



OPEN ACCESS

EDITED BY

Nicola Strisciuglio,
University of Twente, Netherlands

REVIEWED BY

Shujuan Yi,
Heilongjiang Bayi Agricultural
University, China
Xiujun Zhang,
Chinese Academy of Sciences (CAS), China

*CORRESPONDENCE

B. Baranidharan
✉ baranidb@srmist.edu.in

RECEIVED 18 August 2023

ACCEPTED 05 February 2024

PUBLISHED 06 March 2024

CITATION

Tharani Pavithra P and Baranidharan B (2024)
OSPS-MicroNet: a distilled knowledge
micro-CNN network for detecting rice
diseases. *Front. Comput. Sci.* 6:1279810.
doi: 10.3389/fcomp.2024.1279810

COPYRIGHT

© 2024 Tharani Pavithra and Baranidharan.
This is an open-access article distributed
under the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other forums is
permitted, provided the original author(s) and
the copyright owner(s) are credited and that
the original publication in this journal is cited,
in accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

OSPS-MicroNet: a distilled knowledge micro-CNN network for detecting rice diseases

P. Tharani Pavithra and B. Baranidharan*

Department of Computing Technologies, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu, India

More than half of the world's population relies on rice as their primary food source. In India, it is a dominant cereal crop that plays a significant role in the national economy, contributing to almost 17% of the GDP and engaging 60% of the population. Still, the agricultural sector faces numerous challenges, including diseases that can cause significant losses. Convolutional neural networks (CNNs) have proven effective in identifying rice diseases based on visual characteristics. However, CNNs require millions of parameters, resulting in high computational complexity, so deploying these models on limited-resource devices can be difficult due to their computational complexity. In this research, a lightweight CNN model named *Oryza Sativa Pathosis Spotter* (OSPS)-MicroNet is proposed. OSPS-MicroNet is inspired by the teacher-student knowledge distillation mechanism. The experimental results demonstrate that OSPS-MicroNet achieves an accuracy of 92.02% with only 0.7% of the network size of the heavyweight model, RESNET152. This research aims to create a more streamlined and resource-efficient model to detect rice diseases while minimizing demands on computational resources.

KEYWORDS

convolutional neural network (CNN), rice leaf diseases, ResNet152, MobileNet, OSPS-MicroNet, computational complexity

1 Introduction

Rice is a cereal grain that is a staple food for more than half of the world's population, mainly in Asia and Africa. Rice is an essential agricultural commodity, and it is the third most produced crop in the world after sugarcane and maize. It is an essential crop for human nourishment, providing more than one-fifth of humans' calories globally (Tharani and Baranidharan, 2023). In India, it is one of the foremost and dominant cereal collects, and it is grown in various regions of the country. Rice is a staple crop that thrives in rain-fed regions with a temperature of 25°C and a precipitation of over 100 cm. Rice is cultivated more commonly on irrigated lands with lower precipitation (<https://www.statista.com/topics/4868/agricultural-sector-in-india>).

According to the latest estimates, India has achieved a record production of rice during the 2021–2022 crop year, which runs from July to June. The total rice production during this period is estimated to be 130.29 million tons, where the current production exceeds the average production of the past 5 years by 13.85 million tons, which were 116.44 million tons (https://agricoop.gov.in/Documents/annual_report_english_2022_23.pdf).

However, rice production faces many challenges, and one of the significant challenges is the spread of diseases in crops, causing several sufferers. During the year 2021, agricultural practitioners encountered significant challenges, including ~33% loss due to diseases such

as blast, bacterial sheath blight, false smut, tungro, and brown spots that emerged and were caused by heavy rainfall during the Northeast monsoon (Temniranrat et al., 2021).

Rice diseases constitute a significant challenge for rice farmers worldwide, as they can significantly reduce the yields and quality of rice crops. The five rice diseases that are taken into account in this paper are as follows:

1.1 Rice blast

The pathogen infects rice plants in the form of spores, resulting in the formation of lesions or spots on various plant parts, including the leaves, leaf collar, panicle, culm, and culm nodes. The pathogen remains capable of producing spores for over 20 days, posing a severe threat to susceptible rice crops (Liu and Zhang, 2021). This disease affected even “Manuvarna,” a new rice type announced by the Kerala Agricultural University, cultivated on around 250 acres of land in the Wayanad region (Tharani and Baranidharan, 2023).

1.2 Brown spot

It is a fungal ailment that primarily affects rice plants, targeting various plant parts such as leaves, glumes, seedlings, sheaths, stems, and grains of mature plants. The presence of dark coffee-colored spots is notable in the panicle, and severe infestations can lead to spot formation in the grains, resulting in reduced yield and compromised milling quality (Terensan et al., 2022).

1.3 Bacterial sheath blight

Generally, it will be first noticed during the heading stage of the rice plant. Over time, these lesions expanded both in length and width, exhibiting a wavy margin, and gradually turned a straw yellow color, eventually covering the entire leaf. As the disease advanced, the lesions spread across the entire leaf blade, giving it a straw-colored appearance (Singh et al., 2019).

1.4 Tungro

Rice tungro disease is a result of the synergistic action of two viruses, which are transmitted by leafhoppers. This disease manifests through symptoms like leaf discoloration, stunted growth, decreased tiller numbers, and partially filled grains. Tungro disease also affects certain wild rice relatives and other grassy weeds commonly found in rice paddies. Infections can take place at any growth stage of the rice plant, although they are most commonly observed during the vegetative phase. The tillering stage of the plant is particularly susceptible to the disease's impact (Singh et al., 2023).

1.5 False smut

The pathogen responsible for rice false smut, known as *Ustilaginoidea virens*, enters the rice spikelet through a small gap before heading. The main source of infection is the presence of chlamydospores in the soil. During the vegetative stage of rice growth, the fungus establishes itself by colonizing the tissue located at the growing points of the tillers. The impact of rice false smut is primarily qualitative, affecting the visual appearance of the crop. It is crucial to remove the brown “smut balls” to preserve the visual integrity of the harvested rice (https://en.wikipedia.org/wiki/Ustilaginoidea_virens).

The agricultural sector is actively exploring innovative strategies to enhance crop yields due to unpredictable climate changes, rapid population growth, and concerns about food security. Artificial intelligence in agriculture, often referred to as “Agriculture Intelligence,” is becoming an integral part of the industry's technological evolution. It finds applications in precision farming, disease detection, and crop phenotyping, leveraging tools such as machine learning, deep learning, image processing, artificial neural networks, convolutional neural networks, Wireless Sensor Network (WSN) technology, wireless communication, robotics, the Internet of Things (IoT), various genetic algorithms, fuzzy logic, and computer vision. The integration of these technologies enables a reduction in the extensive use of chemicals, leading to decreased expenses, improved soil fertility, and increased productivity (Pathan et al., 2020).

For example, Wang et al. devised a new method to improve the efficiency of the seeding motor control system in electric-driven seeding (EDS). This approach involves the utilization of a genetic particle swarm optimization (GAPSO)-optimized fuzzy PID control strategy. The complexity of determining fuzzy controller parameters was tackled by integrating two quantization factors for the fuzzy controller's input. Moreover, the introduction of three scaling factors for the output of the fuzzy controller further refines the system. This innovative strategy aims to enhance the performance of electric-driven seeding by addressing challenges associated with fuzzy controller parameterization (Wang et al., 2022).

Zhou et al. developed a specialized system for controlling the application of liquid fertilizer, utilizing a fuzzy PID algorithm to integrate precise variable fertilization and targeted deep fertilization technologies. The primary emphasis was on developing the fertilization equipment and implementing an adaptive fuzzy PID control strategy tailored for targeted variable fertilization. Following this, a mathematical model was formulated for the control system of liquid fertilizer used in targeted variable fertilization, specifically designed to meet the needs of intertillage and fertilization in corn crops (Zhou et al., 2023).

This multi-faceted approach aligns with the industry's quest for sustainable and efficient agricultural practices. Convolutional neural network (CNN)-based computer vision shows promising results in identifying the above-mentioned rice diseases early and accurately, which can assist in managing and controlling these diseases. However, most of the existing CNNs have a huge

number of parameters, which makes them less desirable for edge devices or low-power platforms. This is primarily due to hardware limitations and low memory constraints on portable computing devices. To make CNNs viable and practical in such resource-constrained environments, this paper mainly focuses on techniques that aim to minimize the computational and memory requirements of CNN models, enabling them to run efficiently on edge devices and low-power platforms (Zhang et al., 2019).

In this research paper, we introduce OSPS-MicroNet, a lightweight convolutional neural network (CNN) model developed through the application of the teacher-student network paradigm. The model leverages the knowledge distillation process to enable efficient model compression, transferring insights from a larger, computationally intensive teacher model to a more streamlined and resource-efficient student model.

Our choice of teacher models comprises two pre-trained CNNs, namely, ResNet152 and MobileNet. These well-established models serve as the source of knowledge for our proposed OSPS-MicroNet. The utilization of these advanced architectures not only ensures a robust foundation but also facilitates the extraction of valuable insights for the creation of an effective student model.

The resultant student model, referred to as OSPS-MicroNet, is designed with a minimal number of parameters, ensuring a lightweight and computationally efficient architecture. Through the knowledge distillation process from ResNet152 and MobileNet, OSPS-MicroNet achieves notable accuracies of 92.02 and 90%, respectively. This demonstrates the effectiveness of our custom-built model in capturing and retaining the essential knowledge distilled from the more complex teacher models.

In conclusion, OSPS-MicroNet, with its innovative application of the teacher-student network paradigm and knowledge distillation, presents a promising approach to model compression. The successful transfer of knowledge from ResNet152 and MobileNet to OSPS-MicroNet showcases the potential for achieving high accuracy with a significantly smaller and more resource-efficient neural network.

The paper systematically presents its research on the detection of rice leaf diseases through computer vision. Beginning with a thorough literature review in Section 2, it establishes a contextual foundation for its contributions by examining prior research endeavors. Section 3 details the methodology, including dataset collection in Section 3.1 and in-depth exploration of the convolutional neural network (CNN) and the unique teacher-student network dynamics employed in crafting the OSPS-MicroNet model in Sections 3.2 and 3.3, respectively. The convolutional models chosen for knowledge distillation are elucidated in Section 3.4. Section 4 delves into the operational principles of the CNN models, providing a comprehensive analysis of the architecture and functionality, with a focus on the student model, OSPS-MicroNet. The paper concludes that a more streamlined architecture can outperform its larger counterparts in specific domains through effective knowledge transfer, substantiating this claim with experimental results and critical evaluation. Overall, the structured approach ensures a clear presentation of the research process and findings in the realm of rice leaf disease detection through computer vision.

2 Literature survey

Numerous studies have focused on using deep learning models to identify rice leaf diseases. Many of these studies employ fine-tuned, general-purpose models, and some propose adaptations of conventional CNN models to precisely classify diseased leaves.

Upadhyay and Kumar present an efficient method for identifying and recognizing rice plant diseases using lesion size and color in leaf images. The method presented in this study involves using Otsu's global thresholding technique to binarize images and eliminate background noise. The anticipated model employs fully connected layers of convolution to detect three types of rice diseases and has been trained with 4,000 samples of diseased rice leaves and 4,000 healthy leaf images (Upadhyay and Kumar, 2022).

Temniranrat et al. created an automatic and user-friendly LINE Bot System that can spot paddy crop diseases from real paddy field images to aid rice farmers in enhancing the yield and quality of their crops. The implementation of the LINE Bot system utilizes YOLOv3, which was identified as the most effective performance technique in previous research, to analyze the images and detect any diseases present. The results are then displayed in a clear and easy-to-understand format for the users (Temniranrat et al., 2021).

Patil and Kumar developed a rice-fusion framework that involves two stages. First, numerical features are extracted from data related to agriculture and meteorology that is gathered through sensors. Second, visual features are extracted from images of rice captured by cameras. By processing the fused features through the dense layer, the framework produces a single diagnosis for the rice disease (Patil and Kumar, 2021).

Yang et al. designed a mobile crop disease identification system to be adaptive to poor network environments. The system utilizes rice morphological characteristics for offline acknowledgment of rice false smut (RFS) using support vector machine (SVM) models (Yang et al., 2021).

Jiang et al. examined three distinct diseases in the paddy crop and two types of diseases that prevail in the wheat crop. A dataset of 40 images for each leaf disease has been gathered and improved to enhance the performance of the Visual Geometry Group Network-16 (VGG16) archetypal. The technique suggested to enhance the VGG16 model employs the principle of multi-task learning. Furthermore, transfer learning and alternate learning methods are utilized with the pre-existing ImageNet model to further boost the VGG16 model's precision (Jiang et al., 2021).

Wang et al. presented an innovative slant idea called ADSNN-BO for detecting and classifying rice diseases from images of rice leaves. The proposed ADSNN-BO model follows the MobileNet architecture, is enhanced with an attention mechanism, and is optimized using Bayesian optimization. Feature analysis techniques, such as activation mapping and filter visualization, are employed to ensure interpretability. The findings indicate that the attention-based mechanism employed in the ADSNN-BO model leads to more effective learning of informative features (Wang et al., 2021).

In the study by Chen et al., a frame-based imaging spectroscopy device for near-earth remote sensing was used to examine a complex planting environment. To distinguish between healthy and *U. virens*-infected rice, a mixed detection method was utilized.

The study involved the creation of 49 different arrangements and over 196 differentiation models, which included seven sowing data plots, two farm management types, and 23 pattern recognition methods. The accuracy of the models was verified, and in most cases, it is above 95%. The accuracy was exceptionally high when deep learning-based feature sequences were used, with most accuracies reaching 100% after 100 epochs and iterations (Chen et al., 2022).

Three spectroscopic methods were used in a study by Feng et al. to identify three distinct diseases that impact rice plants, namely, leaf blight, rice blast, and rice sheath blight. These methods include visible/near-infrared hyperspectral imaging (HSI) for spectral data, mid-infrared spectroscopy (MIR), and laser-induced breakdown spectroscopy (LIBS). Identification models were developed using each spectroscopic method (Feng et al., 2020).

Yang et al.'s study introduces a novel approach for the effective and precise identification of rice leaf diseases, utilizing a stacking-based integrated learning model. The model incorporates four distinct convolutional neural networks: an enhanced AlexNet, an improved GoogLeNet, ResNet50, and MobileNetV3. These networks serve as the foundational learners within the stacking framework, while a support vector machine (SVM) functions as the sub-learner. The proposed method leverages the combined strengths of the aforementioned convolutional neural networks and the SVM to enhance the overall disease classification performance (Yang et al., 2023).

Fenu and Mallocci explored the use of an ensemble learning paradigm to construct a robust network for predicting four different pear leaf diseases. To achieve this, they considered several well-known neural network architectures. By comparing these architectures, they identified their strengths and weaknesses in disease prediction. The ensemble learning approach involves combining these diverse neural network models to enhance predictive performance. They adopted the bagging strategy, which entails training multiple instances of each network architecture on bootstrapped subsets of the DiaMOS Plant dataset. This process helps to introduce diversity in the ensemble, reducing overfitting and enhancing generalization capability (Fenu and Mallocci, 2023).

Tuncer et al. developed a novel approach to plant leaf disease recognition and classification that has been developed by merging the Inception architecture with depth-wise separable convolutions. This technique significantly reduces the number of parameters and computational costs without compromising the accuracy of the model. The study involved training and testing the hybrid model using k -fold cross-validation to identify healthy and diseased leaves (Tuncer, 2021).

Bhatia et al. proposed a novel technique for feature extraction in plant disease detection. This technique helps in expanding a small data collection into a highly valued feature space. This expansion generates new features that can effectively capture underlying patterns and trends in the data, resulting in improved accuracy and speed of plant disease detection. By integrating the FMTD function, the proposed technique can achieve both efficient and accurate plant disease detection (Bhatia et al., 2022).

Sriwanna developed a system to predict rice blast disease by analyzing weather data. The study aimed to determine the most effective weather features out of fifteen by utilizing

an ensemble method to rank them. The researchers also evaluated the effectiveness of the anticipated feature ranking technique using five diverse classification representations, namely, multilayer perceptron, SVM, Naive Bayes, decision tree, and K -nearest neighbors. They are evaluated with a standard classification valuation system of measurement to assess the models' performance (Sriwanna, 2022).

Lu et al. introduced a novel approach to tackle the challenges associated with differentiating subtle variations for various rice crop ailments and achieving a higher appreciation rate in the presence of noise interference. The method focuses on enhancing the deep residual shrinkage network. To achieve this, the original network is augmented with an InceptionA module, which serves the purpose of reducing network parameters, lowering arithmetic cost, and enhancing the model's non-linearity. Additionally, the original residual structure's convolutional kernels are substituted with multiple smaller-sized convolutional kernels. These modifications aim to optimize the model's performance in terms of accuracy and efficiency while addressing the specific challenges posed by rice disease recognition (Lu et al., 2023).

Sridevi and Kiran Kumar present a novel rice disease prediction model comprising three main phases. First, pre-processing involves the application of median filtering (MF). In the next phase, various features are extracted, including discrete wavelet transform (DWT) and scale-invariant feature transform (SIFT). These extracted features are then fed into a classification system consisting of multi-layer perceptron (MLP) and long short-term memory (LSTM) networks. Finally, the model produces predicted outcomes for rice disease detection (Sridevi and Kiran Kumar, 2022).

Liu and Zhang introduced a novel approach, termed PiTLiD, leveraging the pre-trained Inception-V3 convolutional neural network and transfer learning to detect plant leaf diseases using phenotype data from plant leaves, particularly in scenarios with small sample sizes. To assess the resilience of their proposed method, they conducted experiments on various datasets characterized by a limited number of samples (Liu and Zhang, 2022).

In recent times, researchers have made significant strides in the development of robust deep-learning models designed to autonomously identify and classify rice diseases. However, the inherent challenge lies in the substantial size, complexity, and heightened computational demands of these advanced models. As a consequence, deploying such models in environments characterized by resource constraints becomes a formidable task.

Given this scenario, there arises a compelling necessity and imperative to innovate lightweight models that can circumvent these challenges without compromising on accuracy. The demand for such models is underscored by the need to extend the application of automated disease identification and classification to resource-constrained environments where the deployment of heavy-weight models is unfeasible.

The quest for lightweight models represents a pivotal avenue in the evolution of deep learning applications, emphasizing the importance of striking a delicate balance between model simplicity and predictive accuracy. By addressing the challenges associated with size and computational requirements, these innovative models have the potential to democratize the benefits of automated ailment

identification in rice crops, making it accessible and feasible across a broader spectrum of environments. In essence, the drive toward lightweight models exemplifies a commitment to advancing the practicality and applicability of deep learning solutions in real-world scenarios with limited computational resources.

3 Methodology

In resource-constrained computing environments, the practicality and feasibility of convolutional neural networks (CNNs) hinge on the implementation of efficient model design and optimization techniques. In this context, we employ the teacher-student neural network technique, a pivotal strategy that plays a crucial role in mitigating the computational and memory requirements of CNN models. By doing so, this technique enhances the overall performance of the model, rendering it adept at running efficiently on edge devices and low-power platforms.

The computational cost stands out as a fundamental metric, quantifying the resources that a neural network utilizes during its training or inference phases. These resources encompass processing power, memory allocation, and time consumption (Laudani et al., 2015). In this pursuit, OSPS-MicroNet is meticulously optimized with well-defined workflows, strategically designed to minimize the computational cost associated with both training and deploying the network.

This optimization not only underscores the efficiency of OSPS-MicroNet but also positions it as a pragmatic solution for deployment in scenarios where computational resources are limited. By addressing the critical aspects of computational cost, the implemented teacher-student neural network technique, coupled with the inherent design of OSPS-MicroNet, ensures a judicious

allocation of resources, making it well-suited for operation in resource-constrained environments.

3.1 Dataset collection

This paper focuses on the major five diseases as outlined in Figure 1 that affect rice plants: (i) blast, (ii) brown spot, (iii) tungro, (iv) false smut, and (v) bacterial sheath blight.

In South India, heavy rainfall and favorable conditions led to the damage of rice crops in 2021, which fell into five disease categories, namely, blast, brown spot, false smut, bacterial leaf blight, and tungro. In December 2021, real-time data of affected and healthy rice crop images were collected in the open fields of Melmaruvathur (latitude 12.435330 and longitude 79.832932), Kavaraipeitai (latitude 13.360368 and longitude 80.142750), and Gummidipoondi (latitude 13.409340 and longitude 80.131410) regions in Tamil Nadu. Around 1,500 images were collected in the open field. These images were captured using Xiaomi and Redmi phones with a resolution of 48 megapixels, but they were prone to noise and distortion due to the uncontrolled environment in the open field, as well as lighting effects and backgrounds.

To obtain more images for training a convolutional neural network, additional images were collected from various sources, including Kaggle (Rice Leafs Diseases Dataset), UCI Machine Learning Repository (Rice Leaf Diseases Dataset), and Computers and Electronics in Agriculture (Mendeley data, “Rice Leaf Diseases Image Samples” dataset by Sethi, Prabira Kumar). Around 5,500 images were collected from these sources. However, preparing image data for CNN training is a significant challenge, especially when the images in the training database are different sizes. Therefore, data augmentation altered the images to fit the network’s expected input size of 224×224 . Normalization was also

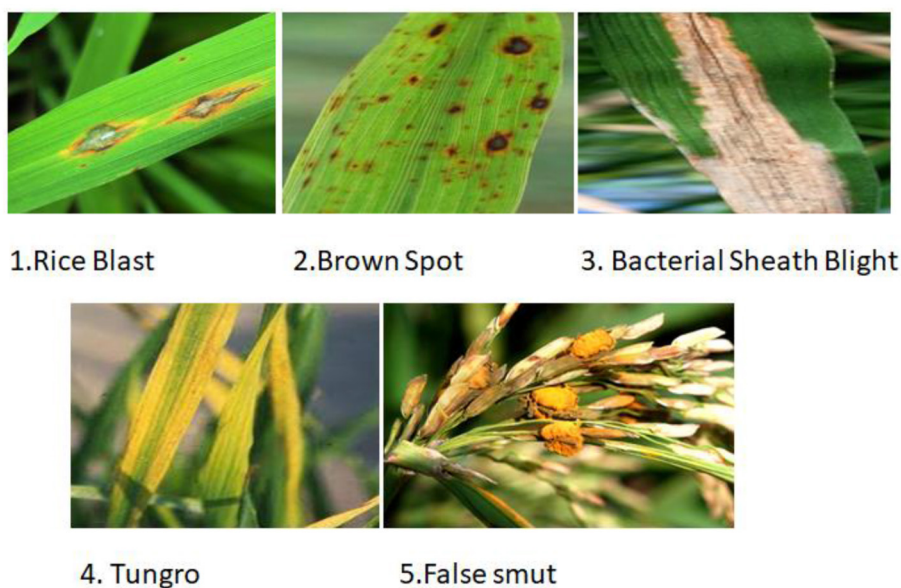


FIGURE 1
Diseases that impinge on rice plants.

performed to avoid gradient propagation issues, and erosion, dilation, opening, and closing operations were used to improve the darker and brighter regions of the input images. The training and test dataset for the MobileNet, ResNet152, and OSPS-MicroNet models consisted of 7,000 pre-processed images.

3.2 Convolutional neural network

Convolutional neural networks (CNNs) exhibit remarkable versatility, making them suitable for processing diverse data types, including images, videos, audio, speech, and natural language. Structurally, a CNN comprises multiple layers, commencing with a convolutional layer and progressing through pooling and ReLU activation, ultimately culminating in a fully connected layer.

The pivotal strength of a CNN lies in its convolutional layers, where it adeptly learns filters tailored for specific tasks, such as detection. This process has a cascading effect, where the output from one convolutional layer serves as the input for the next. After the convolutional layers, the pooling layer assumes a critical role by downsampling the data, leading to substantial reductions in computational demands, memory needs, and parameter counts.

Fully connected layers, aptly named, establish comprehensive connections with their preceding layers. Typically, these layers employ functions like “sigmoid” or “softmax” in the concluding layer to generate predictions regarding classes. At its core, convolutional layers discern features extracted from input data, subsequently condensed by the pooling layers. Leveraging these high-level features, fully connected layers usually perform the classification of input data into predefined categories in the final stages.

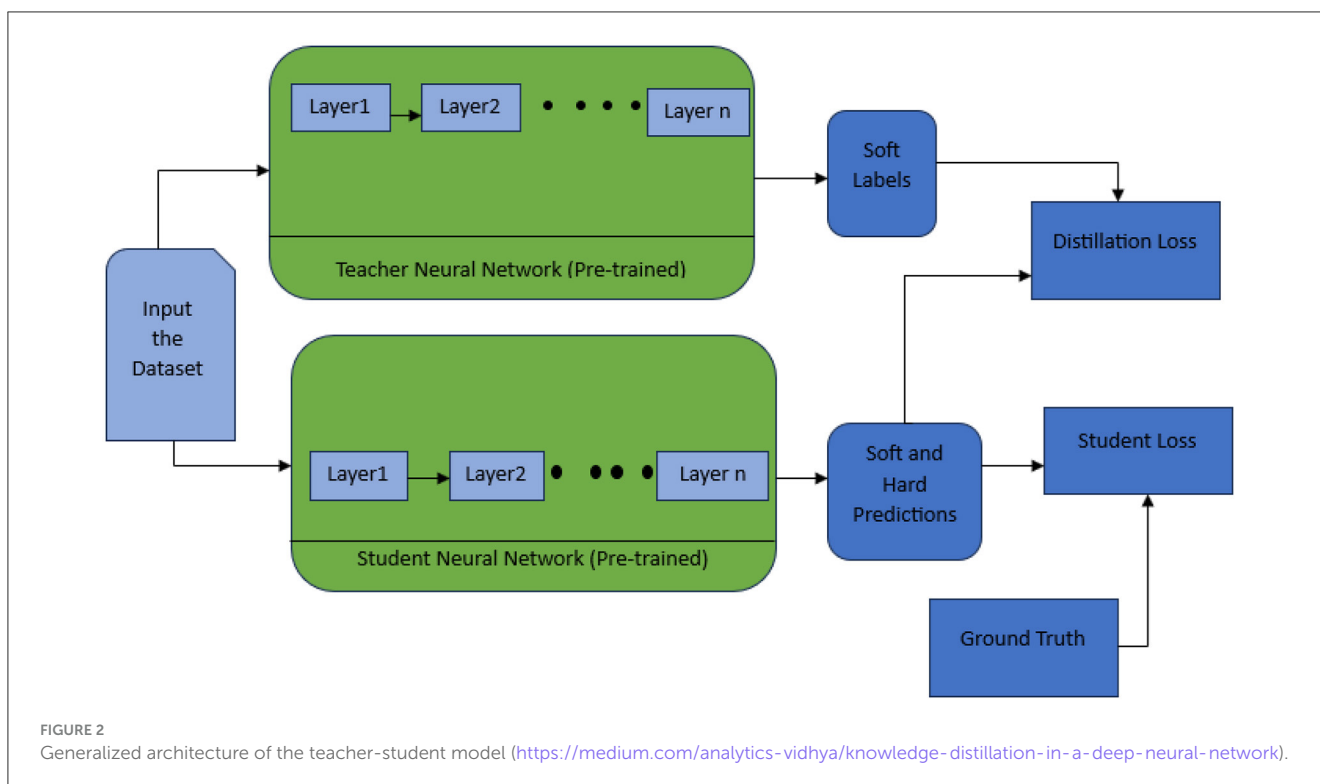
Moreover, the classification layer not only categorizes data but also extracts features essential for both classification and detection activities. This distinctive structural composition and sequential operation of layers empower CNNs to excel in a myriad of data processing tasks (Tugrul et al., 2022).

3.3 Teacher-student network dynamics

A teacher-student model represents a distinctive class of neural networks wherein a sophisticated model, termed the “teacher,” is employed to instruct a simpler, more modest model referred to as the “student” to emulate its behavior. Typically, the teacher model demands more extensive computational resources and boasts higher accuracy, contrasting with the lightweight nature of the student model, rendering it suitable for deployment in environments constrained by resources.

By training the student model to glean knowledge from the teacher model, it becomes adept at undertaking new tasks without necessitating the exhaustive retraining of the teacher model from scratch. This methodology not only accelerates and streamlines the model training process but also fosters efficient knowledge transfer, thereby enhancing the overall adaptability and versatility of the student model. Figure 2 depicts the generalized architecture.

In essence, the Teacher-Student paradigm stands as a strategic approach to harness the strengths of a powerful, computationally intensive model to cultivate a compact and resource-efficient counterpart. This not only facilitates the deployment of models in resource-limited settings but also contributes to the optimization of training procedures, ultimately leading to faster and more effective model training outcomes.



The model's performance is increased by reducing the parameters of a neural network, which in turn minimizes the computational and time complexity of the model. This method is very useful when computational resources or memory size is limited. The working methodology of the teacher-student network is as follows:

1. Training-teacher network: the teacher network is trained with large and complex datasets before using them as a basis for other models or tasks. The training process utilizes significant computational resources, such as high-performing GPUs, and may be performed offline or on dedicated servers. The teacher network learns the complex patterns and relationships in the data, which can be used to improve the accuracy and performance of other models. Due to the high computational demands of this step, it may not be feasible to perform it in real-time or on less powerful hardware.

In this paper, two pre-trained CNN models, namely, ResNet152 and MobileNet, are used as teacher networks.

2. Training-student network: the purpose of a loss function is to quantify the difference between the output of a teacher network and the corresponding output of a student network. The student network is trained through backpropagation, which involves calculating the gradient of the loss function concerning the weights and biases of the student network. This gradient is then used to update the network's parameters in a manner that minimizes the loss function.
3. Error backpropagation: the gradient computation involves a process called error backpropagation, which propagates the error from the output layer of the student network to its input layer. In this process, the gradient of the loss function concerning the outputs of the student network is first computed and then propagated backward through the layers of the network using the chain rule of calculus.

In the backpropagation process, knowledge distillation is employed to establish the correspondence between the teacher and student networks. In addition to replicating the outputs of the teacher network, this method involves training the student network to mimic the internal representations or activations of the teacher network. This helps the student network to better capture the underlying patterns and relationships in the data, leading to improved performance on new tasks. To accomplish this, we can utilize the teacher network's activations as supplementary inputs to the student network and train the student network to generate comparable activations through a distinct loss function (Mirzadeh et al., 2019; Wang et al., 2019; https://kears.io/examples/vision/knowledge_distillation).

4. Knowledge distillation: it is a technique applicable to any model, aiming to compress and transfer knowledge from a computationally intensive large deep neural network (referred to as the teacher) to a smaller neural network (known as the student). By doing so, the student network achieves improved inference efficiency while maintaining the essence of the teacher's expertise. In this technique, the soft output of the teacher model, which includes class probabilities and represents the model's uncertainty about the prediction, is employed to

teach a smaller and simpler model that could not learn the representations on its own. In a classification process, if a teacher model is employed, its final neural network layer would utilize the Softmax function.

The Softmax function takes a vector of real numbers, called logits, and applies the following formula 1 to obtain a probability distribution:

$$y_i(x | t) = \frac{e^{\frac{z_i(x)}{t}}}{\sum_j e^{\frac{z_j(x)}{t}}} \quad (1)$$

where t stands for temperature, which controls the evenness of the probability distribution generated by the Softmax function.

The cross-entropy loss function measures the dissimilarity between predicted and actual probabilities. For each possible class, the loss is the negative logarithm of the predicted probability multiplied by the actual probability. The negative sign tells us that the loss is minimized when the predicted probability is close to the actual probability.

The below formula 2 represents the loss function, namely, cross-entropy, for a classification problem with predicted probabilities.

$$E(X|t) = \sum_i \hat{y}_i(X|t) \log y_i(X|t) \quad (2)$$

where $y_i(x | t)$ is the output of the distilled model and $\sum_i \hat{y}_i(X|t)$ is the output produced by the large model on the same record using a high value of softmax temperature " t " for both models.

The distillation loss function is typically defined as the difference between the "soft" student predictions and the "soft" teacher labels, where "soft" refers to the fact that the predictions and labels are represented as probability distributions over the possible classes rather than as discrete values. By minimizing this difference, the student model learns to approximate the more complex behavior of the teacher model, often resulting in improved performance on the task at hand.

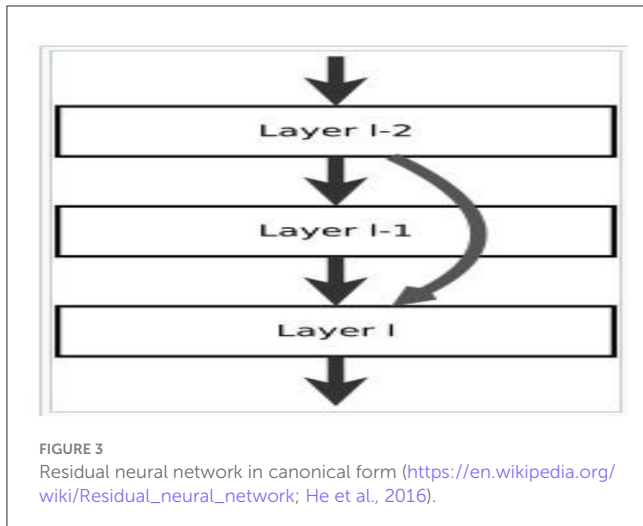
3.4 Convolutional models for knowledge distillation

3.4.1 Teacher network

3.4.1.1 ResNet152

A residual neural network, also known as ResNet, is an artificial neural net that utilizes skip connections or shortcuts to make the network much deeper than previous neural networks. Skip connections jump over some layers in the ResNet, allowing the network to learn residual functions instead of directly trying to learn the underlying mapping. These residual functions make it easier to optimize the network and avoid the vanishing gradient problem. In the below Figure 3 of the residual network, a layer $\ell - 1$ is bounced over initiation from $\ell - 2$.

Forward propagation is performed with weight matrices $W^{\ell-1,1}$ and $W^{\ell-2,1}$ with layers $\ell-2$ to ℓ as in Equation 3



$$a^\ell := g\left(W^{\ell-1,\ell} \cdot a^{\ell-1} + b^\ell + W^{\ell-2,\ell} \cdot a^{\ell-2}\right) \quad (3)$$

where g is the activation function and a^ℓ is the neuron activation for layer ℓ .

Backward propagation with skip paths is given by Equation 4

$$\Delta w^{\ell-2,\ell} := -\eta \frac{\partial E^\ell}{\partial w^{\ell-2,\ell}} = -\eta a^{\ell-2} \cdot \delta^\ell \quad (4)$$

where η is the learning rate and δ^ℓ are the error signals of neurons.

It gives amazing performance when used with transfer learning. However, it can be computationally expensive due to its many layers, which may make it impractical for some applications with limited computational resources, like mobile devices. So, it is selected as one of the teacher models.

The overall structure of ResNet152 used as a teacher model consists of 59,423,110 parameters.

3.4.2 Student network

3.4.2.1 OSPS-MicroNet

OSPS-MicroNet is a compact student network characterized by a modest parameter count of 460,444. This neural network adopts a straightforward architecture, comprising six convolutional 2-dimensional layers (Conv2d). These layers utilize the rectified linear unit (Relu) as the activation function, employ a stride size of 2, and culminate in Maxpooling operations after each layer.

The network architecture further involves the flattening of the processed data, resulting in an output shape of 64. Subsequently, four dense layers follow suit, each featuring output shapes of 600, 300, 150, and 6, respectively. This structured design is summarized in Table 1.

The streamlined and precisely defined structure of OSPS-MicroNet underscores its inherent simplicity, rendering it a highly efficient neural network. Specifically engineered with a deliberate emphasis on a reduced parameter

TABLE 1 Structure of OSPS-MicroNet model.

Layer (type)	Output shape	Size of parameters
Conv2d	(None, 56, 56, 16)	1,216
conv2d_1	(None, 14, 14, 32)	4,640
conv2d_2	(None, 4, 4, 48)	13,872
conv2d_3	(None, 1, 1, 64)	27,712
conv2d_4	(None, 1, 1, 128)	73,856
conv2d_5	(None, 1, 1, 64)	73,792
flatten	(None, 64)	0
Dense	(None, 600)	39,000
dense_1	(None, 300)	180,300
dense_2	(None, 150)	45,150
dense_3	(None, 6)	906
Trainable parameters	460,444	
Non-trainable parameters	0	
Total parameters	460,444	

count, OSPS-MicroNet is strategically designed to excel in resource-constrained environments.

Upon comparing the parameter sizes of prominent models as in Figure 4 such as ResNet152 and MobileNet, OSPS-MicroNet, functioning as a student model, emerges with a significantly minimized number of parameters. This deliberate reduction not only optimizes computational resources but also aligns with the unique demands of environments where constraints on resources are a critical consideration.

In essence, OSPS-MicroNet distinguishes itself as a model of choice for scenarios where efficiency and resource optimization are paramount. The juxtaposition of its minimal parameter counts against more resource-intensive models like ResNet152 and MobileNet accentuates its suitability for applications in diverse settings, particularly those characterized by limitations in computational resources. The data flow of the proposed OSPS-MicroNet is depicted in Figure 5, while Algorithm 1 elucidates the model's functioning.

4 Experimentation results and discussions

In the experimental setup, the training and testing processes were conducted in three distinct phases. During Phase I, RESNET152, MobileNet, and OSPS-MicroNet were individually trained and tested. Subsequently, Phase II involved using ResNet152 as the teacher network and instructing OSPS-MicroNet as the student network. In Phase III, MobileNet assumed the role of the Teacher network, guiding OSPS-MicroNet as the student network. The training dataset consisted of $\sim 7,000$ images, pre-processed by resizing to dimensions of 224×224 pixels.

The experiments were implemented using TensorFlow library functions and Keras on Google Colab Pro Plus. All convolutional neural network (CNN) models underwent training for 25 epochs, employing the Adam optimizer with a learning rate of 0.01 and

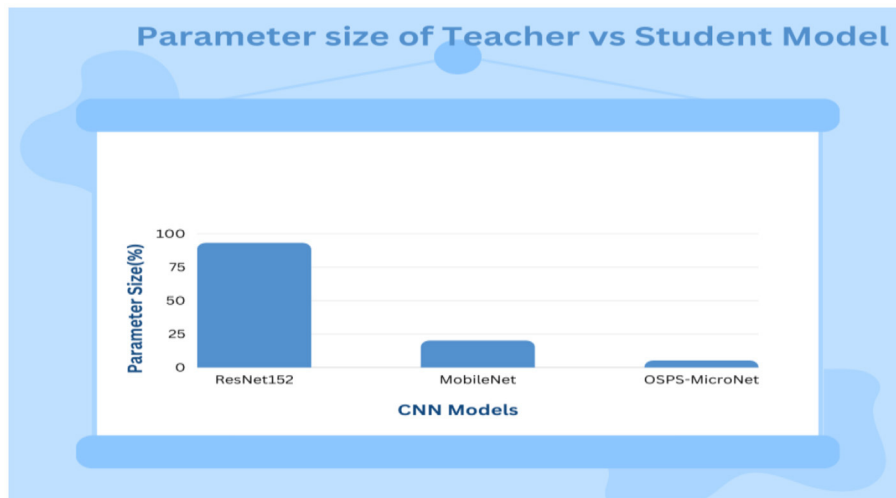


FIGURE 4 Comparison of parameter size.

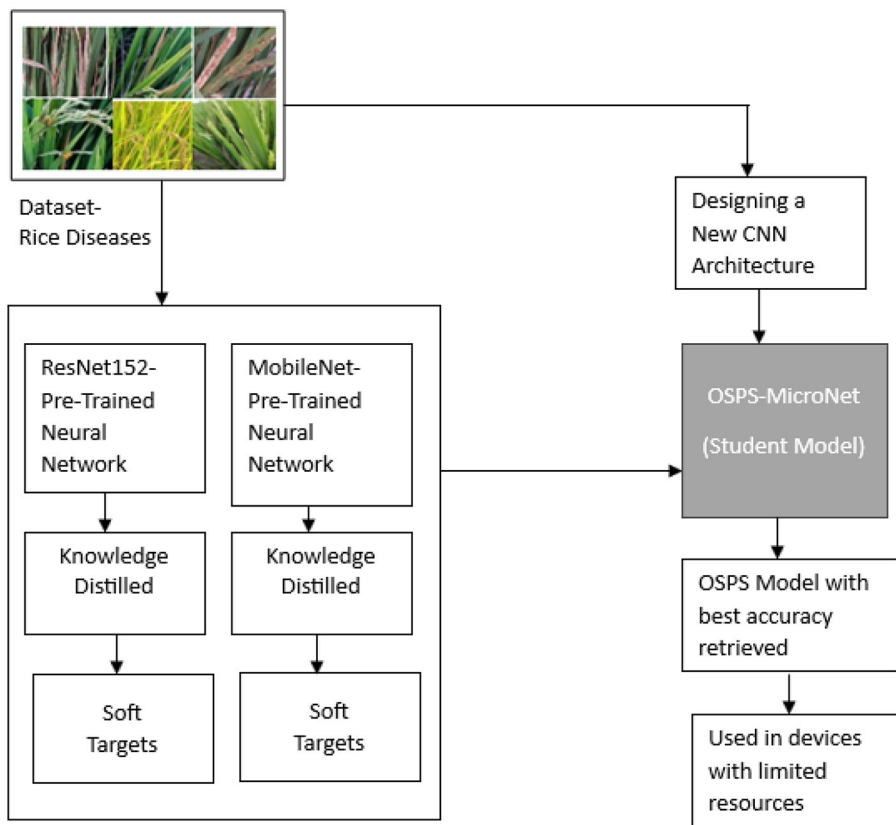


FIGURE 5 Dataflow of the proposed OSPS-MicroNet model.

a batch size of 32. The normalization technique utilized is “Zero scaling,” which involves normalizing or scaling the values of a variable about its mean or median. This process results in a distribution centered around zero. The dataset encompassed six classes, including five types of rice diseases (blast, brown spot,

bacterial sheath blight, false smut, and tungro) and a sixth class for identifying healthy rice leaves.

In Phase I, RESNET152 exhibited the highest accuracy at 98.51%, followed closely by MobileNet at 98.40%. Conversely, OSPS-MicroNet achieved an accuracy of ~20.40%. This notable

difference can be attributed to the inherent characteristics of student models, which are intentionally designed to be smaller and less complex than their teacher counterparts. The reduced size and complexity of OSPS-MicroNet resulted in comparatively lower performance and less accurate predictions, underscoring a trade-off between model size and accuracy.

The superior accuracy of RESNET152 can be primarily attributed to its substantial parameter size. RESNET152 boasts roughly 16 times more parameters than MobileNet and an impressive 129 times more parameters than OSPS-MicroNet. This stark contrast in parameter size

elucidates the critical influence of model complexity on predictive accuracy.

Visual representation of the accuracy trends for RESNET152, MobileNet, and OSPS-MicroNet can be observed in Figures 6–8, respectively. These graphical representations provide a comprehensive overview of the performance dynamics exhibited by each model throughout the experimental phases.

In Phase II, the teacher-student network was strategically employed to enhance the performance of the proposed lightweight CNN model. RESNET152 served as the teacher network, guiding the learning process of OSPS-MicroNet, the designated student network. In this phase, OSPS-MicroNet achieved an accuracy of 92.02%, showcasing a remarkable advancement from its initial accuracy of 20.40%. The Figure 9 demonstrates the accuracy of the OSPS-MicroNet model compared to its teacher, ResNet152. This substantial improvement can be attributed to the process of knowledge distillation, wherein the student model is imparted with an understanding of the relative importance and uncertainties associated with different classes.

The notable accuracy boost was achieved through efficient and focused knowledge distillation from the more complex RESNET152. While RESNET152 boasts 59 million parameters, OSPS-MicroNet operates with a significantly leaner parameter count of only 0.46 million. Despite the accuracy of OSPS-MicroNet lagging behind RESNET152 by 6%, the parameter size of OSPS-MicroNet is merely 0.77% of RESNET152's size. Remarkably, OSPS-MicroNet, occupying <1% of the parameter space of RESNET152, still achieves an impressive accuracy of 92.02%.

This underscores the pivotal role played by teacher models in transferring knowledge and expertise to their student counterparts. Even when student models are smaller and less complex, the distillation of knowledge from teacher models significantly enhances their performance. By distilling the wealth of knowledge held by the teacher model, students can leverage valuable insights, augmenting their capabilities and effectiveness in achieving high accuracy, despite their compact size and reduced complexity.

```

1. Input: D={X} //7000 images of rice
diseases
2. T1//pre-trained CNN Teacher model
ResNet152
3. T2//pre-trained CNN Teacher model
MobileNet
4. // To train the student model S using
knowledge distillation from two pre-trained
teacher models T1 and T2
   for each image xi in the dataset X:
       Set S (Student Model) to the output of
       T1 for the input xi.
       Set S to the output of T2 for the same
       input xi.
       Use Equation (2) to compute the loss
       function with S.
       Apply backpropagation to update the
       parameters of S based on the computed
       distillation loss.
   end for
    
```

Algorithm 1. Training the OSPS-MicroNet student model using pre-trained teacher models ResNet152 and MobileNet.

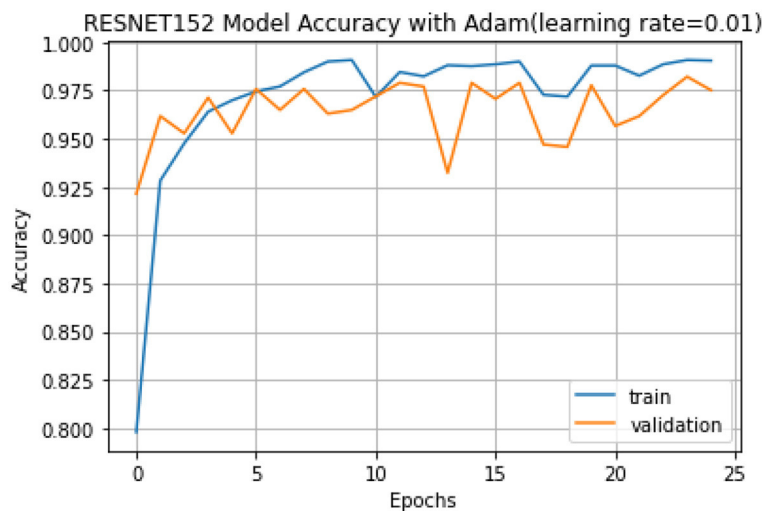


FIGURE 6 ResNet152 model accuracy graph chart.

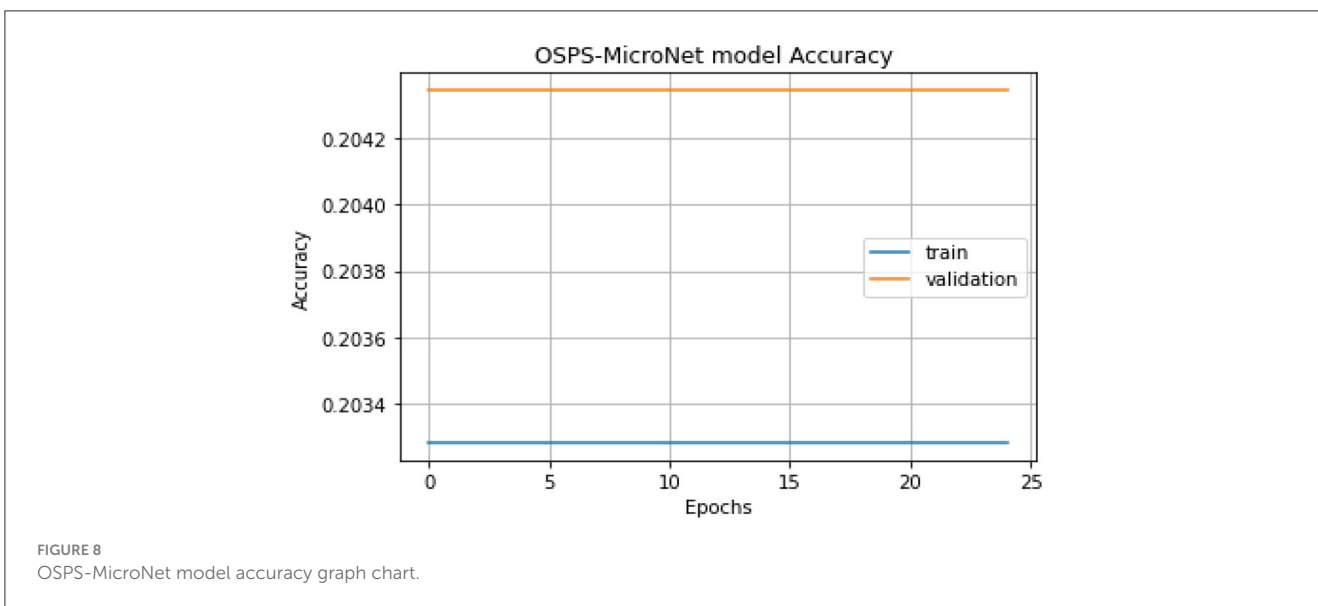
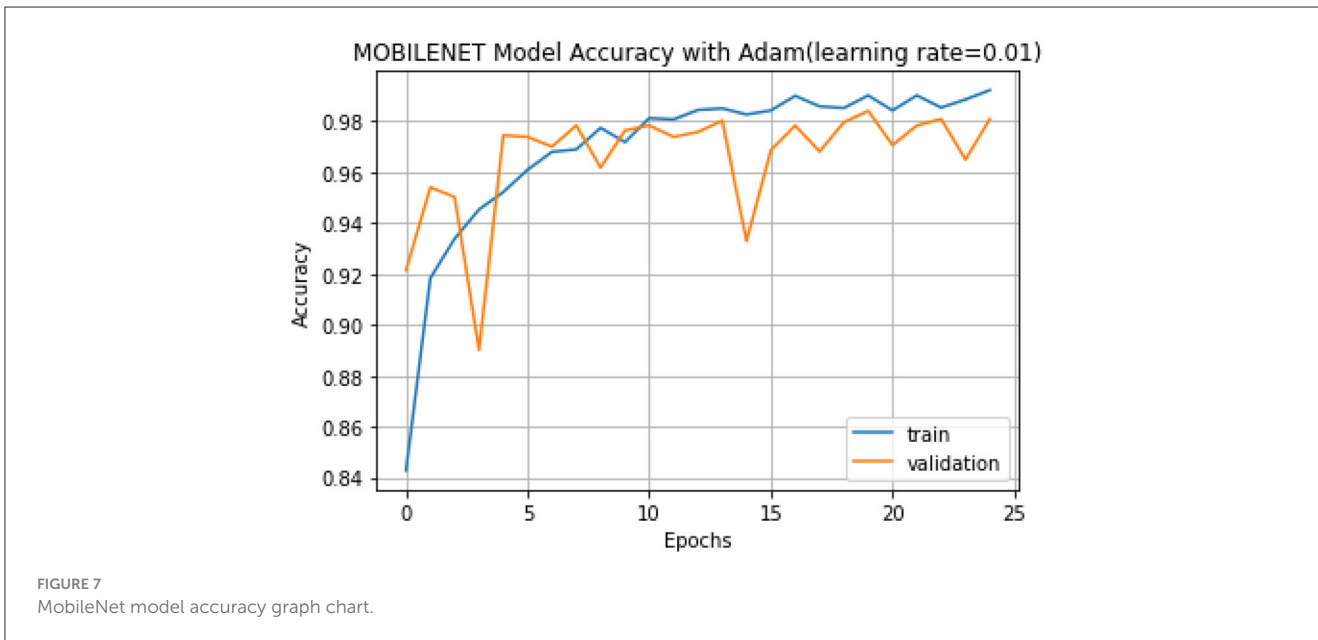
In Phase III of the experiment, MobileNet assumed the role of the teacher network, guiding the learning process of OSPS-MicroNet, the designated student network. Notably, in this phase, OSPS-MicroNet achieved an accuracy of 90%, a significant improvement from its initial accuracy of 20.40% in Phase I. Once again, this substantial increase in accuracy for OSPS-MicroNet was made possible through the process of knowledge distillation. The Figure 10 demonstrates the accuracy of the OSPS-MicroNet model compared to its teacher, MobileNet.

However, when compared to Phase II, where RESNET152 served as the teacher network, the accuracy in this phase experienced a reduction of <2%. It is noteworthy that MobileNet, with 3 million parameters, contrasts with RESNET152, which boasts 59 million parameters. The discernible trend from prior sections underscores that knowledge distillation from a heavyweight network, such as RESNET152, tends to yield superior results.

This observation highlights the impact of the choice of the teacher network on the efficacy of knowledge distillation. The inherent complexity and richness of information in heavyweight networks contribute significantly to the enhanced learning and performance of lightweight student networks. Despite the reduction in accuracy compared to Phase II, the outcome in Phase III still represents a noteworthy advancement for OSPS-MicroNet, emphasizing the intricate interplay between the teacher-student dynamics and the underlying network architecture.

Figure 11 presents a comparative chart depicting the accuracy of the OSPS-MicroNet model in two scenarios: first when executed independently as a stand-alone system, and second when executed with distilled knowledge acquired from both ResNet152 and MobileNet.

To facilitate the deployment of models on low-power or mobile devices, where the utilization of large and computationally intensive teacher models is impractical, smaller student models are



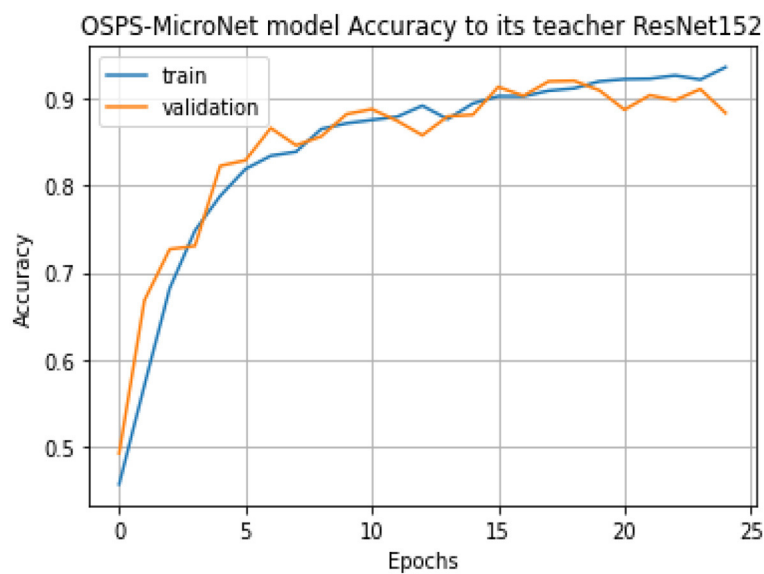


FIGURE 9
OSPS-MicroNet model accuracy to its teacher ResNet152.

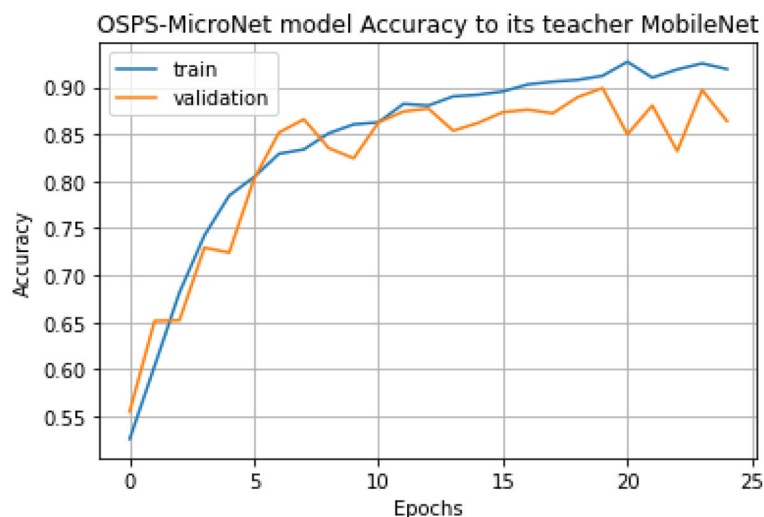


FIGURE 10
OSPS-MicroNet model accuracy to its teacher MobileNet.

employed to emulate the behavior of their teacher counterparts. The OSPS-MicroNet student model, specifically designed to be both precise and compact in comparison to ResNet152 and MobileNet, emerges as an optimal choice for deployment on such resource-constrained devices.

5 Conclusion

Early identification of rice leaf diseases will mitigate its widespread and huge economic loss. In particular, the spread of blast, bacterial sheath blight, false smut, tungro, and brown spot poses a significant challenge to rice yield and leads to significant losses. However, recent developments in CNN algorithms have

made it possible to detect these diseases, but the huge model size is a major concern for its real-time implementation. In this paper, a new micro-CNN model named OSPS-MicroNet was proposed using a teacher-student knowledge distillation process. ResNet152 and MobileNet are used as “teacher” models, and a lightweight OSPS-MicroNet model as the “student” model, which takes up only 0.7% of the teacher model. After training on a dataset of around 7,000 pre-processed images, the OSPS-MicroNet model achieved an accuracy of 92.02% when knowledge is distilled from ResNet152 as its teacher model and 90% when knowledge is distilled from MobileNet as its teacher model. The experimental results have clearly demonstrated the efficiency and compactness of the lightweight OSPS-MicroNet model in detecting diseases when it is trained using a heavyweight teacher model. In the future,

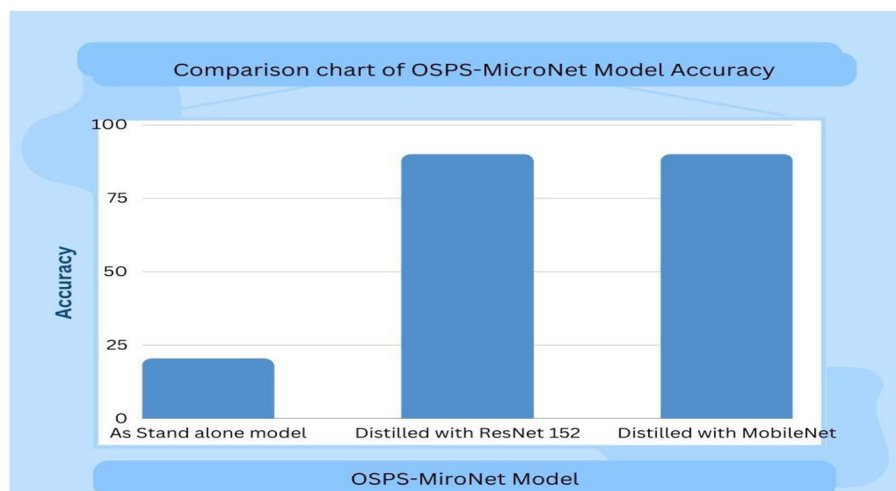


FIGURE 11
Comparison chart of OSPS-MicroNet model accuracy.

the OSPS-MicroNet is intended to be deployed on mobile devices to identify rice leaf diseases in a cost-efficient way. As per the submitted Patent Application No. (202341027943), this model has been formally filed with the Indian Patent Office and is currently in the published stage.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

PT: Conceptualization, Methodology, Visualization, Writing – original draft. BB: Conceptualization, Resources, Supervision, Validation, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

References

- Bhatia, A., Chug, A., Singh, A. P., and Singh, D. (2022). Fractional mega trend diffusion function-based feature extraction for plant disease prediction. *Int. J. Mach. Learn. Cybernet.* 14, 187–212. doi: 10.1007/s13042-022-01562-2
- Chen, F., Zhang, Y., Zhang, J., Liu, L., and Wu, K. (2022). Rice false smut detection and prescription map generation in a complex planting environment, with mixed methods, based on near earth remote sensing. *Remote Sens.* 14:945. doi: 10.3390/rs14040945
- Feng, L., Wu, B., Zhu, S., Wang, J., Su, Z., Liu, F., et al. (2020). Investigation on data fusion of multisource spectral data for rice leaf diseases identification using machine learning methods. *Front. Plant Sci.* 11:577063. doi: 10.3389/fpls.2020.577063
- Fenu, N. G., and Mallocci, F. M. (2023). Classification of pear leaf diseases based on ensemble convolutional neural. *AgriEngineering* 5, 141–152. doi: 10.3390/agriengineering5010009
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). “Deep residual learning for image recognition,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE; Computer Vision Foundation (CVF)), 770–778.

Acknowledgments

We are grateful to the SRM Institute of Science and Technology, Kattankulathur, India, for their motivation and immense support rendered for this research.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Jiang, Z., Dong, Z., Jiang, W., and Yang, Y. (2021). Recognition of rice leaf diseases and wheat leaf diseases based on multi-task deep transfer learning. *Comp. Electron. Agric.* 186:106184. doi: 10.1016/j.compag.2021.106184
- Laudani, A., Lozito, G. M., Fulginei, F. R., and Salvin, A. (2015). On training efficiency and computational costs of a feed forward neural network: a review. *Comput. Intell. Neurosci.* 2015:818243. doi: 10.1155/2015/818243
- Liu, K., and Zhang, X. (2022). PiTLiD: identification of plant disease from leaf images based on convolutional neural network. *IEEE/ACM Transact. Comp. Biol. Bioinform.* 20, 1278–1288. doi: 10.1109/TCBB.2022.3195291
- Liu, X., and Zhang, Z. (2021). A double-edged sword: reactive oxygen species (ROS) during the rice blast fungus and host interaction. *FEBS J.* 289:16171. doi: 10.1111/febs.16171
- Lu, Y., Lin, L., Zhang, X., Liu, W., and Guan, C. (2023). New method for rice disease identification based on improved deep residual shrinkage network. *Syst. Sci. Cont. Eng.* 11:2177770. doi: 10.1080/21642583.2023.2177770
- Mirzadeh, S. I., Farajtabar, M., Li, A., Levine, N., Matsukawa, A., and Ghasemzadeh, H. (2019). Improved knowledge distillation via teacher assistant: Bridging the gap between student and teacher. *Proc. AAAI Conf. Artif. Intell.* 34, 5191–5198. doi: 10.1609/aaai.v34i04.5963
- Pathan, M., Patel, N., Yagnik, H., and Shah, M. (2020). Artificial cognition for applications in smart agriculture: a comprehensive review. *Artif. Intell. Agric.* 4, 81–95. doi: 10.1016/j.aiia.2020.06.001
- Patil, R. R., and Kumar, S. (2021). Rice-fusion: a multimodality data fusion framework for rice disease diagnosis. *IEEE Access* 10, 5207–5222. doi: 10.1109/ACCESS.2022.3140815
- Singh, P., Mazumdar, P., Harikrishna, J. A., and Babu, S. (2019). Sheath blight of rice: a review and identification of priorities for future research. *Planta* 250, 1387–1407. doi: 10.1007/s00425-019-03246-8
- Singh, P., Pritamdas, K., Devi, K. J., and Devi, S. D. (2023). Custom convolutional neural network for detection and classification of rice plant diseases. *Proc. Comput. Sci.* 218, 2026–2040. doi: 10.1016/j.procs.2023.01.179
- Sridevi, S., and Kiran Kumar, K. (2022). Optimised hybrid classification approach for rice leaf disease prediction with proposed texture features. *J. Control Decis.* 11, 2141359. doi: 10.1080/23307706.2022.2141359
- Sriwanna, K. (2022). Weather – based rice blast disease forecasting. *Comp. Electron. Agri.* 193:106685. doi: 10.1016/j.compag.2022.106685
- Temniranrat, P., Kiratiratanaprak, K., Kitvimonrat, A., Sinthupinyo, W., and Patarapuwadol, S. (2021). A system for automatic rice disease detection from rice paddy images serviced via a Chatbot. *Comp. Electron. Agric.* 185:106156. doi: 10.1016/j.compag.2021.106156
- Terensan, S., Nishadi, H., Fernando, S., Nilanthi Silva, J., Chandrika, S. A., and Perera, N. (2022). Morphological and molecular analysis of fungal species associated with blast and brown spot diseases of *Oryza sativa*. *Planta* 250, 1387–1407. doi: 10.1094/PDIS-04-21-0864-RE
- Tharani, P. P., and Baranidharan, B. (2023). “Resource prudent CNN models for disease identification of rice crops,” in *International Conference on Networking and Communications (ICNWC)* (Chennai), 1–8.
- Tugrul, B., Elfatimi, E., and Eryigit, R. (2022). Convolutional neural networks in detection of plant leaf diseases: a review. *Agriculture* 12:1192. doi: 10.3390/agriculture12081192
- Tuncer, A. (2021). Cost-optimized hybrid convolutional neural networks for detection of plant leaf diseases. *J. Ambient Intellig. Hum. Comput.* 12, 8625–8636. doi: 10.1007/s12652-021-03289-4
- Upadhyay, S. K., and Kumar, A. (2022). A novel approach for rice plant diseases classification with deep convolutional neural network in springer. *Int. J. Inf. Technol.* 14, 185–199. doi: 10.1007/s41870-021-00817-5
- Wang, J., Gou, L., Zhang, W., Yang, H., and Shen, H.-W. (2019). DeepVID: deep visual interpretation and diagnosis for image classifiers via knowledge distillation. *IEEE Trans. Vis. Comput. Graph* 25, 2168–2180. doi: 10.1109/TVCG.2019.2903943
- Wang, S., Zhao, B., Yi, S., Zhou, Z., and Zhao, X. (2022). GAPS optimized fuzzy PID controller for electric-driven seeding. *Sensors* 22:6678. doi: 10.3390/s22176678
- Wang, Y., Wang, H., and Peng, Z. (2021). Rice diseases detection and classification using attention based neural network and Bayesian optimization. *Exp. Syst. Appl.* 178:114770. doi: 10.1016/j.eswa.2021.114770
- Yang, L., Yu, X., Zhang, S., Zhang, H., Xu, S., Long, H., et al. (2023). Stacking-based and improved convolutional neural network: a new approach in rice leaf disease identification. *Front. Plant Sci.* 14:1165940. doi: 10.3389/fpls.2023.1165940
- Yang, N., Chang, K., Dong, S., Tang, J., Wang, A., Huang, R., et al. (2021). Rapid image detection and recognition of rice false smut based on mobile smart devices with anti-light features from cloud database. *Biosyst. Eng.* 218, 229–244. doi: 10.1016/j.biosystemseng.2022.04.005
- Zhang, K., Cheng, K., Li, J., and Peng, Y. (2019). A channel pruning algorithm based on depth-wise separable convolution unit. *IEEE Access* 7, 173294–173309. doi: 10.1109/ACCESS.2019.2956976
- Zhou, W., An, T., Wang, J., Fu, Q., Wen, N., Sun, X., et al. (2023). Design and experiment of a targeted variable fertilization control system for deep application of liquid fertilizer. *Agronomy* 13:1687. doi: 10.3390/agronomy13071687