Preprints.org

Article

Not peer-reviewed version

# Lightweight CNN with Data Augmentation and Batch Normalization for Fashion-MNIST Classification

Tao Jingchu , Abdul Salam Shah , Aisha Farooq *

Posted Date: 28 May 2026

doi: 10.20944/preprints202605.1915.v1

Keywords: CNN; Fashion-MNIST; data augmentation; batch normalization; early stopping; lightweight architecture; dropout

Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC, OpenAlex.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

*Preprints.org* (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 May 2026 doi:10.20944/preprints202605.1915.v1

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

*Article*

# Lightweight CNN with Data Augmentation and Batch Normalization for Fashion-MNIST Classification

**Tao Jingchu, Abdul Salam Shah and Aisha Farooq ***

**School of Computer Science, Taylor's University, Subang Jaya, Malaysia**

* Correspondence: aishafarooq2323@gmail.com

**Abstract**

The given research paper is an end-to-end architecture of grayscale clothing image classification with a lightweight Convolutional Neural Network (CNN) with the Fashion-MNIST dataset. Its architecture consists of three convolutional layers with Batch normalization to stabilize training, Dropout to avoid overfitting, MaxPooling to reduce spatial, and data augmentation (random rotation, shifting, zooming, flipping) to increase the effective training set. Early Stopping callback was used to terminate training when the validation performance leveled off. The model obtained 88.63%. test accuracy, which indicates that a tailor-crafted lightweight CNN can be used to perform competitively on Fashion-MNIST without resorting to complex heavyweight architectures. The precision and F1-scores were high when it came to categories that had distinct visual characteristics (trousers, sandals, bags) and categories with similar textures and outlines (T-shirts, pullovers, coat) were likely to be misclassified. The paper also contextualizes these findings concerning the development of CNN architecture of LeNet-5 to AlexNet and VGGNet, and explains the implications of the results to the effective use of AI in resource-restricted settings.

**Keywords:** CNN; Fashion-MNIST; data augmentation; batch normalization; early stopping; lightweight architecture; dropout

---

## 1. Introduction

Image classification is a basic task of computer vision, which tries to categorize images into a set of predefined categories according to visual sensation such as shape, texture and color [1]. It is extensively used in medical image diagnosis, visual search in e-commerce and autonomous driving target recognition. Prior to deep learning, classification was based on manual designed features with HOG and Local Binary Patterns (LBP) which were not able to cope with complex backgrounds, large changes in size and complex human body postures [2–4]. The CNNs can learn end-to-end automatic features of raw pixel data by using multi-layer convolution and pooling operations [5,6]. Deep layers learn the high-level abstract features that are useful to achieve accurate classification whereas shallow layers learn the low-level features (edges, corners, textures) of the image. Nevertheless, such deep architectures as AlexNet and VGGNet do not scale to medium-size grayscale datasets such as Fashion-MNIST because of the number of parameters and computation costs [7]. The study develops a simple custom CNN with Batch Normalization and data augmentation that is tailored to the properties of Fashion-MNIST. Figure 2 shows a cloud–edge AI architecture that allows scalable and real-time fashion image classification.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 May 2026 doi:10.20944/preprints202605.1915.v1



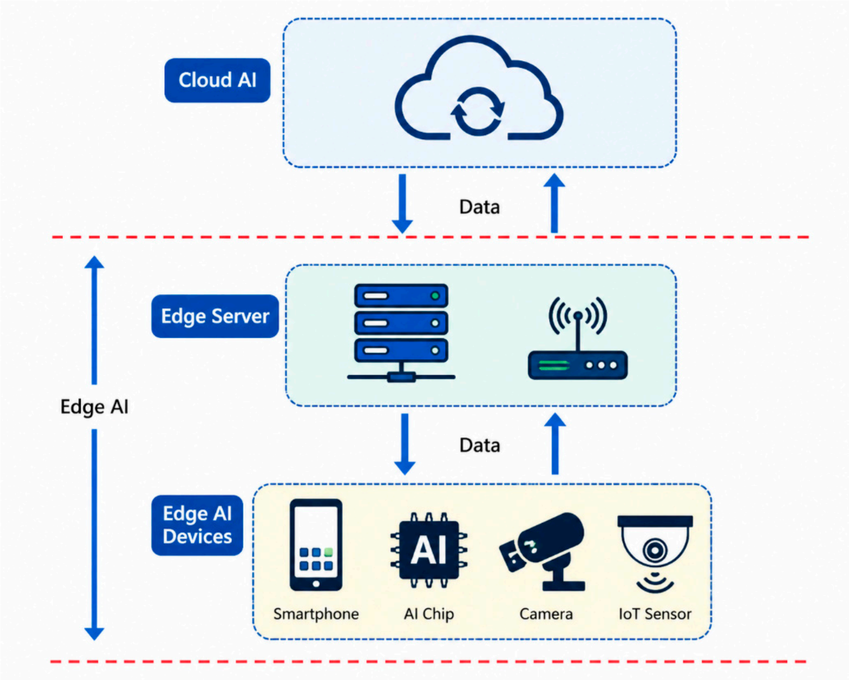


**Figure 1.** Cloud–edge AI design for fashion image classification

*1.1. Challenges in Image Classification*

Even with all the advancements that have taken place in image classification design there are several challenges like variations in lighting conditions, object orientation, occlusion, and intra-class similarity. These all can negatively impact model performance[8].

In datasets like Fashion-MNIST here images are grayscale and low-resolution (28×28 pixels). This helps distinguish between visually similar classes. Such as shirts, pullovers, and coats becomes particularly difficult also overfitting remains a concern when models are too complex relative to dataset size [9].

## 2. Related Work

The classical approaches (SIFT, HOG + SVM/k-NN) reached 7075 percent accuracy on Fashion-MNIST [10]. CNN CNN_5 defined the backbone of workflow in CNN containing approximately 60,000 parameters. AlexNet (60 million parameters) introduced ReLU, training on the GPU, and Dropout but is too big to be deployed in a lightweight system. VGGNet (138+ million parameters) was more accurate with 3x3 kernels, but it involves enormous computations. This paper develops a balanced architecture that is appropriate to Fashion-MNIST on the medium scale.

**Table 1.** CNN Architecture Comparison.

| Architecture | Key Feature | Parameters | Suitable For |
|---|---|---|---|
| LeNet-5 | CNN core workflow | 60K | Small, simple images |
| AlexNet | ReLU, GPU, Dropout | 60M | Large-scale complex classification |
| VGGNet | 3×3 small kernels | 138M+ | High-precision, transfer learning |
| This Study | BN + Dropout + Augmentation | Lightweight | Medium-size grayscale datasets |

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 May 2026 doi:10.20944/preprints202605.1915.v1



### 2.1. Deep Learning for Image Classification

Deep learning help convert image classification through enabling models to by design learn hierarchical feature representations from raw data [11]. Convolutional Neural Networks have become rapidly used mainly due to their ability to capture spatial dependencies and patterns in images. Nonetheless traditional machine learning approaches that rely on handcrafted features, CNNs perform end-to-end learning, resulting in improved accuracy and strength [12]. Recent studies present that CNN-based models reliably outperform classical methods across various domains. This includes medical imaging, remote sensing, and fashion analysis [13].

### 2.2. CNN Architectures for Fashion Image Classification

Yet this faces unique challenges due to subtle differences between categories and low-resolution datasets such as Fashion-MNIST. Several studies showcase that CNN architectures can be used to address these challenges. [14,15] recommends a deep learning-based classification model for apparel images and achieving high accuracy through optimized convolutional layers. Equally [16] presented a hierarchical CNN architecture that increases classification performance by capturing multi-level features. These approaches prove that CNNs can proficiently distinguish between visually similar clothing items. Although performance is still constrained by dataset restrictions [17–19].

## 3. Methodology

Fashion-MNIST (70,000 images, 28x28, 10 classes) was scaled to [0, 1] by Min-Max scaling and divided into train/validation/test sets and one-hot coded. Data augmentation was implemented with random rotation, width/height changes, zoom and horizontal flips. The CNN consists of three convolutional layers, BatchNormalization and Dropout and MaxPooling, then Flatten, Dense layers and Softmax output. Early Stopping observed loss in validation to avoid unwarranted training [21].

### 3.1. Data Preprocessing

To guarantee efficient model training, pixel values of the images were normalized using Min-Max scaling to a range of [0, 1]. This normalization helps in stabilizing gradient updates and accelerating convergence during training. The image data was reshaped to include a channel dimension (28×28×1) to make it compatible with CNN input requirements.

Categorical labels were transformed into one-hot encoded vectors, allowing the model to treat the classification task as a multi-class problem with probabilistic outputs [22].

### 3.2. Data Augmentation

For generalization and to reduce overfitting we use data augmentation techniques to artificially expand the training dataset [23].

The augmentation plan comprises of:

1. We designed random rotations to simulate different orientations
2. Applied width and height shifts to account for positional variations
3. Zoom transformations to capture scale differences
4. Horizontal flipping to introduce mirror variations

These changes enable the model to learn stable features and improve robustness when encountering unseen data [24].

## 4. Results and Discussion

The model was able to test at 88.63 percent accuracy. Trousers, sandals and bags had high precision and F1-scores because they possess unique visual characteristics [25]. T-shirts, pullovers, and coats, which had such similar textures and outlines, were likely to be confused, especially in noisy low-resolution grayscale images. This is indicative of the inadequate power of the model to

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 May 2026 doi:10.20944/preprints202605.1915.v1



identify slightly different clothes instead of data imbalance (Fashion-MNIST is balanced across classes).

**Table 2.** Test Performance Summary.

| Metric | Value |
|---|---|
| Test Accuracy | 88.63% |
| High-Performance Classes | Trouser, Sandal, Bag |
| Challenging Classes | Shirt, Pullover, Coat |

### *4.1. Class-wise Performance Analysis*

A comprehensive evaluation of class-wise performance showcase us that certain categories achieved better higher precision and F1-scores than others. Classes like trouser, sandal, and bag feature strong classification performance due to their distinct shapes and structural features, namely trousers have a unique vertical symmetry whereas bags and sandals possess clearly distinguishable outlines compared to other clothing items [26–28].

On the other hand, classes including shirt, pullover, and coat are more puzzling for the model. These classifications share similar textures, silhouettes, and pixel distributions. This makes it difficult for the CNN to extract sufficiently discriminative features. For this reason, misclassifications commonly occurred among these classes [29–31].

### *4.2. Misclassification Patterns*

The detected misclassification trends present a key limitation of low-resolution grayscale datasets as images in Fashion-MNIST lack color information and fine-grained texture details, the model relies heavily on shape-based features [32]. Results in confusion among visually similar upper-body garments. Especially when subtle differences in sleeve length or fabric type are not easily captured [33–35].

## 5. Future Research and AI Connections

Greater resolution color data (DeepFashion), attention on discriminative garment features, transfer learning using ImageNet-trained lightweight networks (MobileNet, EfficientNet) and multi-modal learning that integrates visual and textual clothes information. To be deployed effectively, model quantization and pruning might facilitate real-time inference in mobile and edge devices in smart retail settings.

### *5.1. Attention Mechanisms for Feature Discrimination*

One encouraging way is the integration of attention mechanisms into CNN architectures. Attention modules allow models to focus on the most discriminative regions of an image, such as sleeves, collars, or textures. These are critical in order to distinguish among similar clothing items and can enhance classification accuracy by reducing confusion between classes with overlapping visual characteristics.

### *5.2. Transfer Learning with Lightweight Architectures*

Transfer learning using pre-trained lightweight models such as MobileNet and EfficientNet offers another avenue for improvement. These architectures are trained on large-scale datasets like ImageNet, can offer robust feature extractors that generalize well to smaller datasets. Fine-tuning such models can considerably increase performance while retaining computational efficiency.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 May 2026 doi:10.20944/preprints202605.1915.v1



## 6. Conclusions

The CNN with data augmentation and Batch Normalization (lightweight) was tested on Fashion-MNIST with 88.63% accuracy, which proves that the strategies of multi-regularization can offset the architectural simplicity. The model is also effective in capturing hierarchical characteristics and also generalization to unobservable data. The main weakness, namely confusion of similarly-textured clothes in grayscale is a factor of dataset shortcomings that can be countered with better quality input and advanced architectures.

This research showcases that the practical relevance of lightweight deep learning models in image classification task mainly in scenarios with limited computational resources also highlights the significance of selecting proper preprocessing and regularization techniques to maximize model effectiveness. Future research should focus on exploring more complex datasets and improved feature representation methods. Also work on hybrid architectures to further enhance performance and address present limitations.

## References


1. Csurka, G., et al. (2004). Visual categorization with bags of keypoints. ECCV Workshop.
2. Dalal, N., & Triggs, B. (2005). HOG for human detection. CVPR, 886–893.
3. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.
4. He, K., et al. (2016). Deep residual learning. CVPR, 770–778. https://doi.org/10.48550/arXiv.1512.03385
5. Krizhevsky, A., et al. (2012). ImageNet classification with deep CNNs. NIPS, 25, 1097–1105.
6. LeCun, Y., et al. (1998). Gradient-based learning. Proceedings of the IEEE, 86(11), 2278–2324. doi: 10.1109/5.726791
7. Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. IJCV, 60(2), 91–110. https://doi.org/10.1023/B:VISI.0000029664.99615.94
8. Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks. ICLR. https://doi.org/10.48550/arXiv.1409.1556
9. Aldughayfiq, B., Ashfaq, F., Jhanjhi, N. Z., & Humayun, M. (2023). A Deep Learning Approach for Atrial Fibrillation Classification Using Multi-Feature Time Series Data from ECG and PPG. *Diagnostics*, *13*(14), 2442. https://doi.org/10.3390/diagnostics13142442
10. Srivastava, N., et al. (2014). Dropout: A simple way to prevent neural networks from overfitting. JMLR, 15, 1929–1958.
11. Xiao, H., et al. (2017). Fashion-MNIST. arXiv:1708.07747.
12. Zhang, J., et al. (2018). Review of traditional feature extraction methods. JVCIR, 55, 123–136. DOI:10.1088/1742-6596/1591/1/012028
13. Shin, S. Y., Jo, G., & Wang, G. (2023). A novel method for fashion clothing image classification based on deep learning. Journal of Information and Communication Technology. DOI:10.32890/jict2023.22.1.6
14. Fatima-tuz-Zahra, N. Jhanjhi, S. N. Brohi, N. A. Malik and M. Humayun, "Proposing a Hybrid RPL Protocol for Rank and Wormhole Attack Mitigation using Machine Learning," *2020 2nd International Conference on Computer and Information Sciences (ICCIS)*, Sakaka, Saudi Arabia, 2020, pp. 1-6, doi: 10.1109/ICCIS49240.2020.9257607.
15. Chamnong, N., Werapun, J., et al. (2023). Hierarchical convolutional neural networks using CCP-3 block architecture for apparel image classification. International Journal of Clothing Science and Technology. DOI:10.14569/IJACSA.2023.0140624
16. Papadopoulos, S. I., Koutlis, C., Sudheer, M., et al. (2022). Attentive hierarchical label sharing for enhanced garment and attribute classification of fashion imagery. In Artificial Intelligence Systems in Fashion and Textiles. Springer. https://doi.org/10.1007/978-3-030-94016-4_7
17. Samia, B., Soraya, Z., & Malika, M. (2022). Fashion images classification using machine learning, deep learning and transfer learning models. IEEE Conference on Image and Signal Processing. doi: 10.1109/ISPA54004.2022.9786364.


© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 May 2026 doi:10.20944/preprints202605.1915.v1




18. Amin, M. S., Wang, C., & Jabeen, S. (2023). Fashion sub-categories and attributes prediction model using deep learning. The Visual Computer. Springer. https://doi.org/10.1007/s00371-022-02520-3
19. Wen, X., Yu, X., Wang, Y., Yang, C., & Sun, Y. (2023). A hybrid 3D–2D feature hierarchy CNN with focal loss for hyperspectral image classification. Remote Sensing. https://doi.org/10.3390/rs15184439
20. Muzammal, S. M., Murugesan, R. K., Jhanjhi, N. Z., Humayun, M., Ibrahim, A. O., & Abdelmaboud, A. (2022). A Trust-Based Model for Secure Routing against RPL Attacks in Internet of Things. *Sensors*, *22*(18), 7052. https://doi.org/10.3390/s22187052
21. Ray, A., & Kolekar, M. H. (2022). Image segmentation and classification using deep learning. International Journal for Signal and Image Processing. Wiley. doi: 10.1002/9781119861850.ch2.
22. Zhou, Z., Liu, M., Deng, W., & Wang, Y. (2022). Clothing image classification algorithm based on convolutional neural network and optimized regularized extreme learning machine. Textile Research Journal. DOI:10.1177/00405175221115472
23. Xhaferra, E., Cina, E., & Toti, L. (2022). Classification of standard fashion MNIST dataset using deep learning-based CNN algorithms. In International Symposium on Electronics and Telecommunications. IEEE. doi: 10.1109/ISMSIT56059.2022.9932737.
24. Jhanjhi, N.Z. (2025). Investigating the Influence of Loss Functions on the Performance and Interpretability of Machine Learning Models. In: Pal, S., Rocha, Á. (eds) Proceedings of 4th International Conference on Mathematical Modeling and Computational Science. ICMMCS 2025. Lecture Notes in Networks and Systems, vol 1399. Springer, Cham. https://doi.org/10.1007/978-3-031-91005-0_43
25. Madhavi, B., Mahanty, M., Lin, C. C., & Lakshmi Jagan, B. O. (2025). SwinConvNeXt: A fused deep learning architecture for real-time garbage image classification. Scientific Reports. DOI:10.1038/s41598-025-91302-7
26. Dakshina, D. S., Jayapriya, P., & Kala, R. (2023). Saree texture analysis and classification via deep learning framework. International Journal of Data Science and Artificial Intelligence.
27. Zhang, M., Wang, Z., Wang, X., Gong, M., Wu, Y., & Li, H. (2023). Feature-preserving generative adversarial network data augmentation strategy for hyperspectral image classification. Pattern Recognition. DOI:10.1016/j.patcog.2023.109701
28. Gupta, J., Pathak, S., & Kumar, G. (2022). Deep learning (CNN) and transfer learning: A review. Journal of Physics: Conference Series. DOI:10.1088/1742-6596/2273/1/012029
29. Nocentini, O., Kim, J., Bashir, M. Z., & Cavallo, F. (2022). Image classification using multiple convolutional neural networks on the Fashion-MNIST dataset. Sensors. https://doi.org/10.3390/s22239544
30. Saeed, S., Abdullah, A., Jhanjhi, N.Z. et al. New techniques for efficiently k-NN algorithm for brain tumor detection. *Multimed Tools Appl* **81**, 18595–18616 (2022). https://doi.org/10.1007/s11042-022-12271-x
31. Absur, M. N., Nasif, K. F. A., Saha, S., et al. (2024). Revolutionizing image recognition: Next-generation CNN architectures for handwritten digits and objects. In IEEE Symposium Proceedings. doi: 10.1109/ISWTA62130.2024.10651815.
32. Mei, S., Song, C., Ma, M., & Xu, F. (2022). Hyperspectral image classification using group-aware hierarchical transformer. IEEE Transactions on Geoscience and Remote Sensing. DOI:10.1109/TGRS.2022.3207933
33. Imtiaz, A., Pathirana, N., Saheel, S., & Khan, M. (2024). A review on the influence of deep learning and generative AI in the fashion industry. Journal of Future Technologies. https://doi.org/10.62411/faith.3048-3719-29
34. Bbouzidi, S., Hcini, G., Jdey, I., & Drira, F. (2024). Convolutional neural networks and vision transformers for fashion mnist classification: A literature review. arXiv preprint arXiv:2406.03478. https://doi.org/10.48550/arXiv.2406.03478
35. Khan, A., Jhanjhi, N. Z., Haji, D. H. T. B. A., & Omar, H. A. H. B. H. (2024). Internet of Things (IoT) impact on inventory management: A review. Cybersecurity measures for logistics industry framework, 224-247. DOI: 10.4018/978-1-6684-7625-3.ch008


**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.