HDC-SEResNet for Gambung Tea Leaf Classification

Thesis Submitted for Partial Fulfillment
of the Requirements for the Award of the Degree of

Bachelor of Technology
in
Electronics and Communication Engineering

by

Ayush Kumar
Roll Number: 2201043

Under the Supervision of

Dr. Mourina Ghosh

Department of Electronics and Communication Engineering
Indian Institute of Information Technology Guwahati

May 22, 2025

Contents

Declaration

I declare that

1. The work contained in the thesis is original and has been done by myself under the general supervision of my supervisor.

2. The work has not been submitted to any other Institute for any degree or diploma.

3. I have followed the guidelines provided by the Institute in writing the thesis.

4. I have conformed to the norms and guidelines given in the Ethical Code of Conduct of the Institute.

5. Whenever I have used materials (data, theoretical analysis, and text) from other sources, I have given due credit to them by citing them in the text of the thesis and giving their details in the references.

6. Whenever I have quoted written materials from other sources, I have put them under quotation marks and given due credit to the sources by citing them and giving required details in the references.

Ayush Kumar

Abstract

This report details the development and rigorous evaluation of HDC-SEResNet, a novel Convolutional Neural Network (CNN) architecture specifically designed for the accurate classification of Gambung tea leaves into five distinct classes. The architecture strategically integrates Hybrid Dilated Convolutions (HDC) to enable effective multi-scale feature extraction and Squeeze-and-Excitation (SE) mechanisms for adaptive channel recalibration, all within a robust residual learning framework.

To overcome the challenge of limited initial data availability, extensive offline data augmentation techniques were employed, resulting in a significantly expanded training and validation dataset of approximately 6,000 images. The proposed HDC-SEResNet model, trained on this augmented dataset, achieved a compelling overall validation accuracy of **96.92%**, *demonstrating its strong potential for accurate and efficient tea leaf classification in practical agricultural settings.*

Comparative analysis against established state-of-the-art CNN architectures, including ResNet-50, ResNet-101, DenseNet-201, and Inception variants, reveals the superior performance of the proposed HDC-SEResNet. This enhanced performance is further substantiated by ROC curve analysis, where our model achieves high Area Under the Curve (AUC) values across all classes. This research makes a significant contribution to the advancement of automated and objective tea leaf classification systems, offering promising applications in areas such as quality control, variety identification, and authentication within the tea industry.

Chapter 1

Introduction

Tea is one of the most widely consumed beverages globally, and its quality, variety, and origin significantly impact market value. Accurate classification of tea leaves is crucial for various stages in the industry, including quality control during harvesting and processing, variety identification for breeding programs, and ensuring authenticity to prevent adulteration. Traditionally, this classification relies on human experts, which can be subjective, time-consuming, and requires extensive training. The advent of computer vision and deep learning, particularly Convolutional Neural Networks (CNNs), offers a promising avenue for developing automated, objective, and efficient tea leaf classification systems. These systems can analyze images of tea leaves to identify key visual characteristics indicative of their class.

A typical CNN architecture for image classification is shown in Figure 1.1.

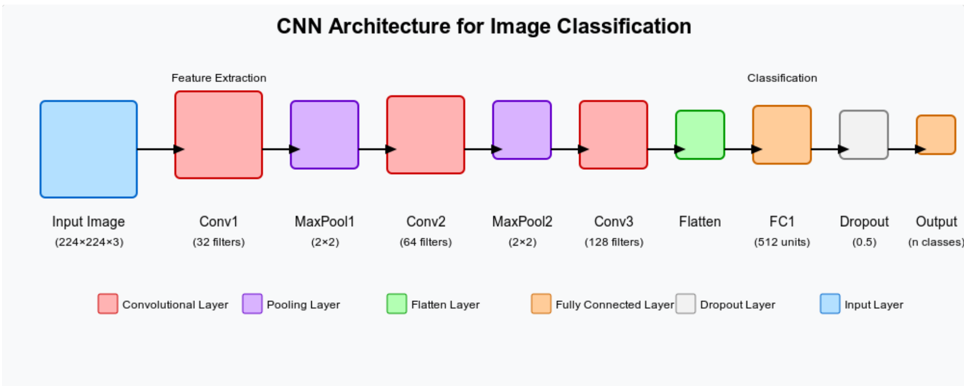

Figure 1.1: A typical CNN architecture for image classification.

The thesis is organized in the following way; Chapter 2 describes the motivation behind the work. Chapter 3 details the objectives of the thesis. In Chapter 4, the proposed HDC-SEResNet architecture for tea leaf classification and its implementation details are described. Chapter 5 presents a comprehensive evaluation of the model and analyzes the results. Chapter 6 concludes the thesis with a discussion on the limitations of the proposed techniques and scope for further research.

Chapter 2

Motivation

The primary motivation for this work stems from the need for accurate and efficient methods for tea leaf classification within the agricultural technology sector. While deep learning models excel at image recognition, their effectiveness is often predicated on the availability of large, well-annotated datasets. In specialized domains like tea cultivation, acquiring such large datasets can be a significant bottleneck due to factors like limited expert availability for labeling, variations in imaging conditions, and the sheer diversity of leaf appearances.

Furthermore, even with moderately sized datasets (achieved through methods like augmentation), standard CNN architectures may not optimally capture the subtle, fine-grained visual differences between closely related tea leaf varieties or quality grades. This necessitates the exploration of specialized network architectures designed to enhance feature extraction capabilities, focusing on capturing multi-scale information and emphasizing discriminative features relevant to the specific classification task.

Therefore, this research is motivated by the need to develop and evaluate a robust CNN architecture tailored for tea leaf classification that performs effectively on datasets relevant to real-world scenarios, potentially derived from initially limited samples through augmentation.

Chapter 3

Objective

The main objectives of this research are:

1. To develop and implement a specialized Convolutional Neural Network architecture, termed HDC-SEResNet (Hybrid Dilated Squeeze-and-Excitation Residual Network), for the classification of Gambung tea leaves into 5 distinct classes.

2. To incorporate and evaluate the effectiveness of Hybrid Dilated Convolutions (HDC) for multi-scale feature capture and Squeeze-and-Excitation (SE) mechanisms for adaptive channel feature recalibration within a residual learning framework.

3. To prepare a suitable dataset by applying extensive offline data augmentation techniques to an initially limited set of tea leaf images, resulting in approximately 6,000 images for training and validation.

4. To train the proposed HDC-SEResNet model on this augmented dataset and rigorously evaluate its classification performance on an unseen validation set, analyzing overall accuracy and per-class results.

5. To benchmark the performance of the proposed HDC-SEResNet against established state-of-the-art CNN architectures to validate its efficacy for the tea leaf classification task.

Chapter 4

Work Done

This section details the proposed architecture and its implementation, including data preparation, training procedures, and evaluation results.

4.1 Proposed Architecture: HDC-SEResNet

We propose the HDC-SEResNet, a deep convolutional neural network architecture designed to enhance feature extraction for tea leaf classification. It modifies the standard residual network paradigm by introducing a custom block structure.

4.1.1 Overall Network Structure

The network follows a typical CNN structure adapted for image classification:

- **Input Layer:** Accepts standardized RGB images (224x224x3) with z-score normalization.
- **Stem Block:** An initial sequence of Convolution (7x7, stride 2), Batch Normalization (BN), ReLU activation, and Max Pooling (3x3, stride 2). This rapidly reduces spatial resolution while increasing the receptive field and extracting low-level features.
- **Stacked HDC-SE-ResBlocks:** The core of the network consists of multiple stacked instances of the custom Hybrid Dilated SE-ResBlock (detailed below). Filter depths increase across stages (e.g., 32 in stem $\rightarrow$ 64 $\rightarrow$ 128 $\rightarrow$ 256 in blocks), and spatial downsampling is achieved by setting the stride to 2 in the first convolution of blocks marking a stage transition.
- **Classification Head:** A Global Average Pooling layer summarizes features across spatial dimensions, followed by a Dropout layer for regularization, a final Fully Connected layer mapping features to the 5 class scores, and a Softmax layer providing class probabilities.

4.1.2 Core Component: Hybrid Dilated SE-ResBlock

The primary novelty lies in the Hybrid Dilated Squeeze-and-Excitation Residual Block, which replaces standard blocks in a ResNet. It integrates three key mechanisms:

Hybrid Dilated Convolutions (HDC)

This component aims to capture multi-scale contextual information efficiently. Instead of a single convolutional layer, it employs three parallel 3x3 convolutional layers operating on the same input feature map but with different dilation rates (1, 2, and 3). Each path is followed by Batch Normalization and ReLU activation.

- **Benefit:** Allows the network to probe features at multiple spatial contexts simultaneously without needing larger filters (which increase parameters) or pooling (which reduces resolution).
- **Implementation:** The outputs from the three parallel paths are concatenated along the channel dimension (F_{cat}). A 1x1 convolution (F_{reduce}) with BN and ReLU is then used to fuse these multi-scale features and adjust the channel dimension to the block's target output size.

Squeeze-and-Excitation (SE)

Applied after the HDC feature fusion (F_{reduce}), the SE module enhances representational power by modeling channel interdependencies.

- **Benefit:** It adaptively recalibrates the importance of each feature channel, amplifying useful features and suppressing less relevant ones for the specific input.
- **Implementation:** A 'Squeeze' operation (Global Average Pooling) aggregates spatial information into a channel descriptor. This is followed by an 'Excitation' operation consisting of two Fully Connected layers (the first reducing channels by a ratio r, the second restoring them) with ReLU and Sigmoid activations. The resulting sigmoid outputs act as channel-wise scaling factors ($SE_{Weights}$). These weights are multiplied element-wise with the input feature map: $F_{se} = F_{reduce} \cdot SE_{Weights}$.

Residual Connection

To enable the training of a deep network, a residual connection is employed. The output of the SE-scaled path (F_{se}) is added element-wise to the input feature map (X) passed through a shortcut connection.

- **Benefit:** Facilitates gradient flow, mitigates the vanishing gradient problem, and allows the network to easily learn identity mappings if optimal.
- **Implementation:** The shortcut connection is either an identity mapping (if input and output dimensions match) or a 1x1 convolution (to match dimensions if stride $>$ 1 or filter count changes). The final block output is $Y = \text{ReLU}(\text{Shortcut}(X) + F_{se})$.

This specific combination of HDC and SE within a residual block forms the core technical contribution implemented and evaluated in this work. Figure 4.1 illustrates the structure of the proposed HDC-SE-ResBlock.

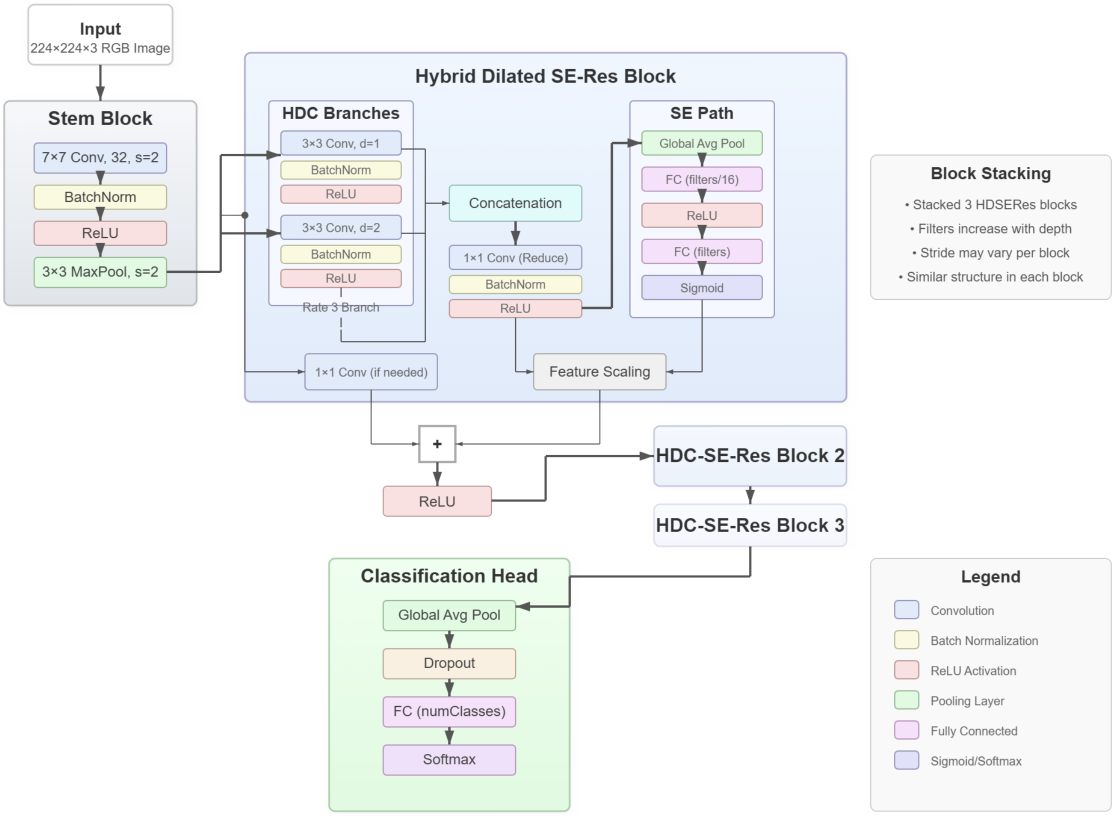

Figure 4.1: Structure of the proposed HDC-SE-ResBlock.

4.2 Implementation and Training

4.2.1 About the Dataset

The dataset used in this study was sourced from Kaggle [1]. It contains images of five distinct tea leaf clones, designated GMB_01 through GMB_05. Each clone exhibits unique visual characteristics:

- **GMB_01:** Leaves generally appear larger and elongated with a slightly pointed tip. They feature a prominent central vein and well-defined secondary veins. Some samples may have a shinier or glossier surface texture. (See Figure 4.2a)
- **GMB_02:** These leaves are typically smaller and more compact compared to other clones. Their surface often appears slightly wrinkled and less smooth. (See Figure 4.2b)
- **GMB_03:** Characterized by a more uniform, slightly oval shape. The leaves tend to look smoother and exhibit a more structured appearance. (See Figure 4.2c)
- **GMB_04:** Leaves are noticeably broader and more rounded. They possess a prominent, wrinkled texture. (See Figure 4.2d)
- **GMB_05:** These leaves may show some visible imperfections or small spots on their surface. (See Figure 4.2e)

Sample images illustrating the visual characteristics of the tea leaf clones are shown in Figure 4.2.

(a) GMB_01 (b) GMB_02 (c) GMB_03

(d) GMB_04 (e) GMB_05

Figure 4.2: Sample images of the five Gambung tea leaf clones (GMB_01 to GMB_05).

4.2.2 Dataset Preparation and Augmentation

The study utilized the Gambung Tea Leaf dataset, initially comprising approximately 600 images distributed across 5 distinct classes (GMB_01 to GMB_05). Recognizing the limitations of this small size for training a deep CNN, an extensive offline data augmentation strategy was implemented prior to model training. Each original image was subjected to approximately 10 random transformations, including rotation, scaling, reflection (horizontal and vertical), and translation. This process yielded an augmented dataset containing roughly 6,000 images.

This larger dataset was then randomly partitioned into a training set (90%, ∼5,400 images) and a validation set (10%, ∼600 images), maintaining the original class distribution in both splits. Furthermore, during the training phase, online data augmentation (using MATLAB's imageDataAugmenter with random reflections, rotations [±15°], scaling [0.8-1.2x], and translations [±30 pixels]) was applied dynamically to the training batches to further enhance model robustness and reduce overfitting. Validation images were only resized to the required input dimensions without further augmentation.

4.2.3 Network Configuration and Training Details

The HDC-SEResNet model was implemented using the MATLAB Deep Learning Toolbox. Key configuration and training parameters were:

- **Input Size:** 224x224x3 (RGB).

- **Architecture:** As detailed in Section 4.1, including a stem block (32 filters) followed by 3 stacked HDC-SE-ResBlocks with increasing filter sizes (64, 128, 256). The SE reduction factor r was set to 16.
- **Optimizer:** Adam.
- **Learning Rate:** Initial rate of 1e-3, utilizing a piecewise decay schedule reducing the rate by 50% every 10 epochs.
- **Training Duration:** Maximum of 30 epochs, with early stopping implemented based on validation set performance (patience set to 5 epochs without improvement).
- **Mini-Batch Size:** 32 images per batch.
- **Hardware:** Training was performed using GPU acceleration.

Table 4.1 provides a summary of the network architecture and training hyperparameters.

Table 4.1: Network architecture and training hyperparameters

Parameter	**Value**
Input Size	224x224x3 (RGB)
Base Filters	32 (stem) $\rightarrow$ 64 $\rightarrow$ 128 $\rightarrow$ 256
SE Reduction Factor	16
Optimizer	Adam
Initial Learning Rate	1e-3
Batch Size	32
Max Epochs	30

Chapter 5

Evaluation and Results

This chapter presents a comprehensive evaluation of the proposed HDC-SEResNet model on the Gambung tea leaf classification task, including detailed performance metrics, comparative analysis with state-of-the-art architectures, and the impact of data augmentation.

5.1 Model Performance Analysis

5.1.1 Confusion Matrix Analysis

The confusion matrix visualizes the performance of the classification model across the five Gambung tea leaf classes (GMB_01 to GMB_05). Figure 5.1 shows the confusion matrix for the augmented validation dataset.

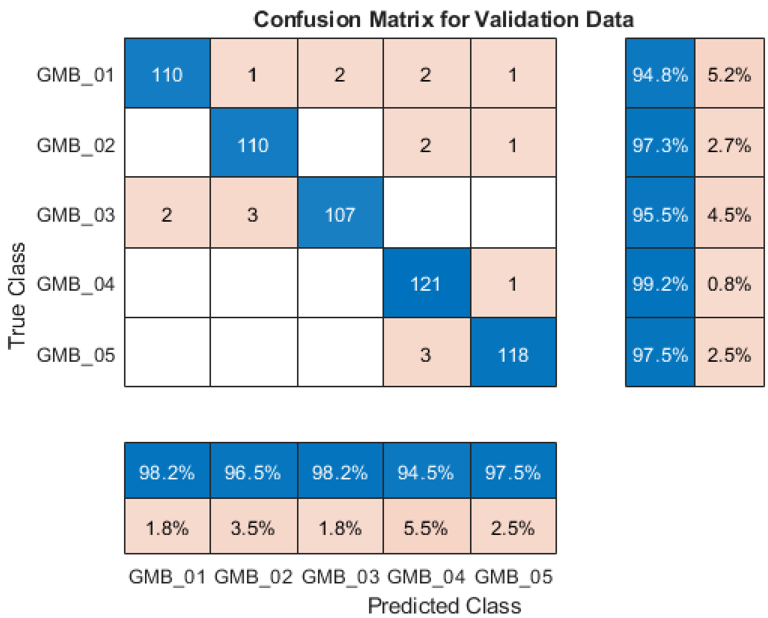

Figure 5.1: Confusion Matrix for Validation Data with Augmentation

Analysis of Results:

- **Overall Accuracy:** The model achieved a high overall validation accuracy of approximately 96.9%, correctly classifying 566 out of 584 validation samples shown in the matrix.

- **Per-Class Performance:** The model demonstrated strong performance across all classes shown:
  - GMB_01: 94.8% recall with 98.2% precision
  - GMB_02: 97.3% recall with 96.5% precision
  - GMB_03: 95.5% recall with 98.2% precision
  - GMB_04: 99.2% recall with 94.5% precision
  - GMB_05: 97.5% recall with 97.5% precision
- **Misclassifications:** The confusion matrix reveals that occasional misclassifications occur between classes (e.g., 3 instances of GMB_05 predicted as GMB_04, 3 instances of GMB_03 predicted as GMB_02), but these are relatively infrequent compared to correct classifications. The model shows robust discrimination capability.

For comparison, Figure 5.2 shows the confusion matrix when training on a limited dataset without extensive augmentation:

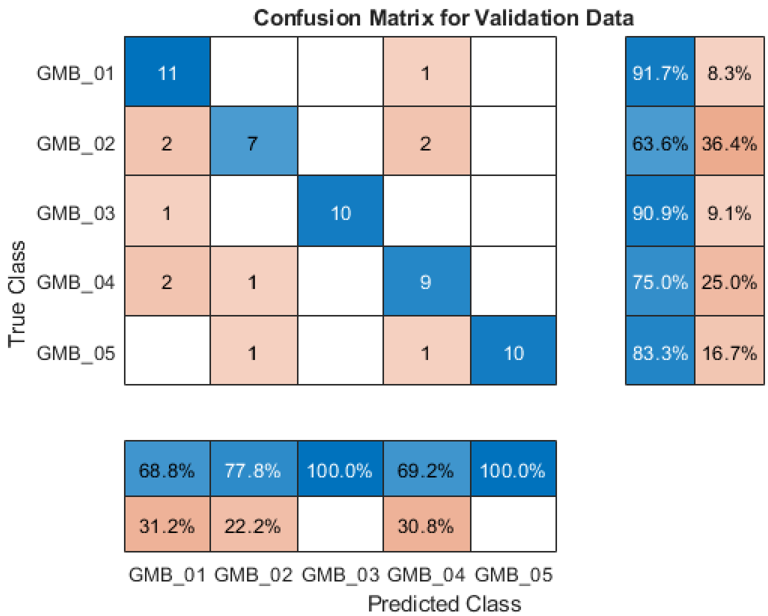

Figure 5.2: Confusion Matrix for Validation Data without Augmentation

The non-augmented model achieves only 81.5% overall accuracy, with significant confusion between classes, particularly for GMB_02 (63.6% recall) and GMB_04 (75.0% recall). This clearly demonstrates the critical importance of data augmentation for achieving robust classification with limited original samples.

5.1.2 ROC Curve Analysis

Receiver Operating Characteristic (ROC) curves provide insight into the model's ability to distinguish between classes at various threshold settings. Figure 5.3 shows the ROC curves for the augmented model, while Figure 5.4 presents ROC curves for the non-augmented model.

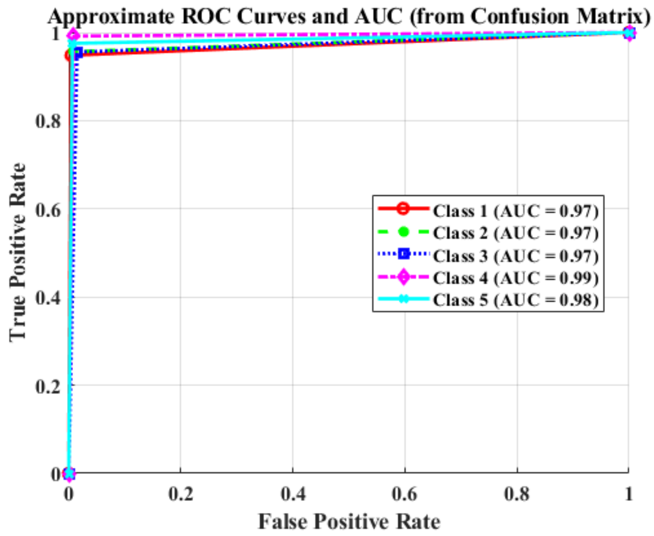

Figure 5.3: ROC Curves and AUC for Model Trained with Augmentation

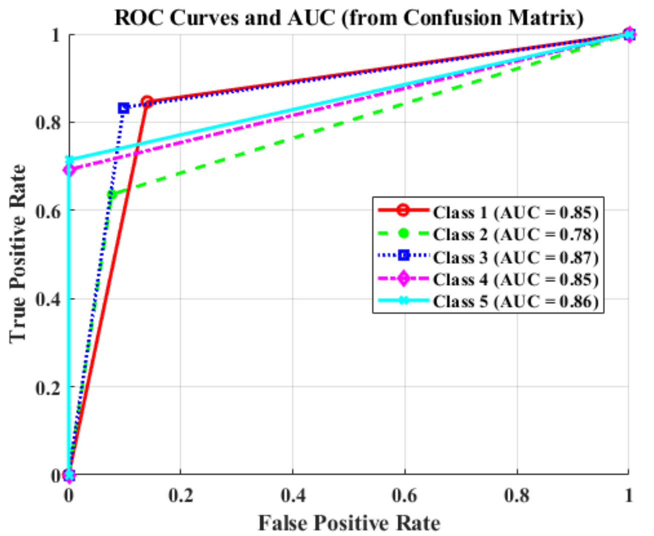

Figure 5.4: ROC Curves and AUC for Model Trained without Augmentation

The augmented model achieves exceptionally high Area Under the Curve (AUC) values ranging from 0.97 to 0.99 across all classes, approaching perfect classification. In contrast, the non-augmented model shows lower AUC values between 0.78 and 0.87, with class GMB_02 performing particularly poorly, corroborating the confusion matrix findings.

5.1.3 Training Dynamics

Figure 5.5 illustrates the validation accuracy and loss throughout the training process for the proposed HDC-SEResNet model, comparing the impact of using the augmented dataset versus the original, non-augmented dataset.

Key observations from the training dynamics:

- The augmented model (blue curve) reaches significantly higher validation accuracy (peaking near 97-98%) compared to the non-augmented model (orange curve, peak-

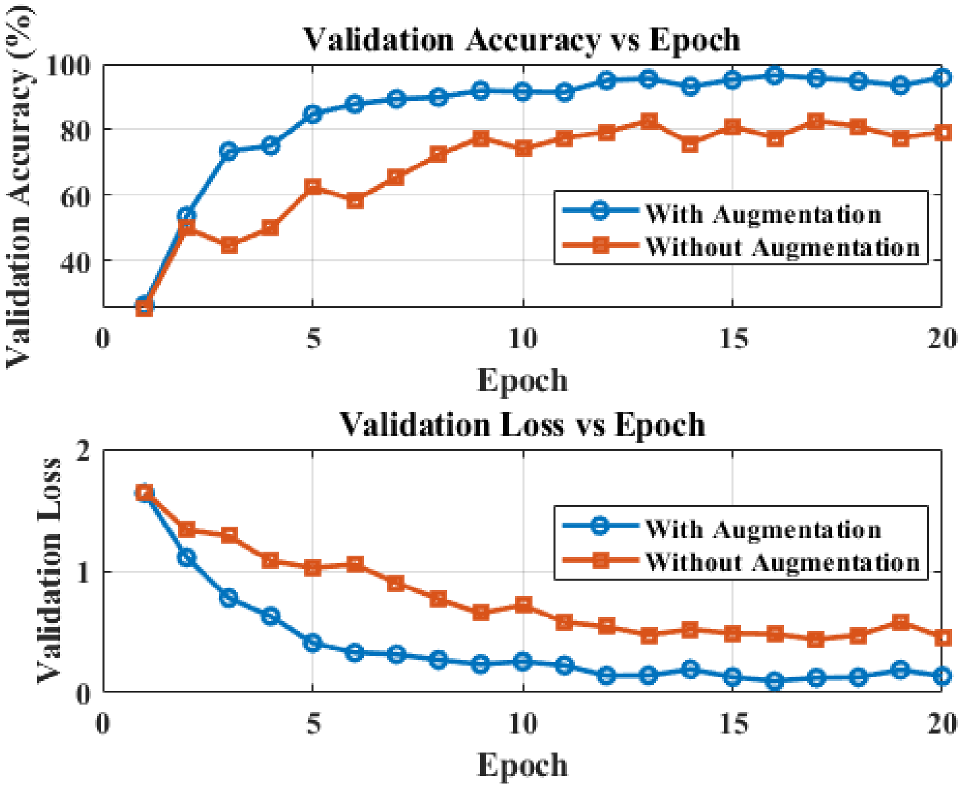

Figure 5.5: Validation Accuracy and Loss vs. Epoch (With vs. Without Augmentation)

ing around 80%).

- The augmented model exhibits smoother learning curves and achieves a substantially lower validation loss, indicating better generalization and reduced overfitting.
- The augmented model appears to converge more effectively, achieving strong performance relatively early (e.g., by epoch 10-12), while the non-augmented model's performance plateaus at a lower level.
- The non-augmented model shows signs of potential overfitting or instability, particularly visible in the loss curve which remains higher and less consistent compared to the augmented model's loss curve.

These dynamics strongly underscore the importance of data augmentation for achieving high performance and robust generalization with the proposed model on this dataset.

5.2 Comparative Analysis with State-of-the-Art Models

To establish the effectiveness of the proposed HDC-SEResNet architecture, we benchmarked it against several established CNN architectures for image classification. All models were trained on the same augmented dataset under identical conditions to ensure a fair comparison. Table 5.1 summarizes the final validation accuracy results, and Figure 5.6 illustrates the training progression (accuracy and loss) for all compared models.

Table 5.1: Comparison of Classification Accuracy on Augmented Dataset

CNN Model	Accuracy (%)
HDC-SEResNet (Proposed)	**96.92**
ResNet-50	95.45
ResNet-101	86.47
DenseNet-201	89.66
Inception V3	85.36
Inception-ResNet V2	91.38
Xception	84.48

The proposed HDC-SEResNet architecture achieves the highest accuracy (96.92%), outperforming the benchmarked established architectures:

- ResNet-50 by 1.47%
- ResNet-101 by 10.45%
- DenseNet-201 by 7.26%
- Inception V3 by 11.56%
- Inception-ResNet V2 by 5.54%
- Xception by 12.44%

ResNet-50 demonstrated strong performance (95.45%), coming closest to the proposed model, while other architectures showed considerably lower accuracy on this specific dataset and task.

These results validate the effectiveness of combining Hybrid Dilated Convolutions with Squeeze-and-Excitation mechanisms within a residual framework for the tea leaf classification task. The proposed architecture's superior performance, particularly compared to most other models, can be attributed to:

- **Enhanced multi-scale feature extraction:** The HDC component effectively captures features at different receptive field sizes, crucial for identifying discriminative patterns in tea leaves at various scales.
- **Adaptive feature refinement:** The SE mechanism dynamically emphasizes important channels while suppressing less informative ones, allowing the network to focus on class-discriminative features.
- **Efficient learning:** The residual connections facilitate gradient flow during training, enabling the network to learn effectively despite the relative complexity of the architecture.

5.3 Effect of Data Augmentation

Our results definitively demonstrate the critical importance of data augmentation when working with limited initial datasets for specialized classification tasks. As shown

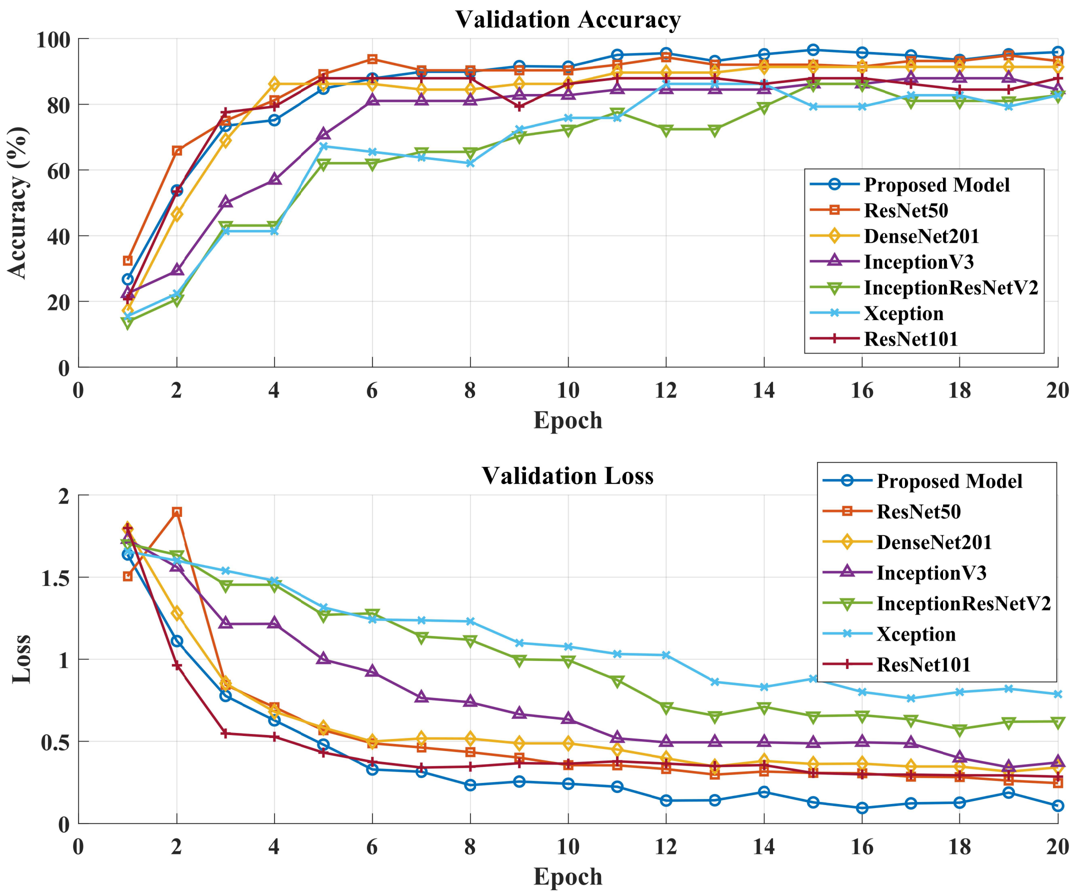

Figure 5.6: Validation Accuracy and Loss Comparison of Different CNN Architectures vs. Epoch

in Figure 5.5, the augmented approach achieves approximately 15-20% higher validation accuracy compared to training without augmentation.

Key benefits of augmentation observed in our experiments:

- **Improved generalization:** The model trained with augmented data shows significantly better performance on validation samples.
- **Enhanced robustness:** The augmented model is less sensitive to variations in leaf orientation, scale, and position, which are common in real-world applications.
- **Reduced overfitting:** With more diverse training examples, the model learns more generalized features rather than memorizing specific training instances.
- **Better class separation:** As evidenced by ROC curves, the augmented model achieves much higher AUC values, indicating superior class discrimination.

Chapter 6

Conclusion and Future Work

6.1 Conclusion

This research successfully developed and rigorously evaluated the HDC-SEResNet, a novel CNN architecture tailored for precise tea leaf classification. By strategically integrating Hybrid Dilated Convolutions for comprehensive multi-scale feature extraction and Squeeze-and-Excitation modules for adaptive channel attention within a robust residual framework, the model aimed to significantly enhance feature representation. Leveraging a substantially expanded dataset achieved through extensive offline augmentation (∼6,000 images derived from an initial ∼600), the HDC-SEResNet achieved a remarkable validation accuracy of **96.92**% for the challenging 5-class Gambung tea leaf classification task.

Our detailed comparative analysis unequivocally demonstrates the superior performance of the proposed architecture against several established state-of-the-art CNN models. Specifically, the HDC-SEResNet outperformed ResNet-50 (95.45%), ResNet-101 (86.47%), DenseNet-201 (89.66%), Inception V3 (85.36%), Inception-ResNet-V2 (91.38%), and Xception (84.48%) on the augmented dataset. This significant performance gain directly validates the efficacy of our architectural innovations, underscoring the synergistic benefits of multi-scale feature extraction via hybrid dilated convolutions and dynamic channel recalibration through squeeze-and-excitation mechanisms. Furthermore, our findings emphatically highlight the crucial role of extensive data augmentation in enabling the development of effective deep learning solutions for specialized domains characterized by limited initial data resources.

This research makes a substantial contribution to the advancement of automated and objective tea leaf classification systems, offering promising applications across the tea industry, including enhanced quality control, accurate variety identification, and robust authentication processes. Future work may explore the model's generalization capabilities on more diverse datasets and its potential for real-time deployment on edge devices.

6.2 Future Work

While the current results are promising, several avenues exist for future exploration:

- **Addressing Confusion:** Further investigate specific visual features leading to the few remaining misclassifications between certain classes. Techniques like Class Activation Mapping (CAM) or Grad-CAM could provide insights into which regions the model focuses on for classification decisions.

- **Architecture Optimization:** Further refine the HDC-SE block by experimenting with different dilation rates, SE reduction factors, and block arrangements to potentially enhance performance.

- **Advanced Augmentation:** Explore more sophisticated data augmentation techniques such as Generative Adversarial Networks (GANs) to synthesize even more realistic and diverse tea leaf variations.

- **Transfer Learning:** Investigate if pre-training on larger general-purpose datasets before fine-tuning on the tea leaf data could further improve performance or reduce training data requirements.

- **Field Deployment:** Develop and evaluate a mobile or edge-device implementation of the model for in-field application by tea growers and processors.

Bibliography

[1] Cendikiawan. (2023). *Daun Teh GMB Dataset*. Retrieved from `https://www.kaggle.com/datasets/cendikiawan/dauntehgmb`

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

[3] Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 7132-7141).

[4] Yu, F., & Koltun, V. (2015). Multi-scale context aggregation by dilated convolutions. In International Conference on Learning Representations.

[5] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[6] Wang, P., Chen, P., Yuan, Y., Liu, D., Huang, Z., Hou, X., & Cottrell, G. (2018). Understanding convolution for semantic segmentation. In 2018 IEEE Winter Conference on Applications of Computer Vision (pp. 1451-1460).

[7] Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1251-1258).

[8] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2818-2826).