ISTP

Leveraging Transfer Learning for Durian Variety Classification

Voo Nyuk Mee,[1] **Lim Tong Ming,**[1] **and Lim Yee Mei**[2]

[1]Tunku Abdul Rahman University of Management and Technology, Kuala Lumpur, Malaysia

[2]OMIS Consulting Sdn Bhd, Kuala Lumpur, Malaysia

Abstract: This study evaluates the effectiveness of transfer learning compared to training a Convolutional Neural Network (CNN) from scratch for classifying eight durian varieties. A CNN with two convolutional layers, two max-pooling layers, and three fully connected layers improves accuracy from 20% to 68.79%, reduces average loss from 2.0 to 0.58, and requires 86 min and 17 s for training. In contrast, transfer learning with ResNet-18 significantly enhances both performance and efficiency. Fine-tuning the pre-trained model increases accuracy from 52% to 94%, stabilizing at 94–96%, reduces average loss from 1.36 to 0.14, and completes training in just 10 min and 20 s. Using ResNet-18 as a fixed feature extractor improves accuracy from 42% to 78%, stabilizing at 78–80%, reduces average loss from 1.68 to 0.68, and completes training in 10 min and 13 s. This study introduces an application-specific adaptation of transfer learning for durian variety classification, bridging a gap in agricultural image analysis research. The findings demonstrate how leveraging pre-trained feature representations enables rapid convergence and robust classification, even with limited data. Additionally, this research provides a foundation for automated durian grading, sorting, and quality assessment, contributing to precision agriculture and smart farming. By benchmarking transfer learning against traditional CNNs, this study highlights the superior efficiency and accuracy of transfer learning for agricultural applications, with broader implications for fruit classification, yield optimization, and post-harvest processing automation.

Keywords: CNN; durian variety; image classification; ResNet; transfer learning

I. INTRODUCTION

Image classification is a fundamental aspect of computer vision, with applications in agriculture [1–7], healthcare [8–13], and autonomous systems [14–16]. The advent of deep learning transforms this field, with CNNs [17] playing a crucial role in achieving remarkable performance in image-based tasks. However, while traditional CNNs are highly effective, they often require large amounts of labeled data and substantial computational resources for training from scratch [18,19]. These requirements pose significant challenges in domains where labeled datasets are limited or expensive to curate.

Transfer learning [20,21] emerges as a promising solution to overcome these limitations. By leveraging pre-trained models on large benchmark datasets such as ImageNet [22], transfer learning allows models to utilize learned feature representations and adapt them to specific tasks with minimal training [23]. This approach is particularly beneficial when the target dataset is small [24], as it reduces training time and computational demands while improving key performance metrics such as accuracy and loss [25,26].

This study compares the effectiveness of a traditional CNN architecture with transfer learning techniques for classifying eight durian varieties. The CNN, consisting of two convolutional layers, two max-pooling layers, and three fully connected layers, serves as the baseline. Transfer learning is applied using the pre-trained ResNet-18 model under two scenarios: fine-tuning the entire model and using it as a fixed feature extractor. The study evaluates both approaches based on accuracy, F1 score, and training time,

providing a comprehensive analysis of their performance and efficiency.

Beyond comparing CNN and transfer learning, this study introduces Application-Specific Transfer Learning (ASTL) for durian variety classification, a novel adaptation that optimizes pre-trained models specifically for agricultural applications. This research tailors transfer learning to capture subtle visual variations between durian varieties, improving classification robustness in real-world agricultural settings. Furthermore, the study explores the impact of transfer learning on automating fruit classification for precision agriculture, with implications for automated grading, sorting, and yield estimation. By demonstrating how transfer learning reduces computational requirements while maintaining high accuracy, this research paves the way for scalable and cost-effective AI-driven solutions in agricultural automation.

This study contributes to advancing deep learning applications in agriculture by addressing dataset limitations, improving classification efficiency, and highlighting the practical deployment of transfer learning in smart farming systems.

II. LITERATURE REVIEW

The growing demand for automated image classification across diverse domains has propelled research in deep learning, particularly in CNNs. CNNs have become the cornerstone of modern image processing, demonstrating their robustness in capturing local and hierarchical features through convolutional layers, activation functions, and pooling operations. Early breakthroughs, such as LeNet [17] and AlexNet [27], laid the foundation for increasingly complex architectures like ResNet [20] and DenseNet [21]. These models have achieved outstanding performance in benchmarks

Corresponding author: Voo Nyuk Mee (e-mail: voonm@tarc.edu.my).

such as ImageNet and COCO. However, their reliance on large labeled datasets and substantial computational resources poses challenges when applied to niche domains with limited data availability.

To address these challenges, transfer learning has emerged as an effective solution. Pre-trained architectures such as ResNet, VGG [28] and Inception [29] have demonstrated exceptional performance in various image classification tasks. Transfer learning leverages models that have been pre-trained on large-scale datasets like ImageNet to initialize weights, which can then be adapted for specific tasks. This method reduces the time and computational cost required for training while improving performance on tasks with insufficient data. As highlighted by Yosinski *et al.* [23], transfer learning enables models to reuse generalizable features learned from source tasks, facilitating efficient fine-tuning for target applications. For instance, He *et al.* [20] demonstrated the effectiveness of ResNet's pre-trained features for diverse downstream tasks, emphasizing the depth and flexibility of residual learning. Studies by Pan and Yang [25] and Tan *et al.* [26] have further highlighted the efficiency of transfer learning in achieving higher accuracy and faster convergence compared to training networks from scratch. Plested and Gedeon [24] explored deep transfer learning for image classification, emphasizing its effectiveness in scenarios with limited labeled data. Their study reinforced the idea that transfer learning not only enhances performance but also significantly reduces the computational resources required for training models from scratch.

In the agricultural domain, where datasets are often limited and class imbalance is common, transfer learning has proven particularly useful. Studies such as those by Kamilaris and Prenafeta-Boldú [30] have shown that pre-trained networks can effectively classify crops, detect diseases, and monitor plant growth with high accuracy. Similarly, research on food classification, including fruit and vegetable identification [31], has demonstrated that transfer learning outperforms both traditional machine learning methods and CNNs trained from scratch. The ability to extract relevant features without extensive retraining aligns well with the constraints of agricultural datasets.

Transfer learning has also had a significant impact on the medical imaging domain. A review by Morid *et al.* [32] examined the application of ImageNet-pretrained models for medical image analysis, underscoring the versatility of transfer learning across various medical imaging tasks. Their study highlighted improvements in both accuracy and computational efficiency. Additionally, Kursun *et al.* [33] introduced Contextually Guided Convolutional Neural Networks (CG-CNNs) for learning transferable representations. Their approach focused on developing broad-purpose representations in shallow CNNs trained on limited datasets, enhancing their applicability to new classification tasks without additional training.

Two primary strategies dominate transfer learning applications: fine-tuning and feature extraction. Fine-tuning allows weight updates across the entire network or specific layers, enabling the model to adapt to the specific task. In contrast, feature extraction freezes the pre-trained weights, leveraging them as a static feature extractor while training only the final classifier. Both methods have demonstrated substantial improvements in image classification tasks, particularly in domains with limited training data, such as medical imaging [34] and agriculture [30]. Raghu *et al.* [34] emphasized that fine-tuning often leads to better performance in complex tasks, while feature extraction offers simplicity and computational efficiency.

While transfer learning has been extensively studied, few works have directly compared its performance against traditional CNNs trained from scratch on domain-specific datasets. This study aims to bridge this gap by evaluating the performance of a simple CNN and transfer learning approaches using ResNet for the classification of eight durian varieties, focusing on accuracy, F1 score, and training efficiency. By analyzing these methods, this research provides insights into the trade-offs between traditional CNNs and transfer learning techniques in practical applications. The reviewed studies reinforce the significance of transfer learning in image classification, particularly in fields constrained by data limitations. The ability to leverage pre-trained models not only accelerates the training process but also enhances the generalization capabilities of models across diverse tasks. In summary, the convergence of CNN architectures and transfer learning continues to drive progress in image classification, offering practical solutions to challenges posed by limited data availability and computational constraints.

III. DATASET PREPARATION

The first step involved collecting images of different durian varieties from farms, stalls, and distribution centers. A total of 13 varieties and 6,488 images were gathered. The collected durian varieties are shown in Table I. Each image was manually labeled with its corresponding durian variety, with the assistance of durian owners to ensure accurate classification. For this study, only eight durian varieties and 1,978 images were selected to maintain an equal number of images per variety, ranging between 100 and 300. The selected durian dataset varieties are presented in Table II. To address class imbalance, 50 images from Kaggle [35] were incorporated into the D101 class, bringing its total to 123 images. This adjustment helped mitigate class imbalance [36–38], which could otherwise lead to bias toward the majority classes (D197 and D24) and result in poor performance on underrepresented classes.

To enhance the diversity of the training data and optimize the model's generalization capability, data augmentation techniques were applied. These techniques included scaling and cropping. Images were resized while preserving their aspect ratio to introduce variations in scale. Additionally, portions of the images were cropped to focus on the fruit, further improving the model's ability to learn distinguishing features.

Table I.   Collected Durian varieties

D197	D160	D7	D144	D145
2895	477	6	7	7
D2	D13	D18	D24	D101
319	7	8	1877	73
D175	D224	D200		
141	214	457		

Table II.   Durian dataset

D2	D24	D101	D160	D175
300	300	123	300	141
D197	D200	D224		
300	300	214		

[D175, D200, D2, D24]

Fig. 1. Sample Durian.

Table III. The ResNet-18 architecture

Layer Name	Kernel Size	Stride	Padding	Output
Conv1	7×7	2	3	64
Layer 1	3×3	1	1	64
Layer 2	3×3	2	1	128
Layer 3	3×3	2	1	256
Layer 4	3×3	2	1	512
AvgPool	–	–	–	-
FC Layer	–	–	–	8

Subsequently, the dataset was partitioned into a training set and a validation set. The training set constituted the largest subset, representing 70% of the total dataset, and was used for model training. The validation set, representing 30% of the total dataset, was employed to fine-tune hyperparameters and assess the model's performance during training. Prior to feeding the images into the model, several preprocessing steps were applied. Normalization was performed by scaling pixel values to a range of [0,1], facilitating faster model convergence. Images were resized to ensure uniform dimensions, a prerequisite for processing by neural networks. Additionally, the dataset was adjusted using the mean and standard deviation to normalize across the training and validation sets.

The dataset was organized to facilitate efficient access and usage. A directory structure was created to clearly separate the training and validation sets, with subdirectories for each durian class. Various tools and platforms were utilized to streamline the dataset preparation process. Data augmentation libraries, including TensorFlow's image preprocessing utilities and torchvision's image transformation utilities, were employed. Cloud storage and a Docker container were used for managing datasets and training the classification models. Sample durian images are illustrated in Fig. 1.

IV. METHODOLOGY

A. Constructing Models with Transfer Learning Using ResNet-18

Two models utilizing transfer learning were constructed to classify eight durian varieties. Transfer learning is a machine learning technique that enables a pre-trained model to be adapted for a new classification task. This approach is particularly beneficial when working with limited datasets, as it allows the model to leverage knowledge acquired from large-scale datasets such as ImageNet. One of the most used architectures for transfer learning in image classification is ResNet-18, a variant of the Residual Network (ResNet) family.

ResNet-18 [20,39] is a deep CNN consisting of 18 layers, including 16 weight layers (convolutional layers with trainable parameters) and 2 fully connected (FC) layers for classification. The architecture is designed to address the vanishing gradient problem in deep networks by incorporating residual connections. These connections allow gradients to flow more effectively during backpropagation, facilitating efficient training of deeper networks.

The network takes an input image of size $224 \times 224 \times 3$, which is resized and normalized before being fed into the model. The first layer consists of:

- A 7×7 convolutional layer with 64 filters and a stride of 2, which helps capture low-level features.

- Batch normalization and ReLU activation, which stabilize and accelerate training.

- A 3×3 max pooling layer (stride = 2) follows, reducing the spatial dimensions and extracting dominant features.

ResNet-18 is structured into four stages of residual blocks, each containing two 3×3 convolutional layers, followed by batch normalization and ReLU activation. The key characteristic of these blocks is the identity mapping, which bypasses one or more layers, ensuring efficient gradient propagation. The number of filters in the convolutional layer doubles at each stage: stage 1 – 64 filters, stage 2 – 128 filters, stage 3 – 256 filters, and stage 4 – 512 filters. These layers extract high-level features from images, capturing patterns, textures, and edges that define different categories.

Following the residual blocks, a global average pooling (GAP) [40,41] layer is applied, reducing the spatial dimensions of the feature maps to a single value per filter. This step prevents overfitting by significantly reducing the number of trainable parameters while retaining crucial information. The feature maps are then flattened and passed through a fully connected layer for classification. The architecture of the model with ResNet-18 is presented in Table III.

1) ResNet-18 as FIXED FEATURE EXTRACTOR. To adapt ResNet-18 for durian variety classification, the final FC layer, originally designed for ImageNet's 1,000 classes, was replaced with a new classification layer for the eight durian varieties. The transfer learning model was constructed by loading the ResNet-18 model pre-trained on ImageNet. The convolutional layers were frozen to retain pre-learned feature extraction capabilities, preventing gradient updates for these layers. This significantly reduced computational cost and accelerated training.

The final fully connected layer of ResNet-18 was removed and replaced with a new one with randomly initialized weights. This new FC layer was trained from scratch, learning specific features for distinguishing between the eight durian varieties. A softmax activation function was applied in the output layer to produce probability distributions for classification. The training was configured with a batch size of 16, 100 epochs for sufficient convergence, Stochastic Gradient Descent (SGD) optimization algorithm, and cross-entropy loss for multi-class classification.

The approach was evaluated in terms of accuracy, F1 score, loss, and training time to assess its performance and efficiency in classifying eight durian varieties. The progress of the results was displayed at each epoch, with sample results presented in Fig. 2. The loss plot is shown in Fig. 3, while the accuracy and F1 score plot is presented in Fig. 4.

The results demonstrated an increase in accuracy from 42% to 78%, stabilizing at 78–80%, with a reduction in average loss from 1.68 to 0.68, and completing training in just 10 minutes and 13 s.

```
Epoch 99/99
----------
train Loss: 0.6883 Acc: 0.7884
Accuracy of the network on the 1385 train imag
Accuracy of ['D101', 'D160', 'D175', 'D197', '
4141, 0.8286, 0.8095, 0.8905, 0.9067, 0.8190],
        device='cuda:0')
f1 score of the network on the 1385 train imag
f1 score of ['D101', 'D160', 'D175', 'D197', '
5000, 0.8131, 0.7925, 0.8558, 0.8635, 0.8056],
        device='cuda:0')
```
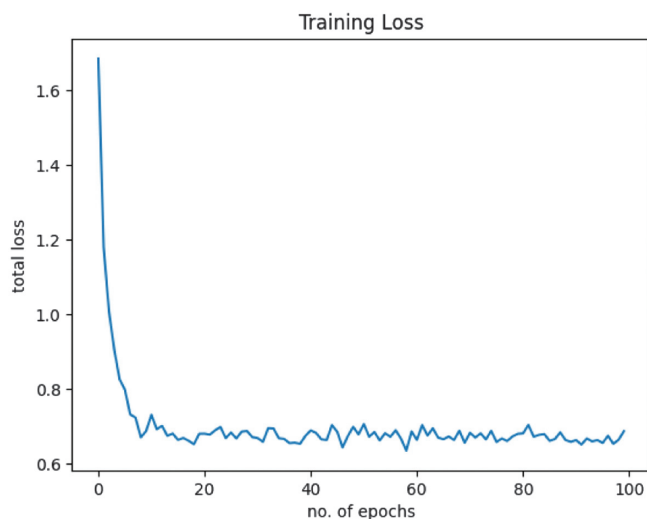
Fig. 2. Sample epoch results.

Fig. 3. Total loss.

Fig. 4. Total accuracy and F1 score.

2) FINE-TUNING ResNet-18 MODEL. The ResNet-18 model was first loaded with its pre-trained weights. These weights contained valuable knowledge obtained from the large-scale ImageNet

```
Epoch 99/99
----------
train Loss: 0.1440 Acc: 0.9581
Accuracy of the network on the 1385 train imag
Accuracy of ['D101', 'D160', 'D175', 'D197', 'I
8687, 0.9619, 0.9810, 1.0000, 0.9867, 0.9524],
        device='cuda:0')
f1 score of the network on the 1385 train imag
f1 score of ['D101', 'D160', 'D175', 'D197', 'I
8821, 0.9665, 0.9717, 0.9976, 0.9801, 0.9569],
        device='cuda:0')
```

Fig. 5. Sample epoch results.

dataset, making them highly useful for extracting general image features such as edges, textures, and shapes. Since the original FC layer in ResNet-18 was designed for 1,000 ImageNet classes, it did not match the number of target categories in the new dataset. Therefore, this final FC layer was removed and replaced with a new one that had the appropriate number of output nodes (corresponding to the eight durian varieties to classify).

Instead of using random initialization, the entire model except for the new fully connected layer was initialized with the pre-trained weights from ResNet-18. This allowed the model to retain general feature extraction capabilities while only learning a new classification mapping. Once the new fully connected layer was added, the rest of the training process followed a standard deep learning workflow. The training was conducted with a batch size of 16, cross-entropy loss is used for multi-class classification, SGD optimization and 100 epochs to ensure convergence.

The approach was evaluated in terms of accuracy, F1 score, loss, and training time to assess its performance and efficiency in classifying eight durian varieties. The progress of the results was displayed at each epoch, with sample results presented in Fig. 5. The loss plot is shown in Fig. 6, while the accuracy and F1 score plot is presented in Fig. 7.

The results demonstrated an increase in accuracy from 52% to 94%, stabilizing at 94–96%, with an average loss reduction from 1.36 to 0.14. The model completed training in just 10 minutes and 20 s.
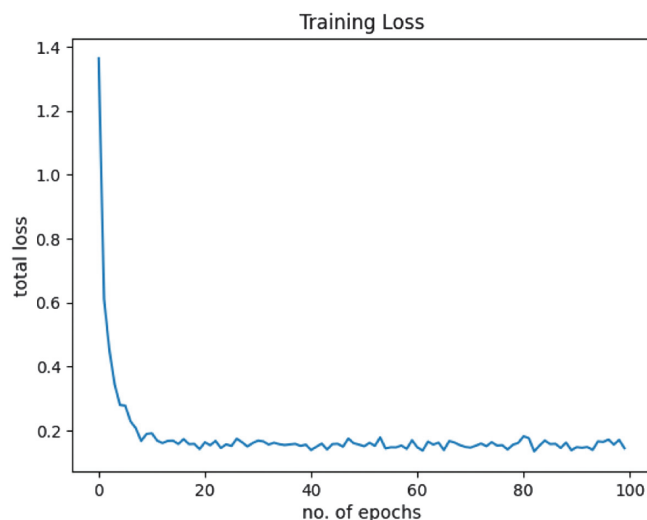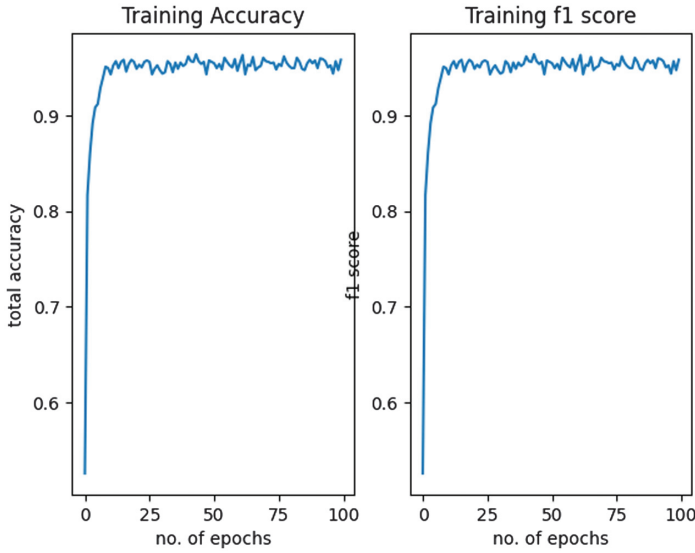

Fig. 6. Total loss.

Fig. 7. Total accuracy and F1 score.

B. CONSTRUCTING A BASIC CNN

A simple CNN was developed as a baseline to compare the effectiveness of a traditional CNN architecture with transfer learning techniques for classifying eight durian varieties. The CNN consisted of two convolutional layers, two max-pooling layers, and three fully connected layers. The network utilized ReLU activation functions in the hidden layers and a softmax classifier in the output layer. The architecture of the CNN is presented in Table IV.

Each convolutional layer contained 24 kernels or filters, which learned the spatial hierarchies of features from the input data and generated a feature map. These kernels were learned during training, enabling the network to recognize specific patterns in images, such as edges, textures, and other relevant features. After convolution, the feature maps passed through ReLU activation functions, introducing non-linearity into the model. This non-linearity allowed the network to learn more complex patterns. Max pooling layers were employed to down-sample the feature maps, reducing the spatial dimensions of the data while retaining the most important information.

Following the convolutional and pooling layers, the feature maps were flattened into a vector and passed through fully connected layers. These layers performed the final classification by combining the high-level features learned in the previous layers. The output layer used the softmax function to generate a probability distribution over the possible classes: D2, D24, D101, D160, D175, D197, D200, and D224. The class with the highest probability was selected as the predicted label.

The model was trained for 100 epochs. The training data was loaded into a training data loader, with a batch size set to 16. The classifier was trained using the SGD optimization technique, which aimed to minimize the cross-entropy loss. During the training phase, the weights W and biases b were adjusted to reduce the discrepancy between the predicted probabilities and the true labels. The ReLU activation functions transformed the summed weighted inputs from each node into the node's activation or output.

The performance of the model was evaluated on a separate test set to determine its accuracy and generalizability. The trained model was assessed based on accuracy, F1 score, loss, and training time to evaluate its effectiveness and efficiency in classifying the eight durian varieties. The results were displayed at each epoch, with corresponding sample outputs presented in Fig. 8. The loss plot is shown in Fig. 9, while the accuracy and F1 score plots are illustrated in Fig. 10. The final results demonstrated an accuracy and F1 score of 68.79%, an average loss of 0.58, and a total training time of 86 min and 17 s.
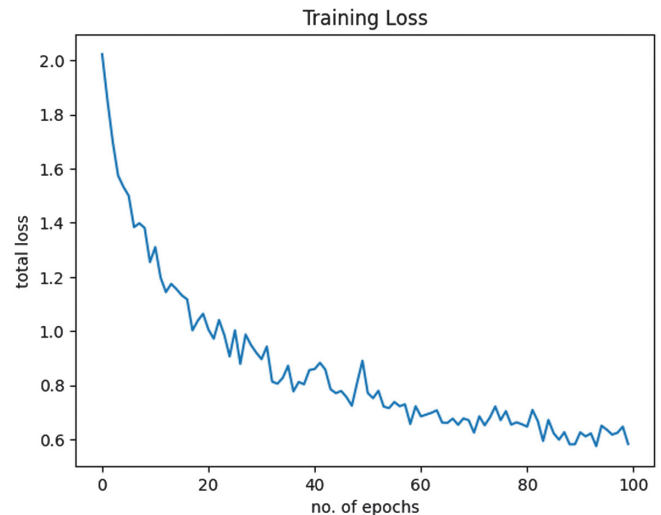
```
Epoch 100
--------------------------------
Accuracy of the network on the 1385 train images:
Accuracy of ['D101', 'D160', 'D175', 'D197', 'D2'
3423, 0.7650, 0.6444, 0.8637, 0.8597, 0.6612])
f1 score of the network on the 1385 train images:
f1 score of ['D101', 'D160', 'D175', 'D197', 'D2'
4287, 0.7564, 0.6257, 0.8222, 0.8096, 0.6570])
Avg loss: 0.583445

Finished Training
Training complete in 86m 17s
Best val Acc: 0.687928
```

Fig. 8. Sample epoch results.

Table IV. The CNN architecture

Layers	Function	Output
Convolution 1	ReLU	16,6,201,201
Subsampling 1	–	16,6,100,100
Convolution 2	ReLU	16,16,77,77
Subsampling 2	–	16,16,38,38
Flatten	–	16,23104
Fully Connected 1	ReLU	16,5776
Fully Connected 2	ReLU	16,722
Fully Connected 3	ReLU	16,8
Output	Softmax	8 classes

Fig. 9. Total loss.
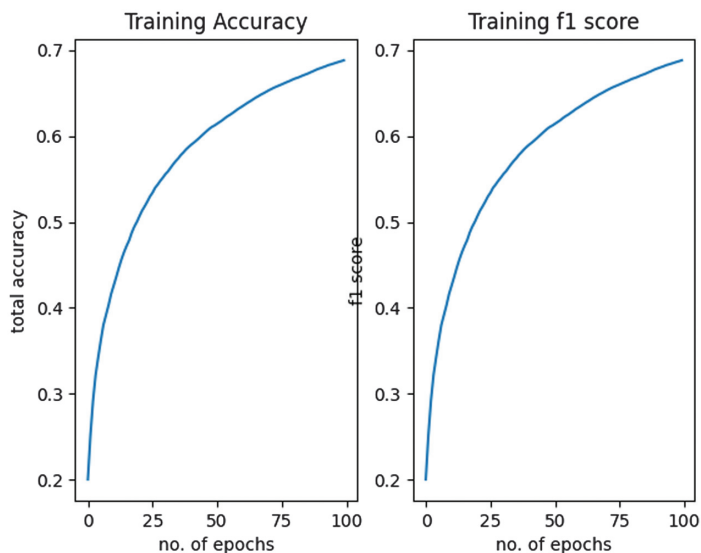
Fig. 10. Total accuracy and F1 score.

V. APPLICATION-SPECIFIC TRANSFER LEARNING FOR DURIAN CLASSIFICATION

Durian has been a high-value tropical fruit with numerous varieties exhibiting subtle visual differences. Traditional classification methods relied on expert knowledge, manual sorting, or standard machine learning techniques, which often required extensive labeled data and significant computational resources. This study leveraged transfer learning, specifically fine-tuning ResNet-18, to enhance durian variety classification, demonstrating how pre-trained models effectively distinguished between visually similar fruit categories.

By transferring knowledge from large-scale datasets like ImageNet, which contained millions of general images, the model initially learned to recognize shape, texture, and color variations. Fine-tuning ResNet-18 enabled the network to specialize in durian-specific features, significantly improving accuracy while reducing training time compared to training a CNN from scratch. This approach bridged the gap between general deep learning applications and specialized agricultural use cases, making AI-powered fruit classification more accessible and efficient.

The implementation of transfer learning in durian classification had broader implications for agricultural automation. It could be integrated into smart sorting systems that classified durians based on variety, ripeness, and quality. Automated conveyor belt systems equipped with AI-powered cameras rapidly sorted large batches of durians, reducing human dependency, improving efficiency, and ensuring consistent grading.

Accurate classification of durian varieties also benefited supply chain management by ensuring that specific varieties were sent to appropriate markets. Some varieties were more desirable for export, while others were preferred for local consumption. By automating variety identification, farmers and exporters optimized pricing and distribution logistics. When combined with drone or robotic vision systems, transfer learning-based classification assisted in real-time fruit detection, helping farmers estimate yield and determine the optimal harvesting time based on variety-specific maturity signs.

The application of transfer learning in durian classification represented a pioneering step toward automated and intelligent agriculture. By demonstrating how a pre-trained model like ResNet-18 could be fine-tuned for specialized fruit classification tasks, this study laid the groundwork for broader applications in automated sorting, supply chain optimization, precision farming, and smart harvesting systems. As AI technology continued to evolve, transfer learning played a crucial role in enhancing agricultural efficiency, reducing labor costs, and improving food quality assurance.

VI. DISCUSSION

This study investigated the effectiveness of leveraging transfer learning compared to training a traditional CNN from scratch for classifying eight durian varieties. The findings revealed significant improvements in both performance and training efficiency when utilizing transfer learning with ResNet-18. The baseline CNN model consisted of two convolutional layers for feature extraction, two max-pooling layers for dimensionality reduction, and three fully connected layers for classification. Throughout the training process, this model improved its accuracy from 20% to 68.79%, reduced the average loss from 2.0 to 0.58, and completed training in 86 min and 17 s. While these results demonstrated that the CNN model could learn meaningful features over time, its accuracy remained lower, and its training time was significantly longer compared to transfer learning-based models.

Transfer learning leveraged a pre-trained ResNet-18 model, which had already learned general image features from ImageNet, enabling efficient adaptation to a new classification task with a smaller dataset. In this study, two transfer learning approaches were evaluated: fine-tuning the pre-trained model and using ResNet-18 as a fixed feature extractor. Fine-tuning involved training the entire ResNet-18 model, allowing both convolutional and fully connected layers to update their weights based on the new dataset. This approach led to a significant improvement in accuracy, increasing from 52% to 94%, and stabilizing at 94–96%, while the average loss decreased from 1.36 to 0.14. The training process was highly efficient, completing in just 10 minutes and 20 s. These results highlighted the effectiveness of fine-tuning in adapting pre-trained models to new tasks, particularly when the target dataset shares similarities with the original ImageNet dataset.

In contrast, using ResNet-18 as a fixed feature extractor involved freezing all convolutional layers and updating only the final fully connected layer. This method improved accuracy from 42% to 78%, stabilizing at 78–80%, while reducing the average loss from 1.68 to 0.68. The training time was slightly shorter at 10 min and 13 s. While computationally more efficient, this approach yielded lower accuracy compared to fine-tuning, suggesting that freezing the convolutional layers limited the model's ability to capture fine-grained distinctions among durian varieties. Nevertheless, this method remained a viable option when computational resources were limited or when rapid deployment was required. The comparison demonstrated that while both approaches benefited from transfer learning, fine-tuning provided superior performance by allowing the model to refine its feature representations based on the new dataset.

These findings suggested that transfer learning, particularly fine-tuning, offered a superior balance of accuracy and efficiency compared to training CNN models from scratch, making it a highly effective approach for durian variety classification. Additionally,

Table V.   Comparison table

Model	CNN	ResNet-18 Fine-tuned	ResNet-18 Fixed Features
Accuracy	20% → 68.79%	52% → 94%	42% → 78%
Stability	No clear stability	94–96%	78–80%
Loss	2.0 → 0.58	1.36 → 0.14	1.68 → 0.68
Training	86 min 17 s	10 min 20 s	10 min 13 s

this study established a foundation for integrating AI-driven fruit classification into broader agricultural automation systems, paving the way for scalable, intelligent solutions in precision farming. The comparison is summarized in the following Table V.

VI. CONCLUSION

The results indicated that transfer learning significantly outperformed training a CNN from scratch in both accuracy and training efficiency. Fine-tuning ResNet-18 achieved the best performance, yielding the highest accuracy and the lowest loss in the shortest training time. The fixed feature extractor approach was more computationally efficient but resulted in lower accuracy, suggesting that some degree of model adaptation is necessary for optimal performance. Therefore, leveraging pre-trained models, particularly through fine-tuning, proved to be highly beneficial for complex classification tasks such as durian variety classification.

Beyond the performance comparison between CNN and transfer learning, this study introduces ASTL for durian classification, a novel approach that optimized deep learning models for agricultural applications. This study tailored transfer learning to distinguish fine-grained durian varieties, addressing a critical need in the durian industry for automated quality control, sorting, and authentication. Furthermore, this study demonstrated that transfer learning can extend beyond classification to support agricultural automation, including intelligent fruit sorting, yield estimation, and supply chain optimization. The findings highlighted the scalability and accessibility of AI-driven solutions in precision agriculture, reducing reliance on large labelled datasets and making advanced classification systems feasible for real-world deployment.

This study established a foundation for integrating AI into durian classification and agricultural automation, offering a scalable, efficient, and cost-effective solution for fruit classification and supply chain optimization. Future research could explore expanding this approach to other agricultural products, multi-modal learning, and real-time deployment in smart farming environments.

ACKNOWLEDGEMENTS

CONFLICT OF INTEREST STATEMENT

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

REFERENCES

[1] K. M. Abubeker, Abhijit, S. Akhil, V. K. A. Kumar, and B. K. Jose, "Computer vision-assisted real-time bird eye chili classification using YOLO v5 framework," *J. Artif. Intell. Technol.*, vol. 4, no. 3, pp. 265–271, Nov. 2023. [Online]. Available: https://ojs.istp-press.com/jait/article/view/251.

[2] P. K. Yadav, T. Burks, Q. Frederick, J. Qin, M. Kim, and M. A. Ritenour, "Citrus disease detection using convolution neural network generated features and softmax classifier on hyperspectral image data," *Front. Plant Sci.*, vol. 13, pp. 1 043 712–1 043 712, 2022.

[3] M. G. Lim and J. H. Chuah, "Durian types recognition using deep learning techniques," in *2018 9th IEEE Control Syst. Grad. Res. Colloquium (ICSGRC)*, IEEE, 2018, pp. 183–187.

[4] J. N. Uy and J. F. Villaverde, "A durian variety identifier using canny edge and CNN," in *2021 IEEE 7th Int. Conf. Control Sci. Syst. Eng. (ICCSSE)*, IEEE, 2021, pp. 293–297.

[5] N. M. Zarifie Hashim, M. H. Azfar Khairil Bahri, S. M. Abd Ghani, M. D. Sulistiyo, K. A. Mohd Kassim, and N. A. Hanin Zahri, "An introduction to a smart durian Musang king and durian Kampung classification," in *2022 2nd Int. Conf. Intell. Technol. (CONIT)*, IEEE, 2022, pp. 1–6.

[6] Z. W. Adedeji Olugboja and Y. Sun, "Parallel convolutional neural networks for object detection," *J. Adv. Inf. Technol.*, vol. 12, no. 4, pp. 279–286, 2021. [Online]. Available: https://www.jait.us/uploadfile/2021/1029/20211029032000786.pdf

[7] Z. W. Michael, J. Norval, and Y. Sun, "Evaluation of image processing technologies for pulmonary tuberculosis detection based on deep learning convolutional neural networks," *J. Adv. Inf. Technol.*, vol. 12, no. 3, pp. 253–259, 2021. [Online]. Available: https://www.jait.us/uploadfile/2021/0719/20210719054403677.pdf.

[8] A.-K. Balve and P. Hendrix, "Interpretable breast cancer classification using CNNs on mammographic images," 2024. [Online]. Available: https://arxiv.org/abs/2408.13154.

[9] S. R. Krishna, N. Cherukuri, Y. D. Kumar, R. Jayakarthik, B. Nagarajan, A. Balaram, and G. D. Jyothi, "Convolutional neural networks for automated diagnosis of diabetic retinopathy in fundus images," *J. Artif. Intell. Technol.*, vol. 3, no. 4, pp. 205–214, Aug. 2023. [Online]. Available: https://ojs.istp-press.com/jait/article/view/264.

[10] F. Gao, B. Li, L. Chen, Z. Shang, X. Wei, and C. He, "A softmax classifier for high-precision classification of ultrasonic similar signals," *Ultrasonics*, vol. 112, p. 106 344, 2021.

[11] H. M. Ahmed and M. Y. Kashmola, "A proposed architecture for convolutional neural networks to detect skin cancers," *IAES Int. J. Artif. Intell. (IJ-AI)*, vol. 11, no. 2, pp. 485–493, 2022. [Online]. Available: https://ijai.iaescore.com/index.php/IJAI/article/view/21135.

[12] S. Jamdar, J. Vaddin, S. B. Nargundkar, and S. Patil, "Deep learning method for lung cancer identification and classification," *IAES Int. J. Artif. Intell. (IJ-AI)*, vol. 13, no. 1, pp. 1119–1128, 2024. [Online].

Available: https://ijai.iaescore.com/index.php/IJAI/article/view/22474.

[13] P. J. Nicholas, A. To, O. Tanglay, I. M. Young, M. E. Sughrue, and S. Doyen, "Using a resnet-18 network to detect features of Alzheimer's disease on functional magnetic resonance imaging: A failed replication. Comment on Odusami et al. analysis of features of Alzheimer's disease: Detection of early stage from functional brain changes in magnetic resonance images using a finetuned resnet18 network. Diagnostics 2021, 11, 1071," *Diagnostics*, vol. 12, no. 5, 2022. [Online]. Available: https://www.mdpi.com/2075-418/12/5/1094.

[14] A. A. Golroudbari and M. H. Sabour, "Recent advancements in deep learning applications and methods for autonomous navigation: a comprehensive review," 2023. [Online]. Available: https://arxiv.org/abs/2302.11089.

[15] S. Pavlitska, N. Lambing, A. K. Bangaru, and J. M. Zöllner, "Traffic light recognition using convolutional neural networks: a survey," 2023. [Online]. Available: https://arxiv.org/abs/2309.02158.

[16] H.-Y. Lin, C.-K. Chang, and V. L. Tran, "Lane detection networks based on deep neural networks and temporal information," *Alexandr. Eng. J.*, vol. 98, pp. 10–18, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1110016824004009.

[17] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to hand-written zip code recognition," *Neural Comput.*, vol. 1, no. 4, pp. 541–551, Winter 1989.

[18] S. Yang, W. Xiao, M. Zhang, S. Guo, J. Zhao, and F. Shen, "Image data augmentation for deep learning: a survey," 2023. [Online]. Available: https://arxiv.org/abs/2204.08610.

[19] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, no. 1, p. 60, Jul. 2019. [Online]. Available: https://doi.org/10.1186/s40537-019-0197-0.

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recogn. (CVPR)*, IEEE, 2016, pp. 770–778.

[21] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recogn. (CVPR)*, IEEE, 2017, pp. 4700–4708.

[22] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conf. Comput. Vis. Pattern Recogn.*, IEEE, 2009, pp. 248–255.

[23] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?," in *Proc. 28th Int. Conf. Neural Inf. Process. Syst. – Vol. 2*, vol. 27, Cambridge, MA: MIT Press, 2014, pp. 3320–3328.

[24] J. Plested and T. Gedeon, "Deep transfer learning for image classification: a survey," ArXiv, vol. abs/2205.09904, 2022. [Online]. Available: https://arxiv.org/abs/2205.09904. Accessed on: January 23, 2025.

[25] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, 2010.

[26] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, "A survey on deep transfer learning," 2018. [Online]. Available: https://arxiv.org/abs/1808.01974.

[27] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. 25th Int. Conf. Neural Inf. Process. Syst. – Vol. 1, Ser. NIPS'12*, Red Hook, NY, USA: Curran Associates Inc., 2012, pp. 1097–1105.

[28] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2015. [Online]. Available: https://arxiv.org/abs/1409.1556.

[29] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recogn.*, IEEE, 2015, pp. 1–9.

[30] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: a survey," *Comput. Electron. Agric.*, vol. 147, pp. 70–90, 2018.

[31] A. Dkhar and R. K. Bania, "Fruit and vegetable recognition system using deep transfer learning approach," in *Proceedings of the NIELIT's International Conference on Communication, Electronics and Digital Technology*, I. Woungang, S. K. Dhurandher, and Y. J. Singh, Eds. Singapore: Springer Nature Singapore, 2024, pp. 29–38.

[32] M. A. Morid, A. Borjali, and G. Del Fiol, "A scoping review of transfer learning research on medical image analysis using ImageNet," *Comput. Biol. Med.*, vol. 128, p. 104115, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482520304467

[33] O. Kursun, A. Patooghy, P. Poursani, and O. V. Favorov, "Cg-CNN: Self-supervised feature extraction through contextual guidance and transfer learning," *IEEE Access*, vol. 12, pp. 155 851–155 866, 2024.

[34] M. Raghu, C. Zhang, J. Kleinberg, and S. Bengio, *Transfusion: Understanding Transfer Learning for Medical Imaging*. Red Hook, NY, USA: Curran Associates Inc., 2019.

[35] E. Barua, "Durian," 2022, data retrieved from Kaggle. Available: https://www.kaggle.com/datasets/edenbarua/dataset?select=DURIAN.

[36] P. Thölke et al., "Class imbalance should not throw you off balance: choosing the right classifiers and performance metrics for brain decoding with imbalanced data," *NeuroImage*, vol. 277, p. 120253, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811923004044.

[37] M. Buda, A. Maki, and M. A. Mazurowski, "A systematic study of the class imbalance problem in convolutional neural networks," *Neural Netwk.*, vol. 106, pp. 249–259, Oct. 2018. [Online]. DOI: 10.1016/j.neunet.2018.07.011.

[38] Z. Chen, J. Duan, L. Kang, and G. Qiu, "Class-imbalanced deep learning via a class-balanced ensemble," *IEEE Trans. Neural Netwk. Learn. Syst.*, vol. 33, no. 10, pp. 5626–5640, 2022.

[39] A. Zhang, Z. C. Lipton, M. Li, and A. J. Smola, *Dive into Deep Learning*. Cambridge University Press, 2023. Available: https://d2l.ai/index.html.

[40] M. Lin, Q. Chen, and S. Yan, "Network in network," 2014. [Online]. Available: https://arxiv.org/abs/1312.4400.

[41] Y. Z. Gürbüz, O. Şener, and A. A. Alatan, "Generalized sum pooling for metric learning," in *2023 IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, IEEE, 2023, pp. 5439–5450.