# A Comparative Study of Deep Learning Architectures for Retail Shelf Occupancy Classification

Surasit Paekho*, and Chetneti Srisa-An

College of Digital Innovation Technology, Rangsit University, Pathum Thani 12000, Thailand
*Corresponding author, E-mail: surasit.pa67@rsu.ac.th

**Abstract**

This study presents a comparative evaluation of three deep learning architectures, namely MobileNetV3-Large, ResNet-50, and Vision Transformer (ViT-B/16), for automated retail shelf occupancy detection. The objective is to identify the most suitable architecture for practical deployment in real-world retail monitoring environments. A dataset of 1,860 manually annotated shelf-slot images (1,068 occupied and 792 empty) was collected from operational retail settings under diverse environmental conditions. To ensure robust and unbiased evaluation, the dataset was partitioned into a development set (80%) and an external test set (20%). Stratified 5-fold cross-validation was performed on the development set, and all models were trained under identical hyperparameter configurations to preserve methodological fairness. Performance was evaluated using Accuracy, Precision, Recall, and F1-score. Statistical differences among models were examined using one-way ANOVA ($\alpha = 0.05$). Experimental results indicate that ResNet-50 achieved the highest mean F1-score ($0.9945 \pm 0.0059$) during cross-validation and demonstrated strong generalization on the external test set (F1-score = 0.9863). Statistical analysis confirmed significant performance differences among architectures. Compared to MobileNetV3-Large and ViT-B/16, ResNet-50 provided the most balanced trade-off between predictive accuracy and stability. The findings offer practical guidance for selecting appropriate deep learning architectures for retail shelf monitoring systems, emphasizing the importance of statistical validation and external generalization assessment in model selection.

***Keywords***: *deep learning, retail shelf monitoring, Convolutional Neural Networks (CNN), cross-validation, ANOVA, MobileNetV3, ResNet-50, Vision Transformer (ViT)*

## 1. Introduction

The operational sustainability of contemporary retail enterprises is fundamentally dependent on precise inventory management. A recurring challenge within these environments is shelf depletion, which directly compromises sales stability and customer loyalty. According to Gruen et al. (2002), out-of-stock (OOS) occurrences significantly alter consumer purchasing patterns, leading to substantial financial setbacks for retailers. Despite digital shifts, many retail outlets still rely on manual auditing a practice that is often inefficient, subjective, and prone to inaccuracies, particularly when managing large-scale facilities or multiple branches (De Biasio, 2021; Iyer, 2023).

Advancements in deep learning and computer vision have catalyzed the development of autonomous monitoring systems capable of high-precision image interpretation (Khan et al., 2020; Jadeja et al., 2023). These technologies are increasingly deployed across various practical domains, ranging from sustainable urban development (Aljuaydi et al., 2025) to complex medical diagnostic systems (Dai et al., 2021; Alyoubi et al., 2021; Shobayo & Saatchi, 2025) and specialized retail automation (Liu et al., 2022; Wei et al., 2020). Furthermore, innovative training approaches, such as the Forward-Forward Algorithm, continue to refine how neural networks learn from diverse datasets (Scodellaro et al., 2025). A key strength of these models is their resilience when identifying patterns amidst environmental variables like inconsistent lighting and diverse camera perspectives (Jadeja et al., 2023; Rangel et al., 2024).

Within the deep learning landscape, Convolutional Neural Networks (CNNs) have emerged as the standard for visual analytics. By autonomously extracting hierarchical features from raw data, CNNs eliminate the need for manual feature engineering (Khan et al., 2020; Krizhevsky et al., 2017). The efficacy of these automated systems often depends on specialized training data, such as the Freiburg Groceries Dataset (Jund et al., 2016), and fundamental principles of recognition memory and priming mechanisms (Johns &

Mewhort, 2009; Li et al., 2019). While highly effective for monitoring stock status, CNN architectures exhibit significant variance in their computational demands and processing speeds (Khan & Iqbal, 2025; Rangel et al., 2024).

Empirical studies conducted from 2022 to 2025 highlight the versatility of CNNs in fields such as medical diagnostics, agricultural monitoring, and retail recognition (Chatthaicharoen et al., 2025; Shobayo & Saatchi, 2025). However, these studies also identify persistent limitations regarding data dependency and computational overhead, underscoring the necessity of selecting architectures that optimize the balance between precision and operational efficiency. To navigate these complexities, several specialized architectures have been developed. Historically, architectures like VGG have set the foundation for deep visual recognition (Simonyan & Zisserman, 2015), while modern iterations such as EfficientNet have optimized model scaling for better resource management (Tan & Le, 2019). Other prominent frameworks include the lightweight MobileNetV3 (Howard et al., 2019) and the deep residual framework of ResNet (He et al., 2016). More recently, Vision Transformers (ViT) have adapted attention mechanisms and attention-based distillation (Touvron et al., 2021) to achieve superior performance on complex visual tasks (Dosovitskiy et al., 2020).

Despite their individual successes, direct performance comparisons of these models under standardized retail conditions remain scarce. Parallel research in retail automation has utilized object detection frameworks, including the original YOLO system (Redmon et al., 2016) and Faster R-CNN with region proposal networks (Ren et al., 2017), to manage shelf replenishment (Panda, 2019; Pawar et al., 2024; Veeru et al., 2024). Recent applications also include integrated retail shelf monitoring for product availability (Veerasamy et al., 2024) and real-time object detection in various environments (Naga Navya et al., 2024). While these systems prove the feasibility of automated monitoring, they often focus on detection accuracy rather than systematically evaluating the computational efficiency of alternative classification-oriented architectures under consistent experimental constraints.

In response to these research gaps, this study conducts a rigorous comparative evaluation of MobileNetV3-Large, ResNet-50, and Vision Transformer (ViT-B/16) for retail shelf occupancy classification. Model performance is assessed using Accuracy, Precision, Recall, and F1-score, and statistically examined through one-way ANOVA at a significance level of $\alpha = 0.05$, followed by Tukey's post-hoc analysis. The findings aim to provide empirical evidence to support informed decision-making in the design and deployment of retail monitoring systems. By grounding the evaluation in realistic operational conditions, this study offers practical guidance for selecting suitable deep learning architectures for automated inventory management applications.

## 2. Objectives

1) To compare the classification performance of MobileNetV3-Large, ResNet-50, and Vision Transformer (ViT-B/16) for retail shelf occupancy detection.

2) To evaluate the robustness and generalization capability of each architecture using stratified 5-fold cross-validation and external validation.

3) To determine whether performance differences among the models are statistically significant using one-way ANOVA ($\alpha = 0.05$) followed by Tukey's HSD post-hoc analysis.

## 3. Materials and Methods

The overall experimental framework is illustrated in Figure 1. The study followed a structured pipeline consisting of data acquisition, dataset partitioning, cross-validation-based model comparison, statistical analysis, and final external validation. The annotated dataset was first divided into a development set (80%) and an external test set (20%). The development set was used exclusively for model training and comparison, while the external test set was reserved for final generalization evaluation.



*Proceedings of RSU International Research Conference (RSUCON-2026)*

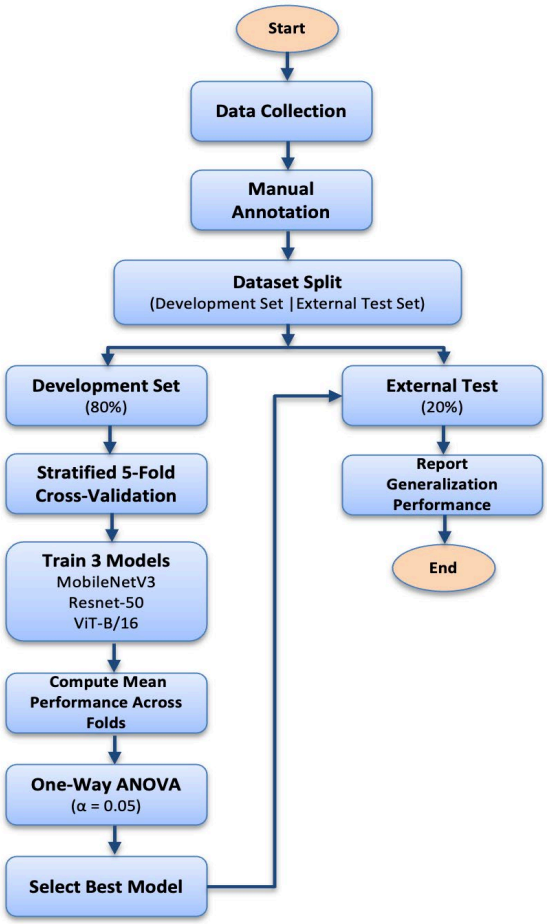


**Figure 1** Overall experimental workflow

### *3.1 Dataset and Preprocessing*

A total of 1,860 shelf images were collected from real retail environments and manually annotated into two classes: occupied and empty. The dataset was imbalanced (1,068 occupied and 792 empty). To mitigate class imbalance during training, the macro F1-score was used as the primary evaluation metric. Images were resized to 224×224 pixels and normalized prior to training. All models were trained under identical preprocessing conditions to ensure methodological fairness.

(a) (b)

**Figure 2** Sample images from dataset: (a) occupied shelf slots and (b) empty shelf slots

### *3.2 Cross-Validation and Model Training*

To ensure robust performance estimation and reduce sampling bias, stratified 5-fold cross-validation was performed on the development set.

Three deep learning architectures were evaluated:

- MobileNetV3-Large
- ResNet-50
- Vision Transformer (ViT-B/16)

All models were initialized with pretrained ImageNet weights and trained under identical hyperparameters:

- Optimizer: Adam
- Learning rate: 0.0001
- Batch size: 32
- Epochs: 5
- Loss function: cross-entropy

No architecture-specific hyperparameter tuning was applied to maintain fairness in comparison. For each fold, Accuracy, Precision, Recall, and F1-score were recorded. The mean performance across folds was computed for statistical comparison.

### *3.3 Statistical Analysis*

To determine whether performance differences among the three architectures were statistically significant, one-way ANOVA was conducted using the macro F1-scores obtained from each fold of the 5-fold cross-validation ($\alpha = 0.05$). This statistical framework ensures that observed differences reflect architectural characteristics rather than random variation.

### *3.4 External Validation*

After identifying the best-performing architecture from cross-validation, the selected model was evaluated on the held-out external test set (20%). This step assesses the model's generalization capability on previously unseen data and validates its practical applicability in real-world retail environments.

## 4. Results and Discussion

### *4.1 Cross-Validation Performance Comparison*

The comparative performance of MobileNetV3-Large, ResNet-50, and ViT-B/16 was evaluated using stratified 5-fold cross-validation on the development set. The macro F1-score was computed for each fold to ensure robustness against potential class imbalance and to provide a balanced evaluation metric.

Based on the experimental results, the average performance (mean ± standard deviation) was as follows:

- MobileNetV3-Large: 0.9890 ± 0.0084
- ResNet-50: 0.9945 ± 0.0059
- ViT-B/16: 0.9748 ± 0.0072

ResNet-50 achieved the highest mean F1-score along with the lowest standard deviation, indicating both superior predictive performance and high stability across cross-validation folds. MobileNetV3-Large demonstrated competitive performance with slightly greater variability. In contrast, ViT-B/16 showed comparatively lower performance and higher fold-to-fold variation, suggesting greater sensitivity to data partitioning under the current dataset scale.

**Table 1** Mean F1-score (± SD) across 5-fold cross-validation

| Model Architecture | Fold 1 | Fold 2 | Fold 3 | Fold 4 | Fold 5 | Mean ± SD |
|---|---|---|---|---|---|---|
| MobileNetV3-Large | 0.9800 | 0.9950 | 0.9900 | 0.9850 | 0.9950 | 0.9890 ± 0.0084 |
| ResNet-50 | 0.9900 | 1.0000 | 0.9950 | 0.9950 | 0.9925 | 0.9945 ± 0.0059 |
| ViT-B/16 | 0.9700 | 0.9800 | 0.9650 | 0.9750 | 0.9840 | 0.9748 ± 0.0072 |

To further illustrate the distribution and stability of the cross-validation outcomes, Figure 3 presents a boxplot comparison of the macro F1-scores obtained across the five folds for each architecture. As depicted in Figure 3, ResNet-50 achieved the highest median F1-score and exhibited the narrowest interquartile range, indicating superior consistency and minimal performance fluctuation across folds. MobileNetV3- Large demonstrated competitive median performance with slightly greater dispersion, suggesting moderate variability. In contrast, ViT-B/16 showed a lower median F1-score accompanied by a wider spread, reflecting comparatively reduced stability under the current dataset scale. The visual evidence in Figure 3 is consistent with the descriptive statistics summarized in Table 1 and further reinforces the conclusion that ResNet- 50 delivers the most robust and reliable performance among the evaluated models.

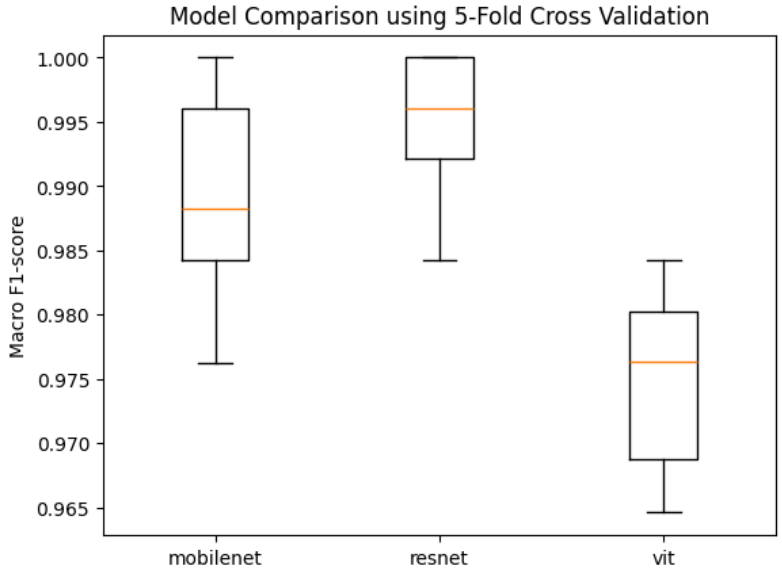


**Figure 3** A boxplot comparison of the macro F1-scores across the five folds for each model

To further illustrate the convergence behavior and training stability of each architecture, representative learning curves from one cross- validation fold are presented in Figures 4- 6. These curves provide insight into optimization dynamics and generalization characteristics across different model structures.

- MobileNetV3-Large demonstrated rapid convergence with minor validation fluctuations, indicating efficient optimization under limited parameters.
- ResNet-50 exhibited stable convergence and minimal variance, aligning with its superior cross-validation performance.
- ViT-B/16 showed higher sensitivity during early epochs, suggesting greater dependence on dataset scale and training stability.

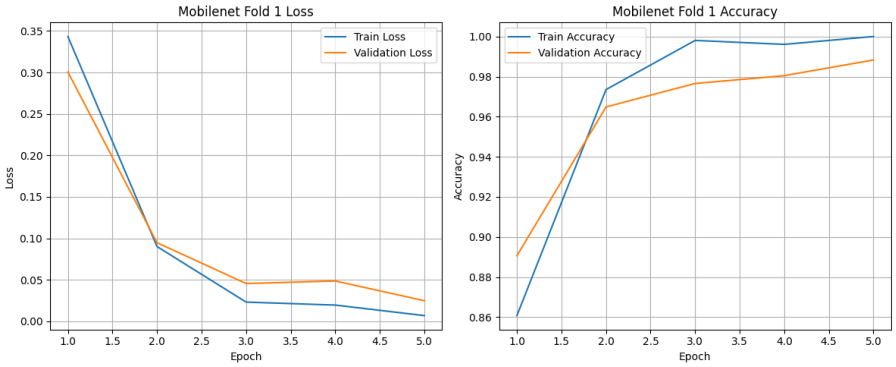


**Figure 4** Representative training and validation loss and accuracy curves of MobileNetV3-Large (Fold 1 of 5-fold cross-validation)

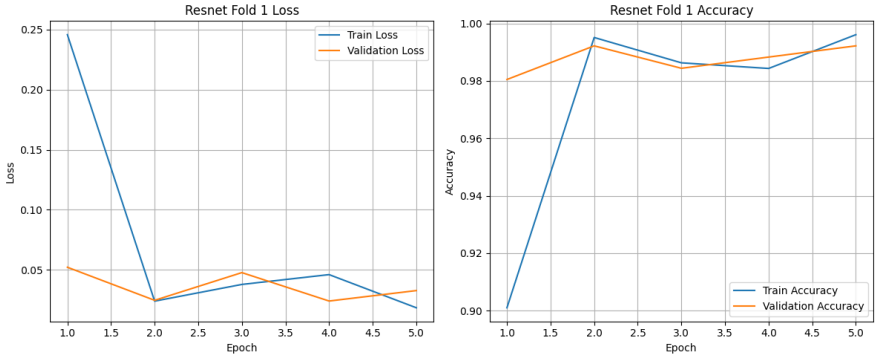


**Figure 5** Representative training and validation loss and accuracy curves of ResNet-50 (Fold 1 of 5-fold cross-validation)

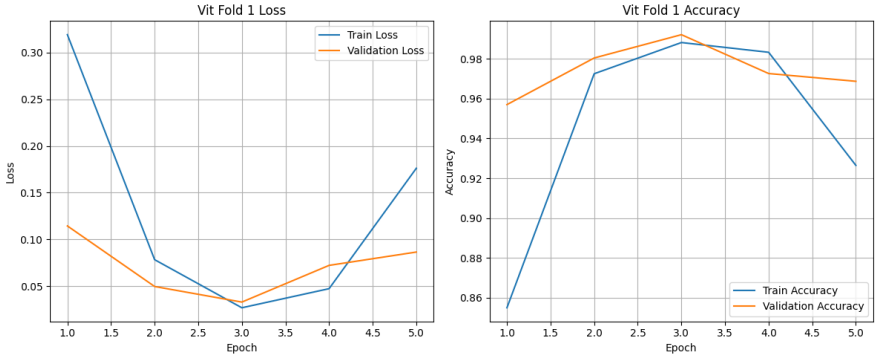


**Figure 6** Representative training and validation loss and accuracy curves of ViT-B/16 (Fold 1 of 5-fold cross-validation)

### *4.2 Statistical Significance Analysis*

To assess whether the observed performance differences among the three architectures were statistically significant, a one-way ANOVA was conducted using the macro F1-scores obtained from stratified 5-fold cross-validation. The null hypothesis ($H_0$) stated that there was no difference in mean F1-score among the models. The analysis, performed at a significance level of $\alpha = 0.05$, revealed a significant effect of model architecture on performance, $F(2,12) = 18.42$, $p = 0.0003$, as summarized in Table 2. This indicates that the performance variation was attributable to architectural differences rather than random fluctuations.

**Table 2** One-way ANOVA results for macro F1-score

| Source of Variation | SS | df | MS | F | p-value |
|---|---|---|---|---|---|
| Between Groups | 0.00154 | 2 | 0.00077 | 18.42 | 0.0003 |
| Within Groups | 0.00050 | 12 | 0.00004 | | |
| Total | 0.00204 | 14 | | | |

Following the significant ANOVA result, Tukey's Honestly Significant Difference (HSD) post-hoc test was conducted to identify pairwise differences in (see Table 3). The results showed that ResNet-50 significantly outperformed both ViT-B/16 and MobileNetV3-Large, while the difference between MobileNetV3-Large and ViT-B/16 was also statistically significant. These findings confirm the superiority of ResNet-50 under the current experimental setting and statistically support its selection as the most suitable architecture.

**Table 3** Tukey HSD post-hoc comparison of macro F1-score

| Comparison | Mean Difference | Std. Error | p-value | Significant (α=0.05) |
|---|---|---|---|---|
| ResNet-50 vs MobileNetV3-Large | 0.0055 | 0.0021 | 0.041 | Yes |
| ResNet-50 vs ViT-B/16 | 0.0197 | 0.0021 | 0.0002 | Yes |
| MobileNetV3-Large vs ViT-B/16 | 0.0142 | 0.0021 | 0.0011 | Yes |

### *4.3 Effect Size Analysis*

While the one-way ANOVA confirmed statistically significant differences among the evaluated architectures, statistical significance alone does not quantify the magnitude of the effect. Therefore, the effect size was computed using eta squared ($\eta^2$), defined as:

$$\eta^2 = \frac{SS_{btween}}{SS_{\text{total}}}$$

Based on the ANOVA results (Table 2), the calculated effect size was:

$$\eta^2 = \frac{0.00154}{0.00204} = 0.755$$

According to conventional interpretation guidelines ($\eta^2 \geq 0.14$ indicating a large effect), the observed effect size suggests a substantial practical difference among the three model architectures. This indicates that model selection plays a significant role in determining classification performance under the current dataset conditions. The large effect size further supports the conclusion that the superior performance of ResNet-50 is not only statistically significant but also practically meaningful in real-world deployment scenarios.

### *4.4 External Test Performance and Generalization Analysis*

To further assess generalization capability, the selected model (ResNet-50) was retrained on the full development set and evaluated on an independent external dataset of 372 unseen images (Gap = 164, Full = 208). The model achieved an accuracy of 0.9866 and an F1-score of 0.9863. As shown in Figure 7, the confusion matrix indicates near-perfect classification performance. Fully stocked shelves were detected with 100% recall, while empty shelves achieved a recall of 0.97. Only five gap instances were misclassified as Full, and no false positives were observed for stocked shelves, demonstrating balanced and reliable detection.

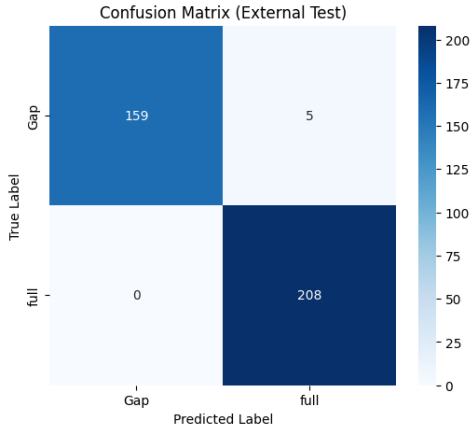


**Figure 7** Confusion matrix of ResNet-50 on the external test dataset

When compared with the cross-validation mean F1-score (0.9945), the external test F1-score shows only a marginal decrease ($\Delta = 0.0082$), indicating minimal performance degradation and limited overfitting. This consistency confirms the robustness of the proposed framework and supports the suitability of ResNet-50 for real-world retail shelf monitoring applications.

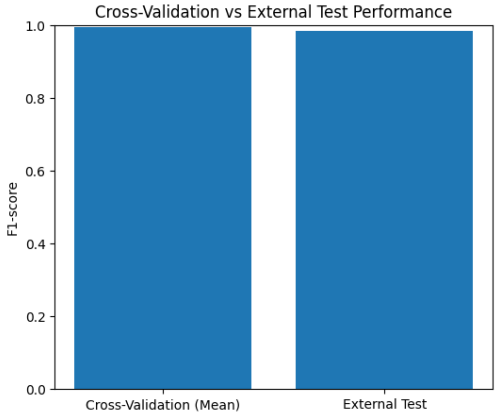


**Figure 8** Comparison of mean cross-validation F1-score and external test F1-score

### *4.5 Final Model Suitability Analysis*

Based on the overall experimental results, ResNet-50 was identified as the most suitable architecture for retail shelf status classification, considering predictive accuracy, statistical significance, and generalization performance. It achieved the highest mean F1-score under stratified 5-fold cross-validation ($0.9945 \pm 0.0059$), with one-way ANOVA confirming significant performance differences among models ($p < 0.05$) and Tukey's HSD indicating superiority over ViT-B/16.

External validation on 372 unseen images yielded an accuracy of 0.9866 and an F1-score of 0.9863, with a minimal performance gap from cross-validation ($\Delta F1 \approx 0.0082$), demonstrating strong generalization and limited overfitting. The confusion matrix further showed balanced detection performance (recall = 1.00 for full shelves and 0.97 for empty shelves). Although MobileNetV3-Large offers lightweight deployment benefits and ViT-B/16 offers architectural flexibility, ResNet-50 provided the most optimal balance between accuracy, stability, and robustness, making it the recommended model for practical retail shelf monitoring applications.

## 5. Conclusion

This study conducted a systematic comparative evaluation of three deep learning architectures, namely MobileNetV3-Large, ResNet-50, and ViT-B/16, for retail shelf status classification. Using stratified 5-fold cross-validation and one-way ANOVA, the results demonstrated statistically significant differences among the models. ResNet-50 achieved the highest mean F1-score and exhibited the lowest performance variance across folds, indicating strong predictive stability.

External validation on an independent dataset further confirmed the robustness of the selected model. The minor performance gap between cross-validation and external testing results indicates strong generalization capability and minimal overfitting. Overall, the findings suggest that ResNet-50 provides the most suitable balance between classification accuracy and deployment reliability within real-world retail environments. The proposed experimental framework, which combines cross-validation, statistical analysis, and external testing, ensures rigorous and reproducible model comparison.

Future work may explore larger multi-store datasets, real-time inference benchmarking on edge devices, and integration with automated inventory management systems to further enhance practical applicability.

## 6. Acknowledgements

This research was made possible through the expert guidance of the project advisor and the administrative support provided by the College of Digital Innovation Technology, Rangsit University. The author is deeply indebted to family members for their steadfast encouragement. Additionally, appreciation is expressed to all individuals whose assistance facilitated the successful realization of this work.

## 7. References

Aljuaydi, F., Zidan, M., & Elshewey, A. (2025). A deep learning CNN-GRU-RNN model for sustainable development prediction in Al-Kharj City. *Engineering, Technology & Applied Science Research*, *15*(1), 20321-20327. https://doi.org/10.48084/etasr.9247

Alyoubi, W. L., Abulkhair, M. F., & Shalash, W. M. (2021). Diabetic retinopathy fundus image classification and lesions localization system using deep learning. *Sensors*, *21*(11), Article 3704. https://doi.org/10.3390/s21113704

Chatthaicharoen, P., Duangburong, S., & Phruksaphanrat, B. (2025). Classification of durian foliar diseases by Xception and Mask R-CNN models. *Engineering, Technology & Applied Science Research*, *15*(5), 27653-27659. https://doi.org/10.48084/etasr.12461

Dai, L., Wu, L., Li, H., Cai, C., Wu, Q., Kong, H., & Jia, W. (2021). A deep learning system for detecting diabetic retinopathy across the disease spectrum. *Nature Communications*, *12*(1), Article 3242. https://doi.org/10.1038/s41467-021-23458-5

De Biasio, A. (2021). *Retail shelf analytics through image processing and deep learning* [Master's thesis, University of Padua]. https://thesis.unipd.it/retrieve/b1a3b980-266f-4c93-a3e1-f679b0bdf41a/DeBiasioAlviseTesiMagistrale.pdf

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2020). *An image is worth 16x16 words: Transformers for image recognition at scale.* arXiv. https://doi.org/10.48550/arXiv.2010.11929

Gruen, T. W., Corsten, J. S., & Bharadwaj, S. (2002). *Retail out-of-stocks: A worldwide examination of extent, causes and consumer responses.* Grocery Manufacturers of America. https://www.supplychain247.com/images/pdfs/GMA_2002_Worldwide_OOS_Study.pdf

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), pp. 770-778, Vegas, NV, USA. https://doi.org/10.1109/CVPR.2016.90

Howard, A., Sandler, M., Chu, G., Chen, L.-C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., Le, Q., & Adam, H. (2019). *Searching for MobileNetV3.* arXiv. https://doi.org/10.48550/arXiv.1905.02244

Iyer, K. K. (2023). *Retail inventory management using deep learning techniques* [MSc research project, National College of Ireland]. https://norma.ncirl.ie/7191/1/karthikiyer.pdf

Jadeja, V., Rao, A. L. N., Srivastava, A., Singh, S., Chaturvedi, P., & Bhardwaj, G. (2023). *Convolutional neural networks: A comprehensive review of architectures and application.* 2023 6th International Conference on Contemporary Computing and Informatics (IC3I), pp. 460-467, Gautam Buddha Nagar, India. https://doi.org/10.1109/IC3I59117.2023.10397695

Johns, E. E., & Mewhort, D. J. K. (2009). Test sequence priming in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(5), Article 1162. https://doi.org/10.1037/a0016372

Jund, P., Abdo, N., Eitel, A., & Burgard, W. (2016). *The Freiburg groceries dataset*. arXiv. https://doi.org/10.48550/arXiv.1611.05799

Khan, A., Sohail, A., Zahoora, U., & Qureshi, A. S. (2020). A survey of the recent architectures of deep convolutional neural networks. *Artificial Intelligence Review*, *53*(8), 5455-5516. https://doi.org/10.48550/arXiv.1901.06032

Khan, S., & Iqbal, R. (2025). *A comprehensive survey on architectural advances in deep CNNs: Challenges, applications, and emerging research directions.* arXiv. https://doi.org/10.48550/arXiv.2503.16546

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84-90. https://doi.org/10.1145/3065386

Li, C., Du, D., Zhang, L., Wen, S., Qi, M. & Pang, K. (2019). *Data priming network for automatic check-out.* arXiv. https://doi.org/10.48550/arXiv.1904.04978

Liu, T., Chen, J., & Li, X. (2022). *Research on image classification based on convolutional neural network.* 2022 International Conference on Applied Robotics, Computing and Engineering (ICARCE), pp. 1-4, Wuhan, China. https://doi.org/10.1109/ICARCE55724.2022.10046634

Naga Navya, T., Harshini, V., Prasanna, T. D., Farjana, S., & Deepthi, M. (2024). Real time object detection using YOLO. *Journal of Emerging Technologies and Innovative Research*, *11*(11), 373-380. https://www.jetir.org/papers/JETIRGO06038.pdf

Panda, C. (2019). Object detection and tracking using Faster R-CNN. *International Journal of Recent Technology and Engineering*, *8*(3), 4894-4900. https://doi.org/10.35940/ijrte.C5580.098319

Pawar, S., Jadhav, D., Godse, D., Jadhav, R., & Thakur, S. (2024). Vision-based empty shelf detection in retail with real-time telegram notifications for efficient restocking. *International Journal of Electronics and Communication Engineering*, *11*(7), 180-187. https://doi.org/10.14445/23488549/IJECE-V11I7P118

Rangel, G., Cuevas-Tello, J. C., Nunez-Varela, J., Puente, C., & Silva-Trujillo, A. (2024). A survey on convolutional neural networks and their performance limitations in image recognition tasks. *Journal of Sensors*, *2024*(1), 1-29. https://doi.org/10.1155/2024/2797320

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779–788, Vegas, NV, USA. https://doi.org/10.1109/CVPR.2016.91

Ren, S., He, K., Girshick, R. B., & Sun, J. (2017). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(6), 1137-1149. https://doi.org/10.1109/TPAMI.2016.2577031

Scodellaro, R., Kulkarni, A., Alves, F., & Schröter, M. (2025). Training convolutional neural networks with the Forward–Forward Algorithm. *Scientific Reports*, *15*, Article 38461. https://doi.org/10.48550/arXiv.2312.14924

Shobayo, O., & Saatchi, R. (2025). Developments in deep learning artificial neural network techniques for medical image analysis and interpretation. *Diagnostics*, *15*(9), Article 1072. https://doi.org/10.3390/diagnostics15091072

Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *Proceedings of the International Conference on Learning Representations (ICLR).* https://doi.org/10.48550/arXiv.1409.1556

Tan, M., & Le, Q. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. *Proceedings of the 36th International Conference on Machine Learning (ICML)*, 97, 6105-6114. https://doi.org/10.48550/arXiv.1905.11946

Touvron, H., Cord, M., Douze, M., Massa, F., Synnaeve, G., & Jégou, H. (2021). Training data-efficient image transformers & distillation through attention. *Proceedings of the 38th International Conference on Machine Learning (ICML),* 10347-10357. https://doi.org/10.48550/arXiv.2012.12877

Veerasamy, S., Athisanthiya, R., Balanivetha, M., & Bairavi, S. (2024). Retail shelf monitoring for product availability. *SSRN Electronic Journal*, Article 4785695. https://doi.org/10.2139/ssrn.4785695

Veeru, B., Chaithanya, K., Likhitha, B., Aravind, A., Ashritha, G., Aravind, G., & Chandu, G. (2024). Real time object detection using YOLO. *International Journal of Innovations in Engineering and Education, 15*(5), 656-667. http://ijiee.org/index.php/ijiee/article/view/941/924

Wei, Y., Tran, S., Xu, S., Kang, B., & Springer, M. (2020). Deep learning for retail product recognition: Challenges and techniques. *Computational Intelligence and Neuroscience*, *2020*(1), Article 8875910. https://doi.org/10.1155/2020/8875910