Comparative Analysis on Image Classification Architectures.

ECEN 758 Data Mining and Analysis Project Fall 2023

Rahul Gautam (334003785)
MS Data Science Grad - Fall 23
Bryan, Texas - 77801

Harshit Garg (834007124)
MS Data Science Grad - Fall 23
Bryan, Texas - 77801

Prince Tibadia (334002581)
MS Data Science Grad - Fall 23
Bryan, Texas - 77801

***Abstract*—Image Classification has found a significant amount of attention since 2012. Many classification algorithms exist at the current date. CNN architectures have gone through rapid evolution and in the current times have reached various possibilities. Apart from that, many machine learning algorithms also exist which have also proven to yield substantial results. Various machine learning algorithms like Support Vector Machine, Decision Tree, Random Forest, etc. exist and if utilized with the correct constraints may prove to give significant results. In this article, we conduct a detailed comparative analysis based on classification of the Fashion-MNIST dataset.**

***Keywords*— Data Mining, PCA, ANN, CNN, Random Forest.**

I. Introduction

The field of computer vision has witnessed remarkable progress in recent years, with advancements in image classification and object recognition algorithms playing a pivotal role. One dataset that has become instrumental in benchmarking and developing such algorithms is the FashionMNIST dataset. Created as a drop-in replacement for the traditional MNIST dataset, FashionMNIST offers a more challenging task by focusing on classifying images of clothing items into distinct categories.

In this report, we embark on an exploratory data analysis and classification journey using the FashionMNIST dataset. The dataset comprises 60,000 training images and 10,000 test images, each belonging to one of the 10 fashion categories. The goal is to delve into the intricacies of the dataset, understand its unique challenges, and employ various machine learning and deep learning algorithms with parameter tuning to get the best accuracy results. The complete code for the project is present in the Github repository [15].

II. Literature Review

The literature review draws upon diverse studies to provide a nuanced background for the proposed report. Yang Li and Xuewei Chao's work [1] explores ANN-based continual classification in agriculture, emphasizing neural network's significance in this domain. Imran Iqbal et al.'s study [2] addresses challenges in small dataset image classification, offering a comparative investigation of learning algorithms.

Studies like Jagadeesh Pujari"s work [3] focus on SVM and ANN-based classification of plant diseases, contributing insights into agricultural disease detection. Greeshma KV and Sreekumar K's research [4] adds value with Fashion-MNIST classification using HOG feature descriptors and SVM.

Vandana Kannan's study [5] introduces feature selection using genetic algorithms, enhancing our understanding of optimization techniques. Yansheng Li, Xin Huang, and Hui Liu's work [6] on unsupervised deep feature learning for urban village detection broadens the scope of image analysis.

Mingyuan Xin and Yong Wang's exploration [7] of image classification models with deep convolutional neural networks contributes to advanced neural network architectures. Iqbal H. Sarker's comprehensive overview [8] of machine learning algorithms provides a broader context for methodology considerations.

Hyungkeuk Lee's paper [9] addresses model optimization for image classification on edge devices, essential for practical deployment. Olivia Nocentini et al.'s work [10] on multiple convolutional neural networks for image classification offers insights into advanced techniques.

Guoshu Zheng and Qiuyu Zhu's research [11] on unsupervised image feature extraction using scattering transform and self-supervised learning provides an alternative approach to feature representation. The foundational work by Han Xiao's [12] introduces the Fashion-MNIST dataset, a crucial benchmark for image classification.

Finally, M. Weber, M. Welling, and P. Perona's study [13] on unsupervised learning of models for recognition provides historical context and foundational principles in unsupervised learning. Together, these studies contribute a comprehensive understanding of methodologies, challenges, and innovations in image classification and related fields.

III. Method

In our study, we initiated the process by obtaining the FashionMNIST dataset from the GitHub repository [14]. The dataset, consisting of training and test images depicting various fashion items, was imported, and organized into distinct sets. Following this, the images were flattened to transform 2D arrays into 1D vectors, optimizing the data for subsequent processing. Standardization was then applied using the Standard Scaler to ensure consistent pixel values across all images, setting each feature to have a mean of 0 and a variance of 1.

To address the challenges posed by high-dimensional data, we employed Principal Component Analysis (PCA) for dimensionality reduction. A careful selection of 50 components aimed to retain essential information while reducing overall dimensionality efficiently.

Our methodology extended to training a diverse set of machine learning algorithms on both the original and PCA-converted datasets. This comprehensive approach included k-Nearest Neighbors (KNN), Logistic Regression, Gradient Boosting, Stochastic, Support Vector Classifier (SVC), Random Forest Classifier, Decision Trees, and Linear Support Vector Machine (Linear SVM).

To ensure a robust assessment of each model's performance and to check for overfitting, 3-fold cross-validation was employed on PCA-converted datasets. This approach provided a comprehensive evaluation through accuracy metrics.

In addition to the extensive exploration of traditional machine learning algorithms, our methodology extended to the implementation of Convolutional Neural Networks (CNN) to further enhance the classification results. After completing the analyses with KNN, Logistic Regression, Gradient Boosting, SVC, Random Forest Classifier, Decision Trees, and Linear SVM, we sought to leverage the capabilities of deep learning.

The CNN model was designed to capture intricate patterns and hierarchical features within the image data. This involved the utilization of convolutional layers for spatial feature extraction, pooling layers for down-sampling and reducing computational complexity, and densely connected layers for high-level feature representation. The model was trained on the original dataset and subsequently on the PCA-converted data to compare its performance against traditional machine learning approaches.

The implementation of CNN introduces a novel dimension to our study, allowing for an exploration of the efficacy of deep learning in comparison to conventional algorithms.

IV. Experiments

A. Dataset Details

The FashionMNIST dataset, consisting of 60,000 training images and 10,000 test images, forms the core of our research. An evolution from the MNIST dataset, it poses a challenging image classification task with ten distinct fashion categories, including T-shirts, trousers, pullovers, dresses, coats, sandals, shirts, sneakers, bags, and ankle boots. Each 28x28 pixel grayscale image intricately captures the details of various fashion items, contributing to the dataset's complexity. Our study revolves around this diverse dataset, employing both traditional and deep learning methods to unveil patterns essential for advancing the field of computer vision.

B. Data Preprocessing

Our data preprocessing pipeline was pivotal in optimizing the original FashionMNIST dataset for subsequent analysis, and it included a series of crucial steps. Initially, we imported the data from the provided GitHub repository, organizing it into training and test sets. To prepare the images for feature extraction, we flattened the 2D arrays into 1D vectors, ensuring compatibility with various machine learning algorithms.

Standardization played a crucial role in homogenizing the pixel values across all images. Leveraging the Standard Scaler, we normalized each feature to have a mean of 0 and a variance of 1, minimizing the impact of differing scales and facilitating more efficient model training.

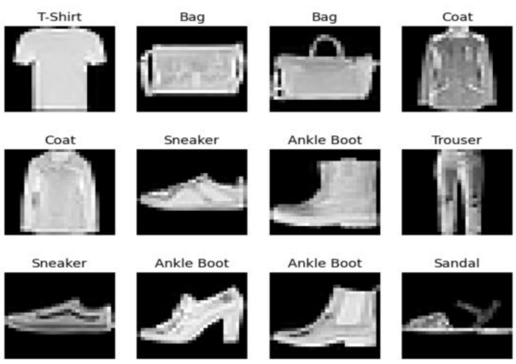

Fig. 1. Sample FashionMNIST Data

To address the challenges posed by the high dimensionality of the original dataset, we incorporated PCA for dimensionality reduction. This step involved capturing the most relevant information while reducing the number of features to 50 principal components. This not only streamlined computational complexity but also retained the essential characteristics of the data.

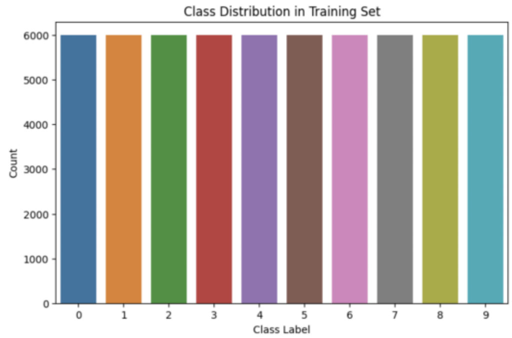

Fig. 2. Class Distribution for Training Data

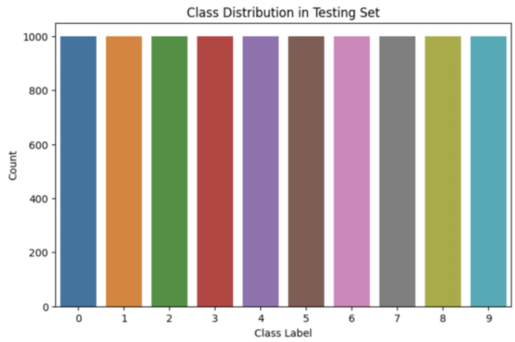

Fig. 3. Class Distribution for Test Data

The application of PCA served as a critical preprocessing step, enabling us to strike a balance between information retention and computational efficiency. These preprocessing measures collectively laid the foundation for our subsequent experiments, ensuring that the data was suitably prepared for both traditional machine learning algorithms and more sophisticated deep learning approaches.

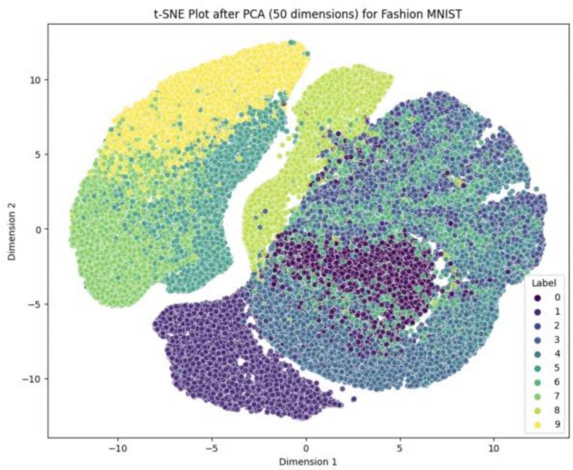

Fig. 4. t-SNE Plot for FashionMNIST

In addition to the preprocessing steps applied to the original FashionMNIST dataset, we incorporated image augmentation specifically tailored for our deep learning models. Recognizing the importance of expanding the diversity of the training set, image augmentation was employed as a strategic technique to introduce variability and enhance model generalization. Through random transformations such as rotations, shifts, and flips, we artificially expanded the dataset, creating a more robust training environment for our deep learning models, including both ANN and CNN. This augmentation not only contributed to the robustness of our models but also helped mitigate the risk of overfitting, allowing for more effective learning from the available data. The integration of image augmentation exemplifies our commitment to optimizing the dataset for the unique requirements of deep learning methodologies.

C. Machine Learning Models

1. K-Nearest Neighbors (KNN) - A simple yet effective classification algorithm where an instance is classified based on the majority class of its k nearest neighbors. KNN is well-suited for image classification tasks, leveraging proximity-based decision-making, making it an intuitive choice for our FashionMNIST dataset. For KNN, we set the number of neighbors (k) to 5, used Euclidean distance (p=2), and assigned equal weights to neighbors (weights="uniform").

2. Logistic regression - A linear model for binary and multiclass classification, logistic regression estimates the probabilities of class membership. This model is chosen for its interpretability, simplicity, and adaptability to multiclass classification, aligning with the diverse nature of our FashionMNIST categories. For Logistic Regression, we set the regularization parameter (C) to 1, employed automatic determination for the multi-class strategy, and used L2 penalty regularization.

3. Linear Support Vector Machine (Linear SVM) - A linear classifier that finds the hyperplane that best separates classes. This model is selected for its effectiveness in binary and multiclass classification, particularly in scenarios where classes are linearly separable or nearly separable. The Linear SVM model utilized an L2 penalty, a hinge loss function, and a regularization parameter (C) set to 1.

4. Decision Trees - Decision Trees make decisions based on feature values to classify instances. This model is employed for its simplicity and interpretability, providing insights into the decision-making process for each classification. For Decision Trees, we set the criterion to "entropy," the maximum depth to 10, and used the "best" strategy for choosing splits.

5. Random Forest - Random Forest builds an ensemble of decision trees to enhance prediction accuracy and control overfitting. The ensemble nature of Random Forest mitigates overfitting and enhances generalization, making it well-suited for our image classification task. In our Random Forest implementation, we set the criterion to "entropy," the maximum depth to 50, and the number of trees (n_estimators) to 100.

6. Gradient Boosting - Gradient Boosting builds a series of weak learners to iteratively correct errors, creating a robust ensemble model. This model excels in capturing complex relationships in data, making it apt for extracting intricate patterns within the FashionMNIST dataset. For Gradient Boosting, we set the loss function to "log_loss" and the maximum depth of each tree to 3.

7. Support Vector Classifier (SVC) - SVC constructs hyperplanes in high-dimensional space to separate classes. This model is effective in capturing complex decision boundaries, making it suitable for the intricate patterns present in the FashionMNIST images. The SVC model utilized an RBF kernel, a regularization parameter (C) set to 1.0, and a gamma value set to "scale," with a fixed random state at 42. These parameter choices were made to optimize the model's performance on the FashionMNIST dataset.

D. Deep Learning Models

1. Artificial Neural Networks (ANN) - After finishing the experiment with machine learning models, we planned to implement neural nets on the dataset. Starting with ANN, we deployed 3 dense layers (512, 256, 128) with the input being a flattened image and predicting the probability of the input image belonging to the 10 class labels. Dropout with probability=0.25 was used after the first fully connected layer and the ReLU activation function was used after the first two layers. The class label with the maximum probability is selected as the final prediction. The training and test dataset is used as defined in the section. The pixels are transformed by dividing by 255 and scaled parameters (mean = 0.1307), (standard deviation = 0.3081). In addition to data pre-processing, we have used a 3-fold cross-validation strategy to prevent model over-fitting and give more precise results. The model is trained with a cross-entropy loss for 50 epochs with an early stopping criterion: if the validation loss does not improve in 10 successive epochs, the training process is stopped and the last model with the lowest validation loss is returned to test on new samples.

2. Convolution Neural Networks (CNN) - When experimenting with ML models and ANN, one major issue observed was it looks at the pixels in a tabular and does not capture the relationship between which may be a major factor for predictive models for images. This motivated us to shift our approach towards convolutional layers as they can extract relationships between neighboring pixels. We built a small CNN module by stacking 2 convolutional layers with Batch normalization followed by max-pooling layers on the top of our ANN. The activation function for convolutional layers is kept the same as ReLU which was used in our ANN model. The architecture for our CNN model is shown in the figure. Using the same training and testing dataset, the input to the CNN model is an image now and we used image augmentation techniques such as randomly flipping training images vertically or rotating with a small angle to add more randomness to training data. A 3-fold cross-validation strategy is used again during training and the best model is obtained using the same stopping criterion used during the training of our ANN model.

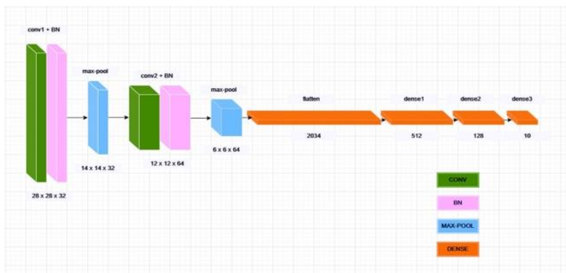

Fig. 5. CNN Architecture

E. Selection of the best model

In the exploration of the Fashion MNIST dataset, the selection of the best model involved a meticulous process where we considered traditional ML models, ANN, and CNN as candidate models. The initial phase included training classic machine learning models like [fill all ML models tested] to establish a baseline performance. Subsequently, ANN was employed to capture complex relationships within the data. Recognizing the image-centric nature of the Fashion MNIST dataset, CNN was finally introduced. The CNNs excelled in feature extraction and hierarchical pattern recognition, showcasing superior performance compared to the other models. We also noticed that CNN was generally more computationally efficient than traditional machine learning models and ANN when it comes to image-related tasks. This efficiency arises from the specific architecture and operations performed by CNNs, which leverage shared weights and hierarchical feature extraction. The selection of the best model was ultimately driven by a comprehensive evaluation of accuracy, computational efficiency, and the model's ability to generalize well to unseen data, with the convolutional neural network emerging as the optimal choice for the task at hand.

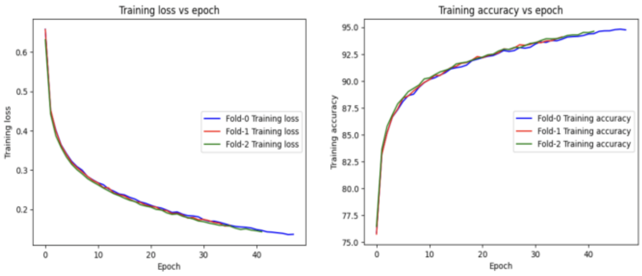

Fig. 6. Training Loss and Accuracy VS Epoch

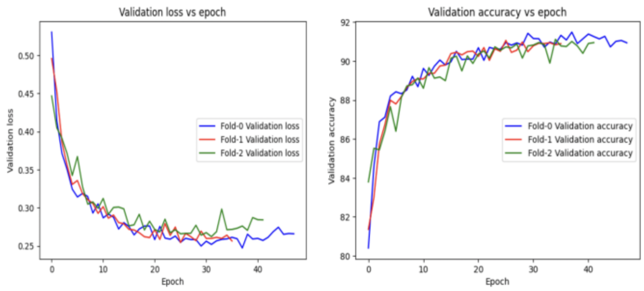

Fig. 7. Validation Loss and Accuracy VS Epoch

F. Hyper-parameter tuning of the CNN model.

The hyperparameter tuning process for our CNN model involved a systematic exploration of key parameters to optimize its performance. Specifically, we focused on tuning the learning rate, considering three candidate values—0.01, 0.001, and 0.0001—each representing different scales of adjustment during the model's training. Additionally, we explored the choice of optimizer, experimenting with both Adam (Adaptive Moment Estimation) and SGD (Stochastic Gradient Descent). The learning rate plays a crucial role in determining the step size during optimization, while the optimizer influences the efficiency of weight updates. Through a rigorous search across these parameter combinations, we sought to identify the optimal configuration that maximizes the CNN model's predictive accuracy and convergence speed. This fine-tuning process is integral to refining the model's performance, ensuring it adapts optimally to the intricacies of the dataset and generalizes well to unseen data.

G. Model Interpretation

Model interpretability is a critical aspect of leveraging deep learning models, and techniques such as Grad-CAM (Gradient-weighted Class Activation Mapping) and guided backpropagation provide valuable insights into the decision-making process of a Convolutional Neural Network (CNN) trained on the Fashion MNIST dataset. Grad-CAM illuminates the regions of an image that significantly contribute to the model's predictions by highlighting relevant feature maps. This visualization aids in understanding which parts of the input image are crucial for the network's classification decisions. On the other hand, guided backpropagation offers a fine-grained perspective by attributing importance to individual pixels through backpropagation of gradients. Combining Grad-CAM and guided backpropagation allows for a comprehensive interpretation of the CNN's decisions, shedding light on the specific features and patterns driving the model's classifications. This enhanced interpretability is crucial for validating model behavior, identifying potential biases, and building trust in the model's predictions, particularly in image-centric tasks like those involving fashion items in the MNIST dataset.

H. Business Insights

Convolutional Neural Network models can be used to identify photographs of clothing, which can yield insightful results and benefit your business in several ways. You may improve the personalization of your recommendations and potentially increase sales by getting to know the kind of clothing that a consumer is interested in. Examine the classifications to learn more about the most popular styles of clothes among consumers. By keeping more of the in-demand commodities stocked and reducing the extra inventory for the less popular items, you can use this knowledge to optimize your inventory. Cost reductions and increased operational effectiveness may result from this. Put tracking and monitoring systems in place to evaluate your CNN model's performance over time. To guarantee precise and pertinent classifications, update the model frequently considering fresh data and changing fashion trends. Your picture classification system's efficacy may be preserved by using this iterative method. Use the model to identify potentially fraudulent activities, such as false product representations. By analyzing images, you can detect discrepancies between the advertised products and the actual items, helping to maintain trust with your customers.

I. Results

In our experimental evaluation using the Fashion MNIST dataset, we assessed the performance of a diverse set of machine learning models, ranging from traditional methods such as k-Nearest Neighbors (KNN), Random Forest (RF), Decision Trees, and Gradient Boosting, to more complex architectures like Artificial Neural Networks (ANN) and Convolutional Neural Networks (CNN). The results revealed notable distinctions in the models' capabilities to recognize and classify diverse fashion items. Notably, within the realm of traditional models, the SVM classifier emerged as the most proficient, as measured by accuracy, closely followed by Random Forest. The ANN, leveraging its ability to capture complex patterns, showcased improved performance compared to traditional methods. Meanwhile, CNN, specifically designed for image-based tasks, outperformed all other models, highlighting its effectiveness in extracting intricate features and spatial hierarchies inherent in fashion images. Overall, the comprehensive evaluation provided valuable insights into the strengths and limitations of each model class in the context of the Fashion MNIST dataset. Fig 3 depicts performance of traditional ML models on validation dataset as well as on the test dataset.

Table 1. Machine Learning Models Results

Model	3-fold mean accuracy on val data (with PCA)	Accuracy on test data (with PCA)
Random Forest	86.31	85.73
Decision Tree	76.88	76.5
SVM (Linear)	84.66	83.62
KNN	85.07	85.12
Logistic Regression	82.82	81.57
GradientBoosting	84.38	83.55
SVC	87.3	86.68

After an exhaustive series of experimental trials, the Convolutional Neural Network (CNN) emerged as the most effective model, prompting a focused exploration into hyperparameter tuning to ascertain the potential for further performance enhancement. The optimal configuration was identified through the utilization of the SGD optimizer with a learning rate set at 0.01. Subsequently, a comprehensive quantitative assessment extended beyond accuracy to encompass key metrics such as precision, recall, and F1-score for each class label. Additionally, a detailed examination via a confusion matrix was undertaken to garner deeper insights into the model's performance across diverse fashion categories. This meticulous analysis serves to enrich our understanding of the CNN's efficacy and provides a nuanced evaluation framework beyond conventional accuracy metrics.

CLASS	PRECISION	RECALL	F1-Score	SUPPORT
T-Shirt	0.84	0.89	0.87	1000
Trouser	0.99	0.98	0.98	1000
Pullover	0.89	0.87	0.88	1000
Dress	0.93	0.9	0.92	1000
Coat	0.85	0.92	0.88	1000
Sandal	0.99	0.98	0.98	1000
Shirt	0.78	0.72	0.75	1000
Sneaker	0.96	0.98	0.97	1000
Bag	0.98	0.98	0.98	1000
Ankle Boot	0.98	0.97	0.97	1000
ACCURACY			0.92	10000
MACRO AVG	0.92	0.92	0.92	10000
WEIGHTED AVG	0.92	0.92	0.92	10000

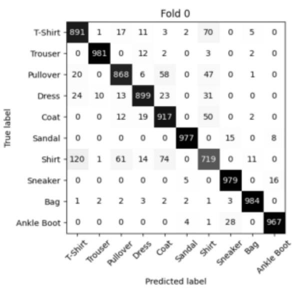

Fig. 8. CNN Fold 0 Results

CLASS	PRECISION	RECALL	F1-Score	SUPPORT
T-Shirt	0.87	0.89	0.88	100
Trouser	0.97	0.99	0.98	100
Pullover	0.84	0.91	0.87	100
Dress	0.91	0.92	0.91	100
Coat	0.87	0.86	0.87	100
Sandal	0.99	0.98	0.99	100
Shirt	0.8	0.72	0.76	100
Sneaker	0.97	0.96	0.97	100
Bag	0.99	0.97	0.98	100
Ankle Boot	0.95	0.98	0.97	100
ACCURACY			0.92	1000
MACRO AVG	0.92	0.92	0.92	1000
WEIGHTED AVG	0.92	0.92	0.92	1000

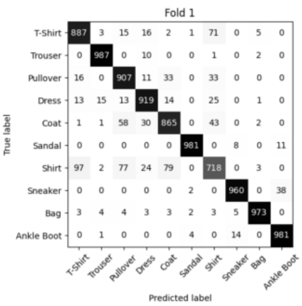

Fig. 9. CNN Fold 1 Results

CLASS	PRECISION	RECALL	F1-Score	SUPPORT
T-Shirt	0.87	0.89	0.88	100
Trouser	0.99	0.98	0.98	100
Pullover	0.87	0.89	0.88	100
Dress	0.91	0.94	0.93	100
Coat	0.87	0.89	0.88	100
Sandal	0.99	0.99	0.99	100
Shirt	0.8	0.73	0.76	100
Sneaker	0.97	0.98	0.97	100
Bag	0.98	0.98	0.98	100
Ankle Boot	0.98	0.97	0.97	100
ACCURACY			0.92	1000
MACRO AVG	0.92	0.92	0.92	1000
WEIGHTED AVG	0.92	0.92	0.92	1000

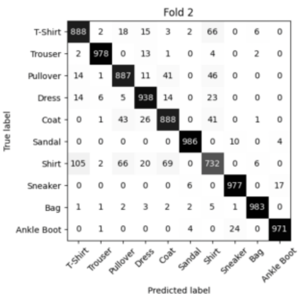

Fig. 10. CNN Fold 2 Results

V. Conclusion

In conclusion, our experiment with the Fashion MNIST dataset employing a spectrum of models spanning traditional machine learning techniques, Artificial Neural Networks (ANN), and Convolutional Neural Networks (CNN) has provided valuable insights into their respective capabilities for fashion item classification. Traditional models, including k-Nearest Neighbors (KNN) and Random Forest, demonstrated competitive performance, particularly the Support Vector Machine (SVM) classifier, which emerged as the top performer. The ANN, with its capacity to capture intricate patterns, exhibited improved accuracy compared to traditional methods, while the CNN, specifically tailored for image-based tasks, outperformed all other models, showcasing its effectiveness in discerning nuanced features inherent in fashion images. The superior performance of the CNN prompted a dedicated exploration of hyperparameter tuning, further enhancing its capabilities. This comprehensive evaluation underscores the nuanced strengths of each model class, providing a foundation for informed model selection in the context of fashion image classification tasks.

References

[1] Yang Li and Xuewei Chao, "ANN-Based Continual Classification in Agriculture," May 2020.

[2] Imran Iqbal, Gbenga Abiodun Odesanmi, Jianxiang Wang, and Li Liu, "Comparative Investigation of Learning Algorithms for Image Classification with Small Dataset," June 2021.

[3] Jagadeesh D. Pujari, Rajesh Yakkundimath, and Abdulmunaf Syedhusain Byadgi, "SVM and ANN Based Classification of Plant Diseases Using Feature Reduction Technique."

[4] Greeshma KV, Sreekumar K., "Fashion-MNIST classification based on HOG feature descriptor using SVM," January 2019.

[5] Vandana Kannan, "Feature Selection using Genetic Algorithms," February 2018.

[6] Yansheng Li, Xin Huang, and Hui Liu, "Unsupervised Deep Feature Learning for Urban Village Detection from High-Resolution Remote Sensing Images."

[7] Mingyuan Xin and Yong Wang, "Research on image classification model based on deep convolution neural network," 2019.

[8] Iqbal H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," 2021.

[9] Hyungkeuk Lee, NamKyung Lee, and Sungjin Lee, "A Method of Deep Learning Model Optimization for Image Classification on Edge Device," September 2022.

[10] Olivia Nocentini, Jaeseok Kim, Muhammad Zain Bashir, and Filippo Cavallo, "Image Classification Using Multiple Convolutional Neural Networks on the Fashion-MNIST Dataset," December 2022.

[11] Guoshu Zheng and Qiuyu Zhu, "Unsupervised Image Feature Extraction Based on Scattering Transform and Self-supervised Learning with Highly Training Efficiency," 2019.

[12] Han Xiao, Kashif Rasul, Rolland Vollgraf, "Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms," September 2017.

[13] M. Weber, M. Welling, P. Perona, "Unsupervised Learning of Models for Recognition," 2000

[14] FashionMNIST - https://github.com/zalandoresearch/fashion-mnist

[15] Code Implementation Github - https://github.com/N122RAHUL/ECEN-758-Data-Mining-and-Analysis-Project/tree/main