Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

RACEGNet for Adaptive Reliability Aware Expert Fusion in Durian Disease Classification

Md. Naimul Islam Nuhash
Daffodil International University
Raja Tariqul Hasan Tusher
Daffodil International University
Ahmed Moustafa
University College of Professional Education
Md. Ali Hossain
ali.cse@diu.edu.bd

Daffodil International University

Research Article

Keywords: Durian disease classification, Heterogeneous ensemble learning, Reliability-aware gating, Mixture-of-experts, Adaptive expert fusion

Posted Date: November 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11063328/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

RACEGNet for Adaptive Reliability Aware Expert Fusion in Durian Disease Classification

Md. Naimul Islam Nuhash[a,b], Raja Tariqul Hasan Tusher[a,b], Ahmed Moustafa[c,d,e], Md. Ali Hossain[a,b]

[a]Department of Computer Science and Engineering, Daffodil International University, Dhaka 1216, Bangladesh

[b]NanoBio Technology Center, Daffodil International University, Dhaka 1216, Bangladesh

[c]Department of Human Anatomy and Physiology, the Faculty of Health Sciences, University of Johannesburg, Johannesburg, South Africa

[d]Faculty of Psychology, University College of Professional Education, Wrocław, Poland

[e]Centre for Data Analytics & School of Psychology, Bond University, Robina, Gold Coast, Queensland, Australia

Corresponding author

Md. Ali Hossain

Corresponding author's email address

ali.cse@diu.edu.bd

Abstract

Plant disease recognition under real-field scenarios is still a challenging task because of high variations in terms of symptom presentation, affected plant organs, lighting, background, and severity of diseases. Existing deep learning methods typically utilize single feature representation or fixed ensemble strategies that limits their ability for adaptation to heterogeneous disease manifestations. In addition, existing durian disease classification studies have mainly focused on organ-specific recognition, such as leaf or trunk disease classification, whereas real-field disease recognition requires handling diverse manifestations across different plant organs and visual conditions. These limitations highlight the need for an adaptive framework that can exploit complementary representations from heterogeneous architectures while dynamically determining their relevance for each input sample. Motivated by these challenges, This study proposes Reliability-Aware Cross-Expert Gating Network (RACEGNet) for durian disease classification. RACEGNet is a heterogeneous expert fusion framework that integrates four architecturally distinct pretrained experts including ResNet-50, ConvNeXt-Tiny, MaxViT-T, and DeiT-Small/16 and consists of three key modules including Cross-Expert Representation Learning (CERL), Reliability Estimation Module (REM), and Adaptive Top-k routing. It offers an adaptive recognition approach through the combination of heterogeneous CNN and Transformer-based representations and the dynamic adjustment of expert contributions based on the reliability of each sample rather than a single model or a fixed fusion method. This framework was evaluated on a real field durian disease dataset consisting of ten classes where individual disease classes include symptoms manifested on multiple plant organs, such as leaves, trunks, and fruits etc. The experimental results show that RACEGNet obtained 96.30% test accuracy and 96.34% Macro F1-scores, which outperformed several individual expert models and traditional ensembling techniques. An ablation study proves the effect of cross-expert contextualization, reliability estimation, and adaptive routing on improving classification results. The statistical analysis confirmed that RACEGNet significantly outperformed its constituent expert models. Since the framework is designed to address the challenges of visual heterogeneity in plant disease recognition, including variations in symptom appearance, affected organs, and field conditions, RACEGNet has potential for extension to other crops. The findings demonstrate that collaboration of heterogeneous DL experts based on reliability is a suitable solution for addressing visual diversity challenges of agriculture disease recognition.

Keywords: Durian disease classification; Heterogeneous ensemble learning; Reliability-aware gating; Mixture-of-experts; Adaptive expert fusion.

1. Introduction

Durian, popularly known as the "King of Fruits," is scientifically referred to as *Durio zibethinus Murr*. From being considered a regional fruit within Southeast Asia, durian fruit has undergone an evolution process from just a fruit into an internationally traded product whose value has been boosted by its unique taste, nutritive benefits, and diverse uses within various industries. Commercial durian planting is done in Thailand, Vietnam, Malaysia, Indonesia, and the Philippines, while the growing demand of durian from international markets, especially China and Western markets, is fueling its fast market growth [1]. The global durian market size was USD 23.23 billion in 2022, with projections indicating that its market size will increase up to USD 46.35 billion in 2032 [2]. Although its economic importance has been on the rise, durian production is quite susceptible to many fungal and bacterial diseases that affect leaves, stem, fruits, and even the tree trunk. Disease infection can lower the productivity of durian

orchards by 10-30% or even more, leading to great economic losses worth millions of dollars per year [3]. Consequently, rapid and accurate disease diagnosis has become essential for improving crop management, reducing yield losses, and ensuring sustainable durian production.

Traditionally, the method used to diagnose diseases in plants involves visual detection performed by professional plant pathologists. Despite the fact that this method can be useful when applied to a small-scale agricultural settings, in case of large-scale commercial plantations it becomes inefficient due to high labor requirements, time consumption, and heavy dependence on the expertise of specialists. In addition, it should be noted that visual estimation is always subjective and can vary greatly from specialist to specialist, thus making the results inconsistent [4]. Other issues, such as the scarcity of agriculture specialists in rural areas, costly examinations, and delayed disease detection after pathogens are already established, also contribute to the ineffectiveness of the manual method [5]. These limitations often result in missed intervention opportunities, inappropriate disease management strategies, and increased yield losses, particularly in resource-constrained agricultural systems where access to diagnostic laboratories is limited [6]. To address these challenges, artificial intelligence (AI), particularly deep learning and computer vision, has emerged as a promising solution for automated, objective, and scalable plant disease diagnosis.

Recent advances in deep learning have substantially improved automated plant disease classification. Convolutional Neural Networks (CNNs) are highly effective in learning hierarchical local representations associated with lesion texture, color variation, boundary structure, and morphological patterns. Vision Transformers (ViTs), in contrast, exploit self-attention mechanisms to model long-range spatial dependencies and broader contextual relationships. More recent hybrid CNN–Transformer architectures attempt to integrate the complementary strengths of local feature extraction and global contextual modeling. Beyond individual architectures, ensemble learning has been employed to improve robustness by combining multiple classifiers using strategies through probability averaging, weighted aggregation, voting, and stacking. Adaptive gating and expert-based learning have further introduced the possibility of varying model contributions according to individual input samples.

Despite these developments, an important challenge remains in practical plant disease recognition: substantial visual heterogeneity may exist even within the same disease class. A considerable portion of plant disease classification research is performed using relatively organ-specific or visually homogeneous imagery, such as datasets dominated by leaves, fruits, or stems considered separately. However, disease manifestations under realistic orchard conditions are considerably more variable. The same disease category may be represented by symptoms occurring on different plant organs, while images may simultaneously differ in lesion morphology, texture, color, scale, symptom severity, surrounding plant tissue, orientation, illumination, viewing distance, and background composition. Consequently, images belonging to a single disease class may exhibit markedly different visual patterns, resulting in substantial intra-class heterogeneity. Conversely, visually similar discoloration, lesions, spots, necrotic regions, or surface deterioration may occur across different disease categories, increasing inter-class visual ambiguity. Representative examples in Fig. 1 illustrate this organ-level intra-class visual heterogeneity, where the same disease category exhibits markedly different appearances across leaves, fruits, branches, flowers, fruitlets, and trunks. Such variability suggests that a single feature extractor may not be equally effective for all disease manifestations, motivating the use of heterogeneous expert representations and sample-specific adaptive fusion.

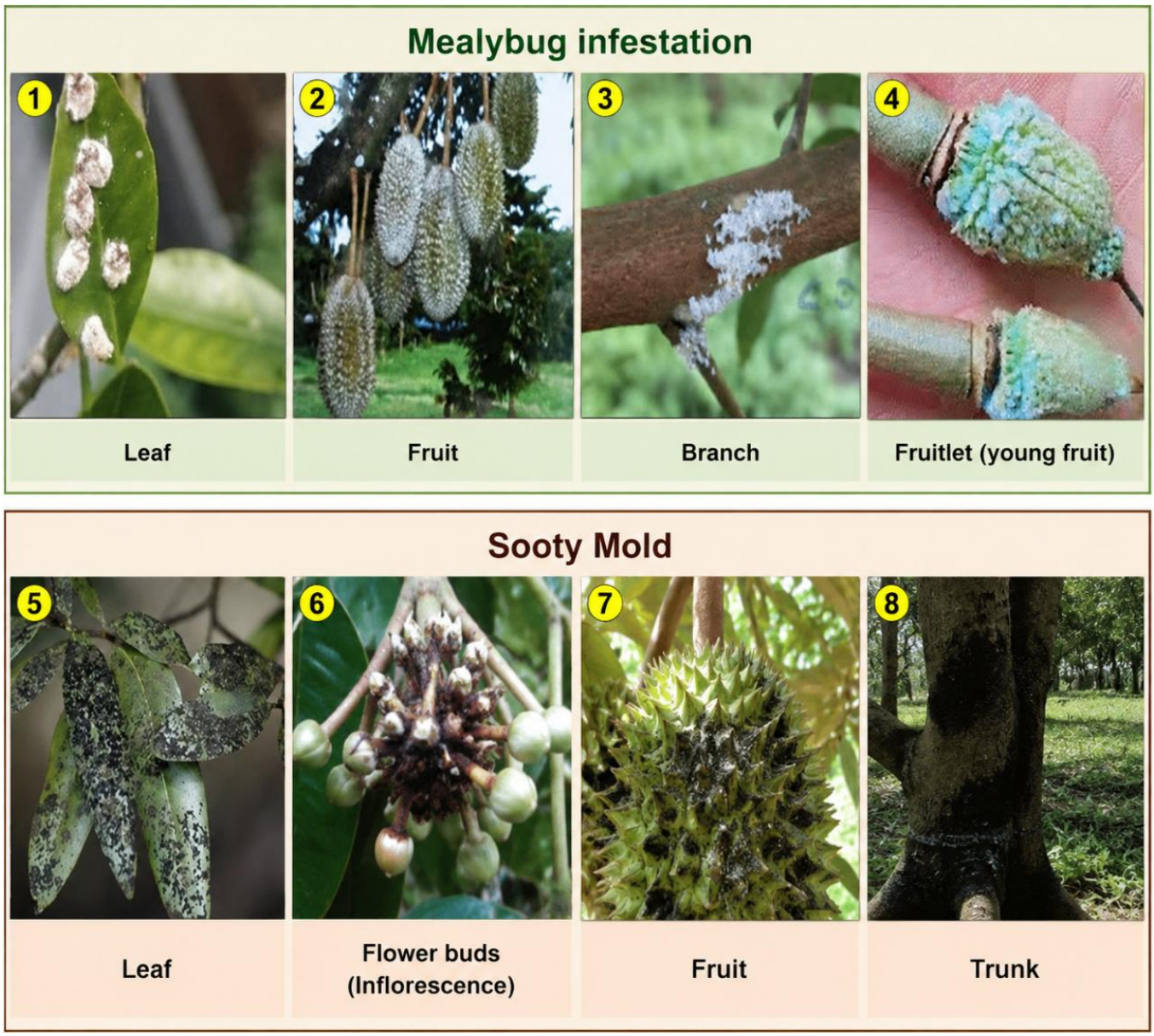


Fig. 1. Representative examples of organ-level intra-class visual heterogeneity in the durian disease dataset.

However, simply combining heterogeneous models does not fully address this challenge. Conventional ensemble methods generally rely on majority voting, uniform probability averaging, predefined weighting, uncertainty-derived weighting, or globally learned aggregation. Such strategies cannot fully account for the possibility that an expert that is reliable for one visual manifestation may be less informative for another. Furthermore, although adaptive expert frameworks can provide sample-dependent routing, relationships among heterogeneous expert representations are not always explicitly modeled before reliability estimation. Expert importance may also be represented only through unconstrained gating scores rather than through an explicit assessment of whether an expert is likely to provide trustworthy evidence for the current input. These limitations become particularly important under heterogeneous field conditions involving changes in illumination, background clutter, viewing angle, occlusion, symptom severity, organ type, and acquisition conditions. Previous studies have demonstrated notable performance degradation when plant disease models trained under controlled or in-domain conditions are evaluated on external data [7], while dataset-specific background cues may also encourage models to rely on acquisition characteristics rather than genuine disease-relevant visual evidence [8]. These challenges motivate an adaptive framework capable of modeling cross-expert relationships while explicitly estimating sample-specific expert reliability.

To address these challenges, this study proposes a Reliability-Aware Cross-Expert Gating Network (RACEGNet) for durian disease classification. The framework is motivated by the premise that heterogeneous disease manifestations may require complementary representations whose relative usefulness varies across individual samples. Four architecturally diverse pretrained experts - ResNet-50, ConvNeXt-Tiny, MaxViT-T, and DeiT-Small/16 - are employed to provide complementary convolutional, modern convolutional, hybrid-attention, and Transformer-based representations. Following expert training and selection, their parameters are frozen before RACEGNet development so that the subsequent gating framework learns to exploit their existing complementary representations without altering the expert models themselves.

RACEGNet integrates three closely connected mechanisms: Cross-Expert Representation Learning (CERL), a Reliability Estimation Module (REM), and Adaptive Top-k routing. CERL maps the heterogeneous deep representations and calibrated class-

probability vectors produced by the four experts into a common latent representation space. REM explicitly estimates the sample-specific reliability of each contextualized expert representation. The resulting reliability scores are used for Adaptive Top-k post-backbone sparse feature fusion. Rather than using a fixed number or fixed subset of experts for every image, RACEGNet dynamically determines how many and which expert representations should participate in the final fusion.

The proposed framework is evaluated on a ten-class durian disease image dataset characterized by heterogeneous disease appearances and imaging conditions. RACEGNet is compared with its four constituent experts, five independently trained deep-learning baseline models, and five conventional ensemble strategies. The contributions of its major architectural components are examined through a hierarchical ablation study evaluating the progressive effects of CERL, explicit reliability estimation, and Adaptive Top-k fusion. Expert-count sensitivity analysis is further performed to investigate the effect of expert-pool size on classification performance and routing behavior. Paired statistical testing is used to assess the performance improvement of RACEGNet over its constituent experts, while computational profiling evaluates the complexity of the complete framework. Model interpretability is further examined using expert-specific Grad-CAM++ visualizations and a RACEGNet route-conditioned composite Grad-CAM++ explanation.

The primary contributions of this study are summarized as follows:

- A heterogeneous expert framework is developed for visually diverse durian disease recognition. The study explicitly considers substantial intra-class variation arising from differences in disease manifestation, affected plant region, lesion appearance, acquisition condition, and background. Architecturally diverse pretrained models are utilized to provide complementary representations rather than relying on a single universal feature extractor.
- A Cross-Expert Representation Learning mechanism is introduced for contextualizing heterogeneous expert representations before reliability estimation. Deep features and calibrated prediction probabilities from experts are projected into a shared latent space and processed through cross-expert self-attention, allowing each expert representation to be interpreted in the context of the other experts rather than independently.
- An explicitly supervised Reliability Estimation Module is developed for sample-specific expert assessment. REM integrates normalized entropy, maximum confidence, peer-contextual feature similarity, and distributional prediction agreement, while its reliability outputs are supervised against frozen-expert correctness. This provides a direct learning objective for distinguishing comparatively reliable and unreliable experts for individual inputs.
- An Adaptive Top-k mechanism is introduced for sample-specific sparse feature fusion. Rather than using a fixed number or fixed subset of experts for every image, RACEGNet dynamically determines both how many and which expert representations should participate in the final fusion according to a cumulative routing-mass threshold.. This enables different disease manifestations to exploit different combinations of complementary expert representations according to their sample-specific reliability. The resulting sparsity operates at the post-backbone feature-fusion stage, with all four expert backbones evaluated for every input.
- A comprehensive evaluation is conducted against both independently trained baseline models and conventional ensemble strategies. The proposed method is examined through individual-expert comparison, conventional ensemble benchmarking, independently trained baseline models comparison, hierarchical component ablation, expert-count sensitivity analysis, paired statistical testing, computational profiling, and route-conditioned Grad-CAM++ interpretation. These experiments collectively assess predictive performance, expert complementarity, sample-specific routing behavior, component contributions, computational characteristics, and model interpretability.

2. Related Work

Deep learning has significantly advanced automated plant disease recognition, progressing from conventional Convolutional Neural Networks (CNNs) to Vision Transformers (ViTs), hybrid architectures, ensemble learning, and adaptive expert-based methods. Each stage of this progression has addressed different aspects of the recognition problem, including local feature extraction, long-range contextual modeling, complementary representation learning, and sample-dependent model collaboration. Accordingly, the following review examines these methodological developments with particular emphasis on their relevance to heterogeneous disease representation and adaptive expert fusion.

2.1. Deep Visual Representation for Plant Disease Classification

CNNs remain widely used because of their ability to extract hierarchical local features associated with lesion texture, color, shape, and morphology. Recent approaches have improved CNN-based disease recognition through multi-scale feature extraction, attention, segmentation, image enhancement, and feature integration [9–15]. Shafik et al. [9], for example, combined Inception and Xception components and reported above 98% accuracy across several plant-disease datasets, while PlantPathNet [10] achieved 99.57% accuracy on PlantVillage using cross-layer feature integration and channel attention. A segmentation-assisted CNN–BiGRU framework similarly achieved 99.64% accuracy for rice leaf disease recognition [11]. Other studies have explored wavelet-domain fusion, image super-resolution, Vision Mamba integration, and few-shot learning [12–15].

Despite their effectiveness, CNNs predominantly emphasize local and hierarchical visual patterns. ViTs provide a complementary representation by using self-attention to capture long-range spatial dependencies. Recent ViT-based studies have incorporated multi-scale patches, uncertainty estimation, and explainability [16–22]. An explainable ViT-B/16 achieved 97.52% validation accuracy for bean leaf disease recognition [16], while MaizeFormerX reported up to 97.8% accuracy using multi-scale

patch embedding and cross-scale attention [17]. RoViT-KAN further reported 99.70% accuracy while jointly addressing disease classification, severity estimation, and uncertainty [19].

The complementary characteristics of CNNs and Transformers have motivated hybrid architectures [23–29]. A CNN–ViT model achieved 95.60% accuracy for mulberry disease classification [23], while other studies have integrated DenseNet, EfficientNet, ResNeXt, Swin Transformer, LSTM, and graph-based modules with Transformer representations [24–29]. Some lightweight hybrid models have reported accuracies approaching 99.9% [26].

Collectively, these studies demonstrate that different architectural families capture different forms of discriminative evidence. Nevertheless, conventional single-backbone and fixed hybrid architectures generally process all samples through the same representation and fusion pathway. Such a formulation may become restrictive when the informative visual cues vary considerably from one image to another. Fine-grained lesion texture may dominate one sample, whereas larger-scale morphology, organ structure, or contextual relationships may be more informative in another. This observation suggests that heterogeneous architectural representations may offer complementary advantages when their relative contribution can be adapted to the characteristics of individual inputs.

2.2. Ensemble Learning and Adaptive Expert Fusion for Plant Disease Classification

Ensemble learning exploits complementary model behavior through voting, probability averaging, weighted aggregation, stacking, or feature fusion [30–36]. A dynamically weighted ensemble combined with CNN-derived features achieved 98.72% accuracy for barley disease recognition [31], while a stacked deep-learning framework reported 98.13% accuracy for multiclass plant disease classification [32]. Transformer-based ensembles have also demonstrated strong performance; Swin-SHARP achieved 98.10% accuracy [33], and a ViT–Swin Transformer ensemble reported 98.16% validation accuracy for sugarcane disease recognition [35].

Although model diversity can improve performance, conventional ensembles generally use fixed or globally learned aggregation. Majority voting gives equal influence to all models, soft voting applies a fixed probability-aggregation rule, and stacking typically learns one global combination function. Such approaches do not explicitly determine whether a particular expert is reliable for an individual sample.

Mixture-of-Experts (MoE) and adaptive gating methods address this limitation by allowing expert contributions to vary with the input. Salman et al. [37] applied a ViT-based MoE framework to in-the-wild plant disease recognition, while Raya-González et al. [38] reported 99.14% accuracy for maize pest and disease recognition using an MoE-based approach. Prototype-aware gating [39], cross-attention and gated fusion [40], adaptive CNN–Transformer fusion [41, 42], policy-guided routing [43], and Bayesian expert fusion [44] have also been explored. EcoTomHybridNet [43], for example, dynamically routed inputs between CNN and Transformer branches and reported an approximately 30% reduction in computational cost.

Nevertheless, several methodological issues remain relevant to heterogeneous expert collaboration. First, adaptive routing scores are frequently optimized as part of an overall classification or gating objective and are not necessarily supervised to represent the probability that a particular frozen expert is correct for the current sample [37, 39–44]. In such cases, a gating score indicates routing preference but does not have an explicit interpretation as expert reliability. Second, heterogeneous experts or branches are often evaluated largely through their own outputs or local gating features, whereas explicit contextualization of one expert representation relative to the remaining experts before reliability estimation is less common [37, 39, 41–44]. Yet relational evidence such as prediction agreement, uncertainty, and representation consistency may be informative when deciding which expert should be trusted. Third, many adaptive approaches retain several branches through continuous weighting or employ a predefined expert-selection cardinality. For example, the framework in [37] evaluates different Top-k settings but ultimately adopts a fixed k, meaning that the number of selected experts does not vary on a sample-by-sample basis.

2.3. Deep Learning for Durian Disease Recognition

Compared with extensively studied crops such as rice, maize, tomato, and potato, deep-learning-based durian disease recognition remains relatively limited. Existing studies have primarily investigated lightweight CNNs, transfer learning, optimization-assisted classification, Transformer-based approaches, and organ-specific disease recognition.

Nguyen-Tat and Nong [45] proposed DurianLSNet with a real-world dataset of 4,437 field images and reported 93.82% accuracy, focusing specifically on durian leaf diseases. Transfer learning combined with Particle Swarm Optimization achieved 92.6% accuracy on a six-class durian leaf dataset [46], while a lightweight YOLOv8-based approach reported 94.9% accuracy using approximately 0.56 million parameters [47]. Mask R-CNN achieved 99.61% accuracy for durian foliar disease classification [48]. Bayesian optimization and ViT-based methods have also reported 91.98% and 94.12% accuracy, respectively, for durian leaf disease recognition [50, 51]. In contrast, Nie et al. [49] focused specifically on durian trunk diseases, where MobileNetV2 achieved approximately 90.4% accuracy.

A notable characteristic of these studies is their predominantly organ-specific formulation. DurianLSNet [45], transfer learning with Particle Swarm Optimization [46], the lightweight YOLOv8-based approach [47], foliar classification using Xception/Mask R-CNN [48], Bayesian-optimized CNN classification [50], and ViT-based classification [51] primarily address leaf or foliar disease imagery, whereas [49] focuses on trunk diseases. However, they do not fully address a unified recognition setting in which samples belonging to the same disease class may originate from different plant organs. Under these conditions, performance obtained within a relatively homogeneous leaf- or trunk-specific domain does not necessarily imply that one architecture will remain equally effective across heterogeneous cross-organ manifestations. Different architectural families may respond preferentially to different aspects of the image for example, localized lesion texture, larger-scale organ morphology, or global spatial context, providing a strong rationale for heterogeneous expert learning in durian disease recognition.

2.4. Research Gap and Positioning of RACEGNet

The reviewed literature reveals four closely related gaps that motivate the proposed framework. First, existing durian disease studies have predominantly investigated organ-specific recognition, whereas a unified cross-organ classification setting can contain substantially greater intra-class appearance variation. Second, heterogeneous expert or branch representations are not commonly contextualized explicitly with respect to one another before expert assessment or routing. Third, adaptive gating generally relies on learned routing, similarity, policy, confidence, or feature-fusion scores rather than an explicitly supervised estimate of whether a frozen expert is correct for the current input. Fourth, adaptive approaches frequently use continuous weighting or a predefined number of selected experts rather than allowing the number of contributing expert representations to vary according to sample-specific reliability.

RACEGNet is designed around these limitations. Rather than assuming that one representation is universally optimal for all disease manifestations, the framework combines four architecturally heterogeneous experts to capture complementary visual information. Accordingly, RACEGNet differs from conventional ensembles and generic adaptive gating by integrating cross-expert contextualization, explicitly supervised reliability estimation, and sample-specific variable-cardinality sparse fusion within a unified framework. Cross-Expert Representation Learning (CERL) projects heterogeneous deep features and calibrated class probabilities into a shared latent space and contextualizes the expert representations through cross-expert self-attention. The Reliability Estimation Module (REM) then estimates sample-specific expert reliability. Adaptive Top-k routing selects a sample-specific subset of expert representations, with both the selected experts and the subset size varying across inputs. The design is intended to exploit heterogeneous architectural strengths while adapting expert participation to the diverse visual manifestations encountered in real-field durian disease recognition.

3. Methodology

This study has been designed in a systematic pipeline that includes data preparation, heterogeneous expert training, probability calibration, reliability-aware ensemble construction, and performance evaluation. The complete methodology of this study is shown in Fig. 2.

For experiment, a publicly available ten-class durian disease dataset was used containing a pre-separated training, validation, and independent test sets, which have 4,154, 649, and 648 images, respectively. The training images were resized to 224 × 224, data augmentation and ImageNet normalization were performed, while the validation and test images were subject to deterministic transformations without data augmentation. Four heterogeneous ImageNet pretrained networks, namely ResNet-50, ConvNeXt-Tiny, MaxViT-T, and DeiT-Small/16 were used as the RACEGNet constituent experts and fine-tuned using the training set. To avoid any overlap between different development phases, the validation set was split into three separate subsets, which served as the subsets for expert checkpoint selection, temperature calibration, and RACEGNet construction. Following checkpoint selection, the four experts were frozen and their output probabilities were temperature-calibrated before RACEGNet training.

RACEGNet combines the frozen and calibrated experts using Cross-Expert Representation Learning (CERL), Reliability Estimation Module (REM), and Adaptive Top-k routing. Notably, the expert backbones were kept frozen during the development of RACEGNet, and all model development choices were made before testing the framework on the independent test set.

Finally, RACEGNet was tested using the independent test set based on the following metrics: accuracy, Macro Precision, Macro Recall, Macro F1-Score, confusion matrix, and class-wise performance. The computational efficiency of the framework was analyzed using the following metrics: number of parameters, model state size, single image inference latency, batch throughput, and GPU memory usage for peak performance. The significance of the difference in accuracy between RACEGNet and baselines was measured using the exact paired McNemar test with Holm correction.

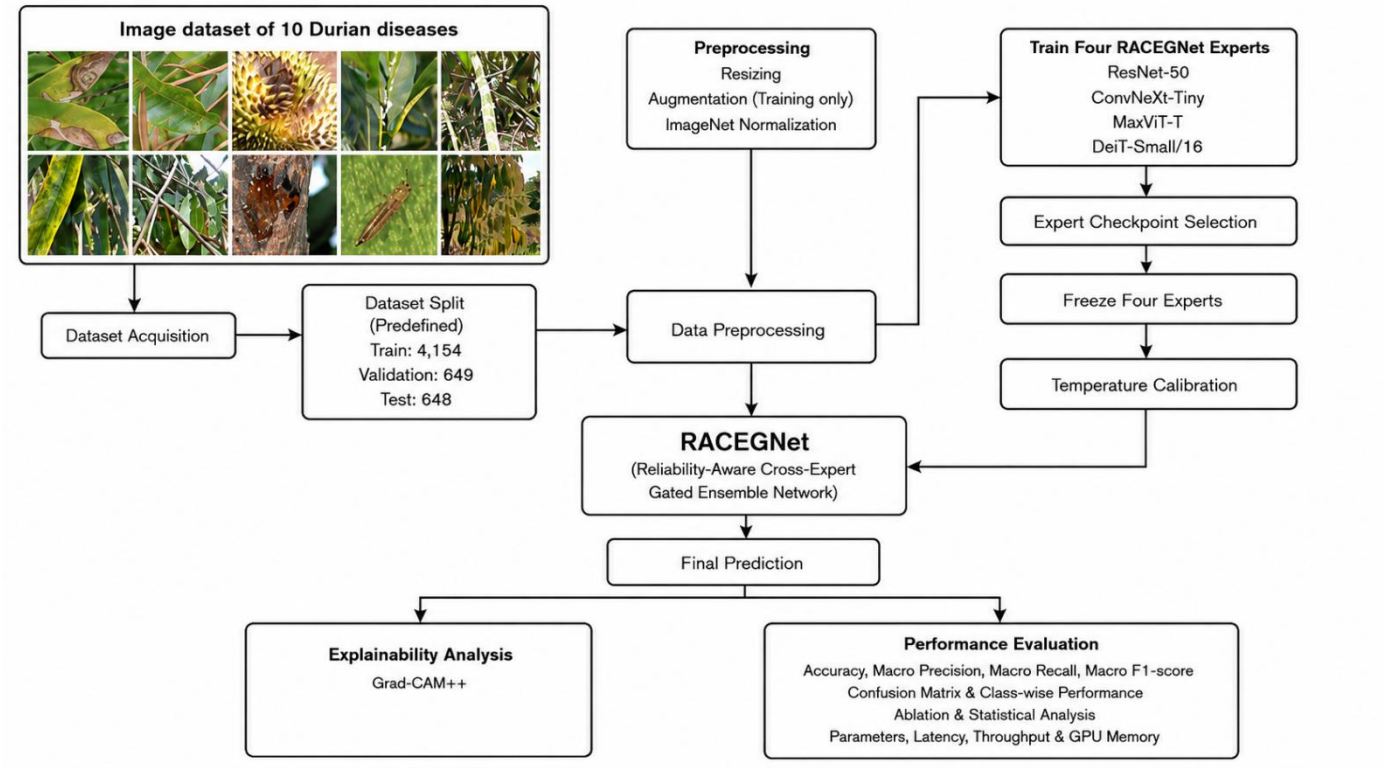


Fig. 2. Overall workflow of the proposed study.

3.1. Dataset Description

This study utilized the publicly available Image Dataset of Ten Durian Diseases Captured in Real-Field Conditions from a Family Orchard in Vinh Long, Vietnam, published on Mendeley Data [3, 52]. The dataset contains 5,451 RGB images belonging to ten durian disease classes: anthracnose disease, canker disease, fruit rot, mealybug infestation, pink disease, sooty mold, stem blight, stem cracking and gummosis, thrips disease, and yellow leaf. The images were collected under natural field conditions using mobile camera. Consequently, the dataset exhibits substantial variations in illumination, viewing angles, backgrounds, and image quality, closely reflecting real-world agricultural environments. The dataset is pre-divided into 4,154 training images, 649 validation images, and 648 test images, with all disease categories represented in each subset. These counts were determined directly from the dataset files used in the experiments. The images were manually verified by plant pathology experts to ensure correct disease labeling [3]. Representative samples from the ten durian disease classes are presented in Fig. 3. The class-wise distribution of images across the training, validation, and test sets is presented in Table 1.

Table 1. Class-wise distribution of images across the training, validation, and test sets

Disease Class	Train	Validation	Test
Anthracnose disease	424	62	63
Canker disease	427	65	65
Fruit rot	406	61	61
Mealybug infestation	406	65	65
Pink disease	416	65	68
Sooty mold	407	68	66
Stem blight	426	64	64
Stem cracking gummosis	424	68	68
Thrips disease	405	66	66
Yellow leaf	413	65	62

To enhance model robustness and minimize overfitting, online data augmentation was applied to the training set, including random resized cropping with a scale range of 0.85–1.0, random horizontal flipping, random rotation of up to ±15°, and color jittering with brightness and contrast adjustments of ±15%, saturation adjustment of ±10%, and hue variation of ±5%. Subsequently, all images were converted into PyTorch tensors and normalized using the ImageNet mean values and standard deviations. For the validation and test sets, only deterministic preprocessing was performed, consisting of resizing, tensor conversion, and ImageNet normalization. No augmentation was applied to the validation and test images, ensuring an unbiased evaluation of the proposed model.

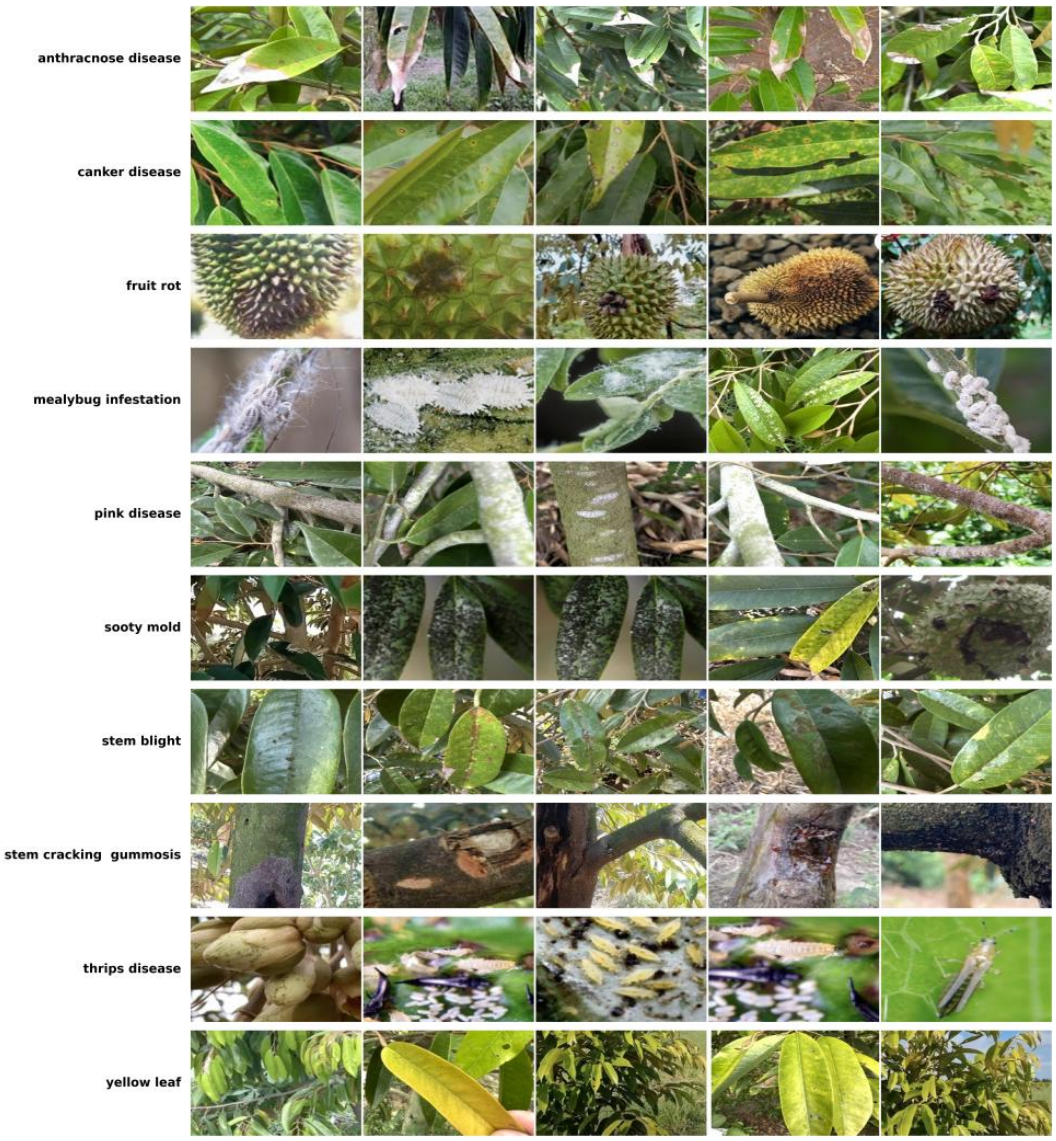


Fig. 3. Sample images of each of the classes.

3.2. RACEGNet Architecture

The proposed Reliability-Aware Cross-Expert Gating Network (RACEGNet) is a heterogeneous ensemble framework designed to improve durian disease classification through sample-specific reliability estimation and adaptive feature fusion. Unlike conventional ensembles that combine expert outputs using equal, fixed, or globally learned weights, RACEGNet estimates the reliability of each expert for an individual input and selectively fuses the corresponding contextualized expert representations. The overall architecture is illustrated in Fig. 4.

RACEGNet integrates four frozen heterogeneous expert networks, ResNet-50, ConvNeXt-Tiny, MaxViT-T, and DeiT-Small/16. Each expert produces a high-level deep feature vector and pre-Softmax logits. The logits are calibrated using expert-specific scalar temperature scaling to obtain calibrated probability vectors. Within Cross-Expert Representation Learning (CERL), each expert's deep feature is concatenated with its calibrated probability vector and projected into a shared latent space using an expert-specific projection and expert-ID embedding. A Cross-Expert Transformer Encoder then contextualizes these representations by modeling interactions among the four heterogeneous experts. The Reliability Estimation Module (REM) subsequently estimates a sample-specific reliability logit for each expert using its contextualized representation together with maximum confidence, normalized entropy, peer-contextual feature similarity, and distributional prediction agreement. REM is explicitly supervised using expert-correctness targets during training.

The resulting reliability logits are converted into routing probabilities using Softmax and passed to Adaptive Top-k routing. For each input, the framework selects the minimum number of experts whose cumulative routing probability reaches the threshold. The selected reliability logits are then re-normalized to obtain sparse routing weights, which are used to compute a reliability-weighted fusion of the selected contextualized expert representations. The fused representation is finally passed through the classification head to predict one of the ten disease classes. Importantly, all four frozen expert backbones are executed before reliability estimation and routing; therefore, Adaptive Top-k introduces sparsity only at the post-backbone feature-fusion stage and does not perform conditional backbone execution. The individual architectural components are described in the following subsections.

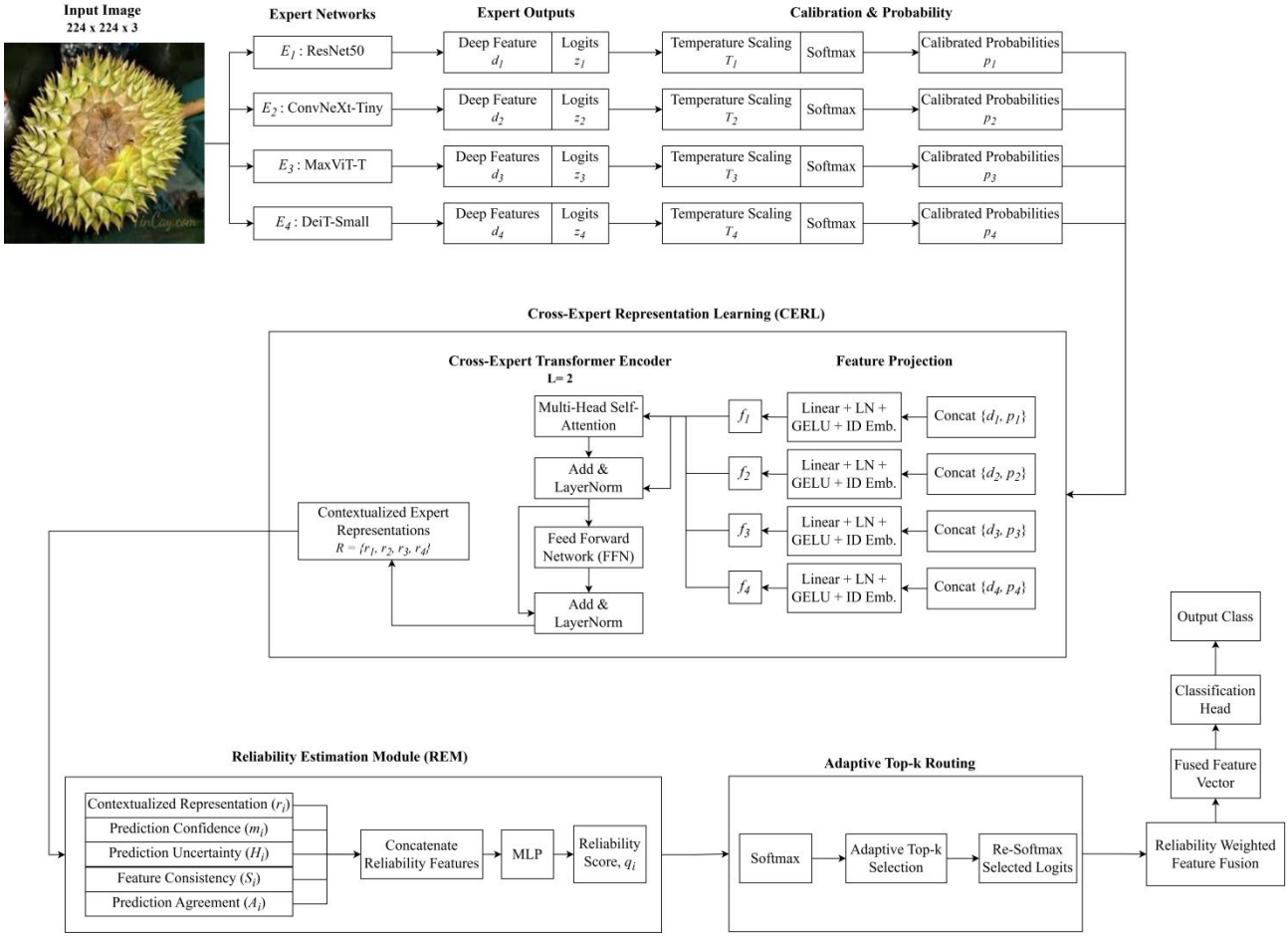


Fig. 4. The architecture of the RACEGNet.

3.2.1 Expert Networks

The first stage of RACEGNet consists of four heterogeneous constituent experts: ResNet-50, ConvNeXt-Tiny, MaxViT-T, and DeiT-Small/16. The pool combines convolutional, Transformer, and hybrid design principles to provide complementary representational behavior. Each expert is initialized from ImageNet-pretrained weights, independently fine-tuned on the durian disease training set, and then frozen before RACEGNet development. Consequently, the expert checkpoints remain unchanged while the trainable RACEGNet modules learn how to contextualize, assess, route, and fuse their outputs.

For an input image x, the i-th frozen expert E_i produces a high-level feature representation $f_i(x)$ and pre-Softmax class logits $\ell_i(x)$, defined in Eq. (1) and Eq. (2).

$$f_i(x) = \phi_i(x) \in \mathbb{R}^{D_i} \tag{1}$$

$$\ell_i(x) = g_i\big(f_i(x)\big) \in \mathbb{R}^{C} \tag{2}$$

where $\varphi_i(\cdot)$ denotes the feature extractor and $g_i(\cdot)$ the original classification head of expert i. The feature dimensions are architecture dependent: 2048 for ResNet-50, 768 for ConvNeXt-Tiny, 512 for MaxViT-T, and 384 for DeiT-Small/16. Here, M = 4 denotes the number of experts and C = 10 the number of disease classes. Both $f_i(x)$ and $\ell_i(x)$ are retained because RACEGNet jointly exploits representation-level information and calibrated prediction behavior.

3.2.2 Temperature Calibration

Because heterogeneous networks can exhibit different confidence calibration even when they achieve similar classification accuracy, each frozen expert is calibrated using scalar temperature scaling before its probabilities are supplied to RACEGNet. A separate set of 100 reserved validation images (10 per class) is used exclusively for post-training calibration; these images are not reused for expert checkpoint selection or RACEGNet development. For expert i, a positive scalar temperature is fitted by minimizing the calibration-set negative log-likelihood given in Eq. (3). The calibrated class probability vector is obtained according to Eq. (4).

$$T_i^* = \arg\min_{T>0} \left[-\frac{1}{N_{\text{cal}}} \sum_{n=1}^{N_{\text{cal}}} \log \left\{ \text{softmax}\left(\frac{\ell_i(x_n)}{T} \right)_{y_n} \right\} \right] \tag{3}$$

The calibrated class-probability vector is then obtained as

$$p_i(x) = \text{softmax}\left(\frac{\ell_i(x)}{T_i^*}\right) \in \mathbb{R}^C \quad (4)$$

Temperature scaling rescales the confidence distribution without changing the ordering of the logits and therefore leaves the expert Top-1 prediction unchanged. The calibrated probabilities are subsequently used in CERL and REM, as well as in probability-based ensemble baselines. This step places confidence- and entropy-derived reliability cues from heterogeneous experts on a more comparable scale.

3.2.3 Cross-Expert Representation Learning

The four experts produce feature vectors with different dimensionalities, which prevents direct interaction in their original representation spaces. Cross-Expert Representation Learning (CERL) resolves this heterogeneity by first projecting each expert descriptor into a common latent space and then contextualizing the resulting expert tokens with a Cross-Expert Transformer. Explicit reliability cues are excluded from the projection descriptor and are reserved for REM so that representation learning and reliability estimation retain distinct roles.

For expert i, the raw CERL descriptor is formed by concatenating the deep feature and calibrated probability vector as defined in Eq. (5). The expert projection and token embedding process follow Eq. (6), while the token matrix construction is given in Eq. (7) and the contextualized representation is obtained using Eq. (8).

$$d_i(x) = [f_i(x) \,\|\, p_i(x)] \quad (5)$$

An expert-specific projection $\psi_i(\cdot)$, implemented as Linear → LayerNorm → GELU, maps $d_i(x)$ to the shared latent dimension. A learnable expert-identity embedding e_i is then added:

$$z_i(x) = \psi_i\big(d_i(x)\big) + e_i \in \mathbb{R}^d \quad (6)$$

The projected tokens are stacked as

$$Z(x) = [z_1(x); z_2(x); \dots; z_M(x)] \in \mathbb{R}^{M\times d} \quad (7)$$

with shared latent dimension d = 64. The token matrix is processed by a two-layer Cross-Expert Transformer encoder with four attention heads, a feed-forward dimension of 128, and dropout of 0.10. The resulting contextualized expert representations are

$$H(x) = \text{CET}\big(Z(x)\big) = [h_1(x); h_2(x); \dots; h_M(x)] \quad (8)$$

Multi-head self-attention allows each expert token to condition its representation on evidence from the remaining experts before reliability is estimated. The contextualized vectors $h_i(x)$ retain expert-specific information while incorporating cross-expert context, and they are used by both REM and the subsequent feature-fusion stage.

3.2.4 Reliability Estimation Module

The Reliability Estimation Module (REM) estimates the competence of each expert separately for every input sample. Rather than relying on confidence alone, REM combines the contextualized representation with four complementary reliability cues: normalized entropy, maximum confidence, peer-contextual feature similarity, and distributional prediction agreement. The first two describe prediction certainty, whereas the latter two quantify consistency with the remaining experts. Reliability cues including normalized entropy and maximum confidence are calculated using Eq. (9) and Eq. (10). The peer contextual representation follows Eq. (11), and the feature similarity is calculated according to Eq. (12). The peer probability distribution and midpoint distribution are defined in Eq. (13) and Eq. (14). Jensen-Shannon divergence and agreement score are obtained from Eq. (15) and Eq. (16). The REM input construction and reliability prediction follow Eq. (17) and Eq. (18).

For expert i, normalized entropy and maximum confidence are computed from the calibrated probability vector as

$$u_i(x) = -\frac{1}{\log C}\sum_{c=1}^{C} p_{ic}(x)\log(p_{ic}(x) + \varepsilon) \quad (9)$$

$$c_i(x) = \max_{1\le c\le C} p_{ic}(x) \quad (10)$$

where $\varepsilon = 10^{-8}$ is used for numerical stability. To quantify feature-level consistency, a leave-one-out peer contextual representation is first formed:

$$\overline{h}_{-i}(x) = \frac{1}{M-1}\sum_{j\neq i} h_j(x) \quad (11)$$

The peer-contextual feature similarity is then obtained from the normalized cosine similarity

$$\gamma_i(x) = \text{clip}\left(\frac{1+\cos\left(h_i(x), \overline{h}_{-i}(x)\right)}{2}, 0,1\right) \quad (12)$$

Prediction agreement is computed from the calibrated probability distributions. The leave-one-out peer probability distribution and its midpoint with p_i(x) are

$$\overline{p}_{-i}(x) = \frac{1}{M-1}\sum_{j\neq i} p_j\,(x) \tag{13}$$

$$m_i(x) = \frac{1}{2}\left[p_i(x) + \overline{p}_{-i}(x)\right] \tag{14}$$

Their Jensen-Shannon divergence is

$$J_i(x) = \frac{1}{2}D_{\mathrm{KL}}\left(p_i(x)\|m_i(x)\right) + \frac{1}{2}D_{\mathrm{KL}}\left(\overline{p}_{-i}(x)\|m_i(x)\right) \tag{15}$$

and the normalized distributional agreement score is

$$\alpha_i(x) = \mathrm{clip}\left(1 - \frac{J_i(x)}{\log 2}, 0{,}1\right) \tag{16}$$

The contextualized representation and the four scalar cues are concatenated into the REM input

$$r_i(x) = [h_i(x) \,\|\, u_i(x) \,\|\, c_i(x) \,\|\, \gamma_i(x) \,\|\, \alpha_i(x)] \tag{17}$$

which is processed by a multilayer perceptron with dimensions 68 → 64 → 64 → 1 to produce the sample-specific reliability logit

$$s_i(x) = \mathrm{REM}\left(r_i(x)\right) \tag{18}$$

The reliability logit is explicitly supervised during training using frozen-expert correctness targets, as described in Section 3.3. Thus, REM learns sample-specific expert competence rather than treating raw confidence as a direct proxy for reliability.

3.2.5 Adaptive Top-k Reliability Routing

The REM logits determine how many and which contextualized expert representations participate in the final fusion. Dense routing probabilities are generated using Eq. (19). Adaptive Top-k selection follows Eq. (20), with the selected expert set and mask defined in Eq. (21). Non-selected logits are masked using Eq. (22) and sparse routing weights are calculated using Eq. (23). The reliability logits are first converted into dense routing probabilities across the M experts:

$$a_i(x) = \frac{\exp\left(s_i(x)\right)}{\sum_{j=1}^{M}\exp\left(s_j(x)\right)} \tag{19}$$

After sorting the routing probabilities in descending order, RACEGNet uses a fixed cumulative routing threshold $\tau = 0.50$ and selects the smallest sample-specific cardinality whose cumulative routing mass reaches that threshold:

$$k(x) = \min\left\{k \in \{1,\ldots,M\}: \sum_{j=1}^{k} a_{(j)}\,(x) \geq \tau\right\}, \qquad \tau = 0.50 \tag{20}$$

The selected expert set and corresponding hard membership mask are

$$\delta_i(x) = \mathbf{1}[i \in \mathcal{S}(x)], \qquad \mathcal{S}(x) = \mathrm{Top}_{k(x)}\left(a(x)\right) \tag{21}$$

All non-selected reliability logits are then masked to negative infinity:

$$\tilde{s}_i(x) = \begin{cases} s_i(x), & \delta_i(x) = 1, \\ -\infty, & \delta_i(x) = 0, \end{cases} \tag{22}$$

Finally, the selected logits are re-normalized to obtain sparse routing weights:

$$w_i(x) = \frac{\exp\left(\tilde{s}_i(x)\right)}{\sum_{j=1}^{M}\exp\left(\tilde{s}_j(x)\right)} \tag{23}$$

This mechanism makes both expert identity and routing cardinality sample dependent. Importantly, all four frozen backbones are executed before REM and routing. Adaptive Top-k therefore introduces sparsity only at the post-backbone feature-fusion stage and should not be interpreted as conditional backbone execution.

3.2.6 Reliability-Weighted Feature Fusion and Final Prediction

RACEGNet performs feature-level rather than probability-level fusion in its full configuration. Reliability-weighted feature fusion is performed according to Eq. (24). The final posterior probability distribution is obtained using Eq. (25), and the final disease prediction is determined using Eq. (26). The sparse routing weights are applied to the contextualized expert representations to form a 64-dimensional fused feature vector:

$$h_{\mathrm{fused}}(x) = \sum_{i=1}^{M} w_i\,(x)h_i(x) \tag{24}$$

The fused representation is passed to the final trainable linear classifier, followed by Softmax, to obtain the posterior distribution over the C disease classes:

$$\hat{p}(x) = \text{softmax}(W_o h_{\text{fused}}(x) + b_o) \tag{25}$$

The final disease prediction is

$$\hat{y}(x) = \underset{1 \le c \le C}{\text{argmax}}\ \hat{p}_c(x) \tag{26}$$

Because $w_i(x) = 0$ for every unselected expert, Eq. (24) automatically reduces to a single contextualized representation when $k(x) = 1$ and to a reliability-weighted combination when multiple experts are selected. In all cases, the representations entering the fusion stage have already incorporated cross-expert context through CERL.

3.3 Model Optimization

Only the RACEGNet-specific projection, contextualization, reliability-estimation, routing/fusion, and final-classification parameters are optimized during RACEGNet development; the four expert backbones and fitted temperatures remain fixed. The routing threshold is fixed at $\tau = 0.50$ throughout development, final retraining, and independent test evaluation. The binary expert correctness target is defined in Eq. (27). The reliability BCE loss is formulated in Eq. (28), with positive-target estimation and balancing weight calculated using Eq. (29) and Eq. (30). The reliability loss is given in Eq. (31). Expert routing statistics are computed using Eq. (32) and Eq. (33), followed by the utilization regularizer in Eq. (34). The classification loss is defined in Eq. (35), and the complete RACEGNet optimization objective is formulated in Eq. (36).

For sample n and expert i, the binary expert-correctness target is defined from the calibrated frozen-expert prediction as

$$t_{ni} = \mathbf{1}\left[\underset{1 \le c \le C}{\arg\max}\ p_{i,c}(x_n) = y_n\right] \tag{27}$$

REM is supervised against these targets using a batch-balanced binary cross-entropy objective. For reliability logit s_{ni}, the elementary BCE term is

$$\ell_{ni}^{\text{BCE}} = -t_{ni}\log\sigma(s_{ni}) - (1 - t_{ni})\log[1 - \sigma(s_{ni})] \tag{28}$$

The positive-target fraction is estimated across all $B \times M$ sample-expert pairs as

$$\pi_{\text{pos}} = \frac{1}{BM}\sum_{n=1}^{B}\sum_{i=1}^{M} t_{ni} \tag{29}$$

For numerical stability, π_{pos} is clipped in implementation to $[10^{-4}, 1 - 10^{-4}]$. The corresponding target-frequency balancing weight is

$$\omega_{ni} = \begin{cases} \dfrac{0.5}{\pi_{\text{pos}}}, & t_{ni} = 1, \\ \dfrac{0.5}{1 - \pi_{\text{pos}}}, & t_{ni} = 0. \end{cases} \tag{30}$$

and the reliability loss is

$$\mathcal{L}_{\text{rel}} = \frac{1}{BM}\sum_{n=1}^{B}\sum_{i=1}^{M} \omega_{ni}\,\ell_{ni}^{\text{BCE}} \tag{31}$$

This weighting gives the correct and incorrect reliability-target groups approximately equal aggregate influence within each mini-batch, reducing domination by the more frequent group.

An auxiliary utilization regularizer is used to discourage persistent routing collapse toward a single expert. Let a_{ni} be the dense pre-selection routing probability and δ_{ni} the hard selection indicator for sample n. The mean dense routing mass and normalized selection frequency for expert i are

$$P_i = \frac{1}{B}\sum_{n=1}^{B} a_{ni} \tag{32}$$

$$F_i = \frac{\sum_{n=1}^{B}\delta_{ni}}{\sum_{j=1}^{M}\sum_{n=1}^{B}\delta_{nj}} \tag{33}$$

The utilization term is then

$$\mathcal{L}_{\text{util}} = M\sum_{i=1}^{M} \text{sg}\,(F_i)P_i \tag{34}$$

where $\text{sg}(\cdot)$ denotes the stop-gradient operator. The observed selection frequency is treated as constant during backpropagation, whereas gradients pass through the dense routing probabilities. This term regularizes expert usage at the fusion stage; it is not a computational load-balancing objective because all four backbones are evaluated for every input.

The primary task loss is the cross-entropy of the final RACEGNet posterior:

$$\mathcal{L}_{\text{cls}} = -\frac{1}{B}\sum_{n=1}^{B} \log \hat{p}_{y_n}(x_n) \tag{35}$$

The complete training objective is

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{cls}} + 0.20\,\mathcal{L}_{\text{rel}} + 0.01\,\mathcal{L}_{\text{util}} \tag{36}$$

The trainable RACEGNet parameters are optimized with Adam using a learning rate of 1×10^{-3}, weight decay of 1×10^{-4}, and batch size 32. Training is allowed for a maximum of 50 epochs with early-stopping patience of five epochs, using meta-validation classification loss as the stopping criterion. The checkpoint with the lowest meta-validation classification loss determines the best epoch E*. RACEGNet is then reinitialized and retrained from scratch on all 200 development images for exactly E* epochs with $\tau = 0.50$ fixed. Only after this final retraining stage is the independent test set instantiated and evaluated.

3.4 Experimental Setup

All experiments were implemented and executed in the Kaggle Notebook cloud-computing environment using Python 3.12.13, PyTorch 2.10.0, Torchvision 0.25.0, and timm 1.0.26. Model training, inference, and computational analyses were performed on an NVIDIA Tesla T4 GPU with CUDA 12.8 and cuDNN support. Automatic mixed-precision computation was enabled where applicable to improve computational efficiency and reduce GPU memory usage. Data loading was performed using two worker processes with pinned memory. Unless stated otherwise, a fixed random seed of 42 was used to support reproducibility throughout data partitioning, model development, and evaluation.

The durian disease dataset comprised 10 classes and was organized into fixed training, validation, and independent test partitions containing 4,154, 649, and 648 images, respectively. The same class-to-index mapping was preserved across all data partitions and experiments. For model development, the original 649-image validation set was further partitioned deterministically using seed 42 into three mutually exclusive subsets. Of these, 349 images were used exclusively for early stopping and checkpoint selection of the four expert networks, 100 images (10 per class) were reserved for post-training temperature calibration, and 200 images (20 per class) were reserved for RACEGNet development. The 200-image RACEGNet subset was further divided by stratified sampling into 160 meta-training and 40 meta-validation images. Thus, expert checkpoint selection, probability calibration, and RACEGNet development were conducted using separate, non-overlapping subsets.

A common input-processing protocol was maintained across the expert networks, RACEGNet, independent comparison models, and ablation experiments. All images were processed at a spatial resolution of 224 × 224 pixels, and a batch size of 32 was used for standard training and evaluation unless otherwise specified. During training, online augmentation included random resized cropping with a scale range of 0.85–1.00, random horizontal flipping, random rotation of up to ±15°, and color jitter with brightness, contrast, saturation, and hue factors of 0.15, 0.15, 0.10, and 0.05, respectively. Images were then converted to tensors and normalized using the ImageNet mean [0.485, 0.456, 0.406] and standard deviation [0.229, 0.224, 0.225]. Validation and test images were processed deterministically by resizing to 224 × 224 pixels, tensor conversion, and the same ImageNet normalization, without data augmentation. The same predefined data partitions, class mapping, preprocessing procedure, and evaluation protocol were retained throughout the study to ensure fair comparison.

Four heterogeneous architectures were employed as RACEGNet experts: ResNet-50, ConvNeXt-Tiny, MaxViT-T, and DeiT-Small/16. The Torchvision-based experts were initialized with their corresponding ImageNet-pretrained weights, while DeiT-Small/16 was initialized using its pretrained timm implementation. Their original classification heads were replaced with 10-class output layers, and all model parameters were fine-tuned on the training set. To maintain a controlled training protocol, all experts were optimized using AdamW with an initial learning rate of 1×10^{-4} and weight decay of 1×10^{-5}. Cross-entropy loss with label smoothing of 0.1 was used as the training objective, together with a CosineAnnealingWarmRestarts learning-rate scheduler with $T_0 = 10$. Training was allowed for a maximum of 50 epochs, with early stopping using a patience of five epochs. For independent architectural comparison, five additional models not used as RACEGNet experts—MobileNetV3-Large, EfficientNet-B0, DenseNet121, RegNet-Y-800MF, and Swin-Tiny—were trained using the same data split, preprocessing pipeline, and optimization protocol as the four expert networks.

After training, the four expert networks were frozen and used without further parameter updates during RACEGNet development. The RACEGNet fusion network was optimized using the Adam optimizer with a learning rate of 1×10^{-3}, weight decay of 1×10^{-4}, a maximum of 50 epochs, and an early-stopping patience of five epochs. The Adaptive Top-k routing threshold was τ=0.50 throughout RACEGNet development and evaluation. For comparison, five conventional ensemble approaches—majority voting, confidence-weighted voting, entropy-weighted voting, soft voting, and logistic stacking—were evaluated using the same frozen and calibrated expert outputs. All model-development and ensemble-selection decisions were completed before evaluation on the independent test set.

A hierarchical ablation study was conducted to assess the incremental contribution of the main RACEGNet components. The evaluated configurations included calibrated soft voting, complete CERL, CERL with REM, and the full RACEGNet incorporating Adaptive Top-k routing. Additional control experiments examined the effects of removing expert-correctness supervision and

utilization regularization. All ablation models followed the same data-partitioning and development protocol as RACEGNet. Statistical significance was assessed using exact paired McNemar tests on the common test set, with Holm correction applied within each predefined family of pairwise comparisons to account for multiple testing.

Model interpretability was examined using Grad-CAM++ visualizations for the four frozen expert networks. For each test image, all expert maps were generated with respect to the same target class to enable direct spatial comparison. For RACEGNet, a route-conditioned composite Grad-CAM++ visualization was generated by combining the expert-specific Grad-CAM++ maps according to the final sparse routing weights, with unselected experts assigned zero contribution. Representative test images from all ten disease classes were analyzed using the frozen final model configuration.

Computational complexity was evaluated for the individual expert models, the full four-expert soft-voting ensemble, and RACEGNet under the same NVIDIA Tesla T4 GPU environment and 224 × 224 input resolution. The analysis included model parameters, serialized model size, single-image inference latency, batch-32 throughput, and peak GPU memory consumption. For RACEGNet, computational measurements included the execution of all four expert backbones together with the subsequent fusion modules. Adaptive Top-k performs sparse selection only at the post-backbone fusion stage and therefore does not reduce backbone computation.

4. Results and Analysis

4.1. Performance Analysis of RACEGNet

The performance of the proposed RACEGNet was compared to its four constituent expert networks on the independent test dataset of 648 images. In this case, latency was recorded with a batch size of 1, while throughput was measured with a batch size of 32. The computational timings were computed without counting disk I/O, image decoding, preprocessing, and host to device transfers. As shown in Table 2, RACEGNet demonstrated the best performance in all measures, obtaining 96.30% test accuracy, 96.49% Macro Precision, 96.33% Macro Recall, and 96.34% Macro-F1. Among the individual experts, MaxViT-T obtained the best performance results with 93.67% accuracy and 93.69% Macro-F1 and DeiT-Small/16 with 93.21% accuracy and 93.35% Macro-F1. Thus, RACEGNet outperformed individual experts by 2.62 to 4.01% in test accuracy and by 2.66 to 4.08% in Macro-F1. These improvements indicate that reliability-aware integration of heterogeneous expert representations provides more effective classification than relying on any single constituent architecture. This improvement is additionally represented in Fig. 5 by comparison of performance of the four constituent experts and the proposed framework with test Accuracy and Macro-F1.

Table 2. Predictive performance and computational characteristics of the four expert models and RACEGNet

Model	Test Accuracy (%)	Macro Precision (%)	Macro Recall (%)	Macro-F1 (%)	Parameters (M)	Model-state footprint (MiB)	Latency (ms/image)	Throughput (images/s)	Peak GPU footprint (MiB)
ResNet-50	92.28	92.35	92.43	92.27	23.53	89.96	7.84	744.71	103.03
ConvNeXt-Tiny	92.90	93.27	93.00	92.99	27.83	106.16	7.92	630.06	114.53
MaxViT-T	93.67	93.96	93.78	93.69	30.41	116.60	33.84	251.18	134.18
DeiT-Small/16	93.21	93.88	93.23	93.35	21.67	82.66	6.99	1101.84	86.24
RACEGNet	96.30	96.49	96.33	96.34	103.76	396.59	56.49	128.92	414.18

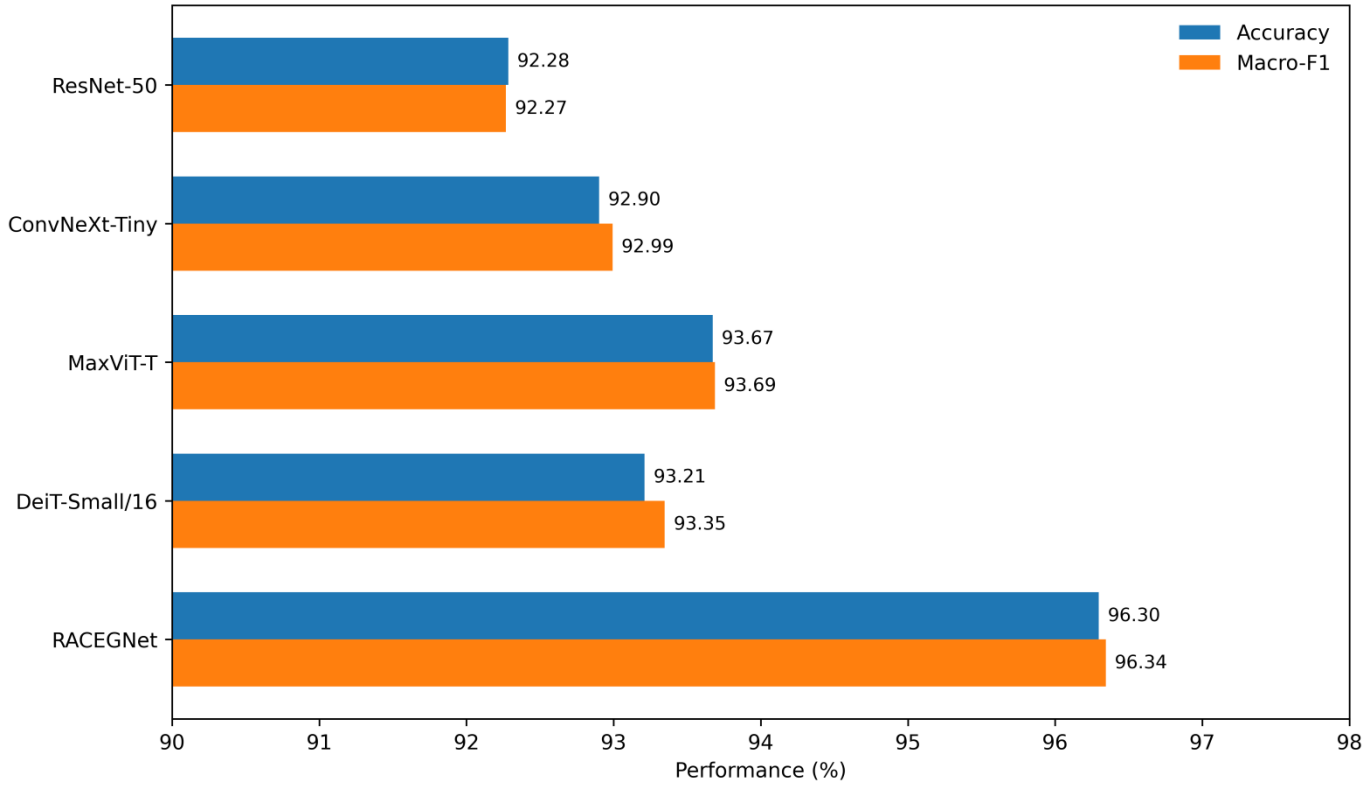


Fig. 5. Comparison of test Accuracy and Macro-F1 between the four constituent expert models and the proposed RACEGNet.

The confusion matrix of the individual experts and RACEGNet illustrated the class-wise distribution of correct and incorrect predictions, highlighting which disease categories are most frequently confused by the model. As shown in Fig. 6, compared to the four individual experts, RACEGNet showed a stronger concentration along the main diagonal, showing less misclassification. Regarding the total number of correctly classified samples, RACEGNet correctly identified 624 test images, while ResNet-50, ConvNeXt-Tiny, MaxViT-T, and DeiT-Small/16 correctly classified 598, 600, 607, and 604 images, respectively. As a result, RACEGNet improved the total number of misclassification by 26, 24, 17, and 20 samples. Especially when considering visually similar disease categories, the improvement was remarkable. For the mealybug infestation, RACEGNet correctly classified 59 samples, which improved by 3 samples more than ResNet-50 and 5 samples more than DeiT-Small/16 while giving comparable results to the other models. In the case of pink disease, the biggest improvement was found, as RACEGNet correctly classified 65 samples, which was 14 samples more than ResNet-50, 12 samples more than ConvNeXt-Tiny, 10 samples more than MaxViT-T, and equally to DeiT-Small/16. For the stem blight, RACEGNet correctly classified 61 samples, which improved by 7 samples more than ConvNeXt-Tiny and DeiT-Small/16, and 3 samples more than ResNet-50. Similarly, for sooty mold, RACEGNet achieved 57 correct predictions, improving over ResNet-50, MaxViT-T, and DeiT-Small/16 by 3, 4, and 1 samples, respectively. For fruit rot, RACEGNet also showed a small improvement by correctly classifying 61 samples, compared with 60, 59, 59, and 59 samples from ResNet-50, ConvNeXt-Tiny, MaxViT-T, and DeiT-Small/16, respectively. In contrast, classes with highly distinctive symptoms, such as canker disease, stem cracking gummosis, and yellow leaf, were already recognized effectively by individual experts, resulting in smaller improvements from fusion. Overall, these results demonstrate that RACEGNet provides the greatest benefit for challenging disease categories where individual models exhibit greater confusion, suggesting that reliability-aware adaptive fusion effectively combines complementary expert representations to reduce class-specific recognition errors.

The results of class-wise performance of the RACEGNet architecture are shown in Table 3. The proposed network presented strong generalization and balanced classification results on all ten durian disease classes based on the high values of Macro Precision (96.49%), Macro Recall (96.33%) and Macro-F1 scores (96.34%). Perfect precision, recall and F1-score values were obtained for canker disease and fruit rot, which means that all samples from those classes were classified without any errors. Classification of stem cracking gummosis was also very accurate because this class achieved perfect recall and high F1-score (98.55%), and a small drop in precision (97.14%) was caused by few false-positive predictions. Thrips disease and yellow leaf diseases were classified with a high level of precision with F1-scores 99.24% and 99.19% respectively. Classification of anthracnose disease was characterized by robust performance with precision 96.88%, recall 98.41% and F1-score 97.64%. On the contrary, there were different results of classification for visually challenging classes. Sooty mold was the most difficult class with lowest recall (86.36%) and lowest F1-score (89.76%), indicating that several true sooty mold samples were misclassified as other disease categories. Stem blight showed the lowest precision (87.14%) despite maintaining a high recall of 95.31%, suggesting that some samples from other classes were incorrectly assigned to this category. Mealybug infestation achieved perfect precision (100.00%) but a lower recall of 90.77%, indicating that although predictions assigned to this class were highly reliable, some true mealybug samples were missed. Pink disease showed intermediate difficulty, with precision, recall, and F1-score values of 90.28%, 95.59%, and 92.86%, respectively. Despite these class-specific variations, the close agreement among Macro Precision, Macro Recall, and Macro-F1 demonstrates that RACEGNet achieved well-balanced performance across the dataset rather than obtaining high overall performance through dominance of only a limited number of disease categories.

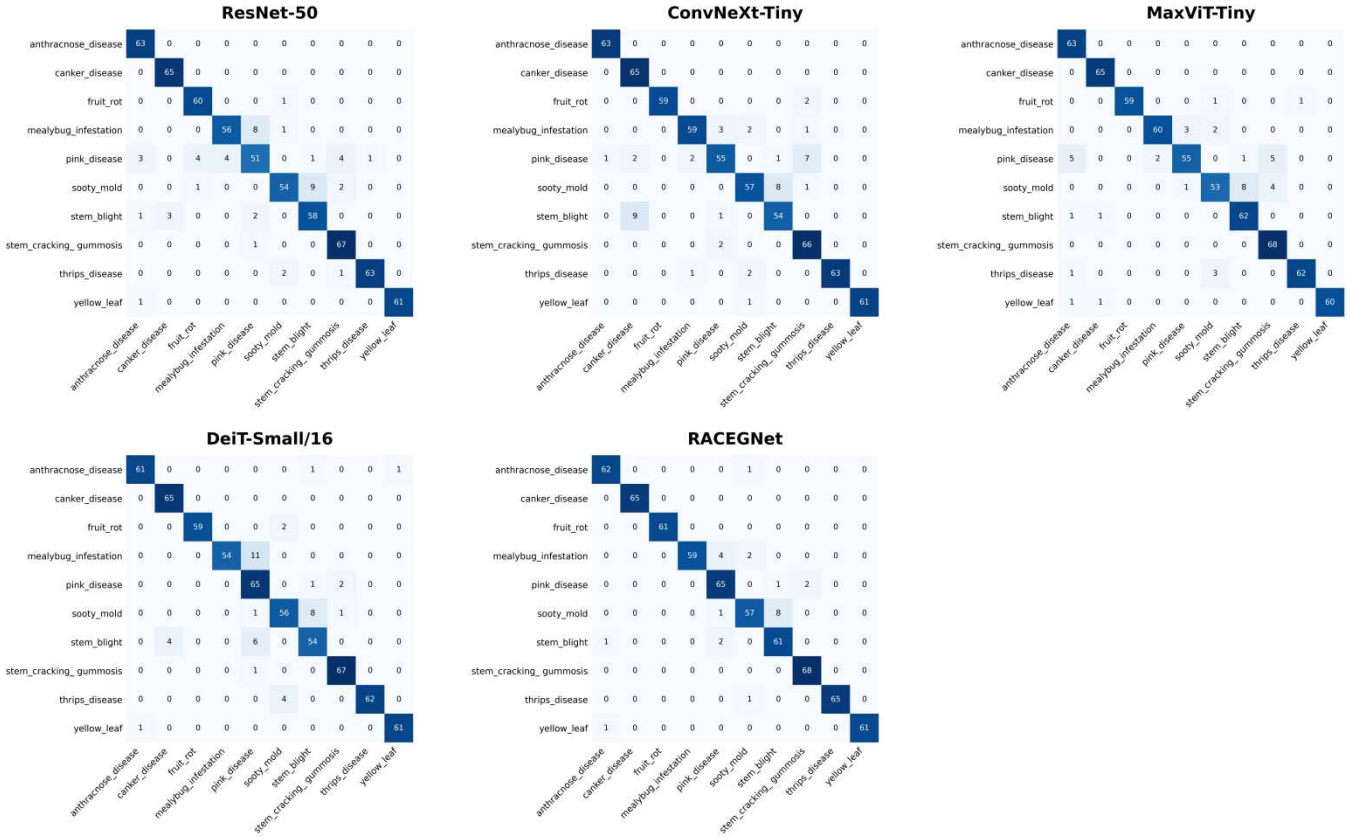


Fig. 6. Confusion matrices of ResNet-50, ConvNeXt-Tiny, MaxViT-T, DeiT-Small/16, and RACEGNet.

Table 3. Class-wise classification performance of RACEGNet

Disease class	Precision (%)	Recall (%)	F1-score (%)	Support
Anthracnose disease	96.88	98.41	97.64	63
Canker disease	100.00	100.00	100.00	65
Fruit rot	100.00	100.00	100.00	61
Mealybug infestation	100.00	90.77	95.16	65
Pink disease	90.28	95.59	92.86	68
Sooty mold	93.44	86.36	89.76	66
Stem blight	87.14	95.31	91.04	64
Stem cracking gummosis	97.14	100.00	98.55	68
Thrips disease	100.00	98.48	99.24	66
Yellow leaf	100.00	98.39	99.19	62

In terms of computational performance, the enhanced prediction accuracy of RACEGNet comes at the expense of increased computational demand at the level of the whole system when compared to the individual expert. The overall number of system parameters in RACEGNet is 103.76 million, its model-state memory usage is 396.59 MiB, and the required median inference time per one image is 56.49 ms, with 128.92 images/s achieved throughput. The majority of the mentioned computational costs come from the execution of four frozen experts' backbones, whereas the RACEGNet fusion modules add only 0.317 million extra parameters. Adaptive Top-k method chose one expert's representation for 96 out of 648 test images (14.81%), and two experts' representations for the rest of 552 images (85.19%), which yielded the mean value of k=1.85. Hence, even though all four experts are used for every input, only a fraction of their representations is involved in the post-backbone fusion process.

To assess whether the improvements compared to the individual experts were statistically significant, exact paired McNemar tests were conducted on the predictions on the same 648 test images. Holm correction was performed for all four prespecified comparisons. n_{10} is the number of test images that were correctly predicted by RACEGNet but not by the competing method, while

n_{01} is the opposite. As can be seen from Table 4, RACEGNet was significantly better than all four individual expert networks after the adjustment for multiple comparisons. The largest improvement was made against ResNet-50, where RACEGNet demonstrated a 4.01 percentage-point difference in the accuracy, with Holm corrected p-value equal to 2.47×10^{-5}. RACEGNet was also significantly better than ConvNeXt-Tiny (+3.40 percentage points), DeiT-Small/16 (+3.09 percentage points), and MaxViT-T (+2.62 percentage points), with all Holm adjusted p-values below 0.001.

The discordant-pair counts further support these findings. For ResNet-50, RACEGNet correctly classified 30 samples that ResNet-50 misclassified, whereas the reverse occurred for only four samples. Similar asymmetries were observed for ConvNeXt-Tiny (29 vs. 7), MaxViT-T (21 vs. 4), and DeiT-Small/16 (25 vs. 5). These results provide paired statistical evidence that the accuracy improvement of RACEGNet is systematic rather than resulting solely from differences in aggregate accuracy. It should be noted that the McNemar test evaluates paired classification correctness and therefore the reported significance applies to accuracy, not directly to Macro-F1.

Table 4. Exact paired McNemar significance tests comparing RACEGNet with the four constituent expert models

Comparator	Comparator Accuracy (%)	RACEGNet Accuracy Gain (pp)	n_{10}	n_{01}	Exact McNemar p-value	Holm-adjusted p-value	Significant
ResNet-50	92.28	+4.01	30	4	6.16×10^{-6}	2.47×10^{-5}	Yes
ConvNeXt-Tiny	92.90	+3.40	29	7	3.13×10^{-4}	9.38×10^{-4}	Yes
MaxViT-T	93.67	+2.62	21	4	9.11×10^{-4}	9.38×10^{-4}	Yes
DeiT-Small/16	93.21	+3.09	25	5	3.25×10^{-4}	9.38×10^{-4}	Yes

The results in Table 2, Fig. 4, and Table 3 demonstrate that RACEGNet provides consistently stronger overall and class-wise classification performance than its individual constituent experts. The significance analysis in Table 4 further confirms that these accuracy improvements are statistically robust. The improvement is achieved at the expected computational cost of evaluating all four heterogeneous expert backbones, while Adaptive Top-k restricts the subsequent fusion stage to a small, sample-specific subset of expert representations.

4.2. Ablation Study

A hierarchical complete-block ablation study was conducted to examine the contribution of the major components of RACEGNet. The analysis progressively incorporated the complete Cross-Expert Representation Learning (CERL), Reliability Estimation Module (REM), and Adaptive Top-*k* routing blocks. As summarized in Table 5, the calibrated four-expert soft-voting model (A0) served as a conventional ensemble baseline and should not be considered a RACEGNet configuration.

Table 5. Hierarchical complete-block ablation of the proposed RACEGNet framework

Configuration	CERL	REM	Adaptive Top-k	Accuracy (%)	Macro Precision (%)	Macro Recall (%)	Macro-F1 (%)	Mean k
A0	✗	✗	✗	94.75	94.92	94.83	94.81	4.00
A1	✓	✗	✗	94.91	95.31	94.95	95.00	4.00
A2	✓	✓	✗	95.37	95.75	95.42	95.42	4.00
A3	✓	✓	✓	96.30	96.49	96.33	96.34	1.85

The results show a progressive improvement as the complete RACEGNet components were incorporated. While replacing the standard method of averaging the probabilities from the calibrated model with the full-stage CERL-based feature fusion brought only a minor improvement in terms of Macro-F1 from 94.81% to 95.00%, which corresponds to the increase of 0.20 percentage points. The introduction of REM boosted Macro-F1 to 95.42%, providing another 0.42 percentage points gain.

However, the most substantial increase was observed upon Adaptive Top-k routing application. In comparison to the dense CERL+REM architecture, the increase in Macro-F1 for the complete RACEGNet was 0.92 percentage points – from 95.42% to 96.34%, and accuracy increased from 95.37% to 96.30%. It should be noted that architectures A2 and A3 contained the same number of post-backbone parameters (0.317 M); thus, the performance improvement was not achieved due to parameter growth. However, the introduction of Adaptive Top-k resulted in a shift of the fusion method from the full participation of all four expert representations to the sample-wise sparse one; hence, the mean number of representations taking part in the fusion process decreased from 4.00 to 1.85. Overall, the complete RACEGNet framework improved Macro-F1 by 1.54 percentage points over calibrated soft voting.

To further examine whether RACEGNet performance depended on the number and composition of available experts, an expert-count sensitivity analysis was conducted using all prespecified expert combinations. The results are summarized in Table

6. The one-expert condition represents the four individual calibrated experts, whereas the two- and three-expert conditions contain all six possible pairs and all four possible triplets, respectively. The four-expert condition corresponds to the exact final RACEGNet configuration. Mean and standard deviation were calculated across all available configurations at each expert count to avoid selecting only the best-performing subset.

Table 6. Expert-count sensitivity of RACEGNet across all prespecified expert compositions

Expert count	No. of configurations	Accuracy mean ± SD (%)	Macro-F1 mean ± SD (%)	Macro-F1 range (%)	Mean selected k	Mean k/M
1	4	93.02 ± 0.58	93.07 ± 0.61	92.27-93.69	1.00	1.000
2	6	94.78 ± 0.89	94.84 ± 0.86	93.94-95.72	1.00	0.500
3	4	94.83 ± 0.78	94.90 ± 0.77	94.37-96.03	1.71	0.570
4	1	96.30	96.34	96.34	1.85	0.463

The mean Macro-F1 increased from 93.07% for the individual experts to 94.84% across all six two-expert configurations and 94.90% across all four three-expert configurations. The full model with four heterogeneous experts achieved the best Macro-F1 score of 96.34%. Importantly, despite the availability of four different expert representations for fusion after the backbone, on average, only 1.85 expert representations participated in the process (on average, the ratio of selected to total experts' number equals to k/M = 0.463).

As a part of the expert composition analysis, it was noticed that the influence of the individual experts was not equal to each other. During the leave-one-expert-out experiment, the deletion of the MaxViT-T expert caused the largest decrease of the Macro-F1 metric in comparison with the full model (1.97 percentage points). Then came ConvNeXt-Tiny with 1.84 percentage points and ResNet-50 with 1.68 percentage points. The deletion of DeiT-Small/16 had a much lower influence – 0.31 percentage points. Such a discrepancy supports the need for the estimation of the sample-specific reliability of experts.

The ablation results demonstrate that RACEGNet benefits from the progressive integration of contextual cross-expert representation learning, reliability estimation, and adaptive sparse fusion. CERL provided a modest initial improvement, REM introduced an additional gain through sample-specific reliability modeling, and Adaptive Top-*k* produced the largest incremental improvement. The expert-count analysis further showed that the complete heterogeneous expert pool yielded the strongest performance while RACEGNet retained sparse, sample-dependent participation at the final fusion stage.

4.3. Explainability Analysis

To assess the spatial evidence employed by individual experts and the proposed RACEGNet model, Grad-CAM++ visualization technique was used on representative test images of all ten durian diseases classes. From Fig. 5, it is observed that the activation patterns in the four heterogeneous experts are quite different for the same input image, indicating that the experts have complementary spatial evidence and not identical visual evidence. ResNet-50 generated activation areas that were more localized in general as compared to those generated by ConvNeXt-Tiny, MaxViT-T, and DeiT-Small/16 which depended on the manifestation of the disease.

The RACEGNet maps showed strong spatial correlation with the symptomatic regions of various disease classes. For the anthracnose disease class, the activation map was more prominent around the visibly infected regions of the leaves. The activation pattern of canker disease focused on the infected region of the leaves while mealybugs showed activation around the visibly infected parts of the fruits. The pink disease and stem cracking gummosis classes showed activations around the abnormal parts of the bark and stems. For the stem blight and thrips disease classes, RACEGNet activations were around the visibly infected leaf regions. In the yellow-leaf example, the activation was distributed over the discolored foliage. For fruit rot and sooty mold, the highlighted regions were comparatively broader, consistent with the more spatially extensive visual manifestations present in the corresponding images.

The variations between the experts' maps are equally revealing. In some instances, an expert has mapped out a relatively concise area while another has mapped out an expansive region or part of the background region. RACEGNet has not replicated the activation map of any particular expert over all examples. On the contrary, it has generated a composite map based on the routing weights corresponding to each sample. This is in line with the design and function of REM and Adaptive Top-k such that the framework is able to weigh different experts based on each input. Fig. 7 gives qualitative proof of how each expert maps out complementary spatial information and how RACEGNet fuses their discriminative information through reliability-aware sparse routing.

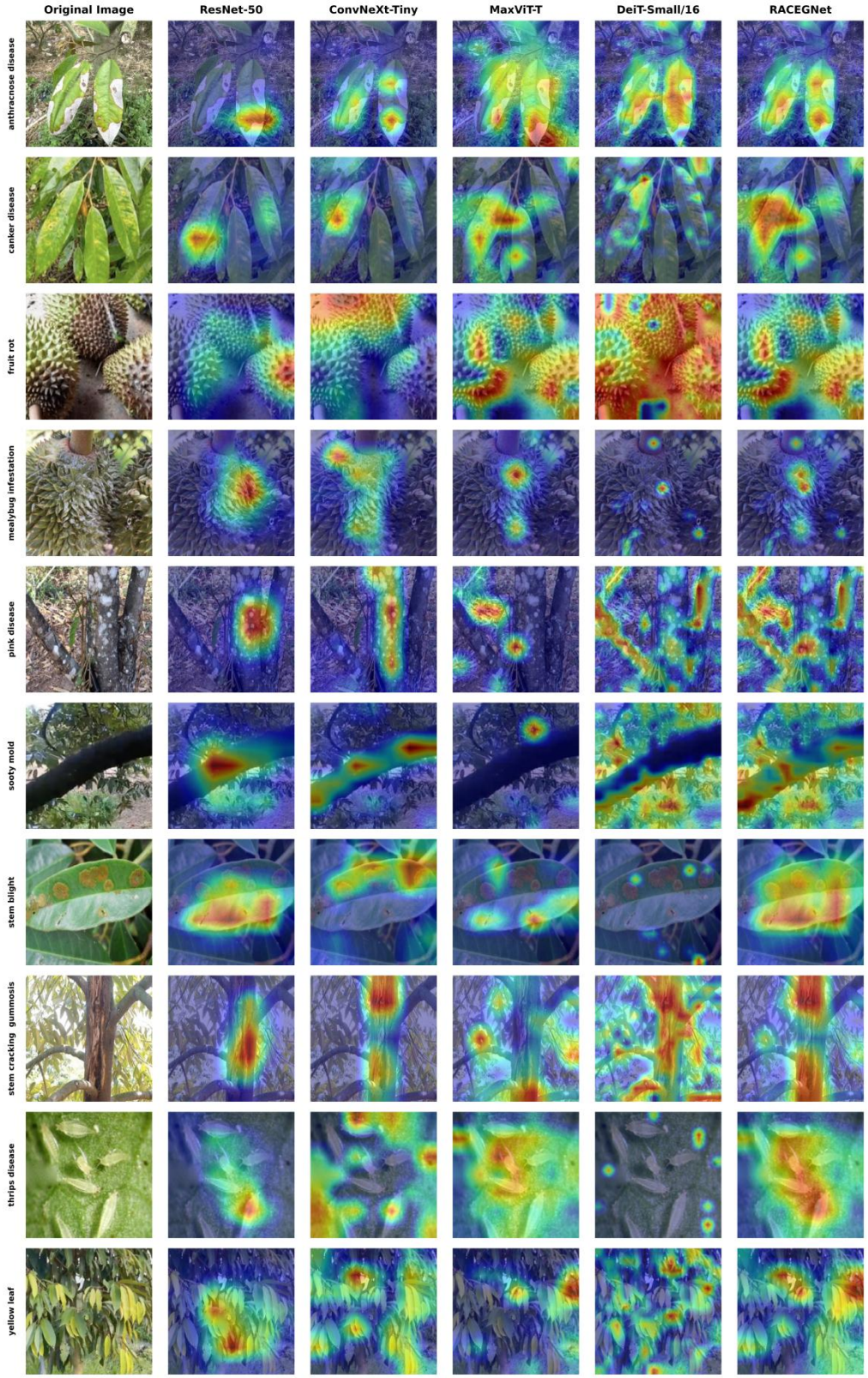


Fig. 7. *Grad-CAM++ visualization of the four constituent expert models and the proposed RACEGNet across the ten durian disease classes.*

4.4. Performance Comparison with Conventional Ensembles and Independent Architectures

To provide a broader assessment of the proposed framework, RACEGNet was compared with five conventional ensemble methods constructed from the same four frozen experts and with five independently trained deep-learning architectures. The complete comparison is presented in Table 7. Only predictive-performance metrics are reported in this section to maintain a consistent comparison across all methods.

Table 7. Performance comparison of RACEGNet with deep-learning baseline architectures and conventional ensemble methods

Model	Accuracy (%)	Macro Precision (%)	Macro Recall (%)	Macro F1-Score (%)
MobileNetV3-Large	92.75	92.86	92.84	92.80
EfficientNet-B0	87.65	87.98	87.85	87.71
DenseNet121	90.43	90.53	90.62	90.40
RegNet-Y-800MF	91.05	91.04	91.17	91.04
Swin-Tiny	92.44	92.72	92.56	92.43
Majority Voting	94.29	94.49	94.37	94.37
Confidence-Weighted Ensemble	94.44	94.68	94.54	94.50
Entropy-Weighted Ensemble	94.60	94.82	94.69	94.65
Linear Logistic Stacking	94.60	95.00	94.66	94.69
Soft Voting	94.75	94.92	94.83	94.81
RACEGNet	96.30	96.49	96.33	96.34

Among the traditional ensemble techniques considered, soft voting showed the best results, with accuracy at 94.75% and Macro-F1 at 94.81%. Next were linear logistic stacking and entropy weighting with Macro-F1 scores of 94.69% and 94.65%, respectively. While all the five traditional ensemble techniques showed improvement due to combining multiple experts' predictions, RACEGNet outperformed them with accuracy and Macro-F1 at 96.30% and 96.34%, respectively. Compared to the strongest traditional ensemble method, which was soft voting, RACEGNet had better accuracy and Macro-F1 by around 1.54 percentage points.

The above improvement compared to traditional ensembling is especially significant considering that both RACEGNet and traditional ensembling use the same experts' predictions, but the difference lies in their aggregation procedures. Majority voting aggregates experts' prediction on the basis of the predicted labels, whereas confidence and entropy-based aggregation takes into account the certainty of expert predictions. In turn, soft voting aggregates calibrated probabilities equally and logistic stacking does so by learning to aggregate on the probability level. As opposed to these traditional aggregation techniques, RACEGNet estimates the reliability of every input sample and adapts the combination of experts' predictions based on this information. The observed performance improvement therefore supports the use of reliability-aware and sample-adaptive fusion over conventional fixed aggregation.

RACEGNet also showed superior performance compared to all five independently trained deep-learning baseline architectures. Out of these architectures, the best one was MobileNetV3-Large with 92.75% accuracy and 92.80% Macro-F1, while Swin-Tiny provided 92.44% accuracy and 92.43% Macro-F1. RegNet-Y-800MF, DenseNet121, and EfficientNet-B0 provided 91.04%, 90.40%, and 87.71% of Macro-F1, respectively. When compared with the best independently trained architecture MobileNetV3-Large, RACEGNet showed superiority by providing 3.55 percentage points higher accuracy and 3.55 percentage points higher Macro-F1. The superiority of the proposed approach was even more pronounced when compared with other independent architectures.

As follows from the results presented in Table 7, RACEGNet achieved the best predictive performance when compared with two groups of approaches. Superiority over conventional ensemble methods shows that simple aggregation of several expert predictions may not be sufficient to obtain such a performance level, while the comparison with independent architectures shows that there is no dependence on specific expert models. The results obtained show the usefulness of reliability-aware cross-expert representation learning and adaptive sample-specific fusion. Comparative performance is also visualized in Fig. 8. The independently trained baseline models generally exhibited lower Accuracy and Macro-F1 than the conventional ensemble approaches, while RACEGNet achieved the highest values for both metrics. This visual trend reinforces the quantitative results reported in Table 7.

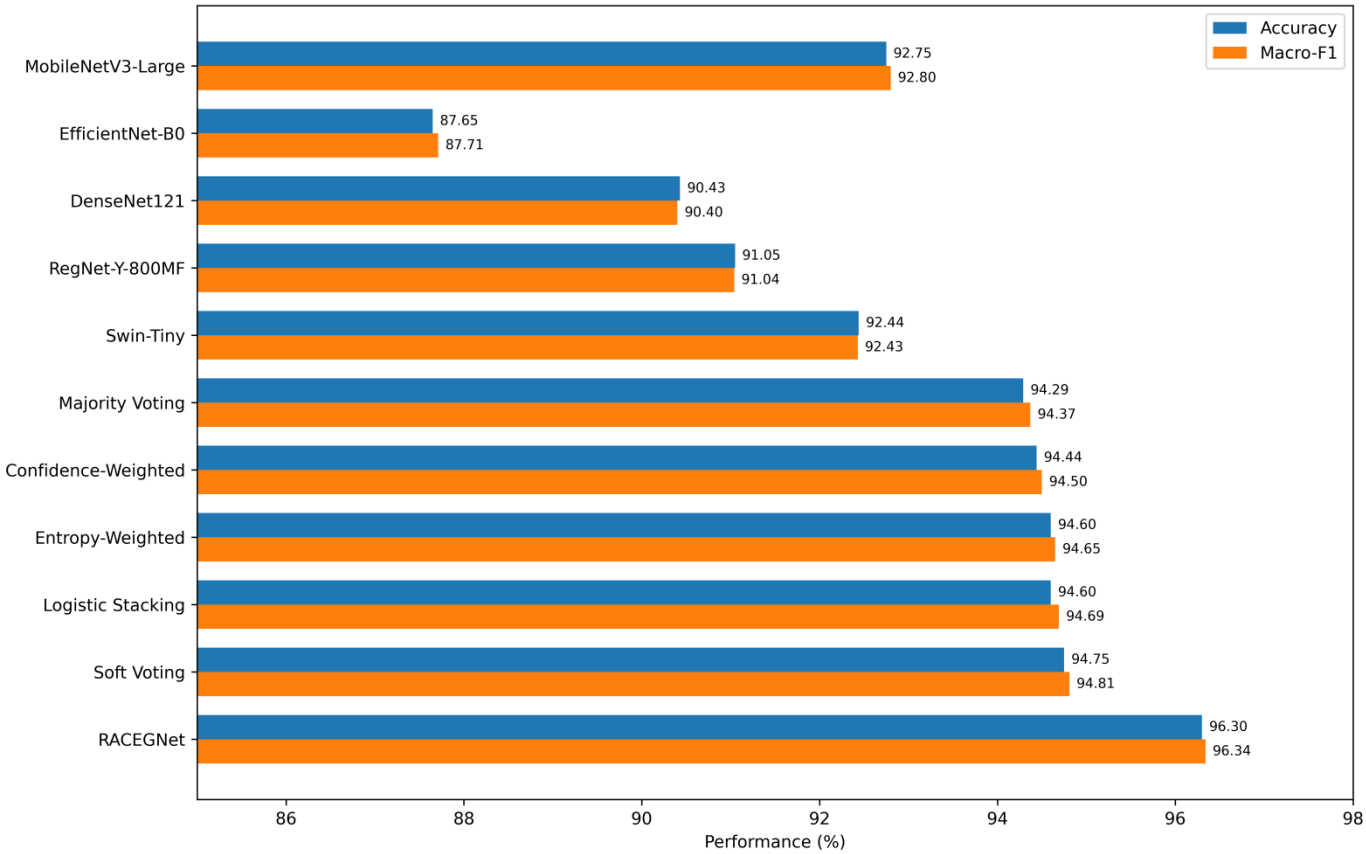


Fig. 8. Comparison of independent test accuracy among deep learning baseline models, conventional ensemble methods, and the proposed RACEGNet.

5. Dicussions

The results from the experiments reveal that RACEGNet is a suitable approach to overcome difficulties in heterogenous durian disease detection. Unlike traditional models that use only one feature representation, RACEGNet utilizes the strengths of multiple experts that are ResNet-50, ConvNeXt-Tiny, MaxViT-T, and DeiT-Small/16. The superior performance of RACEGNet over all individual experts indicates that disease recognition under realistic field conditions benefits from adaptive integration of multiple visual representations rather than dependence on a single universal feature extractor. This observation is important for detecting diseases in durian as the same disease can be observed on different parts of a plant with different characteristics of lesions, lighting, background, etc.

The proposed RACEGNet yielded accuracy of 96.30% and Macro-F1 score of 96.34%, surpassing the highest accuracy individual consitituent DL model, MaxViT-T, by 2.62% in terms of accuracy. The statistically proven better performance via McNemar test also indicates that the obtained better performance did not come about just through chance in the prediction but from better decisions at the sample level. This implies that the reliability-based fusion method is able to select the relevant complementary expertise that may not be detected by single architecture. The improvement over CNN-based, modern convolutional, and Transformer-based experts highlights the advantage of combining different representational inductive biases for complex agricultural image recognition.

The ablation study reveals more about how much contribution comes from each module in RACEGNet. Moving from traditional calibrated soft-voting to Cross-Expert Representation Learning brought noticeable performance boost which proves that contextualization of expert representations before fusion helps. This suggests that expert outputs should not be considered as independent predictions because relationships among experts, including representation similarity and prediction agreement, contain additional information useful for final classification. Introducing the Reliability Estimation Module helped to get even better results proving that estimation of the trustworthiness of expert for a certain input is crucial. The biggest gain was achieved after introducing Adaptive Top-k routing. When compared to the dense CERL+REM architecture, the full RACEGNet resulted in an increase in Macro-F1 score from 95.42% to 96.34%. Since the amount of trainable parameters was equal in both architectures, the gain is not related to the increased complexity of the architecture but is associated with the changed way of fusion. Adaptive Top-k allows the framework to dynamically determine which expert representations should contribute to the final prediction according to sample-specific reliability. This is especially important in heterogeneous disease scenarios where one image may contain discriminative local lesion patterns better captured by convolutional networks, whereas another may require broader contextual information captured by Transformer-based architectures.

The routing procedure further proves that RACEGNet does not simply replicate conventional ensemble averaging. The model uses only one expert representation for some of the samples and two or more expert representations for other samples, with a mean of 1.85 experts chosen per sample. The adaptability of this procedure means that the framework does not have to engage in unnecessary feature fusion if there is sufficient reliable evidence from one expert alone, but can still take advantage of collaboration if multiple experts provide

complementary information. Nevertheless, since all four expert backbones are used before routing, the current Adaptive Top-k scheme works on feature selection and not computation savings.

The analysis of the class-wise results also prove the effectiveness of using reliability-aware fusion for heterogeneous disease recognition. In particular, RACEGNet showed perfect performance in terms of classification of canker disease and fruit rot, whereas a number of other disease types also had F1-scores above 97%, which shows that the proposed approach allows extracting reliable visual features of distinct diseases. Nevertheless, the results have demonstrated some discrepancies in performance of RACEGNet for recognizing certain diseases with relatively lower scores achieved for sooty mold, stem blight, and pink disease. Such results suggest that there exist certain types of diseases which are more difficult to recognize because of their visual similarity, common symptomatology, and variations in appearance. It should be noted that differences in environmental conditions, symptom development stage, and image acquisition settings can cause problems of distinguishing visually similar diseases. Thus, the analysis of class-wise results shows that the recognition difficulty depends on the visual characteristics of a disease and the most difficult cases are those when diseases are visually similar. The ability of RACEGNet to adapt to such class-dependent visual variability demonstrates its relevance to a broader challenge in agricultural image recognition, where disease appearance may vary across crop species, affected plant organs, environmental conditions, and stages of disease development.

Considering the application in agriculture,the proposed framework offers several advantages compared to traditional manual diagnosis and fixed deep-learning classifiers. A successful automated diagnostic model does not just need to have a high accuracy, but it also needs to be robust to the variation in the appearance of the disease and the imaging conditions. As discussed above, RACEGNet takes advantage of multiple expert representations and sample-specific reliability estimation to provide a more adaptable tool for field-oriented disease diagnosis. This could help detect diseases early and aid decision-making in commercial orchards where it is not possible to get expert-level diagnosis quickly. Beyond durian disease recognition, the underlying challenge addressed by RACEGNet is common across many agricultural crops, where disease symptoms can vary due to differences in plant organs, disease progression stages, environmental conditions, and image acquisition settings. Since RACEGNet is designed for adapting with the challenge of visual heterogeneity through adaptive collaboration among diverse experts, the framework has potential for extension to other plant disease recognition tasks.

While having these advantages, there are some limitations that need to be addressed. Firstly, the experiment has been performed on one dataset only of durian diseases gathered in a particular geographical location. Even though the dataset included some field variation, further experiments are needed in order to validate the ability of the model to generalize through additional data sets of other cultivation locations and cameras. Secondly, the framework implies the employment of several expert backbones, which is resource-intensive compared to separate lightweight models. This means that future research should address ways of more effective model implementation, such as expert compression, knowledge distillation, or conditionally executing backbone experts. Third, although Grad-CAM++ analysis provides visual interpretation, further investigation using larger-scale explainability studies and expert validation would strengthen confidence in the learned routing behavior.

Future work can further develop RACEGNet's applicability from image classification to related tasks in agriculture, such as disease severity prediction, lesion detection, and multi-modal assessment of plant health. A key direction of future work is to explore the generalizability of RACEGNet to different plant species on the basis of multi-crop disease datasets. The motivation for such cross-crop evaluation arises from the fact that many agricultural disease recognition problems share common challenges. These include differences in the visual appearance of symptoms, disease development stages, affected plant parts, and image acquisition conditions. Some of the diseases and pests taken into account for the recognition of durian diseases, such as anthracnose, Phytophthora-related diseases, mealybug infestation, thrips damage, and sooty mold symptoms, can also be found in other economically valuable crops, providing a biological basis for evaluating the framework beyond durian. However, even if there is a common group of diseases or pests in several plants, the causal organisms, symptoms, and the process of manifestation might vary in different host plants. Consequently, the ability of RACEGNet to generalize must be determined experimentally in cross-crop validation tests rather than by assuming that due to the similar names of diseases or because of the visual similarities. Adding more information about weather, environment, and orchard management can make RACEGNet more robust in practice. In conclusion, the results show that the idea of adaptive reliability-aware fusion of heterogeneous experts is a promising solution for development of robust agricultural AI-based diagnostic systems, particularly in challenging environments characterized by high visual variability.

6. Conclusions

This study proposed the Reliability-Aware Cross-Expert Gating Network (RACEGNet), which is a heterogeneous expert fusion framework to address the challenge of visually diverse durian disease recognition in realistic field settings. Compared with traditional ensembling methods that adopt either a static or global pooling operation to fuse multiple experts' prediction, RACEGNet allows sample-level expert interaction through the mechanism of Cross-Expert Representation Learning, Reliability Estimation, and Adaptive Top-k selection. This framework is intended to take advantage of complementary representations of various CNN, modern convolutional networks, and Transformer based models as well as adaptively control the experts' contributions to individual images.

Experimental evaluation on a real-field durian diseases dataset showed that the proposed framework constantly outperformed several individual model and traditional ensembling methods. The framework achieved improved classification performance by effectively combining complementary visual representations and reduce errors due to heterogeneous disease appearance. Ablation study demonstrated that each part of the framework progressively contributes to the final performance and the Adaptive Top-k operation additionally helps to perform sparsity and reliability-driven feature fusion. Statistical analysis further verified that the observed improvements over individual experts were significant.

The findings demonstrate the potential of employing reliability aware heterogeneous expert collaboration as an effective technique for automated plant disease detection, especially within agricultural settings that are associated with wide variability in disease manifestation, plant organ affected and imaging conditions. The flexibility of RACEGNet due to its adaptive selection and combination of experts makes it an extendable solution for solving various recognition problems in agriculture.

While more validation in different geographical areas, crops and operating conditions is required, the results of this work clearly demonstrates that the integration of complementary deep features with reliability modeling is beneficial for increasing efficiency of AI-powered agricultural diagnosis systems.

Funding

This research received no external funding.

Ethics and Consent to Participate declarations: not applicable.

Consent to Publish declaration: not applicable.

Data Availability

The dataset analysed during the current study is available in the Mendeley Data repository,

https://doi.org/10.17632/mhjwyb5p48.1

References

[1] Chomdej, O., Tongyoo, P., Nualsri, C., Duangsodsri, T., Yamjabok, J., Attavanich, W., Poonpolgul, S., Poovarodom, S., Sahaya, S., Nakkanong, K. and Plodpai, P., 2026. Durian Varieties, Their Breeding and Intellectual Property Perspectives in Southeast Asia with Special References to Thailand. *Horticulturae*, *12*(7), p.831.
[2] Ali, N.A.W.A., Wong, G.R., Tan, B.C., Lum, W.S. and Mazumdar, P., 2025. Unleashing the potential of durian: Challenges, opportunities, and the way forward. *Applied Fruit Science*, *67*(1), p.3.
[3] Nguyen, T., 2025. Image dataset of ten durian diseases captured in real-field conditions from a family orchard in Vinh Long, Vietnam. *Data in Brief*, p.112244.
[4] Ghimire, S., Lamsal, R., Sankuratri, A., Lakkarsu, P. and Vutla, S., 2026. Artificial intelligence-driven plant disease detection and diagnosis: A comprehensive review of deep learning approaches, multimodal sensing technologies, and future perspectives in precision agriculture. *The Plant Pathology Journal*, *42*(2), p.121.
[5] Lowe, A., Harrison, N. and French, A.P., 2017. Hyperspectral image analysis techniques for the detection and classification of the early onset of plant disease and stress. *Plant methods*, *13*(1), p.80.
[6] Singh, D., Jain, N., Jain, P., Kayal, P., Kumawat, S. and Batra, N., 2020. PlantDoc: A dataset for visual plant disease detection. In *Proceedings of the 7th ACM IKDD CoDS and 25th COMAD* (pp. 249-253).
[7] Mohanty, S.P., Hughes, D.P. and Salathé, M., 2016. Using deep learning for image-based plant disease detection. *Frontiers in plant science*, *7*, p.215232.
[8] Noyan, M.A., 2022. Uncovering bias in the PlantVillage dataset. *arXiv preprint arXiv:2206.04374*.
[9] Shafik, W., Tufail, A., Liyanage De Silva, C. and Awg Haji Mohd Apong, R.A., 2025. A novel hybrid inception-xception convolutional neural network for efficient plant disease classification and detection. *Scientific Reports*, *15*(1), p.3936.
[10] Sönmez, D.P. and Kılıç, Ş., 2026. PlantPathNet: a novel network based on cross-layer feature integration for plant disease classification. *Frontiers in Plant Science*, *17*, p.1783576.
[11] Pai, P., Amutha, S., Patil, S., Sridharan, S., Shobha, T. and Arjunan, R.V., 2025. Slice Residual U-Net Based Rice Plant Disease Classification Using Convolutional Attentional BiGRU. *J. Internet Serv. Inf. Secur.*, *15*(2), pp.387-408.
[12] Panchananam, L.S., Chandaliya, P.K., Akhtar, Z., Upla, K. and Ramachandra, R., 2025. WaveletFusion: enhancing plant leaf disease classification with multi-scale feature extraction and explainable AI. *Expert Systems with Applications*, *285*, p.127947.
[13] Sasikala, V., Sistla, V., Edara, D.C. and Kolli, V.K.K., 2026. Enhancing plant leaf disease classification through self-attention super-resolution GAN and dual attention model. *The Visual Computer*, *42*(1), p.71.
[14] Xu, H., 2026. Corn or maize leaf disease classification based on CNN and vision mamba model. *Scientific Reports*.
[15] Naveed, F., Masih, A., Mahmood, J., Ahmed, M., Ali, A., Saddiqa, A., Abdulnabi, M.S.H. and Agbozo, E., 2025. Sustainable AI for plant disease classification using ResNet18 in few-shot learning. *Array*, *26*, p.100395.
[16] Potharaju, S., Singh, A., Singh, D., Tambe, S.N., Kantipudi, P.M. and Kiranmai, B., 2026. An explainable vision transformer model with transfer learning for accurate bean leaf disease classification. *Scientific Reports*, *16*(1), p.10402.
[17] Rahman, M.M., Gony, M.N., Ullah, M.S., Shuvra, S.M.K., Haque, R., Ahmed, M.R., Masfequier, S.M., Swapno, R., Murugappan, M., Chowdhury, M.E. and Nair, G., 2026. MaizeFormerX: a lightweight vision transformer with cross-scale attention for explainable maize leaf disease diagnosis. *Scientific Reports*.
[18] Batool, A., Kim, Y.W. and Byun, Y.C., 2026. Vision transformer-based approach for plant leaf disease classification with multi-patch selection. *PeerJ Computer Science*, *12*, p.e3544.
[19] Bohra M, N., Patil, A.R., Thorve, O.H., Shahade, A.K., Deshmukh, P.V., Patil, S. and Kotecha, K., 2026. Rose leaf disease classification and severity estimation using an interpretable vision transformer-based multi-task framework. *BMC Plant Biology*.
[20] Baek, E.T., 2025. Attention score-based multi-vision transformer technique for plant disease classification. *Sensors*, *25*(1), p.270.
[21] Isinkaye, F.O., Olusanya, M.O. and Akinyelu, A.A., 2025. A multi-class hybrid variational autoencoder and vision transformer model for enhanced plant disease identification. *Intelligent Systems with Applications*, *26*, p.200490.
[22] Padshetty, S., 2025. A novel twin vision transformer framework for crop disease classification with deformable attention. *Biomedical Signal Processing and Control*, *105*, p.107551.

[23] Hasan, M.A., Haque, F., Sarker, H., Abdullah, R., Roy, T., Taaha, N., Arafat, Y., Patwary, A.K., Ahsan, M. and Haider, J., 2025. Mulberry leaf disease detection by CNN-ViT with XAI integration. *PLoS One*, *20*(6), p.e0325188.
[24] Mondal, A., Chatterjee, A. and Avazov, N., 2026. A hybrid CNN-transformer model with adaptive activation function for potato leaf disease classification. *Scientific Reports*, *16*(1), p.4282.
[25] Ali, B., Talib, R., Hanif, M.K. and Sarwar, M.U., 2026. A hybrid deep learning approach for plant disease classification: Integrating efficient net and vision transformers. *Journal of Animal and Plant Sciences*, *36*(1), pp.79-96.
[26] Hasanah, U., Liu, C.M. and Azmi, I., 2026. A Lightweight Single-Token CNN-Transformer Architecture for Robust Multi-Crop Plant Disease Classification. *IEEE Access*.
[27] Gupta, P.S. and Zade, A.V., 2026. An Intelligent Hybrid Deep Learning Model Integrating CNN, Transformer, and LSTM for Precision Cotton Disease Diagnosis. *International Journal of Computer Information Systems and Industrial Management Applications*, *18*(8s), pp.1001-1012.
[28] Thakur, V., Bhatt, R., Durai, R.R. and Kumar, S., 2026. Attention-Enhanced Hybrid CNN–Transformer Approach Using ResNeXt-50 and Swin Transformer for Apple Leaf Disease Classification. *Applied Computational Intelligence and Soft Computing*, *2026*(1), p.9966874.
[29] Billah, M.M., Sagor, S., Hossain, S., Alahmadi, T.J., Moni, M.A. and Uddin, M.S., 2026. Explainable hybrid multi-branch CNN–ViT–GNN framework for robust hibiscus leaf disease classification. *Scientific Reports*.
[30] Viswanathan, V. and Murugasamy, K., 2026. Modified ensemble machine learning-based plant leaf disease detection model with optimized K-Means clustering. *Network: Computation in Neural Systems*, *37*(1), pp.161-205.
[31] Liu, X., Javidan, S.M., Ampatzidis, Y. and Zhang, Z., 2026. A Hybrid Model of Deep Feature Extraction and Weighted Ensemble Classifiers for Accurate Barley Disease Detection. *Journal of the ASABE*, *69*(1), pp.13-24.
[32] Sharma, A., Dalmia, R., Saxena, A. and Mohana, R., 2025. A stacked deep learning approach for multiclass classification of plant diseases. *Plant and Soil*, *506*(1), pp.621-638.
[33] Khalid, W., Alkhrijah, Y., Khalid, S., Usman, S.M. and Alshuhri, S.S., 2026. Swin-SHARP: a novel approach to wheat disease classification using boosted MAML and weighted ensembling with deep learning classifiers. *Scientific Reports*.
[34] Naveena, S. and Kavitha, K., 2026. EnsembleNet model: A hybrid approach for classifying Gossypium Herbaceum leaf diseases using image feature analysis. *Engineering Applications of Artificial Intelligence*, *164*, p.113225.
[35] Saritha, M. and Rasane, K., 2026. An ensemble of vision and swin transformers with LLM-based explanations for sugarcane leaf disease diagnosis. *Scientific Reports*.
[36] Sharma, P., Sharma, D.P. and Bansal, S., 2025. Optimum RBM encoded SVM model with ensemble feature Extractor-based plant disease prediction. *Chemometrics and Intelligent Laboratory Systems*, *258*, p.105319.
[37] Salman, Z., Muhammad, A. and Han, D., 2025. Plant disease classification in the wild using vision transformers and mixture of experts. *Frontiers in Plant Science*, *16*, p.1522985.
[38] Raya-González, L.E., Alcántar-Camarena, V.A., Cepeda-Negrete, J., Bustos-Gaytán, A., del Rosario Abraham-Juárez, M. and Saldaña-Robles, N., 2025. Application of mixture of experts models for the recognition of pests and diseases in maize. *Array*, p.100502.
[39] Yan, S., Lou, L., Lin, J., Cernava, T., Migheli, Q., Oufensou, S., Chen, X. and Zhang, X., 2026. PAG-MSBNet: Prototype-aware gated multi-scale branch network for open-set crop disease recognition. *Ecological Informatics*, p.103703.
[40] Liu, W., Wang, H., Wu, G. and Wu, Y., 2026. Enhancing crop disease recognition framework via vision-language model with cross-attention and gated fusion. *Scientific Reports*.
[41] Bao, Y., Qi, X., Wang, H., Liu, T. and Bai, Y., 2026. CE-Fusion Botanic: A Lightweight Leaf Disease Detection Model via Adaptive Local–Global Information Fusion. *Applied Sciences*, *16*(7), p.3177.
[42] Haqiq, N., Zaim, M., Sbihi, M., El Alaoui, M., El Amraoui, K., El Kazini, Y., Roukhe, H. and Masmoudi, L., 2026. Lightweight Hybrid Deep Learning for Strawberry Disease Recognition and Edge Deployment Using Dynamic Multi-Scale CNN–Transformer Fusion. *AgriEngineering*, *8*(2), p.75.
[43] Nabil, O. and Leghris, C., 2026. EcoTomHybridNet: Policy-Guided Adaptive CNN–Transformer Inference for Resource-Aware Edge-Based Tomato Leaf Disease Classification. *Future Internet*, *18*(5), p.271.
[44] Zhao, P., Duan, S., Wang, S., Wang, A., Meng, L., Wang, Z. and Dai, Y., 2025. The Bayesian mixture expert recognition model for tobacco leaf curing stages based on feature fusion. *Plant Methods*, *21*(1), p.86.
[45] Nguyen-Tat, T.B. and Nong, D.T., 2026. Lightweight deep learning framework and real-world dataset for durian leaf disease detection in smart agriculture. *Displays*, p.103451.
[46] Hung, T.N.P. and Minh Tuan, N., 2026. Transfer Learning with Particle Swarm Optimization for Durian Leaf Disease Image Classification. *Applied Fruit Science*, *68*(3), p.140.
[47] Van, Q.N., Quang, H.H., Truong, T.P., Minh, T.V., Ha, D.C. and Minh, T.P., 2026. An efficient and lightweight YOLOv8-based model with multi-scale texture representation for durian leaf disease classification. *CTU Journal of Innovation and Sustainable Development*, *18*(Special issue: Plant Pathology), pp.151-162.
[48] Chatthaicharoen, P., Duangburong, S. and Phruksaphanrat, B., 2025. Classification of Durian Foliar Diseases by Xception and Mask R-CNN Models. *Engineering, Technology & Applied Science Research*, *15*(5), pp.27653-27659.
[49] Nie, C.Y., Osman, M.A. and Sarvghadi, M.A., 2025. Durian Trunk Diseases Identification Using Deep Learning. *IEEE Access*.
[50] Söylemez, İ., Nalici, M.E. and Ünlü, R., 2025. High-accuracy identification of Durian leaf diseases: a convolutional neural network approach validated with K-fold cross-validation and Bayesian optimization. *Applied Fruit Science*, *67*(6), p.464.
[51] Daud, M.M., Abualqumssan, A., Nor Rashid, F.A., Md Saad, M.H., Diyana Wan Zaki, W.M. and Mohd Satar, N.S., 2023. Durian Disease Classification using Vision Transformer for Cutting-Edge Disease Control. *International Journal of Advanced Computer Science & Applications*, *14*(12), p.446.
[52] Nguyen, Tan (2025), "Image Dataset of Ten Durian Diseases Captured in Real-Field Conditions from a Family Orchard in Vinh Long, Vietnam", Mendeley Data, V1, doi: 10.17632/mhjwyb5p48.1.