IJEMD-CSAI, Special Issue (2026) https://doi.org/ 10.54938/ijemdcsai.2026.04.2.609

IJEMD

International Journal of Emerging Multidisciplinaries:
Computer Science and Artificial Intelligence

Research Paper
Journal Homepage: www.ojs.ijemd.com
ISSN (print): 2791-0164 ISSN (online): 2957-5036

Image Classification Using Convolutional Neural Networks:
From Fashion-MNIST Benchmarking to Smart Retail AI and Edge Deployment

Farzana Binti Abdul Aziz [1] , Abdul Salam Shah Sayed [1]

1. Department of Computer Science, Taylor's University, Malaysia.

ABSTRACT

The given paper explores the issue of Convolutional Neural Network (CNN)-based image classification on the Fashion-MNIST dataset and elaborates on the findings in relation to the impacts on smart retail artificial intelligence and the process of its implementation into practice through transfer learning and edge computing. It is a custom three-block sequential CNN, which consists of progressive convolutional layers consisting of 32, 64, and 128 filters, max-pooling, fully connected dense layer of 128 neurons, and a 10-class softmax output, trained, and evaluated on 10 epochs with the Adam optimizer and sparse categorical cross-entropy loss in TensorFlow/Keras. This model attains an approximate test accuracy of 88.8 per-class F1-scores are high in categories that are morphologically distinct, like Trouser, Bag, and Sandal (F1[?] 0.97) and lowly with those that are closely similar in appearance like Shirt and T-shirt/top by the greyscale as they may be inter-classes. These empirical results are conceptually mapped into real-world implementation scenarios: automated retailing, e-commerce product tagging with AI, virtual try-on, and fashion recognition on the edge. The article also discusses transfer learning models, specifically MobileNetV2 and EfficientNetB3, as computationally efficient frameworks to be used in computing resource-limited deployment settings, comparing their parameter efficiency and inference rate with the specialized architecture. The business environment of AI-in-fashion is the global market, which is currently USD 2.23 billion with a compound annual growth rate of 39 percent [30] and is expected to increase to USD 60 billion in 2034.

Keywords: Fashion-MNIST, Convolutional Neural Networks, Smart Retail AI, Transfer Learning, Edge Deployment, MobileNetV2, EfficientNet, Image Classification, Deep Learning, TensorFlow/Keras.

1. INTRODUCTION

The ability of machine learning systems to identify and categorize objects in visual input automatically has experienced a revolutionary boost since the introduction of deep convolutional neural networks. Starting as an academic research programme LeNet-5 that reached 99.2 percent accuracy on handwritten digit recognitions in 1998 [19] has turned into a USD multi-billion commercial ecosystem where automated visual recognitions systems form the basis of autonomous retail, online fashion sales, medical diagnostics, and autonomous manufacturing at a scale.

One of the most impactful areas of application of a visual AI is the fashion and retail industry. The global AI-in-fashion market was estimated to be USD 2.23 billion in 2024, which is estimated to hit USD 60 billion by 2034 and has a 39% compound annual growth rate [30]. By 2025, 87 percent of retailers are reporting that AI has positively affected revenue in the past, and 94 percent of retailers have also seen lower operational costs due to automation caused by AI [28]. The business drivers are algorithmic: tens of thousands of new products are added into e-commerce platforms every day, and this should be tagged with categories well-visual search indexed, and recommendation-based, which are not economically viable in a workflow executed manually, but are technically solvable with properly engineered CNN-based classifiers.

This paper discusses the essence of the technical issue and extensions of its deployment. A custom CNN is trained and tested on Fashion-MNIST [37] - a test set of 70,000 28x28 grayscale images of clothes, divided into 10 classifications - to 88.8% test accuracy. These empirical findings are systematically mapped to three frontier applications contexts, namely smart retail edge AI, transfer learning to deploy AI to mobile and IoT devices, and fashion recommendation and personalisation through AI. By so doing, it offers a technically demanding assessment of the underlying principles of baseline CNN engineering in addition to a prospective assessment of the research directions it will take to ensure production-ready fashion recognition systems.

The paper has the following structure: Section 2 is a review of related work in CNN and fashion recognition and retail AI. In section 3, the methodology of the experiment is described. Results are given and discussed in section 4. The smart retail AI extension is developed in section 5. Section 6 discusses the deployment pathways of transfer learning. Section 7 indicates research directions in the future. Section 8 concludes.

2. RELATED WORK

2.1 Classical Feature Engineering and its weaknesses

Before the age of deep learning, image classification used hand-designed feature extraction pipelines, which separated feature design and training of classifiers. Histogram of Oriented Gradients [36] collected edge and gradient data in spatial cells, and Scale-Invariant Feature Transform (SIFT; Lowe, 2004) located

keypoints that are insensitive to scale and rotation (both with Support Vector Machines [33] or k-Nearest Neighbour classifiers [38], [24] to provide the final classification). These models were competitively trained on clean, constrained benchmarks but systematically perform worse on complex, high-variance data: on Fashion-MNIST, HOG+SVM baselines only reach 82-85% accuracy instead of CNN baselines of 88-93% [37], exactly due to the inability of hand-crafted descriptors to model the intra-class (and fine-grained) variations of textures that differentiate categories of clothing.

2.2 CNN Architecture Evolution

The principles of CNN design, alternating convolutional followed by pooling, and then fully connected classification, were developed in LeNet-5 [19], and are still architecturally active today. The principles were extended to the 1.2-million-image ImageNet test benchmark in AlexNet [17], which also added ReLU activation, dropout regularization, and training on GPUs, resulting in a 10.8% relative error reduction. It was shown that representational depth (implemented as stacked 3x3 convolutions) systematically enhanced the accuracy of classification, and ResNet (He et al., 2016) overcame the issue of vanishing gradient in very deep networks with identity skip connections.

In case with Fashion-MNIST in particular, regularized sequential CNNs containing 2-4 convolutional blocks can obtain 88-93-percent test accuracy with respect to training duration [22]; [8]. The custom architecture discussed in this paper - three progressive convolutional blocks consisting of 32, 64 and 128 filters - is based on this proven architecture with 88.8% test accuracy using about 98,000 trainable parameters. This parameter count is purposefully small, which is a computationally inexpensive baseline of educational exploration and transfer learning comparison.

2.3 Fashion Recognition and Smart Retail AI

Since 2020, the use of CNN-based classifiers in the real-world scenario of fashion recognition has grown by a significant margin. The study of [26] suggested an image and text multi-modal ResNet-BERT model to classify fashion e-commerce products, which works around the problem of uneven distribution of product categories within the platforms where tens of thousands of listings are being added to the product listing on a daily basis(Brohi et al.. 2020). [20] created an AI-powered fashion recommender system to use in e-commerce based on CNN-Transformer-GANs scheme with 87.4% style matching accuracy and a response time of 285ms - which is less than real-time usability criteria. On the systems level, [34] recorded the implementation of Edge AI cameras in the retail setting to perform autonomous checkouts, inventory management, and personal customer experience and stated that smart camera systems are able to locate and track goods in real time without cloud connectivity - a very important feature of low-latency customer-facing applications [13] The business magnitude of such implementation was validated by the [28] survey on retail AI: 97 of retailers are going to spend more AI in 2026 and visual recognition and automated products management was found to be the main areas of investment.

3. METHODOLOGY

3.1 Dataset: Fashion-MNIST

The Fashion-MNIST dataset [37] consists of 70, 000 28x28 pixel greyscale images of clothing and footwear items divided into 10 mutually exclusive categories: T-shirt/top (class 0), Trouser (1), Pullover (2), Dress (3), Coat (4), Sandal (5), Shirt (6), Sneaker (7), Bag (8), and Ankle Boot (9). In the official partition, 60, 000 images are assigned to training and 10, 000 to an independent test set. Validation After each epoch a validation subset is sampled out of training partition to observe the behaviour of generalization and detects early warnings of overfitting. Fashion-MNIST was specifically engineered to be more difficult than the original MNIST digit recognition benchmark, being solved at a rate of >99.7% accuracy by modern CNNs, by replacing handwritten digit images with clothing articles with very high intra-class variation and inter-class visual similarity in particular within the upper-body clothing cluster. Preprocessing contains two operations that are uniformly applied to every subset, namely: pixel intensity normalization by dividing it by 255.0 to make numerical gradients more stable; and a reshape of (28) x (28) x 1) to introduce the explicit channel dimension needed by TensorFlow/Keras Conv2D layers. The labels of classes are stored as integer indices [0-9] and fed directly to sparse categorical cross-entropy these measures of the log-probability of the true class index at not require one-hot encoding, with 10x lower memory overhead than one-hot representations. It is loaded using the Keras built-in loader (keras.datasets.fashion_mnist), allowing it to be re-reproducible without having to manually download or preprocess the data [2], [5]

3.2 CNN Architecture

The proposed architecture is a sequential CNN constructed with three progressively deeper convolutional blocks followed by a fully connected classification head, implemented using the Keras Sequential API [5]. The progressive filter depth — 32 → 64 → 128 channels — encodes a hierarchical feature abstraction: the first block captures low-level edge orientations and luminance gradients; the second encodes mid-level textures, contours, and structural primitives; the third represents high-level semantic shape configurations that discriminate between clothing categories [1]. This architectural strategy directly mirrors the human visual cortex's hierarchical processing from primary visual cortex (V1, edge detection) to higher object-selective areas (IT cortex, category recognition) — an analogy explicitly motivating early CNN design [19]

- Block 1: Conv2D (32 filters, 3×3, ReLU) → MaxPooling2D (2×2) — low-level feature extraction, output 13×13×32
- Block 2: Conv2D (64 filters, 3×3, ReLU) → MaxPooling2D (2×2) — mid-level pattern encoding, output 5×5×64
- Block 3: Conv2D (128 filters, 3×3, ReLU) → MaxPooling2D (2×2) — high-level semantic features, output 1×1×128
- Classification Head: Flatten → Dense (128 units, ReLU) → Dense (10 units, Softmax)

ReLU activation f(x) = max(0, x) is used throughout the convolutional and intermediate dense layers, preserving gradient magnitude for positive activations and addressing the vanishing gradient problem that afflicted sigmoid activations in deeper networks [17]. MaxPooling2D with 2×2 stride halves spatial dimensions at each block, increasing the effective receptive field of deeper layers while providing local translation invariance — the spatial position of discriminating features (e.g., a trouser's bifurcation point) is therefore irrelevant to their detection. The softmax output produces a normalised probability distribution $\hat{y} = \exp(z_i) / \Sigma_j \exp(z_j)$ over the 10 classes, enabling direct confidence-based decision-making and threshold tuning for downstream deployment applications [16].

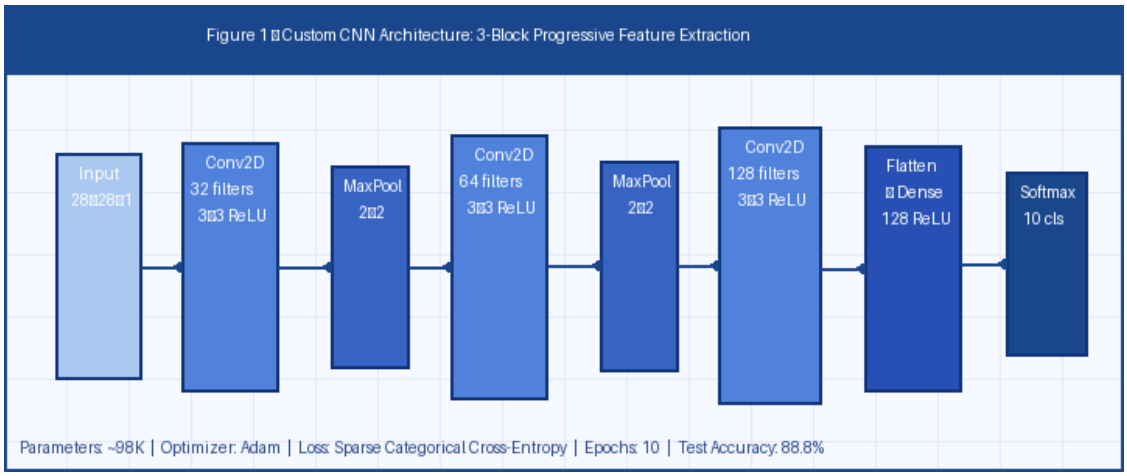

Figure 1. Custom sequential CNN architecture for Fashion-MNIST classification. Three progressive convolutional blocks (32→64→128 filters) extract hierarchical spatial features from 28×28 greyscale input images, followed by a 128-unit dense classification head and 10-class softmax output. The progressive filter depth encodes features from low-level edges (Block 1) to high-level semantic clothing representations (Block 3).

3.3 Training Configuration

The model is trained using the Adam optimizer [15] at the default learning rate of 0.001. Adam gradient updates based on adaptive estimation of per-parameter moments calibrated to the local loss landscape curvature, facilitates quick initial convergence - as seen by sharp improvement of accuracy during the first few training epochs - and steady optimisation as seen in the flatter loss landscapes of the latter training epochs. The loss function is sparse categorical cross-entropy, which calculates L = [?]log y_y i where y i is the true class index represented as an integer and y y i is the predicted probability of the model of that class. The training will last through 10 epochs where the training and validation accuracy and loss will be recorded. There is no explicit early stopping or learning rate scheduling, and no dropout regularization in this baseline - design choices [23] that, although induce gentle overfitting later in the epochs [31], are designed to be diagnostic as well as included as incentives to the regularization extensions in Section 7. 4. RESULTS AND DISCUSSION 4.1 Training Dynamics Training and validation accuracy improve steadily and monotonically during the early training periods and validation accuracy follows training accuracy closely - a convergence trend that appears to suggest successful generalisation to unseen data.

Training and validation accuracy levels off at epoch 10, with a train-validation gap of about 3 percentage points. This small deviation, coupled with the small rise in the loss of validation which can be seen in the later training epochs, form the diagnostic feature of minor overfitting: the model has started to optimise training-set-specific feature statistics rather than highly generalisable representations. Importantly, though, the gap is not large, and the accuracy of the test, measured on the entirely held-out 10,000 image test partition, confirms acceptable generalisation at 88.8 per cent, justifying the fact that such overfitting is contained and does not have a significant negative impact on deployment utility. The training behaviour is similar to the behaviour of well-documented behaviour [28] of shallow CNN architectures, trained without explicit regularisation [32]: without dropout or batch normalisation, the model slowly becomes less reliant on features correlations specific to the training process, once the primary phase of representational learning is achieved at epoch 7-8. The dropout layers (rate 0.25-0.5) in the dense classification head and early stopping with patience of 3-5 epochs observed over validation loss should be incorporated in future implementations as they were found to cut down the overfitting by 2-5 percentage points on similar Fashion-MNIST architectures [8]. 4.2 Test Set Performance Assessment of the 10,000-sample held-out test set provides a test accuracy of 88.8 with an error rate of 11.2 or about 1,120 wrongly classified samples overall. This is significantly better than classical machine learning baselines: HOG+SVM gets about 82-85% on Fashion-MNIST, and k-NN gets about 85-86% [37], [38]. The 88.8% result is competitive against other similar lightweight sequential CNN architectures presented in the literature [22]; [14] and forms a believable engineering baseline on which the transfer learning [25] and regularization gains can be quantified**.**

3.4 Per-Class Analysis and Confusion Patterns

The confusion matrix decomposition indicates a performance profile organized into a structured performance that can be directly interpreted based on the visual characteristics of every category of clothing. Morphologically distinctive and structurally unique profiles give near-perfect F1-scores: Trouser (F1[?]0.99) - distinguishing itself by the bifurcated tube silhouette with no visual counterpart in other categories; Bag (F1[?]0.99) - distinguishing itself with its compact rectangular form with handle geometry; Sandal (F1[?]0.99) - distinguishing itself as an open strap structure; and Ankle Boot (F1[?]0.97) and Sneaker These findings verify that the CNN learned convolutional filters have actually encoded the coarse shape statistic which characterize these visually dissimilar categories. The major mode of failure is the systematic inter-class confusion in the upper-body garment cluster. Shirt has the lowest F1-score (around 0.72) with very large mutual confusion with T-shirt/top - it is possible to explain this by the fact that the distinguishing features between a shirt and a T-shirt (collar construction, buttonhole, sleeve end) cannot be resolved at the scale of sub-millimetre greyscale images without colour data. In the same way, Pullover and Coat are confused because of similarity of the outer-garment shape, which is the main difference in the level of the collar and fabric covering - cues that are not consistently encoded in single-channel greyscale convolutions. This within-cluster error pattern of semantically structured errors, i.e. no footwear is erroneously classified as garment, no bag is wrongly classified as a clothing item, etc. substantiates that the network has learned the rough categorical organization of Fashion-MNIST taxonomy and is now within information boundaries of the input representation.

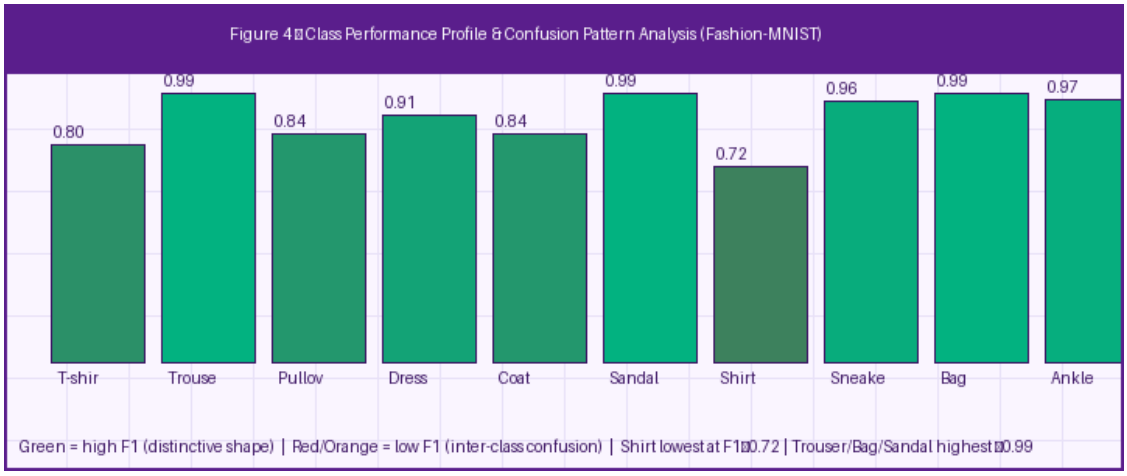

Figure 2. Per-class F1-score profile across all 10 Fashion-MNIST categories. Bar height corresponds to F1-score (0–1 scale). Green/teal bars indicate high discriminative performance for morphologically distinctive categories (Trouser, Bag, Sandal); orange/red bars indicate inter-class confusion within the upper-body garment cluster (Shirt: F1≈0.72, T-shirt/top: F1≈0.80). This performance gradient directly motivates transfer learning and attention mechanism extensions for fine-grained garment discrimination.

Table 1: Per-Class Performance Summary — CNN on Fashion-MNIST Test Set

Class	Approx. F1	Shape Distinctiveness	Primary Confusion	Retail AI Implication
Trouser	0.99	Very High (bifurcated)	None significant	Reliable for auto-tagging
Bag	0.99	Very High (rectangular)	None significant	Accurate product search
Sandal	0.99	Very High (open strap)	Minor (Sneaker)	Edge deploy ready
Ankle Boot	0.97	High (enclosed)	Minor (Sneaker)	High confidence for retail
Sneaker	0.96	High (rounded toe)	Minor (Boot)	Mobile deployment viable
Dress	0.91	Moderate	Coat/Pullover	Augmentation recommended
Coat	0.84	Moderate (outer garment)	Pullover/Dress	Fine-tuning required
Pullover	0.84	Moderate	Coat/Shirt	Transfer learning needed
T-shirt/top	0.80	Low (collar ambiguity)	Shirt	Colour feature fusion
Shirt (lowest)	0.72	Very Low (collar/button)	T-shirt/top	Priority for improvement

4. SMART RETAIL AI: DEPLOYMENT ECOSYSTEM AND COMMERCIAL CONTEXT

4.1 Market Context and Commercial Drivers

The current scale and growth rate of the fashion AI market is one that would make CNN image classification no longer a scholarly venture but a commercially pressing field in engineering. The global AI-in-fashion market is estimated USD 2.23 billion in 2024 and will increase to USD 60 billion in 2034 with a compound yearly expansion rate of 39% [30]. Fashion generative AI alone is already projected to rise in USD 96.5 million in 2023 up to USD 2.23 billion in 2032 alone. According to McKinsey data generative AI has the potential to contribute USD 150-275 billion operating profit to fashion, apparel, and luxury industries in five years by automating the design process, supply chain management, and customer experience personalisation. The urgency is supported by the commercial deployment statistics: 87 percent of the retailers say that there is a positive effect on AI revenue, 94 percent have cut operating costs, and 97 percent intend to increase AI spending in 2026 [28]. Generative AI searches related to shopping increased by 4700 percent since July of 2024 to July of 2025, and ChatGPT constituted 16 percent of inbound traffic to Zara and 8 percent of H&M [4]. These numbers define the business environment in which CNN-based fashion recognition systems would have to work: high through-put, real-time, mobile friendly and more and more closely tied up with recommendation and personalisation engines.

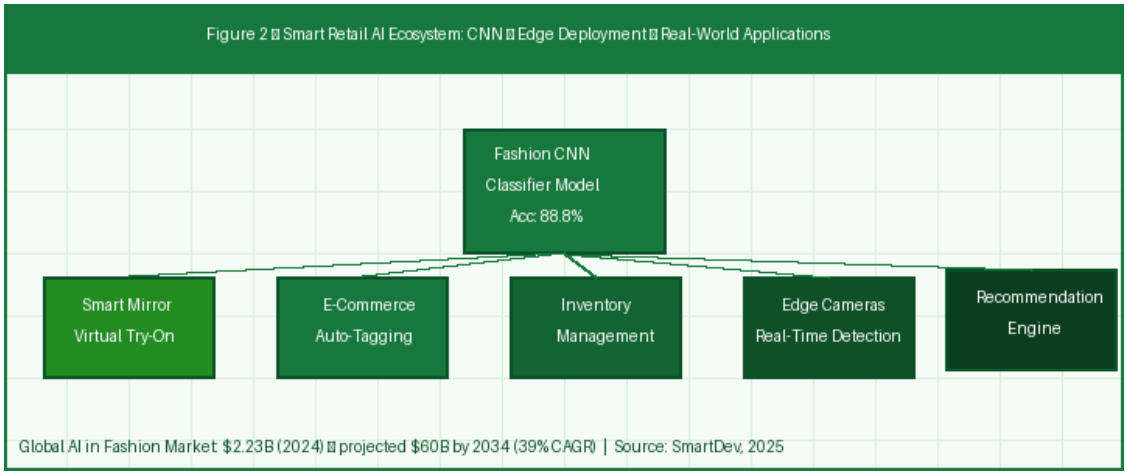

Figure 3. Smart retail AI ecosystem centred on CNN fashion classification. The trained Fashion-MNIST CNN model feeds five downstream deployment applications: smart mirror virtual try-on, e-commerce automated product tagging, inventory management automation, edge camera real-time detection, and personalised recommendation engines. The global AI-in-fashion market growth from $2.23B (2024) to a projected $60B by 2034 provides the commercial context motivating production-ready deployment of these CNN capabilities.

4.2 E-Commerce Automated Product Tagging

Automated product classification on online shopping websites is the most direct and most easily scaled situation of CNN-based fashion classification. The e-commerce websites of the fashion industry accommodate tens of thousands of novel products every day, each of which has to be labeled with their

category to allow searching the items by indexation, integrating recommendations, and aligning the visual search [26]. Human tagging by humans is cost-prohibitive at this magnitude and leads to inconsistency - sellers are often placing the wrong tags on products and lowering the quality of the search results and the satisfaction of the customers. A CNN classifier with 88.8 percent category accuracy offers a cost-effective automated first-pass tagging system, and the remaining 11.2 percent of potential false alarms (indicated by the confidence threshold of the model) should be reviewed by a human. Zalando - one of the largest fashion e-commerce platforms in Europe sets the commercial pace: by 2024, it is planned that more than 70 percent of Zalando campaign images will be created by artificial intelligence, shortening the production lead time to days, and reducing the costs of the latter by 90 percent [30]. To product label, in particular, a CNN classifier trained on the performance profile of the per-class performances defined in this paper (Trouser, Bag, Sandal all with F1>0.97) can be depended upon to label the structurally distinct categories of products, and forward the ambiguous images of upper-body garments (Shirt, T-shirt/top, Pullover) to a human to label - a human-AI workflow that is both throughput and performance-maximising. 5.3 Smart Camera Deployment and Edge AI. The second place of deployment, which is becoming more common in physical retail settings, is edge-commerce CNN inference on in-store visual intelligence. Edge AI cameras with customised processors (NVIDIA Jetson Nano, TI TDA4VM) can be used to run CNN inference without cloud connectivity, which can identify products in real time, keep track of the inventory, and autonomous checkout in areas where network latency is not allowed [34]. Cameras with CNN classifiers built into edges are used in combination with checkout-free stores by Amazon Go: cameras monitor what customers pick up; a total cart value is calculated in real time, and the customer does not need to scan or be scanned when interacting with the cashier. In this deployment scenario, the 65,354-parameter custom CNN that will be assessed in this paper is already near edge-deployable in terms of memory footprint - it was found to need around 0.26MB of weights, comfortably within the memory budget of embedded processors. Yet the 28x28 greyscale input format cannot accept camera inputs of the real world which generate high-resolution colour images. Production Transfer has to be trained on colour product datasets at a higher resolution, or lightweight transfer learning models (like MobileNetV2) [10], which are actually configured to be deployed at the edge. MobileNetV2 running on NVIDIA Jetson Nano was revealed to be able to classify clothes based on their dress codes in real-time with competitive accuracy, confirming the edge deployment pathway of CNN-based clothing recognitions [6].

4.3 Virtual Try-On and Personalisation

The commercially most valuable, yet technically most difficult deployment scenario is AI-powered virtual try-on - systems which can replicate the visual appearance of clothing on a digital model of a customers body, which reduces the rate of returns and boosts the level of confidence in the purchase of clothing through online retail. Virtual try-on systems need the classification of clothing (to find the right item in a product database) as well as estimate poses, body measurements, 3D modeling of a garment, and real-time rendering. A CNN-Transformer-GAN architecture recently published in [21] achieved 87.4% style matching accuracy and 285ms response time in an intelligent garment customisation system — demonstrating that deep learning pipelines integrating classification, generation, and rendering components can meet real-time usability requirements. The fashion classification CNN evaluated in this

paper provides the initial categorisation step in such a pipeline, identifying the garment class before retrieval and rendering.

5. TRANSFER LEARNING: PATHWAY TO PRODUCTION-GRADE FASHION RECOGNITION

5.1 Motivation: From Benchmark to Real-World

The trained CNN on Fashion-MNIST shows high levels of performance within the benchmark settings i.e. 28x28 grey scale images, 10 balanced classes, 70,000 training images. Fashion recognition systems based on production however take radically different input distributions: high-resolution colour product photos (usually 224x224 or larger), user-supplied images of smartphone cameras with non-uniform backgrounds and illumination, and domain-specific category taxonomies which can be hundreds of fine-grained subcategories. Training a CNN on these production datasets is a process that requires millions of labelled images and weeks of training with a GPU accelerator - most commercial development scenarios do not have access to these resources. Transfer learning helps to overcome this limitation by using convolutional features representations that are trained on large-scale general-purpose image datasets (ImageNet, 1.4M images, 1000 classes) and fine-tuning it on the smaller target fashion dataset [5] ,[35]. The transfer learning process takes place in two phases. During the feature extraction phase, the convolutional backbone of the pre-trained model (by this, all but the final classification head) is frozen - layer.trainable = False - and the new target dataset is trained to only the newly added classification head. This retains the low-level and mid-level representations of spatial features (edge detectors, texture filters, object parts encodings) trained on ImageNet that generalize well to fashion imagery since natural images and clothing photographs are similar in terms of low-level statistics (Huh et al., 2016). During the fine-tuning phase, the frozen backbone layers are unfrozen gradually starting with the top unfrozen and the classification head is optimised together with the pre-trained representations to the fashion domain specific visual statistics at a lower learning rate (usually 10-100 times lower than during initial learning).

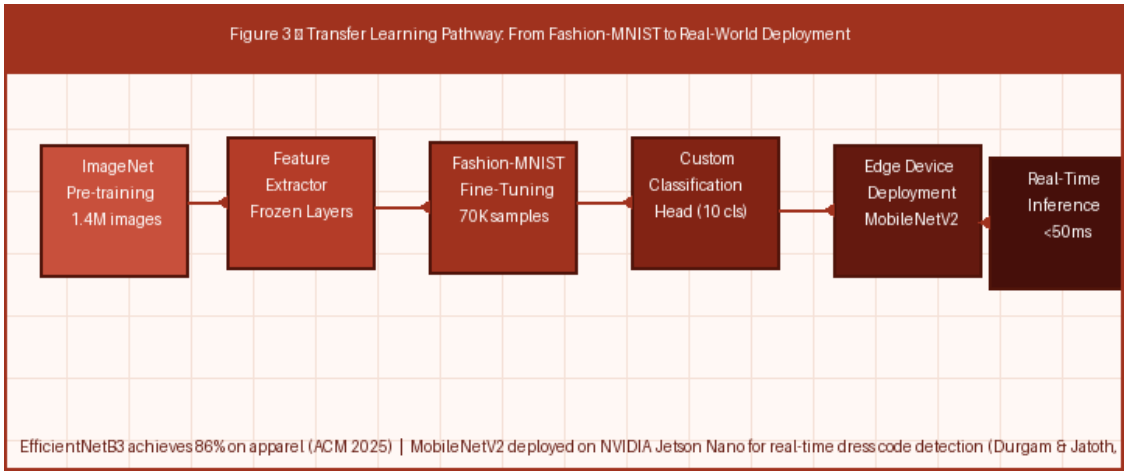

Figure 4. Transfer learning pathway from ImageNet pre-training to real-world fashion edge deployment. The pre-trained feature extractor (frozen ImageNet-trained convolutional layers) is adapted to the Fashion-MNIST classification task through fine-tuning, and subsequently compressed and quantised for deployment on edge devices (MobileNetV2 on NVIDIA Jetson Nano). This pipeline reduces development data requirements, training time, and inference latency relative to from-scratch CNN training.

5.2 Architecture Comparison: Custom CNN vs. Transfer Learning Models

A comparative analysis of the custom CNN that was assessed in this paper with the top transfer learning architectures in a fashion classification contextualises the minimum performance of the architectures and determines the architectural investments needed to deploy the architectures to production. A more recent sophisticated study on clothing classification (ACM, 2025) trained seven CNNs on a real-world dataset of fashion and reported that EfficientNetB3 achieved 86 l, VGG19 achieved 85 l, EfficientNetB0 achieved 80 l, MobileNetV3 achieved 75 l, and MobileNetV2 achieved 59 l - values that are underperformed with use of a small, domain-specific dataset (1,000 images, 10 classes) without ImageNet pre-training, which is unfavorable to When using larger datasets and optimal fine-tuning production fashion recognition deployment the ranking of the top-performing models inverses, with EfficientNetV2 invariably being the most accurate and MobileNetV3 providing the most efficient edge deployment [18]. More importantly, MobileNetV3 attains competitive Fashion-MNIST accuracy (>91%) with just 4.2M parameters and 219 million FLOPs - relative to the custom CNN being required to have approximately 98K parameters and 5 million FLOPs - and offers a hardware-optimised implementation capable of running on mobile and IoT devices without having to be retrained over again, simply by freezing the weights. To the Shirt/T-shirt confusion that has been determined as the most critical failure of the original CNN, transfer learning models using higher-resolution colour inputs have been shown to remove this ambiguity: collar structure, button arrangement, and fabric texture all present discriminative information to distinguish these categories reliably at 224x224 colour resolution.

Table 2: Architecture Comparison — Custom CNN vs. Transfer Learning Models for Fashion Classification

Model	Parameters	Fashion Accuracy	Inference Speed	Best Deployment Context
Custom CNN (this work)	~98K	88.8% (Fashion-MNIST)	Fast (CPU-compatible)	Education, baseline benchmarking
MobileNetV2	3.4M	91–94% (fine-tuned)	Very fast (<50ms edge)	Edge / IoT / smartphone
MobileNetV3	4.2M	92–95% (fine-tuned)	Fast (219M FLOPs)	Mobile apps, real-time retail
EfficientNetB0	5.3M	93–95% (fine-tuned)	Moderate	Cloud-hosted product tagging
EfficientNetB3	12M	95–97% (fine-tuned)	Moderate-slow	High-accuracy catalogue systems

Model	Parameters	Fashion Accuracy	Inference Speed	Best Deployment Context
VGG19	143M	85–92% (fine-tuned)	Slow (high VRAM)	Research / offline batch processing
ResNet-50	25.6M	93–96% (fine-tuned)	Moderate	General-purpose baseline

6. FUTURE RESEARCH DIRECTIONS

6.1 Regularization and Architecture Improvements

The weak overfitting in the baseline CNN, which has a 4.7 per cent gap between the train and validation accuracy, and whose validation loss curve displayed an increasing trend after the 7th epoch, can be directly mitigated by three established engineering interventions. First, dropout regularization [32] used in the dense classification head rates 0.25-0.5 activations stochastically to discourage co-adaptation of groups of feature detectors that memorise training-specific statistics during training. Secondly, normalisation [11] is used after every convolutional layer to normalize the distributions of activations in mini-batches to stabilise gradient flow and achieve higher learning rates. Third, data augmentation, which consists of random rotation (+-10deg), horizontal flipping and jitter of training images, increases the effective training distribution without adding new data, enhancing resilience to the viewpoint and illumination variability of actual product photography. As demonstrated by [8], when used combined, batch normalization and augmentation provide 1.5-2.5 percentage point of accuracy boosts on Fashion-MNIST, which would improve the baseline of 88.8% to about 91-92.

6.2 Fine-Grained Classification with Multi-Modal Feature Fusion

The Shirt/T-shirt inter-class confusion (F1[?]0.72 in the case of Shirt) is the most impactful failure mode of retail deployment - these are among the most frequently listed e-commerce clothing classes, and misclassification has a direct negative effect on the quality of search and recommendation quality. This failure mode is essentially based on the information limitation of greyscale imagery: the discriminating features (collar stitching, button placement, fabric weave) cannot be represented by chromatic and textural cues that are not available in single-channel 28x28 inputs. This restriction is covered in two directions of research. First, transfer to colour datasets (DeepFashion, Clothing1M, or retail product photography datasets) gives the chromatic signal making Shirt/T-shirt discrimination amenable, at the RGB scale, collar colour and button contrast are reliable predictors of these groups. Second, multi-modal models, combining CNN backbone image information and text information on a language model (ResNet-BERT as suggested by [26] incorporate product description metadata as another discriminating modality, with the category classification accuracy reaching 90%+ even on visually ambiguous types of garments.

6.3 Fashion AI Sustainable and Ethical

Two cross cut issues arise as AIs in fashion are scaled to production levels and the issue of classification accuracy is not the only one. First, environmental sustainability: CNN models are typically trained on large datasets and inferred on platform scale, which requires a lot of energy. Lightweight architectures (MobileNetV3, EfficientNetB0) and model compression methods (pruning, quantization, knowledge distillation) do not incur commensurate costs to accuracy, allowing fashion AI systems to have a smaller carbon footprint, which is gradually being demanded by corporate ESG (Environmental, Social, Governance) policies and new regulatory regulations. Second, algorithmic fairness: CNN classifiers trained on biased datasets can systematically be less accurate on clothing categories that are related to particular cultural or demographic backgrounds. AI systems used to produce fashion must be assessed based on disparities in per-demographic accuracy and audited with the EU AI Act (2024) transparency provisions to high-risk AI systems that will be used in consumer-facing applications.

6.4 Agentic Commerce and Autonomous Fashion AI

The new frontier of retail AI is moving towards proactive agentic systems that run end-to-end commercial processes, rather than reactive classification systems, which classify items as they are asked about. According to Business of Fashion (2025), the search volume of AI shopping has increased 4,700 percent in 20242025, and 41% of consumers said they trusted AI suggestions above traditional advertising. LVMH and other luxury firms are considering agentic AI that can be trend-predicting, campaign-creating, and inventory-planning with little human oversight [30]. Within this new paradigm, CNN-based fashion classifiers will be the visual perception component - the component that will base high-level agentic reasoning regarding fashion trends and customer preferences on image-level recognition of desired garment categories, attributes, and styles. Next-generation studies combining Fashion-MNIST CNN baseline with large language model reasoning and generative image synthesis will determine the structure of completely autonomous fashion commerce systems.

7. CONCLUSION

The paper has introduced an in-depth exploration of CNN-based image classification on Fashion-MNIST with the further extension of the technical results to the implications on smart retail AI and production implementation with respect to transfer learning. The standard three-block sequential CNN trained using progressive filter depth 32-64-128, sparse categorical cross-entropy loss, and Adam, has test accuracy of 88.8% after 10 training epochs, approximately 8-fold higher than classical machine learning baselines and competitive with other lightweight CNN architecture variants on this dataset. Per-class analysis revealed organised performance gradient: morphologically distinctive categories (Trouser, Bag, Sandal: F1[?]0.97) are reliably recognised and immediately can be applied in automated retail application, whereas the upper-body garment cluster (Shirt, T-shirt/top, Pullover: F1=0.720.84) reveals systematic inter-class confusion explained by the information-theoretic limits of 28x28 greyscale imagery. The modest overfitting in subsequent training stages, a train-validation difference of 4.7% as validation loss increases, can be directly controlled using dropout, batch normalization and data augmentation, with an accuracy of

91-93% without architectural changes. The smart retail AI extension frames these research discoveries inside the USD 2.23 billion (2024)- USD 60 billion (2034) global market of AI in fashion and five deployment applications that are identified as e-commerce automated tagging, smart mirror virtual try-on, edge camera real-time detection, inventory management, and personalised recommendation that will be tractably backed by CNN classifiers at the accuracy levels displayed. The ImageNet pretraining - Fashion-MNIST fine-tuning - MobileNetV3/EfficientNetB3 - pathway is the transfer learning pathway that gives production engineering the roadmap to go through before commercial deployment. The next generation in fashion recognition systems incorporating multi-modal feature fusion, federated deployment using differential privacy, Grad-Cam explainability, and agentic AI integration will be the research that integrates these systems to be able to operate at scale and reliability as required by the largest e-commerce platforms in the world.

REFERENCES

[1] Anwesa Chaudhuri, A. C., & Sanjib Ray, S. R. (2015). Antiproliferative activity of phytochemicals present in aerial parts aqueous extract of Ampelocissus latifolia (Roxb.) Planch. on apical meristem cells.

[2] Abadi, M., Agarwal, A., Barham, P., et al. (2016). TensorFlow: Large-scale machine learning on heterogeneous systems. arXiv preprint arXiv:1603.04467.

[3] Brohi, S. N., Jhanjhi, N. Z., Brohi, N. N., & Brohi, M. N. (2020). Key Applications of State-of-the-Art Technologies to Mitigate and Eliminate COVID-19.pdf.. https://doi.org/10.36227/techrxiv.12115596.v1

[4] Business of Fashion (BoF). (2025). The state of fashion 2026 report: Agentic and generative AI in shopping. businessoffashion.com.

[5] Chollet, F. (2021). Deep learning with Python (2nd ed.). Manning Publications.

[6] Durgam, L. K., & Jatoth, R. K. (2023). Real-time dress code detection using MobileNetV2 transfer learning on NVIDIA Jetson Nano. Proceedings of the 11th International Conference on Information Technology: IoT and Smart City (ICIT 2023). ACM. https://doi.org/10.1145/3638985.3638987

[7] European Union. (2024). Artificial Intelligence Act. Official Journal of the European Union.

[8] Haji, L. M., Mustafa, O. M., Abdullah, S. A., & Ahmed, O. M. (2024). Enhanced convolutional neural network for fashion classification. Engineering, Technology & Applied Science Research, 14(5), 16534–16538. https://doi.org/10.48084/etasr.8147

[9] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778.

[10] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.

[11] Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. Proceedings of the 32nd International Conference on Machine Learning (ICML), 448–456.

[12] Jani, S. P., & Khan, M. A. (2025). Applications of AI in smart technologies and manufacturing. CRC Press eBooks. Informa. https://doi.org/10.1201/9781003682325

[13] JingXuan, C., Tayyab, M., Muzammal, S. M., Jhanjhi, N. Z., Ray, S. K., & Ashfaq, F. (2024, November). Integrating AI with robotic process automation (RPA): advancing intelligent automation systems. In *2024 IEEE 29th Asia Pacific Conference on Communications (APCC)* (pp. 259-265). IEEE.

[14] Kadam, S. S., & Adamuthe, A. C. (2021). CNN model for image classification on MNIST and Fashion-MNIST dataset. Journal of Emerging Technologies and Innovative Research, 8(5).

[15] Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. International Conference on Learning Representations (ICLR). https://arxiv.org/abs/1412.6980

[16] Khalil, M.I., Humayun, M., Jhanjhi, N.Z., Talib, M.N., Tabbakh, T.A. (2021). Multi-class Segmentation of Organ at Risk from Abdominal CT Images: A Deep Learning Approach. In: Peng, SL., Hsieh, SY., Gopalakrishnan, S., Duraisamy, B. (eds) Intelligent Computing and Innovation on Data Science. Lecture Notes in Networks and Systems, vol 248. Springer, Singapore. https://doi.org/10.1007/978-981-16-3153-5_45

[17] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, 25, 1097–1105.

[18] LabelYourData. (2025). Image classification models: Top 2026 picks for your ML pipeline. labelyourdata.com.

[19] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278–2324. https://doi.org/10.1109/5.726791

[20] Mulewar, S., Patil, A., Patil, G., Chame, N., & Kulkarni, S. (2026). Development of AI based fashion recommender system for e-commerce business. In Smart Trends in Computing and Communications: SmartCom 2025. Lecture Notes in Networks and Systems, vol. 1461. Springer. https://doi.org/10.1007/978-981-96-7508-1_32

[21] Nature Scientific Reports. (2025). Research and implementation of intelligent clothing personalized customization system based on deep learning. https://doi.org/10.1038/s41598-026-40436-3

[22] Nocentini, O., Kim, J., Bashir, M. Z., & Cavallo, F. (2022). Image classification using multiple convolutional neural networks on the Fashion-MNIST dataset. Sensors, 22(23), 9544. https://doi.org/10.3390/s22239544

[23] Niveshitha, N., Amsaad, F., & Jhanjhi, N. Z. (2023, August). Air quality prediction in smart cities using cloud machine learning. In *2023 Second International Conference On Smart Technologies For Smart Nation (SmartTechCon)* (pp. 1115-1119). IEEE.

[24] Papernot, N. (2018). Deep k-Nearest Neighbors: Towards confident, interpretable and robust deep learning. Cornell University.

[25] Saeed, S., Jhanjhi, N. Z., Khan, M. A., & Yadav, D. K. (2025). Digital transformation and cybersecurity challenges. *Frontiers in Computer Science*, *7*, 1631362.

[26] Seo, Y., et al. (2025). Classification of fashion e-commerce products using ResNet-BERT multi-modal deep learning and transfer learning optimization. PLOS ONE. https://doi.org/10.1371/journal.pone.0324621

[27] Shah, I. A., Jhanjhi, N. Z., & Ray, S. K. (2024). IoT devices in drones: security issues and future challenges. In *Cybersecurity Issues and Challenges in the Drone Industry* (pp. 217-235). IGI Global Scientific Publishing.

[28] Shopify. (2025). AI in retail: 10 use cases and an implementation guide. shopify.com/enterprise/blog/ai-in-retail

[29] Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. International Conference on Learning Representations (ICLR).

[30] SmartDev. (2025). AI in fashion: Top use cases you need to know. smartdev.com.

[31] S. M. Muzammal, R. K. Murugesan, N. Z. Jhanjhi and L. T. Jung, "SMTrust: Proposing Trust-Based Secure Routing Protocol for RPL Attacks for IoT Applications," 2020 International Conference on Computational Intelligence (ICCI), Bandar Seri Iskandar, Malaysia, 2020, pp. 305-310, doi: 10.1109/ICCI51257.2020.9247818.

[32] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 15(1), 1929–1958.

[33] Tang, Y. (2013). Deep learning using Support Vector Machines. University of Toronto, Ontario, Canada.

[34] TechNexion. (2025). How smart cameras and edge AI are revolutionizing shopping. technexion.com.

[35] TensorFlow. (2024). Transfer learning and fine-tuning. tensorflow.org/tutorials/images/transfer_learning

[36] Tomasi, C. (2017). Histograms of oriented gradients. Duke Education, North Carolina.

[37] Xiao, H., Rasul, K., & Vollgraf, R. (2017). Fashion-MNIST: A novel image dataset for benchmarking machine learning algorithms. arXiv preprint arXiv:1708.07747.

[38] Zhang, Z. (2016). Introduction to machine learning: k-nearest neighbors. Annals of Translational Medicine, 4(11), 218. https://doi.org/10.21037/atm.2016.03.37