ORIGINAL RESEARCH **Open access**

Convolutional Neural Network Approaches for Multi-Symptomatic Rice Blast Identification

Wei Zhang[1*], Chen Hui[2], Michael Tan[1], David Lee[3]

Abstract

An airborne pathogen causes rice blast, which can swiftly extend from tiny infection pockets and lead to disastrous harvest reductions. To achieve rapid, reliable detection of these pockets in paddy fields, deep learning was applied to classify dual-scenario imagery featuring diverse manifestations of rice leaf blast. This investigation assembled and publicly released a benchmark dataset covering chronic and acute variants of rice leaf blast across multiple developmental phases of the plant, as well as two other prevalent leaf ailments and healthy foliage. The study began by contrasting how assorted training regimes affect imbalanced datasets. Following that, six state-of-the-art convolutional neural network architectures were trained on the dataset via transfer learning, and the hyperparameters of the superior models were then fine-tuned to improve classification accuracy. The evidence demonstrated that image quantity, alongside quality, heavily governed model behavior, and augmenting images greatly eased the burden of skewed recognition across classes. Per the experimental data, InceptionV3 outstripped the other five architectures overall, with its top validation score hitting 99.78% after parameter refinements and its peak test score reaching 99.64%. The work illustrated that harnessing computer vision and deep learning to spot crop disease signs and locate infection pockets by analyzing the feedback rate of diseased images would constitute a potent route toward automated disease surveillance in the years ahead.

Keywords Deep learning, Transfer learning, Rice blast, Crop diseases monitoring, Data imbalance, Image augmentation

*Correspondence:
Wei Zhang
wei.zhang@unimelb.edu.au
[1] Department of AI-Enabled Smart Irrigation and Evapotranspiration Modeling, Faculty of Agriculture, University of Melbourne, Melbourne, Australia
[2] Department of Agricultural Data Science for Tropical Fruits, Faculty of Agriculture, National University of Singapore, Singapore, Singapore
[3] Department of Machine Learning for Crop Water Stress, Faculty of Agriculture, University of Sydney, Sydney, Australia

Introduction

The fungus *Magnaporthe oryzae* causes rice blast, disseminating from minute infection pockets across extensive areas via wind-borne spores, with annual yield penalties ranging from 10% to 30% [1]. Pinpointing leaf blast pockets without delay is vital for launching immediate countermeasures that block wider spread.

Nevertheless, manual field inspection to track these pockets is unworkable. Scanning and tallying infection pockets throughout large acreages demands excessive labor and time [2]. On top of that, in the far-flung areas where rice cultivation typically occurs, there are too few grassroots crop protection workers, and most growers have little formal guidance [3].

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0), which permits non-commercial use, distribution, reproduction, and adaptation in any medium or format, provided that appropriate credit is given to the original author(s) and source, a link to the license is provided, and any changes made are indicated. Adapted material must be distributed under the same license.

© 2023 The Author(s). | Published by Agro Intelligence Research Press

Remote sensing now serves as the primary means of observing crop afflictions. It yields fine-grained hyperspectral or multispectral captures at various scales using sensors mounted on satellites or drones [4]. Sickness-triggered biological shifts alter how crops respond to light [5]. Decoding those subtle deviations lets authorities trace disease spread. Still, procuring such remotely collected data remains both difficult and expensive, slowing its uptake in developing regions. Besides that, the data provide no straightforward snapshot of actual field conditions and thus call for expert analysis, creating a significant technical hurdle [6]. As a result, an acute call exists for rapid, workable, and easily shareable disease-monitoring tools.

Given advances in computer vision and deep learning, picture-based recognition might offer a new approach to disease surveillance [4]. In this model, unmanned aircraft or farmers themselves snap random leaf photos as they move through a field. Deep learning classifiers then decide whether those pictured leaves carry disease [7], and by coupling the geotags stored in each picture with the frequency at which sick leaves appear, one can map how disease incidence shifts from spot to spot. Rice blast fits neatly with that surveillance idea. Leaf blast primarily disrupts leaf-driven photosynthesis and respiration before flowering. Its yield toll pales next to panicle blast, yet subduing leaf blast matters enormously because it furnishes the spore supply that infects panicles later in the season [8]. Catching leaf blast pockets early and applying proper controls can substantially reduce overall losses.

In recent years, convolutional neural networks have been deployed ever more widely to flag disease across crop types [9]. The literature suggests that deep machine learning pipelines for rice disease identification have matured, leading to a gradual increase in accuracy [10]. Nonetheless, a gap remains in fielding actual diagnostic systems built on that technology to surveil rice blast pockets. The great majority of published work focuses on recognizing classic (chronic) leaf blast marks, yet acute marks are equally common and can evolve into chronic ones over time [11]. Weaving together both mark types promises a fuller picture for training blast-spotting models.

Earlier inquiries into rice disease recognition largely focused on fine-tuning or adapting model architectures to boost overall accuracy [12-14]. But prior work has shown that image numbers and image quality both leave deep imprints on model performance [15, 16], and various training recipes can meaningfully improve accuracy [17-19]. That said, very few investigations have dug into dataset design or the lopsided-data puzzle.

To address those gaps, six emblematic CNN architectures were trained on a home-built, class-skewed dataset for rice disease. The strongest contender among the six got singled out, and the impact of toggling three hyperparameters was then scrutinized for its bearing on recognizing leaf blast alongside three other categories. This paper's offerings were:

(1) A benchmark compilation of three common leaf afflictions and healthy leaves was compiled, featuring images of leaf blast that capture both chronic and acute marks across varied growth phases and stem from two distinct photographic approaches.

(2) The divergent influences of two data-handling techniques—picture augmentation and across-class balancing—were mapped on lopsided datasets.

(3) The best-fitting model among the six CNNs for the home-built collection was selected and, through effective training strategies and tuning three hyperparameters, achieved top-tier recognition rates.

Related Works

A substantial body of work on image-based rice disease identification has prioritized the development and refinement of deep learning frameworks. For example, Liu *et al.* [12] engineered a new CNN architecture that combined effective loss functions and optimization methods grounded in VGG and ResNet architectures to pinpoint rice pathologies. After tuning hyperparameters such as batch size and the learning rate, their bespoke model delivered a final recognition rate of 98.64%. As off-the-shelf generic architectures increasingly proved their mettle on large-scale benchmarks, a trend

© 2023 The Author(s). | Published by Agro Intelligence Research Press

emerged toward introducing subtle modifications to these networks for specialized domain classification. Jiang *et al.* [13] used the DenseNet backbone and augmented it to discriminate among five rice afflictions by incorporating channel attention, depth-wise separable convolutions, and the AdaBound optimizer. These additions strengthened feature representation, parameter efficiency, and convergence speed, culminating in a mean classification mark of 99.4%. Along similar lines, Liu *et al.* [14] formulated an upgraded PSOC-DRCNet derived from ResNet101 for rapid and reliable rice disease detection in field environments. This architecture incorporated Dual-Mode Attention (DMA), a Residual Adaptive Block (RAB), and the CerLoss function, used the CPSA optimization procedure to identify the optimal learning rate, and yielded an average classification figure of 93.88%.

Beyond polishing conventional CNN paradigms, an expanding line of inquiry has sought to repurpose proven architectures from outside the vision domain—chiefly from natural language processing (NLP)—for image analysis. Salamai *et al.* [20] proposed a lesion-aware Visual Transformer (ViT) to discriminate multiple rice pathologies by focusing on distinctive lesion characteristics. Their ViT-based solution fused a multi-scale contextual feature extraction module with weakly supervised lesion localization and registered a mean accuracy of 98.74% alongside an F1-score of 98.18% on publicly accessible rice disease repositories. Around the same time, Li *et al.* [21] proposed a computationally frugal deep learning architecture, termed Plant-based MobileViT (PMVT), in which an inverted residual structure replaced MobileViT's standard convolutional block with a 7 × 7 kernel to model long-range pixel dependencies in plant disease snapshots. An integrated convolutional block attention module (CBAM) was further incorporated into the ViT encoder, yielding a classification accuracy of 93.1% for rice leaf diseases. Another flagship NLP model, the Bidirectional Encoder Representations from Transformers (BERT), was used by Chakrabarty *et al.* [22] to diagnose rice disease photographs. They coupled it with a patch segmentation front-end and the Local Interpretable Model-agnostic Explanations (LIME) technique to improve interpretability, ultimately achieving 97.5% accuracy on rice disease test sets.

Several efforts have focused on devising training protocols or fine-tuning hyperparameters to achieve higher performance. Aggarwal *et al.* [17], for instance, introduced a lightweight federated learning methodology for categorizing rice leaf maladies and verified its effectiveness on both independent and identically distributed (IID) and non-IID data partitions. Their outcomes showed that the framework, underpinned by EfficientNetB3, attained 99% accuracy under IID settings. Gogoi *et al.* [18] outlined a triphasic CNN training scheme built upon EfficientNetB7 as the pretrained anchor. The regimen involved progressively increasing the size of rice disease images across successive phases and applying PReLU activations in the dense layer to combat overfitting, achieving an accuracy of 93.99%. Kaur *et al.* [19] used SqueezeNet, VGG16, and InceptionV3 as feature-extraction backends and adjusted hyperparameters, including the learning rate and batch size. Their comparison revealed that SqueezeNet, backed by a neural network classifier, garnered the strongest result of 93.3% for rice leaf disease sorting.

The strategy by which input data are prepared and presented likewise exerts a major influence on model outcomes. Latif *et al.* [23] preprocessed their dataset by removing background noise, resizing, and enhancing images, and then investigated the effects of image normalization and augmentation on recognition performance. By adopting a non-normalized yet augmented dataset under a transfer learning paradigm, VGG19 scored a peak average recognition accuracy of 96.08%. Taking an alternative route, Sudhesh *et al.* [24] deployed a Dynamic Mode Decomposition (DMD)-based attention-guided preprocessing pipeline to isolate and segment infected tissue regions before classification training, eschewing the practice of feeding full images directly into pretrained models. Their findings showed that, among the five pretrained candidates evaluated, the Xception architecture outperformed its peers, achieving 94.33% classification accuracy on DMD-processed inputs.

The studies surveyed above did not address the role of symptom heterogeneity or variation in photographic collection techniques in the task of spotting rice leaf blast. Furthermore, although these contributions posted state-of-the-art results on their respective datasets, whether their methods can replicate similarly dominant results on a purpose-built dataset for rice leaf disease surveillance remains an open question. Against this backdrop, the present study sought to assemble a benchmark corpus for monitoring rice leaf blast and to ascertain the ceiling performance of leading algorithms on it, thereby providing a comparative foundation and a springboard for subsequent investigations.

© 2023 The Author(s). | Published by Agro Intelligence Research Press

Materials and Methods

Collection of rice disease images

This study gathered photographs representing the two dominant symptom categories of rice leaf blast (RLB): the acute manifestation (RLB_A) and the chronic lesion type (RLB_C). To further fortify the models against confounding factors, images of two other prevalent rice leaf disorders—rice brown spots (RBS) and rice sheath blight (RSB)—and healthy rice foliage (HLT) were also collected. The photographic collection was conducted at a variety of locales, spanning Dandong City and Panjin City in Liaoning Province, Wuyuan City in Jiangxi Province, Wugang City in Hunan Province, Lu'an City in Anhui Province, as well as the Shangzhuang experimental station and greenhouse facilities of China Agricultural University in Beijing. The bulk of image capture occurred between July and October across 2018-2021, ensuring the dataset encompassed disease manifestations across a range of rice cultivars and growth phases. To sidestep the bleaching effect of harsh direct light, photo sessions were mostly confined to the windows from 7:00 a.m. to 11:00 a.m. and 3:00 p.m. to 6:00 p.m. The complete image acquisition pipeline is illustrated in **Figure 1**.

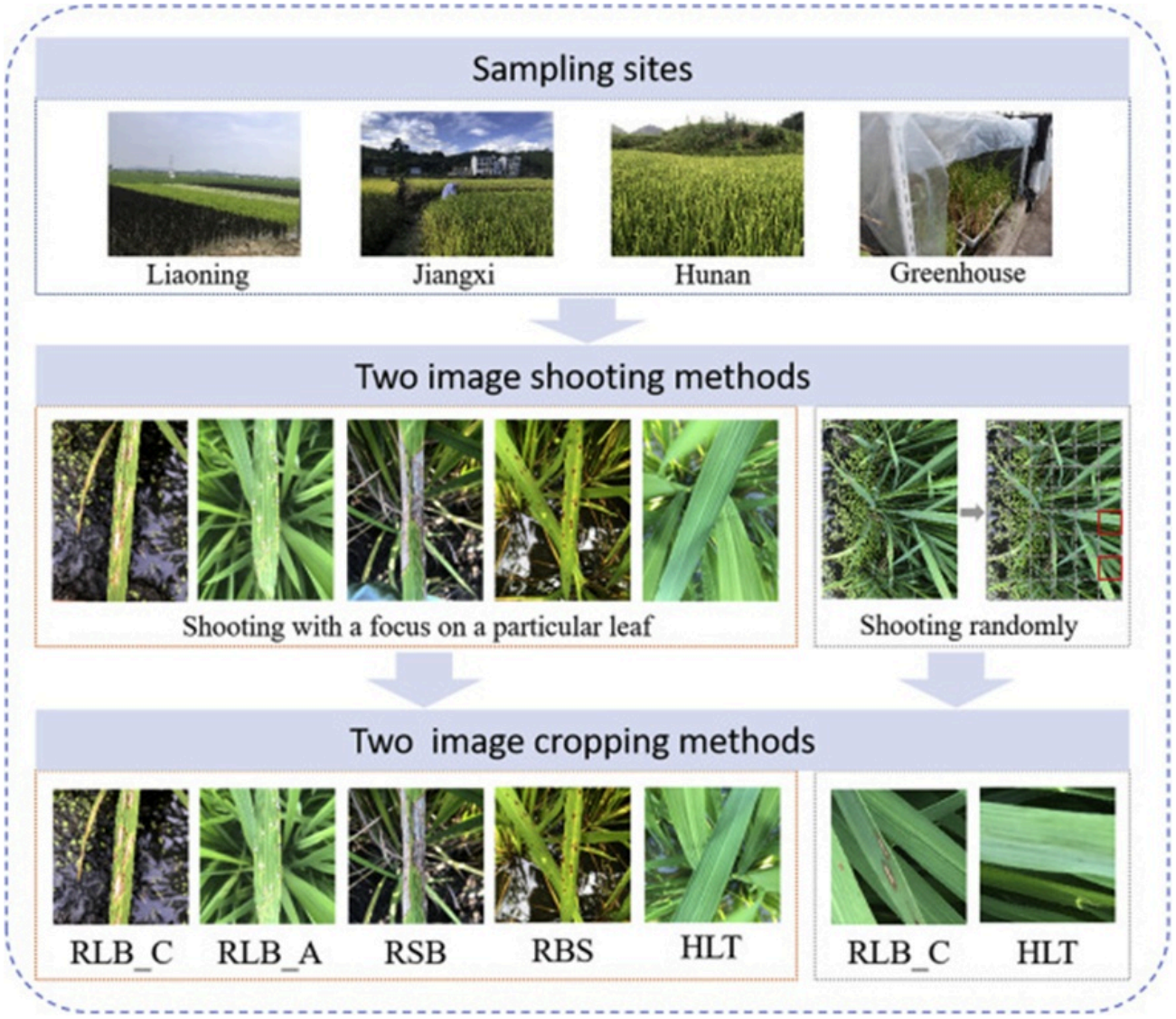


Figure 1. Overview of the dataset construction (RLB_C: Chronic type of rice leaf blast lesions; RLB_A: Acute type of rice leaf blast lesions; RBS: Rice brown spots; RSB: Rice sheath blight; HLT: Healthy rice leaves).

© 2023 The Author(s). | Published by Agro Intelligence Research Press

The collected pictures were subdivided by capture distance and camera angle **(Figure 1)**. The first subset comprised shots taken haphazardly across the field, approximating the image-acquisition style of UAVs or fixed monitoring cameras. This subset was dominated by rice leaf blast instances, in which the lesion regions accounted for only a minuscule fraction of the total pixel count. The second subset, in contrast, consisted of images deliberately framed to mimic a human scout's viewpoint during ground-level surveys; diseased leaves were arranged near the center of the frame and filled roughly one-fifth of the total image area. The randomly acquired field-level photographs were initially partitioned into compact square tiles dictated by the aspect ratio of the source image. Tiles that encompassed rice leaf blast lesions were assigned to the RLB class, whereas tiles containing exclusively unblemished foliage were designated as HLT. Close-proximity photographs were cropped to a square format, with the side length corresponding to the shorter edge of the original image, while the image's center point was held constant.

Dataset construction

Four visual classes made up the collection: RLB, RSB, RBS, and HLT **(Figure 1)**. Raw image tallies for these four groupings varied from a low of 1,174 to a high of 7,798 **(Table 1)**.

Table 1. This study uses an image dataset of three common rice leaf diseases and healthy rice leaves.

Empty cell	Total	HTL	RSB	RBS	RLB
Original dataset	17,811	7798	1413	1174	7426
Augmented dataset	30,578	7798	7762	7592	7426

Abbreviations: RLB = Rice leaf blast; RBS = Rice brown spots; RSB = Rice sheath blight; HLT = Healthy rice leaves.

Data balance: To probe whether uneven class representation negatively biased the overall metrics, synthetic samples were generated for the two groupings—RSB and RBS—that started with markedly fewer photographs **(Table 1)**. The set of transformations drawn upon included vertical and horizontal translation, mirroring across both axes, random scaling, and rotation confined to a narrow angular band **(Figure 2)**. Seeing that the pair of categories slated for augmentation were outnumbered considerably by the other two, setting aside an untouched holdout set with equal class sizes would have starved the augmentation and training pipelines of sufficient material. The remedy adopted was to pool the raw and augmented collections, then randomly split the combined pool into training, validation, and test sets in an 8:2:1 proportion.

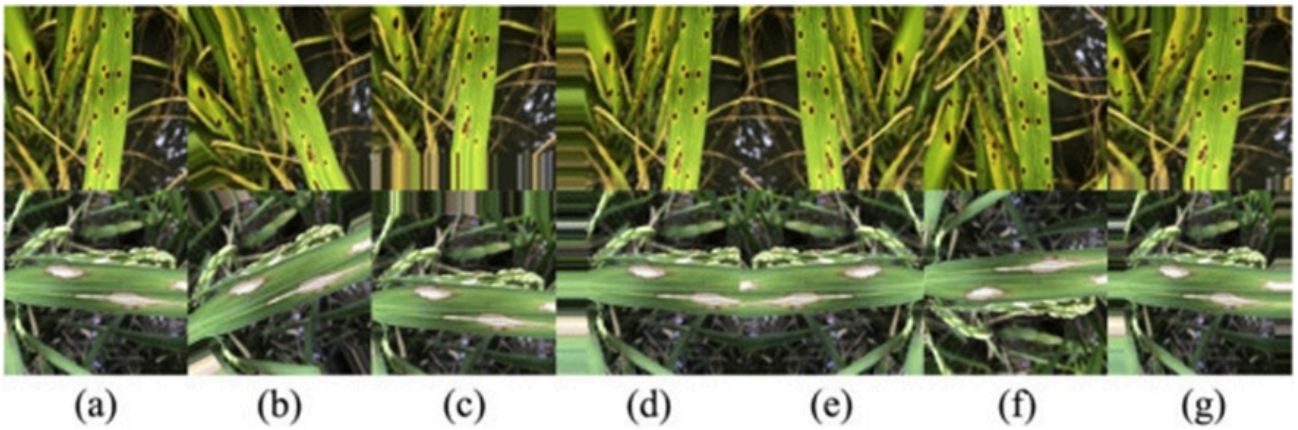


Figure 2. Multiple data augmentation methods applied to images of rice brown spot (up) and sheath blight (down): (a) Original images, (b) Rotation, (c) Height shift, (d) Width shift, (e) Horizontal flip, (f) Vertical flip, and (g) Zoom.

© 2023 The Author(s). | Published by Agro Intelligence Research Press

The complete dataset comprises two portions: the original and the augmented sets, both available on the Kaggle platform (https://www.kaggle.com/huiruz/datasets).

Convolutional neural network models

Because convolutional neural networks (CNNs) stand as the most pervasively deployed paradigm across computer vision tasks within the deep learning landscape [25], half a dozen canonical CNN architectures were singled out for evaluation: VGG16 [26], ResNet50 [27], InceptionV3 [28], Xception [29], NASNet_Mobile [30], and MobileNetV2 [31].

VGG [26] ranks among the most recognizable architectures, composed through an unembellished stacking of convolutional stages, pooling operations, and dense (fully connected) layers. It overcame the weaknesses of AlexNet [32]—namely oversized convolutional filters, aggressive stride settings, and an absence of padding—by trading large receptive fields for a succession of compact 3 × 3 kernels. From the VGG family of models, VGG16 was adopted for this work as the archetype of a conventional deep neural network.

ResNet [27] offers a compelling remedy to the problem of vanishing gradients that haunts very deep architectures through its signature residual connections, permitting networks to scale up to 152 layers without degradation. ResNet50, the most prevalent instantiation in the literature, was taken here as the representative of the residual network philosophy.

InceptionV3 [28] marks the third evolutionary step of the GoogLeNet dynasty [33]. Its foundational idea hinges on expanding the network horizontally by running filters of different dimensions side by side within the same layer, thus enabling the fusion of features captured at varied scales through the orchestrated assembly of distinct modules. The refinements introduced in InceptionV2 [34] and InceptionV3 include splitting larger kernels into smaller components, incorporating batch normalization (BN) layers, and switching to the RMSProp update rule. These adjustments were motivated by the dual goals of reducing the parameter budget and improving accuracy. For the present study, InceptionV3 functioned as the exemplar of a width-oriented design strategy.

Xception [29] is fundamentally erected upon the Inception backbone, interweaving depth-wise separable convolutions with skip connections between constituent blocks. It can therefore be viewed as an amalgam that borrows from multiple foundational architectures.

NASNet [30] was not handcrafted by human designers but instead emerged from an AutoML search procedure. It ships in two sizes, one mobile-grade and one large-scale. For this investigation, NASNet_Mobile, which has a substantially lighter parameter footprint, was selected as the delegate for algorithmically discovered network topologies.

MobileNetV2 [31] was expressly engineered for deployment on mobile and embedded hardware, prioritizing drastic reductions in parameter storage and computational overhead without unduly sacrificing predictive power. MobileNetV2 was used here to represent the lightweight network category.

Hyperparameters

Optimizer

The back-propagation protocol in deep learning causes the loss function to shrink according to a prescribed update regime. The optimizer is the component that dictates the trajectory along which each model parameter is adjusted. Three optimizers that appear most frequently in CNN training workflows—stochastic gradient descent (SGD) [35], root mean square propagation (RMSProp) [36], and Adam [37]—were placed head-to-head in this study.

Learning rate

In deep learning practice, the learning rate emerges as a critical tuning knob that sets the magnitude of weight revisions applied after each loss gradient computation. Two default starting points routinely encountered when initializing model

© 2023 The Author(s). | Published by Agro Intelligence Research Press

training are 0.01 and 0.001. Hence, this study examined the effects of adopting these two learning rate values—0.01 and 0.001—on model quality.

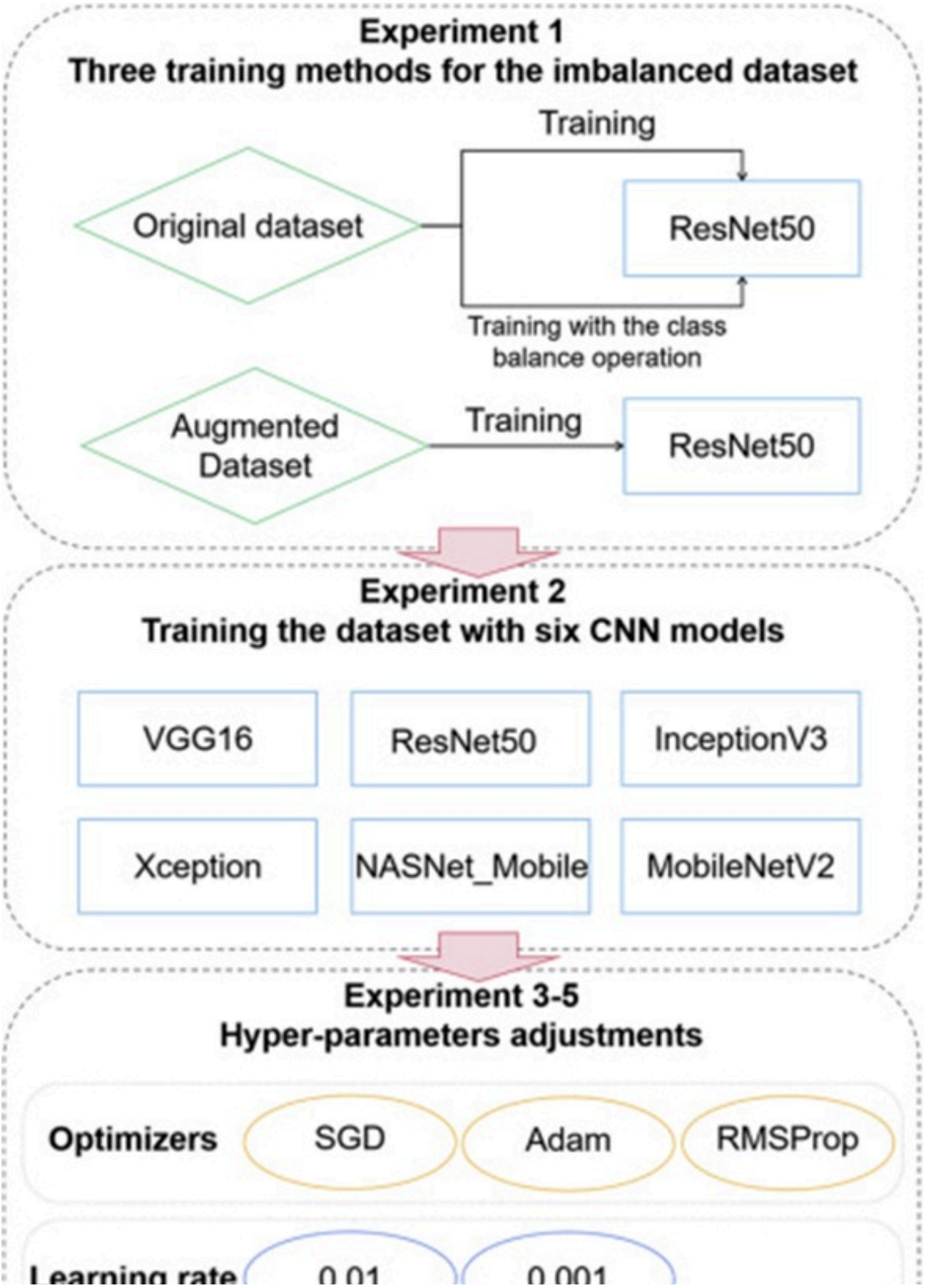
Number of neurons in the fully connected layer

For CNN pipelines, the width of the penultimate dense layer defines the size of the condensed feature vector that survives the upstream hierarchy of multi-dimensional extraction. Selecting a representation capacity that is either too generous or too stingy can leave a substantial imprint on downstream performance. The ideal neuron count is thus contingent on the dataset and the model's behavior. For the experiments reported here, the chosen architecture was trialed with three population sizes—512, 1024, and 2048—allocated to the final fully connected tier before the output layer.

Training parameter settings

A suite of five experiments was laid out to develop a capable model for detecting infection foci of rice leaf blast **(Figure 3)**, and each experimental condition was run three times to ensure reliability. Before hyperparameter tuning, the starting configuration was kept uniform across all architectures under test **(Table 2)**. During the tuning phase, only the designated parameters were permitted to vary while all remaining settings were locked to their baseline values.

© 2023 The Author(s). | Published by Agro Intelligence Research Press



© 2023 The Author(s). | Published by Agro Intelligence Research Press



Momentum	0.9
Optimizer	SGD
Mini-batch size	20

Transfer learning

Transfer learning involves borrowing insights or representational knowledge gleaned from one problem domain and applying them to a separate but thematically adjacent task, thereby accelerating the learning trajectory [38]. In visual computing contexts, this typically entails taking weight configurations derived from massive publicly available picture banks and adapting them to a narrowly focused dataset. The current investigation adopted transfer learning to compress the training timeline, initializing model weights from results previously obtained on the ImageNet large-scale benchmark [39].

Class balance

A class-level weighting intervention was performed on the raw image collection during the first experimental run. Under this scheme, distinct multiplicative factors were assigned to each category, so that groups with limited representation were assigned proportionally larger importance scores, steering the optimization process to penalize errors in those sparse classes more heavily.

The early-stopping mechanism

Every training cycle was governed by an early-termination rule that tracked the validation loss rate as the metric for when to discontinue. If ten successive epochs elapsed with no downward movement in that rate, the learning step size was scaled to one-tenth of its current magnitude. When an additional 10 epochs passed without the step size decreasing further, the training loop was terminated, and the parameter snapshot capturing the lowest validation loss observed at any point during the run was stored for final testing.

Software and hardware configuration

The entire suite of experiments was executed on the Anaconda3 platform, relying on a Keras/TensorFlow backend (Keras 2.2.4, TensorFlow 1.15.0), with all scripts written in Python 3.7. The host machine was equipped with 32 GB of system memory and an NVIDIA GeForce GTX 1080 Ti graphics adapter, running the 64-bit version of Windows 10 Professional.

Heatmap

To illuminate which segments of the input—for example, the periphery of a lesion, its core, or surrounding tissue—carried the most weight in the network's ultimate classification call, class activation heatmaps were constructed for exemplar images that frequently fell victim to mislabeling. This was achieved via the GRAD-CAM technique [40]. In the resulting visualizations, pixels that strongly swayed the decision are colored on a gradient from yellow through to red, whereas

© 2023 The Author(s). | Published by Agro Intelligence Research Press

those exerting minimal influence shift from green toward purple. The approach provides an interpretability lens into the internal feature-extraction mechanisms of CNN-based architectures.

Confusion matrix

To measure how well the deep learning pipeline performed and to diagnose class pairs most susceptible to confusion, confusion matrices were compiled by cross-tabulating, class by class, the number of correctly identified instances against those incorrectly shunted into each alternative category. Within a matrix covering c distinct classes, the figure seated at row j ($j \leq c$) and column i ($i \leq c$) denotes the volume of samples whose ground-truth membership is j yet whose machine-assigned label is i. To strip away distortions caused by uneven class sizes, normalized variants of these matrices were employed, substituting raw counts with percentages.

Results and Discussion

The impact of data imbalance on model accuracy

The evidence from experiment #1, which compared different training regimens using ResNet50, confirmed that augmenting the data pool with woven copies improved both training and validation dynamics, resulting in higher accuracy and lower loss across both phases **(Table 3)**. Although imposing class-level weighting on the original dataset did raise training and validation accuracy to some degree, the uplift paled in comparison with what was achieved by simply training on a balanced dataset augmented with synthetic data.

Table 3. The results of model training using original data and augmented data.

Training methods	Val_loss	Val_acc	Train_loss	Train_acc
Aug_n	0.0346	0.9893	0.0067	0.9981
Aug_n_b	0.0360	0.9901	0.0046	0.9988
Aug_y	**0.0136**	**0.9960**	**0.0016**	**0.9996**

Aug_n: Training with the original imbalanced dataset without any additional operation; Aug_n_b: Training with the original imbalanced dataset with the class balance operation. Aug_y: Training with the augmented balanced dataset without any additional operation. All results, including training accuracy, training loss, validation accuracy, and validation loss, are the mean values across three repeated experiments.

Inspecting the normalized confusion matrices revealed a consistent pattern: when models were exposed solely to the original imbalanced collection—with or without class weighting turned on—the categories RSB and RBS, which suffered from lower sample counts, routinely lagged behind HLT and RLB in test accuracy **(Figure 4)**. Introducing class weights on the original data did raise test performance for RBS and RLB, yet it simultaneously eroded accuracy for the RSB category. On the other hand, once synthetically expanded samples were added to the mix, test accuracy for RSB and RBS improved dramatically, and all four categories cleared the 0.9900 threshold. For HLT and RLB, the two groups already backed by abundant exemplars, test accuracy remained consistently high across augmentations; the HLT class, in fact, achieved a flawless 100% in every data scenario evaluated.

© 2023 The Author(s). | Published by Agro Intelligence Research Press

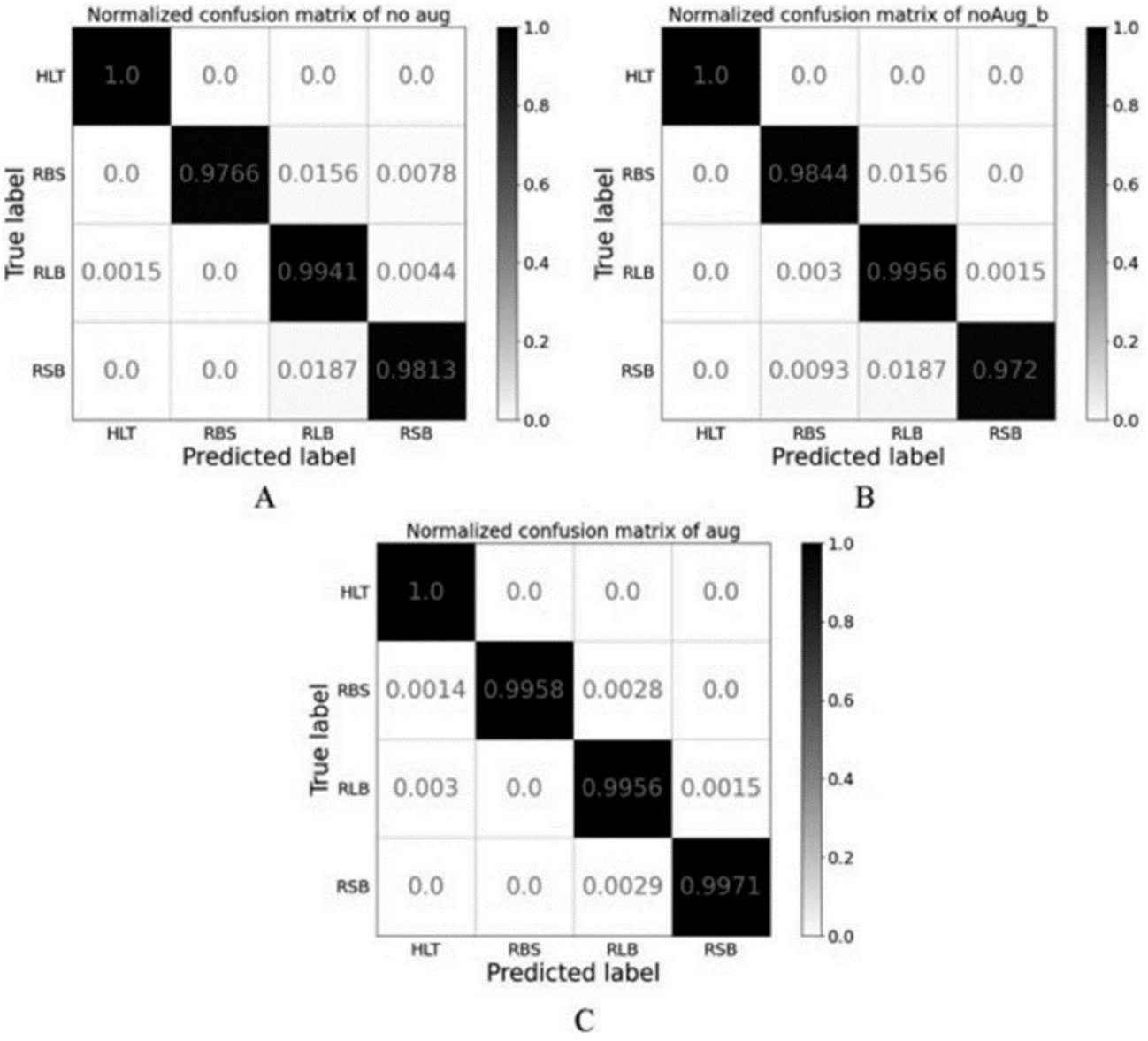
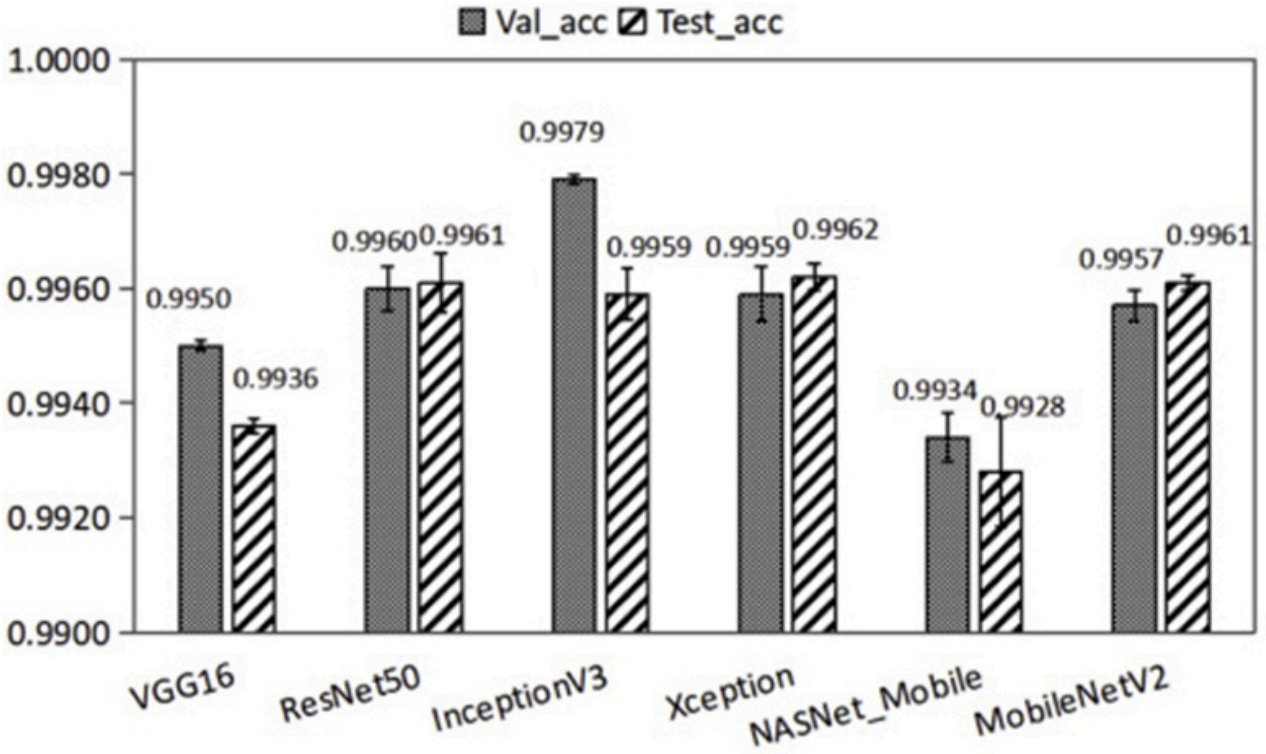


Figure 4. Normalized confusion matrix of test results using the original data and data after augmentation.

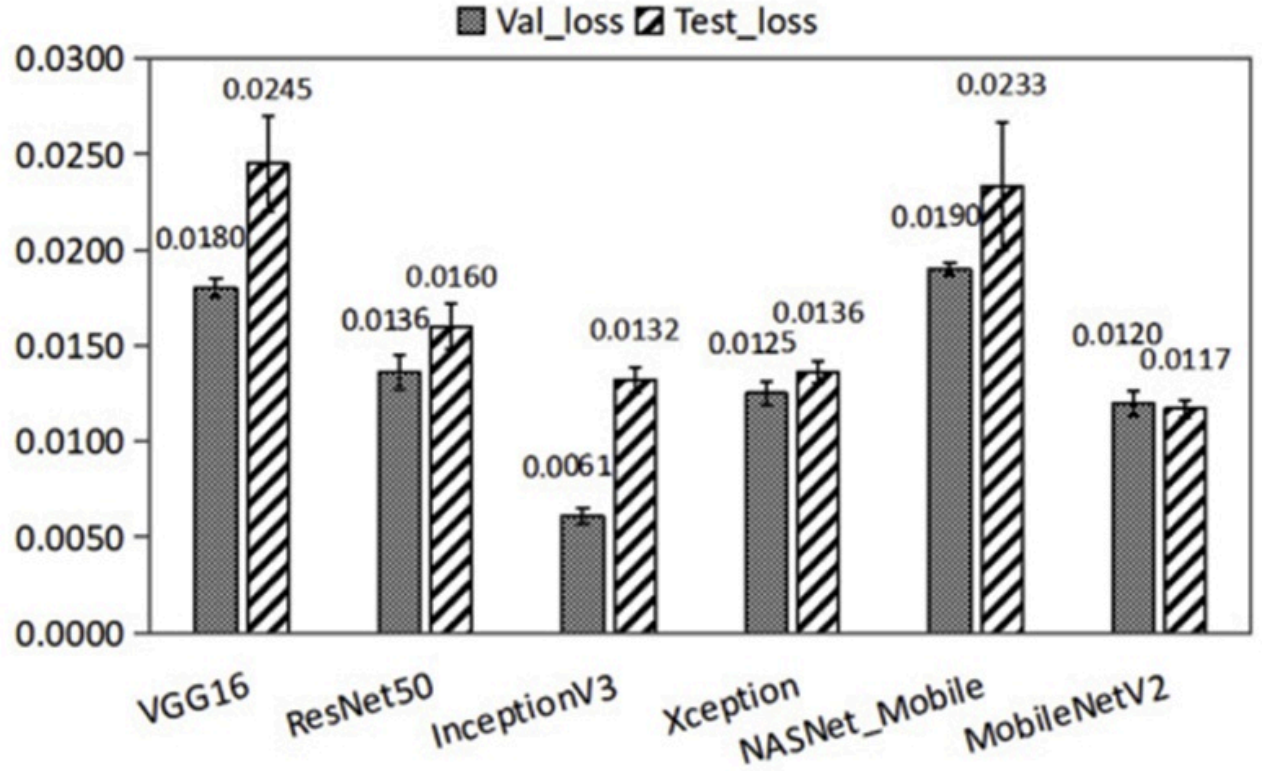
Performance of different CNN models

Experiment #2, which subjected multiple CNN designs to the augmented rice leaf disease repository, established that while every architecture under consideration surpassed 99% on both validation and test measures, their relative standing was far from uniform **(Figure 5)**. VGG16 and NASNet_Mobile constituted the lower tier, neither breaching the 0.9950 bar for test accuracy; additionally, their validation behavior appeared steady, but test-set output proved more erratic. InceptionV3 claimed the highest average validation accuracy among all six entrants, peaking at 0.9979 with tightly clustered replicates, and simultaneously posted the lowest mean validation loss. Despite a hint of variability during testing, its test accuracy was virtually indistinguishable from that of Xception—separated by a negligible 0.0003 gap **(Figure 5a)**—while its test loss undercut five rivals, bested only by MobileNetV2 **(Figure 5b)**. Consequently, InceptionV3 was designated as the candidate for the hyperparameter refinement phase.

© 2023 The Author(s). | Published by Agro Intelligence Research Press



(a)



(b)

© 2023 The Author(s). | Published by Agro Intelligence Research Press

Figure 5. The validation and test results of six CNN models using the rice disease image dataset.

The influence of hyperparameter adjustment on model accuracy

Experiment #3 demonstrated that the SGD optimizer yielded the most tightly grouped outcomes over the three replicate trials, with training and validation coordinates concentrated in a narrow zone hugging the lower-right quadrant, where precision approaches unity. Error approaches zero **(Figure 6)**. The Adam variant could not match SGD's consistency but outperformed RMSProp, the solver originally shipped with InceptionV3.

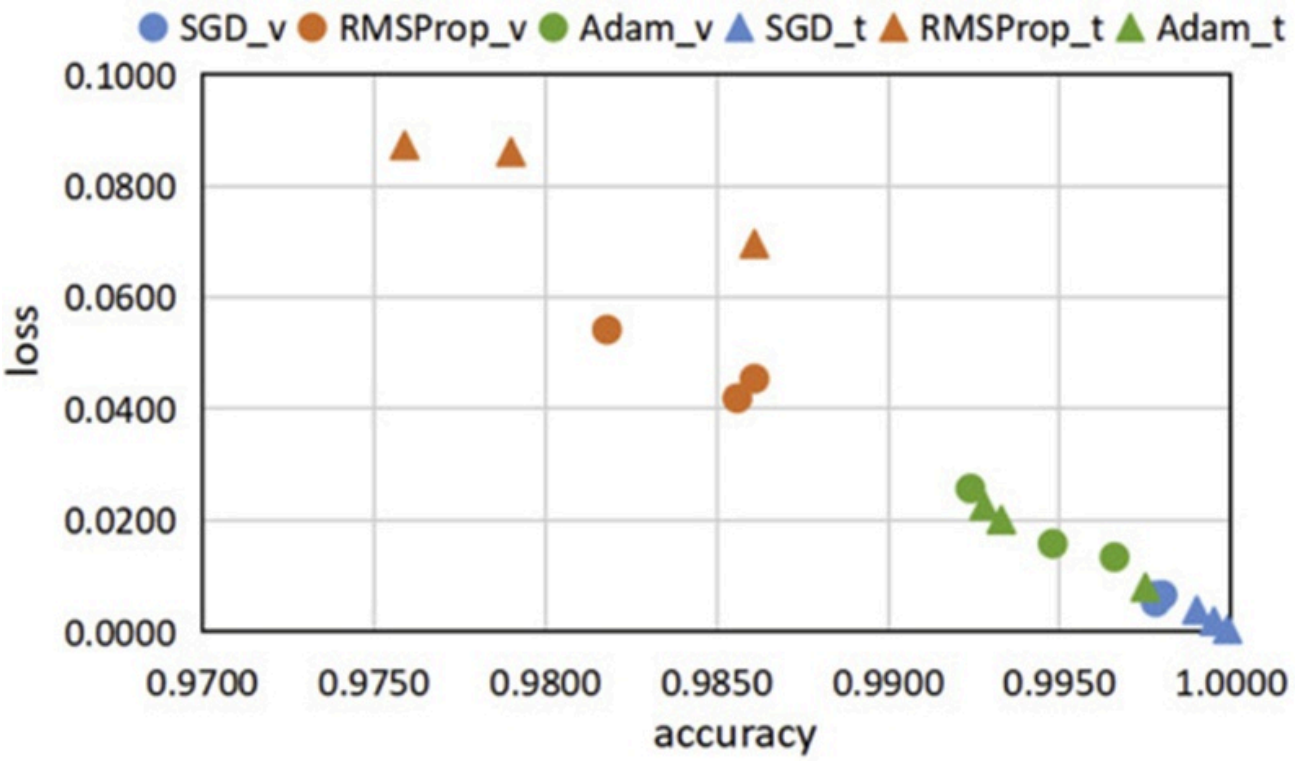


Figure 6. The training results of InceptionV3 models with different optimizers. (SGD_v, RMSProp_v, and Adam_v mean the results on the validation set by InceptionV3 models with three different optimizers; SGD_t, RMSProp _t, and Adam_t mean the results on the training set by InceptionV3 models with three different optimizers. The closer to the lower-right corner, the higher the parameter accuracy, the lower the parameter loss, the better the model performance.

The fourth experiment, which compared two different starting learning rates for InceptionV3, showed a clear advantage for the lower value. Setting the initial step size to 0.001 yielded superior validation and test accuracy relative to 0.01 **(Figure 7)**, and the accompanying loss values were substantially reduced under the 0.001 configuration **(Figure 7)**.

© 2023 The Author(s). | Published by Agro Intelligence Research Press

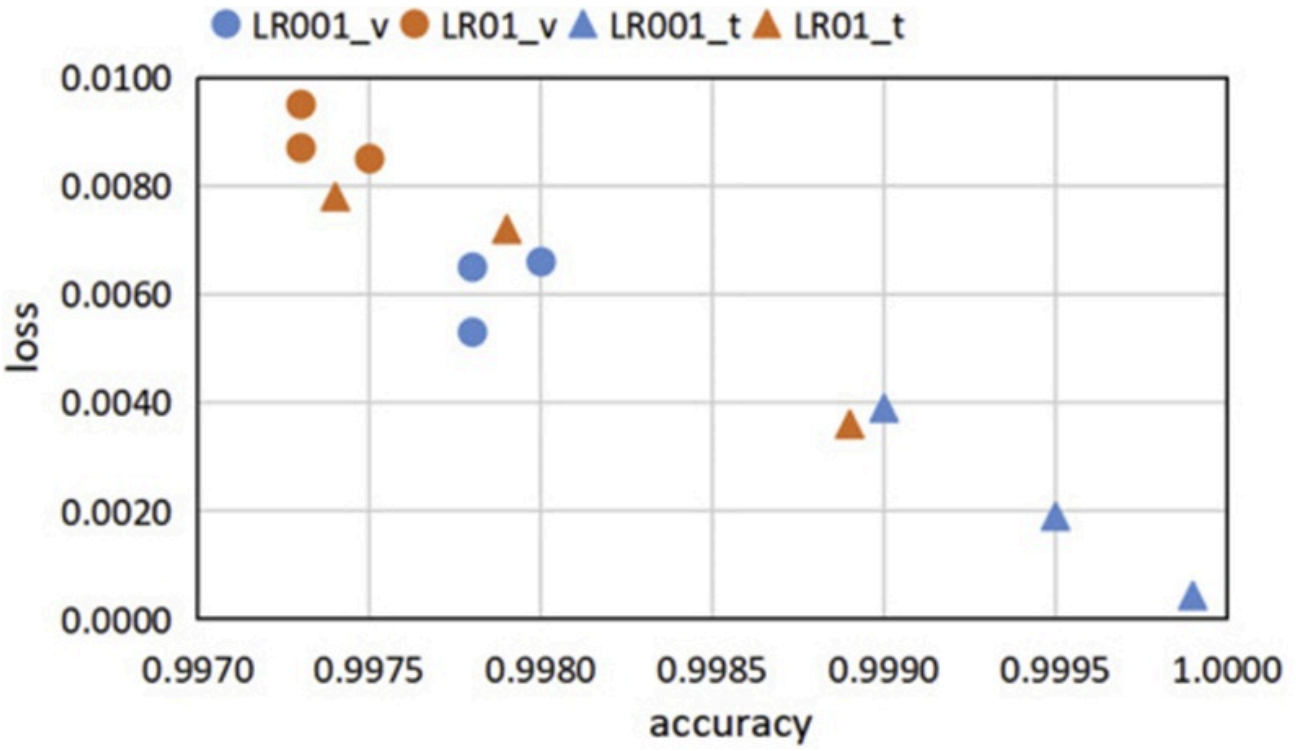


Figure 7. The training results of InceptionV3 models with different initial learning rates. (LR001_v and LR01_v mean the result of InceptionV3 models on the validation set when the initial learning rate is set to 0.001 and 0.01, respectively; LR001_t and LR01_t mean the result of InceptionV3 models on the training set when the initial learning rate is set to 0.001 and 0.01, respectively. The closer to the lower-right corner, the higher the parameter accuracy, the lower the parameter loss, the better the model performance.

The fifth and final experiment indicated that altering the neuron count in the penultimate dense layer did exert some influence on model behavior. However, the effect was not systematic **(Figure 8)**. No obvious segregation emerged between the validation and test distributions for the three neuron capacities—512, 1024, and 2048.

© 2023 The Author(s). | Published by Agro Intelligence Research Press

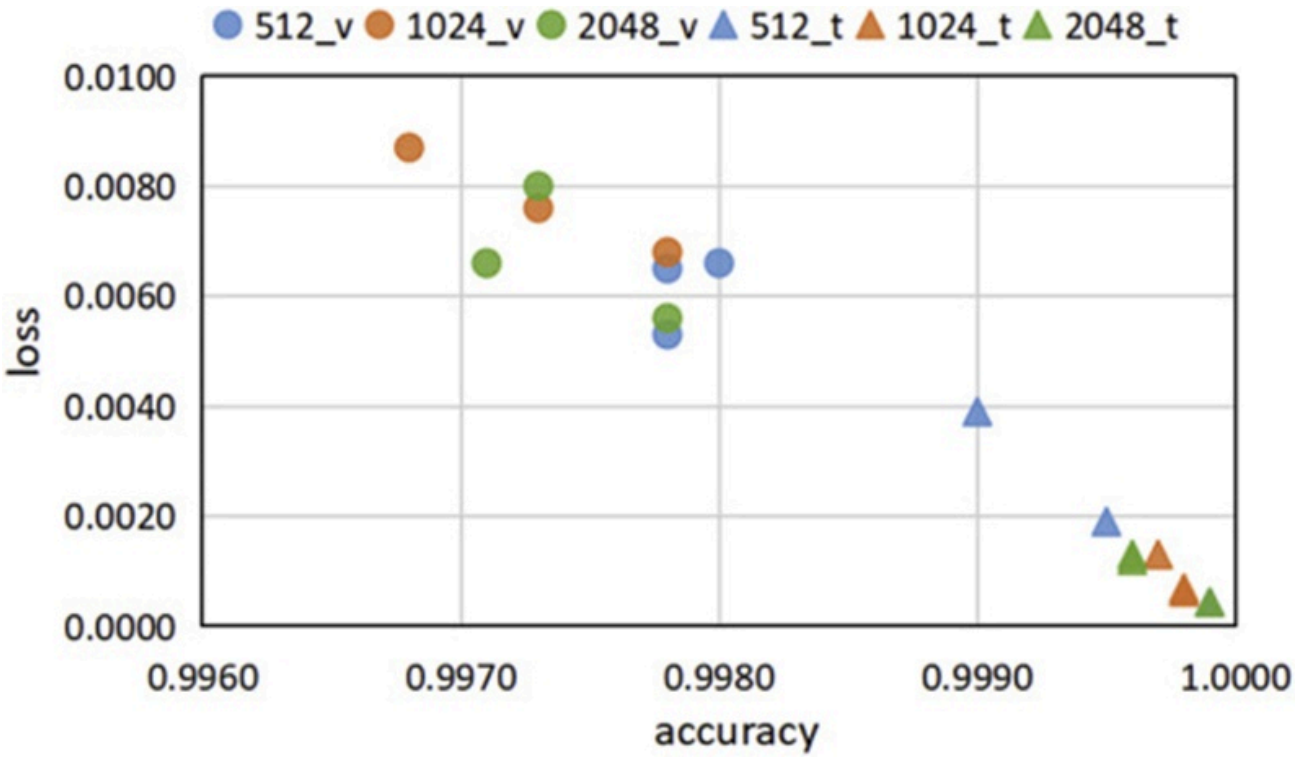


Figure 8. The training results of InceptionV3 models with different neuron numbers in the fully connected layer. (512_v, 1024_v, and 2048_v mean the result of InceptionV3 models on the validation set when the neuron number of the fully connected layer is set as 512,1024 and 2048, respectively; 512_t, 1024_t, and 2048_t mean the result of InceptionV3 models on the training set when the neuron number of the fully connected layer is set as 512,1024 and 2048, respectively. The closer to the lower-right corner, the higher the parameter accuracy, the lower the parameter loss, the better the model performance.

Results on the held-out test set indicated that setting the neuron capacity to 2048 in the penultimate dense layer yielded the highest test accuracy across the trio of configurations, achieving 0.9964 **(Table 4)**. Breaking down the parameter budgets showed roughly 2.60×10^7 learnable weights for the 2048-neuron setup, about 2.39×10^7 for the 1024-neuron variant, and nearly 2.28×10^7 for the 512-neuron alternative. While the total parameter footprint directly affects model storage demands and inference latency, the three budgets remained within a single order of magnitude for the neuron counts trialed in this investigation.

Table 4. The best results were obtained with the InceptionV3 models using different neuron numbers of fully connected layers.

FC_NUM	Test_loss	Test_acc	Val_loss	Val_acc	Train_loss	Train_acc
512	**0.0137**	0.9957	0.0066	**0.9980**	0.0019	0.9995
1024	0.0179	0.9946	0.0068	0.9978	0.0007	0.9998
2048	0.0152	**0.9964**	**0.0056**	0.9978	**0.0004**	**0.9999**

The best results from the three repeated experiments were selected based on validation accuracy.

Examination of the confusion matrices confirmed strong test performance across the board, irrespective of how many neurons populated the penultimate dense layer; out of 2,780 test specimens, misclassification was confined to roughly

© 2023 The Author(s). | Published by Agro Intelligence Research Press

ten instances across all variants **(Figure 9)**. Errors were disproportionately concentrated in the RLB group, whose members were confused with each of the three remaining categories on multiple occasions. Compared with its 512- and 1024-neuron siblings, the 2048-neuron model demonstrated a modest edge in correctly resolving RLB cases **(Figure 9)**. Accordingly, the test accuracy achieved by this configuration was adopted as the final result of the hyperparameter optimization process.

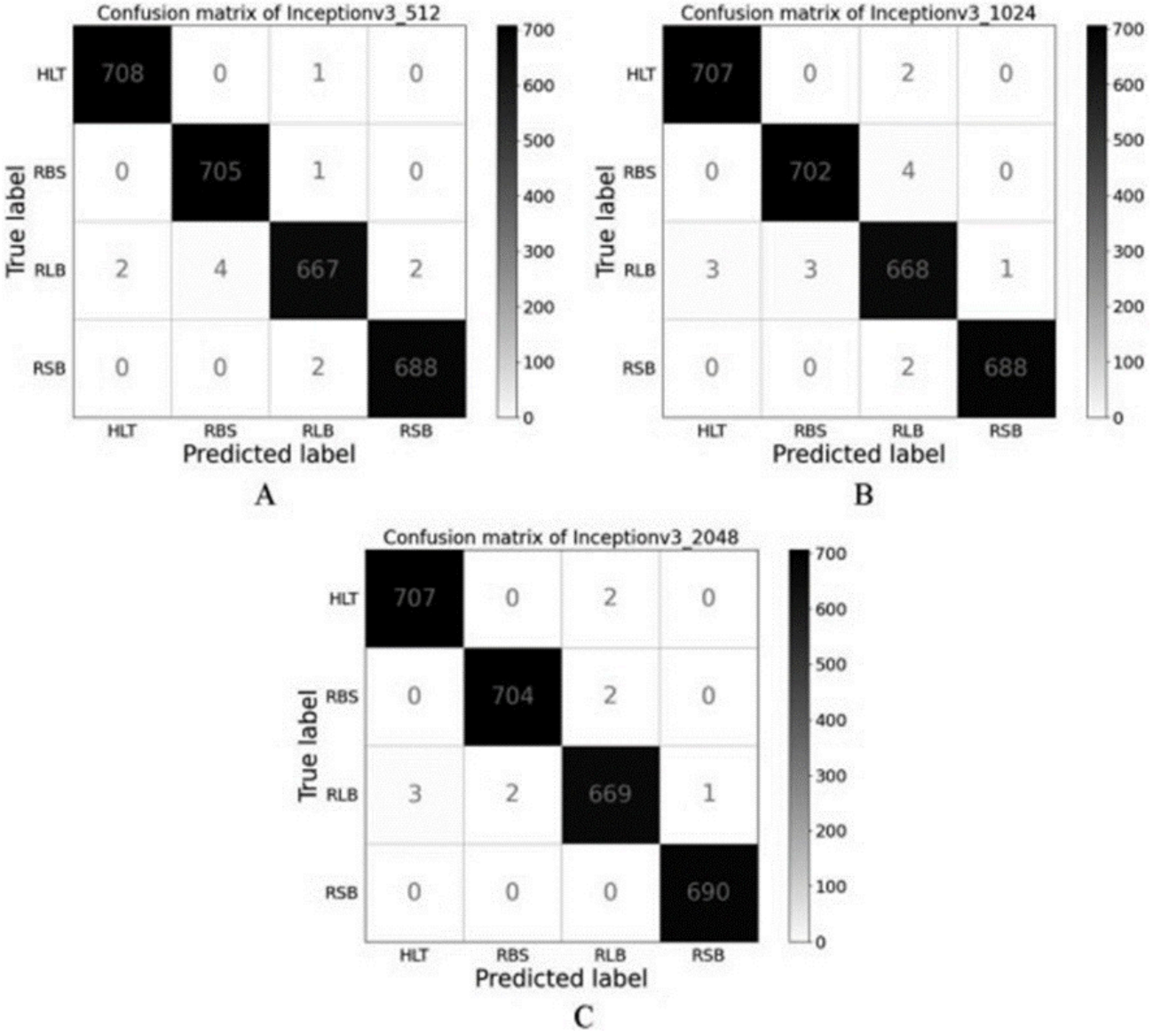


Figure 9. Confusion matrix of best test results using InceptionV3 models with different neuron numbers of fully connected layers.

Analysis of misclassified samples

A total of ten photographs received incorrect predictions from the InceptionV3_2048 model. To understand the nature of these failures, heatmap visualizations were generated for each mislabeled instance. As shown in **Figure 10**, the problematic examples fell into two broad groupings. The first cluster consisted of pictures where the spatial localization of lesion-related features appeared reasonably on-target. Yet, the final class assignment was still erroneous (e.g., **Figures 10b-10d, 10F, 10g**). Inspection of the corresponding unprocessed images revealed that the affected plants were either at a very early disease stage, bearing lesions of minuscule dimensions, or had progressed to a late stage, where lesions

© 2023 The Author(s). | Published by Agro Intelligence Research Press

had fused into expansive necrotic zones—both scenarios posing heightened diagnostic difficulty. The second cluster comprised images in which the model's attention was directed toward incorrect image regions (e.g., **Figures 10e, 10h, 10i, 10j**). Review of the original captures revealed that these were from the UAV-simulated acquisition mode, characterized by low pixel quality and ill-defined lesion perimeters.

Furthermore, the stochastic cropping procedure pushed the diseased tissue away from the central image area, suggesting a predisposition of the network to prioritize features near the frame midpoint. A particularly noteworthy observation concerns **Figure 10a**: a small fragment of leaf displaying blast lesions was present behind the healthy foreground leaf. This detail escaped the notice of human curators during dataset preparation but was correctly flagged by the model, lending further weight to its robust disease-spotting competence.

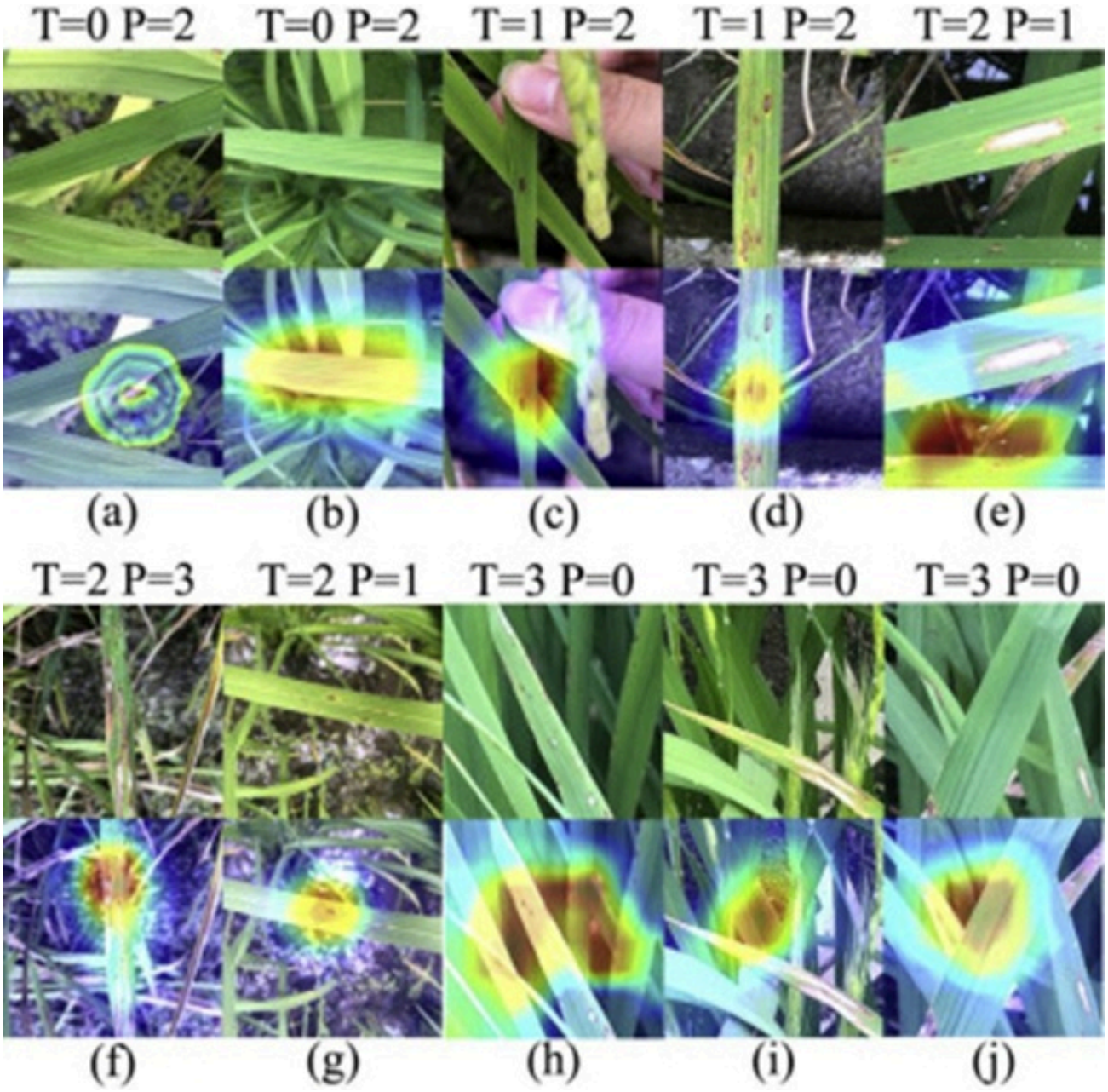


Figure 10. Heatmaps of samples misclassified by the InceptionV3_2048 model (T: True label, P: Predicted label, 0: Healthy rice leaves, 1: Rice brown spot, 2: Rice leaf blast, 3: Rice sheath blight). The red part has the highest contribution

© 2023 The Author(s). | Published by Agro Intelligence Research Press

to the final prediction results. On the contrary, the purple part makes the least contribution to the final image prediction.

Comparison with the latest related works

To lend additional support to the claim of superior performance by InceptionV3 on this rice pathology image bank, results from the most recent relevant publications were included alongside ours **(Table 5)**. The survey found that unequal class representation was prevalent across most earlier datasets, though few researchers proposed dedicated remedies to address it. Moreover, when augmentation was deployed, it tended to be applied uniformly across all classes rather than targeted to boost recognition in specific underperforming categories. Lastly, it bears emphasizing that the dataset compiled for this study originated from intricate real-field settings and included north of 7,000 examples belonging to the rice blast class alone; against this backdrop, our optimized model registered a remarkable validation accuracy of 99.78% **(Table 4)**, and the accuracy restricted to the multi-symptom rice blast subset climbed to an exceptional 99.86% **(Figure 9)**.

Table 5. Comparison of our results and those recently reported on rice disease identification using deep learning models.

Reference	Year	Data source	Lab/ Field	Classes	Total/RLB	Augmentation	Balance	Model	Accuracy (%)
Liu *et al.* [12]	2023	Open source	Field	3	1546/555	Yes	Yes	Self-built CNN	98.64
Jiang *et al.* [13]	2023	Open source	Field	5	1353/298	Yes	No	Improved DenseNet	99.4
Liu *et al.* [14]	2024	Open source	Field	5	2562/-	Yes	No	PSOC-DRCNet	93.88
Aggarwal *et al.* [17]	2024	Open source	Both	4	5932/1440	No	No	EfficientNetB3	99
Gogoi *et al.* [18]	2023	Self-built	Both	5	8883/2325	No	No	EfficientNetB7	93.99
Kaur *et al.* [19]	2024	Open source	Both	6	8572/1440	No	No	SqueezeNet	93.3
Salamai *et al.* [20]	2023	Open source	Field	10	10,407/1738	No	No	Lesion-aware ViT	98.74
Li *et al.* [21]	2023	Shared	Both	2	1007/-	No	No	PMVT	93.1
Chakrabarty *et al.* [22]	2024	Open source	Field	5	337/74	Yes	No	BEiT	97.5
Latif *et al.* [23]	2022	Open source	Lab	6	2167/363	Yes	Yes	VGG19	96.08
Sudhesh *et al.* [24]	2023	Open source	Field	5	3416/1570	No	No	XceptionNet	94.33
Ours	2025	Self-built	Field	4	17,811/7426	Yes	Yes	Optimized InceptionV3	99.78

The effort described herein resulted in the creation of a visual data repository covering multiple symptom variants of rice leaf blast, two additional widespread leaf disorders, and disease-free foliage. Raw image counts reached 17,811,

© 2023 The Author(s). | Published by Agro Intelligence Research Press

ballooning to 30,578 after synthetic copies were incorporated, translating to a class-wise average exceeding 7,000 exemplars. To contextualize this scale, one can look to ImageNet [39], arguably the best-known open-access visual recognition benchmark; its ILSVRC 2012 training partition spans 1,000 object categories and approximately 1.4 million pictures, yielding a per-class mean near 1,400. Narrowing the domain to plant pathology, the PlantVillage dataset [41] is widely acknowledged as the foremost repository of its kind, containing 87,848 images across 28 plant taxa and 58 health or disease states—roughly 1,515 images per label. Earlier investigations have documented that within-class sample abundance is intimately linked to how accurately that class can be learned [16]. The exceedingly high accuracy figures we report, derived from training in which more than 7,000 exemplars backed every class, therefore carry substantial weight and point toward the genuine field applicability of the resulting model.

Guarding against overfitting rests as much on data variety and quality as on sheer volume. Amassing a first-rate dataset requires models to encounter richly heterogeneous inputs. Photographing diseased rice in situ introduces an array of confounding background elements that can derail automated classification pipelines. With the practical aim of monitoring rice blast in actual paddies, this study used simulated surveillance hardware that captured frames at a moderate distance. Under such conditions, symptomatic tissue covered markedly fewer pixels and constituted a much smaller fraction of the overall scene than in close-range imagery. Boulent *et al.* [42] proposed a tripartite taxonomy of crop disease images: those shot in controlled environments, those centered on a particular organ in uncontrolled environments, and those neither centered on a specific organ nor shot in controlled environments. The present dataset spans the latter two divisions, both of which entail elevated classification difficulty. Models fitted on such data can thus serve a dual purpose, functioning both as on-the-ground disease diagnostics tools and as components of a broader blast surveillance framework.

A further insight we gained was that naively training on a class-skewed dataset without applying any mitigating measures led to depressed accuracy in categories with thinner representation. In contrast, augmenting those lean classes with additional examples bolstered the model's discriminatory power for them and, as a downstream effect, nudged overall accuracy higher. These trends aligned in part with the work of Buda *et al.* [43], who established that disproportionate class sizes can hamper both the convergence trajectory and the generalization capacity of learned models. A separate observation was that the class-weighting intervention appeared to exert independent, and sometimes opposing, effects across different categories; accuracy improvements in some classes occasionally coincided with accuracy erosion in others, hinting that this strategy, on its own, does not provide a reliable mechanism for improving holistic recognition performance.

Each of the six classic CNN blueprints tested in this study achieved comparatively strong disease classification rates, with InceptionV3 showing a slight performance advantage. The distinguishing characteristic of the InceptionV3 architecture, relative to the other candidates, is its simultaneous exploitation of network depth and width; the deployment of convolutional filters spanning multiple kernel sizes within the same layer enables cross-scale feature aggregation, a property that manifested in more consistent and favorable training dynamics. Furthermore, because the dataset assembled here contained more than 7,000 instances per class, training on such a plentiful sample set likely amplified its representational capacity. A methodical sweep over optimizer algorithms and initial learning rate magnitudes for InceptionV3 showed that deliberate selection among these hyperparameters can squeeze out further accuracy improvements within a confined margin.

The version of InceptionV3 trained on our custom-built image collection delivered a standout performance, achieving validation accuracy of 0.9978 and test accuracy of 0.9964 upon completion of hyperparameter tuning—metrics that hold their own against the highest values published to date [13]. On top of that, the dataset underpinning this research outstripped those used in the majority of prior investigations in terms of scale. Further, it incorporated multiple symptomatic expressions of a single pathology, exemplified by the inclusion of both acute and chronic lesion types of rice leaf blast. Equipping all six CNN models with this dataset was sufficient to achieve validation accuracies above 99%. This pattern suggests a potentially important takeaway: when a dataset is sufficiently large and carefully curated, drastic

© 2023 The Author(s). | Published by Agro Intelligence Research Press

architectural innovations may not be obligatory; instead, well-chosen training protocols together with thoughtful hyperparameter tuning may suffice to deliver recognition performance that meets the demands of real-world deployment.

Conclusion

The study reported here involved the collection of multi-symptom rice blast imagery spanning various developmental stages, acquired using two separate photographic approaches, culminating in a benchmark field dataset purpose-built for blast monitoring. Image augmentation was subsequently deployed as a countermeasure against imbalanced class distributions, outperforming traditional class-weighting techniques. Ultimately, precise rice disease categorization was achieved using an InceptionV3 model trained via transfer learning and hyperparameter refinement; the final model surpasses numerous previously reported systems and is poised for further evaluation under operational field conditions.

Acknowledgements

This work was supported by the National Science Foundation for Young Scientists of China (grant numbers 32402464 and 31471727). We want to thank Donggang Demonstration Farm and Panjin Saline Alkali Land Utilization and Research Institute for providing experimental fields for the collection of rice disease images.

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 07 Mar 2022 Revised: 12 May 2022 Accepted: 30 May 2022
Published online: 25 May 2023

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at https://creativecommons.org/licenses/by-nc-sa/4.0/.

References

1. Skamnioti P, Gurr SJ. Against the grain: safeguarding rice from rice blast disease. Trends Biotechnol. 2009;27(3):141-50.

2. Jiang Z, Dong Z, Jiang W, Yang Y. Recognition of rice leaf diseases and wheat leaf diseases based on multi-task deep transfer learning. Comput Electron Agric. 2021;186:106184.

3. Deng RL, Tao M, Xing H, Yang XL, Liu C, Liao KF, et al. Automatic diagnosis of rice diseases using deep learning. Front Plant Sci. 2021;12:701038.

4. Dang M, Wang H, Li Y, Nguyen T, Tightiz L, Xuan-Mung N, et al. Computer vision for plant disease recognition: a comprehensive review. Bot Rev. 2024;90:251-311.

5. Zheng Q, Huang W, Xia Q, Dong Y, Ye H, Jiang H, et al. Remote sensing monitoring of rice diseases and pests from different data sources: a review. Agronomy. 2023;13:e1851.

6. Guo W, Gong Z, Gao C, Yue J, Fu Y, Sun H, et al. An accurate monitoring method of peanut southern blight using unmanned aerial vehicle remote sensing. Precis Agric. 2024;25:1857-76.

7. Anam I, Arafat N, Hafiz MS, Jim JR, Kabir MM, Mridha MF. A systematic review of UAV and AI integration for targeted disease detection, weed management, and pest control in precision agriculture. Smart Agric Technol. 2024;9:e100647.

8. Bastiaans L. Effects of leaf blast on growth and production of a rice crop: 1. Determining the mechanism of yield reduction. Neth J Plant Pathol. 1993;99:323-34.

9. Lu J, Tan L, Jiang H. Review on convolutional neural network (CNN) applied to plant leaf disease classification. Agriculture. 2021;11:e707.

10. Sharma M, Kumar CJ, Bhattacharyya DK. Machine/deep learning techniques for disease and nutrient deficiency disorder diagnosis in rice crops: a systematic review. Biosyst Eng. 2024;244:77-92.

11. Kumar J, Chaube HS, Singh US, Mukhopadhyay AN. Plant Diseases of International Importance. Prentice Hall Inc. New Jersey. 1992.

12. Liu H, Cui Y, Wang J, Yu H. Analysis and research on rice disease identification method based on deep learning. Sustainability. 2023;15:e9321.

13. Jiang M, Feng C, Fang X, Huang Q, Zhang C, Shi X. Rice disease identification method based on attention mechanism and deep dense network. Electronics (Basel). 2023;12:e508.

14. Liu Z, Zhou G, Zhu W, Chai Y, Li L, Wang Y, et al. Identification of rice disease under complex background based on PSOC-DRCNet. Expert Syst Appl. 2024;249:e123643.

15. Yuan Y, Chen L, Ren Y, Wang S, Li Y. Impact of dataset on the study of crop disease image recognition. Int J Agric Biol Eng. 2022;15:181-6.

16. Zhou H, Huang D, Wu BM. Deep recognition of rice disease images: how many training samples do we really need? J Sci Food Agric. 2024;104:8070-8.

17. Aggarwal M, Khullar V, Goyal N, Prola TA. Resource-efficient federated learning over IoAT for rice leaf disease classification. Comput Electron Agric. 2024;221:e109001.

18. Gogoi M, Kumar V, Begum SA, Sharma N, Kant S. Classification and detection of rice diseases using a 3-stage CNN architecture with transfer learning approach. Agriculture. 2023;13:e1505.

19. Kaur A, Guleria K, Trivedi NK. A deep learning-based model for biotic rice leaf disease detection. Multimed Tools Appl. 2024;83:83583-609.

© 2023 The Author(s). | Published by Agro Intelligence Research Press

20. Salamai AA, Ajabnoor N, Khalid WE, Ali MM, Murayr AA. Lesion-aware visual transformer network for paddy diseases detection in precision agriculture. Eur J Agron. 2023;148:e126884.

21. Li G, Wang Y, Zhao Q, Yuan P, Chang B. PMVT: A lightweight vision transformer for plant disease identification on mobile devices. Front Plant Sci. 2023;14:e1256773.

22. Chakrabarty A, Ahmed ST, Islam MFU, Aziz SM, Maidin SS. An interpretable fusion model integrating lightweight CNN and transformer architectures for rice leaf disease identification. Ecol Inform. 2024;82:e102718.

23. Latif G, Abdelhamid SE, Mallouhy RE, Alghazo J, Kazimi ZA. Deep learning utilization in agriculture: detection of rice plant diseases using an improved CNN model. Plants. 2022;11:e2230.

24. Sudhesh KM, Sowmya V, Kurian PS, Sikha OK. AI based rice leaf disease identification enhanced by dynamic mode decomposition. Eng Appl Artif Intell. 2023;120:e105836.

25. Voulodimos A, Doulamis N, Doulamis A, Protopapadakis E. Deep learning for computer vision: a brief review. Comput Intell Neurosci. 2018:1-13.

26. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv. 2014;arXiv:1409.1556.

27. He KM, Zhang XY, Ren SQ, Sun J. Deep residual learning for image recognition. Proc IEEE Conf Comput Vis Pattern Recogn. 2016:770-8.

28. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. Proc IEEE Conf Comput Vis Pattern Recogn. 2016:2818-26.

29. Chollet F. Xception: deep learning with depthwise separable convolutions, Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu; 2017. p. 1800-7.

30. Zoph B, Vasudevan V, Shlens J, Le QV. Learning transferable architectures for scalable image recognition, Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City ; 2018. p. 8697-710.

31. Sandler M, Howard A, Zhu ML, Zhmoginov A, Chen L.MobileNetv2: inverted residuals and linear bottlenecks, Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City ; 2018. p. 4510-20.

32. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Commun ACM. 2017;60:84-90.

33. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. Proc IEEE Conf Comput Vis Pattern Recogn. 2015:1-9.

34. Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. https://doi.org/10.48550/arXiv.1502.03167.

35. Robbins H, Monro S. A stochastic approximation method. Ann Math Stat. 1951;22:400-7.

36. Tieleman T, Hinton G. Lecture 6.5-RMSProp: divide the gradient by a running average of its recent magnitude. Coursera Neural Netw Learn. 2012;4:26-31.

37. Kingma DP, Ba J. Adam: a method for stochastic optimization. https://doi.org/10.48550/arXiv.1412.6980.

38. Chen J, Chen J, Zhang D, Sun Y, Nanehkaran YA. Using deep transfer learning for image-based plant disease identification. Comput Electron Agric. 2020;173:e105393.

© 2023 The Author(s). | Published by Agro Intelligence Research Press

39. Deng J, Dong W, Socher R, Li LJ, Li K, Li FF. ImageNet: a large-scale hierarchical image database, Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach; 2009. p. 248-55.

40. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. GRAD-CAM: visual explanations from deep networks via gradient-based localization. Int J Comput Vis. 2020;128:336-59.

41. Hughes DP, Salathé M. An open access repository of images on plant health to enable the development of mobile disease diagnostics. 2015. 1511.08060.
https://doi.org/10.48550/arXiv.

42. Boulent J, Foucher S, Theau J, St-Charles P. Convolutional neural networks for the automatic identification of plant diseases. Front Plant Sci. 2019;10:e941.

43. Buda M, Maki A, Mazurowski MA. A systematic study of the class imbalance problem in convolutional neural networks. Neural Netw. 2018;106:249-59.

© 2023 The Author(s). | Published by Agro Intelligence Research Press