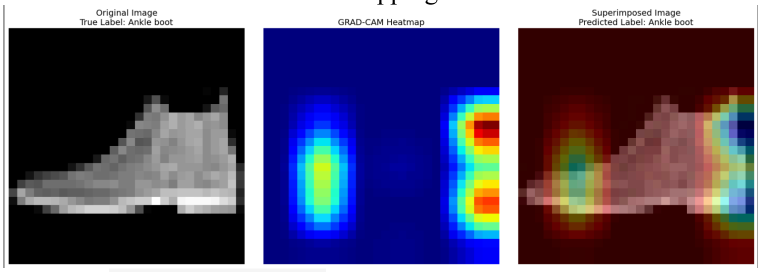
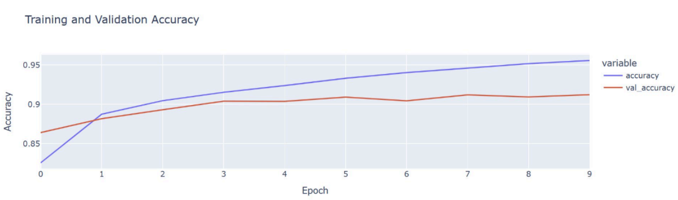
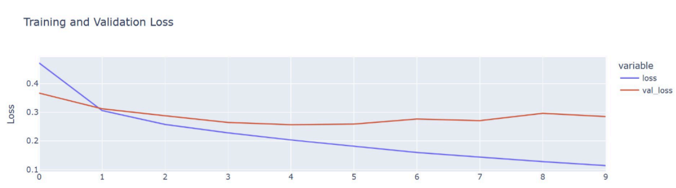


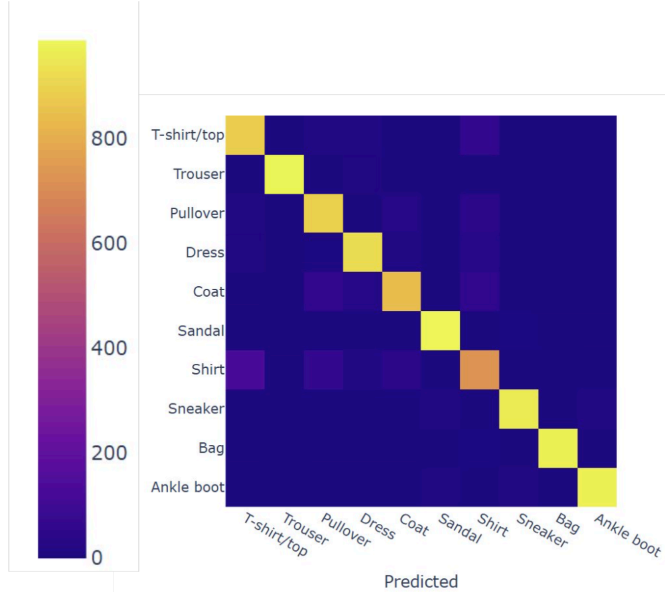
Image Breakdown:
-**Left**: Original Image
Displays the grayscale input image of the item classified as "Ankle Boot."
Ground truth label: **Ankle Boot**.
-**Middle**: Grad-CAM Heatmap
Highlights the regions most influential for the model's prediction.
Bright areas (yellow/red) indicate where the model focuses to identify the object.
-**Right**: Superimposed Image
Combines the original image and the Grad-CAM heatmap.
Shows the model correctly focused on the shoe region for its prediction.

CNN model performance analysis

Metric	Training	Validation
Final	96%	91%
Final Loss	0.12	0.28
Convergence	Continuous	Stable with fluctuation







PREDICTION ON UNLABELLED DATASETS USING TRAINED CNN MODEL

ISSUE FACING

- Dataset bias: Grayscale images with white backgrounds limit real-world applicability
- Model struggles with colored images and complex backgrounds, misclassifying common items

FUTURE WORK

Incorporates more complex datasets with colored images and diverse backgrounds.Address specific class misclassifications (e.g., T-shirt vs. Shirt).Explore hybrid models combining traditional and deep learning techniques for improved performance.

CONCLUSION

This project demonstrates the effectiveness of machine learning in classifying Fashion MNIST images, with deep learning models outperforming traditional methods in accuracy and scalability. While traditional models offer simplicity, they lack the robustness required for complex image patterns. The study underscores the potential of machine learning in real-world applications, from e-commerce to manufacturing.