Ecological Informatics 75 (2023) 102068

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

Deep transfer learning model for disease identification in wheat crop

Sapna Nigam [a], Rajni Jain [b,1,*], Sudeep Marwaha [a], Alka Arora [a], Md. Ashraful Haque [a], Akshay Dheeraj [a], Vaibhav Kumar Singh [c]

[a] ICAR-Indian Agricultural Statistics Research Institute, New Delhi, India
[b] ICAR- National Institute of Agricultural Economics and Policy Research, New Delhi, India
[c] ICAR-Indian Agricultural Research Institute, New Delhi, India

ARTICLE INFO

Keywords:
Wheat rusts
EfficientNet
Deep transfer learning
Convolutional neural networks

ABSTRACT

Wheat rusts, caused by pathogenic fungi, are responsible for significant losses in Wheat production. Leaf rust can cause around 45–50% crop loss, whereas stem and stripe rust can cause up to 100% crop loss under suitable weather conditions. Early treatment is crucial in reducing yield loss and improving the effectiveness of phytosanitary measures. In this study, an EfficientNet architecture-based model for Wheat disease identification is proposed for automatically detecting major Wheat rusts. We prepared a dataset, referred to as WheatRust21, consisting of 6556 images of healthy and diseased leaves from natural field conditions. We attempted several classical CNN-based models such as VGG19, ResNet152, DenseNet169, InceptionNetV3, and MobileNetV2 for Wheat rust disease identification and obtained accuracy ranging from 91.2 to 97.8%. To further improve accuracy, we experimented with eight variants of EfficientNet architecture and discovered that our fine-tuned EfficientNet B4 model achieved a testing accuracy of 99.35%, a result that has not been reported in the literature so far to the best of our knowledge. This model can be easily integrated into mobile applications for use by stakeholders for image-based wheat disease identification in field conditions.

1. Introduction

The traditional approaches to plant disease diagnosis and management practices are pathologist-oriented, subjective, costly, time-consuming, and labor-intensive (Mohanty et al., 2016). Due to the shortage of availability of expertise and labor, AI-based computer vision approaches need to be explored for cost-effective and timely identification of disease occurrence for each piece of land under cropping (Kamilaris and Prenafeta-Boldú, 2018; Abade et al., 2020; Arnal Barbedo, 2013; Haque et al., 2022a). These methods can be adapted to monitor plant health regularly, give information about the minimum required dosage of pesticides to treat crop diseases, and reduce the undesirable use of chemicals (Ferentinos, 2018; Nigam and Jain, 2020). Currently, deep learning algorithms are at the forefront of AI-based computer vision approaches. These algorithms allow computer models to understand data by independently extracting features, unlike the machine learning algorithms of the previous decade that relied on human-engineered feature extraction (LeCun et al., 2015). Well-known deep learning architectures such as AlexNet (Krizhevsky et al., 2012), VGGNet, InceptionV3 (Szegedy et al., 2016), ResNet, and DenseNet (Huang et al., 2017) were used by many researchers in plant disease classification. Most researchers used the publicly available PlantVillage dataset for horticultural crops (Hughes and Salathé, 2015). On this dataset, recent literature reports disease identification accuracies up to 98% using deep learning models (Naik et al., 2022; Sutaji and Yıldız, 2022).

Classical deep learning architectures enable the building of new disease identification models using a transfer learning approach. As described by Too et al. (2019), transfer learning is a machine learning technique in which a model trained on one task is fine-tuned for a related but different task. This approach is simple and effective, avoids overfitting when training data is scarce and helps reduce computational complexity while improving the learning capacity of the new model. Mohanty et al. (2016) conducted experiments with 60 configurations based on the transfer learning model and also trained the model from scratch for comparison. They demonstrated that fine-tuning a pre-trained network is better for plant disease classification than training a model from scratch. Lee et al. (2020), Chen et al. (2020), Thangaraj

* Corresponding author.
E-mail addresses: sapna.nigam@icar.gov.in (S. Nigam), Rajni.jain@icar.gov.in, rajnijain67@gmail.com (R. Jain).
[1] Corresponding author, email address: Rajni.jain@icar.gov.in (Rajni Jain)First author, email address: sapna.nigam@icar.gov.in (Sapna Nigam)

https://doi.org/10.1016/j.ecoinf.2023.102068
Received 1 November 2022; Received in revised form 13 February 2023; Accepted 10 March 2023
Available online 13 March 2023

et al. (2021), Zhao et al. (2022), and Haque et al. (2022b) also successfully demonstrated the potential of transfer learning models for identifying plant diseases.

Despite its importance for global food security, there have been few studies on computer vision for disease identification in cereal crops due to a lack of datasets. Among cereals, Wheat is a crucial staple food crop that can suffer a 45–100% reduction in yield due to rusts, depending on the level of infestation at the time of diagnosis and treatment. Three types of rust can affect wheat: stripe or yellow, stem or black, and a leaf or brown rust. Currently, the most common method of diagnosing Wheat disease is through a visual assessment by experts using the naked eye. Although recent advancements in deep learning algorithms have been made, few researchers have attempted to develop a wheat disease identification model, likely due to the lack of a wheat disease dataset. Recognizing the need for automated wheat rust identification models, we have attempted to compile a wheat disease dataset for three seasons from 2019 to 2021. We used the same for our experiments in this work.

At the same time, several other researchers have also undertaken similar work for wheat disease identification (Picon et al., 2019; Arnal Barbedo, 2019; Mi et al., 2020; Bao et al., 2021). One notable successful attempt was made by Lu et al. (2017), who improved VGGNet models, a classical deep learning architecture, to develop an in-field wheat disease detection and achieved an accuracy of 97.95%. They used their own compiled dataset, which is not publicly available for further research. Several similar attempts on wheat crops using their own compiled datasets and classical deep learning architectures have also been reported. For instance, Nigam et al. (2021a) worked on the automated classification of infected yellow rust disease from healthy leaves. Schirrmann et al. (2021) developed a ResNet-based image classifier for detecting stripe rust in winter wheat canopies. Nigam et al. (2021b) also carried out the detection and severity assessment of stem rust in wheat crops. The commonality between these three studies is that they focused mainly on: (i) identifying a single diseased leaf and achieving accuracy in the range of 94–97%, (ii) using their own compiled dataset, and (iii) classical deep learning architecture.

Following the success of classical deep learning architectures in identifying plant diseases, Tan and Le introduced a new deep learning architecture called EfficientNet (Tan and Le, 2020). The effectiveness of EfficientNet was evaluated on the PlantVillage dataset by Atila et al. and achieved high accuracy levels, ranging from 98.4 to 99.97%, for different horticultural crops (Atila et al., 2021). Inspired by this success, Genaev et al. applied the EfficientNet architecture to wheat disease identification and achieved an accuracy of 94.1% (Genaev et al., 2021). They utilized the Wheat Fungi Disease dataset (WFD2020), which consisted of 2414 images for five different fungal diseases. One reason for the lower accuracy compared to horticultural crops could be the limited number of images for each type of Wheat rust disease. To the best of our knowledge, no other studies have applied the EfficientNet architecture to Wheat rust disease identification.

The present study experimentally explores the performance of the EfficientNet architecture on the Wheat dataset, referred to as WheatRust21, throughout this paper, for three kinds of rust identification. Although there have been few attempts in the literature to use transfer learning models, serious attempts are necessary to establish a model for Wheat disease in real-life conditions. The “No free lunch theorem” exists, meaning that a model developed on some benchmarking datasets may not be suitable for other datasets. Therefore, rigorous experimentation was necessary to develop a high-performing Wheat disease identification model. As the dataset was exhaustive and covered an equal proportion of three diseased and one healthy class, our model is expected to be directly used on any other Wheat dataset for disease identification. We experimented with eight variants of the EfficientNet architecture and five available architectures of classical deep learning algorithms and fine-tuned each algorithm to get its best performance, which is reported in this paper on our WheatRust21 dataset. We achieved the highest testing accuracy of 99.35% on this dataset.

Our paper presents three significant contributions to the field: (1) The Wheat Rust Dataset (WheatRust21), which comprises 6556 rust images gathered from actual field conditions for two crop seasons (2019–2021); (2) Wheat disease identification models built using classical CNN architecture, and (3) An efficient Wheat disease identification model based on the EfficientNet B4 architecture.

2. Methodology

2.1. Dataset

In this study, we investigated three types of Wheat rust diseases: Stripe rust (also known as yellow rust), Stem rust (black rust), and Leaf rust (brown rust). We collected images of plants affected by each of these diseases and images of healthy plants (refer to Table 1 and Fig. 1).

The images of the WheatRust21 dataset were captured from the Plant Pathology Division of ICAR-Indian Agricultural Research Institute in New Delhi and its regional centre in Madhya Pradesh over the period of 2019–2021. The dataset contains 6556 images, which have been divided into four categories, as indicated in Table 1. The images were collected on sunny days between 12:00 pm and 2:00 pm, at regular intervals of 10 days after the appearance of the disease symptoms, to ensure that the leaves were in similar growth stages. The images were taken using a handheld mobile camera with 20 megapixels and a 25 mm wide lens, as this type of device is commonly available to farmers.

Compared to the PlantVillage dataset, WheatRust21 has some notable differences:

a) The class label is associated with each image in the PlantVillage dataset.
b) Each image in PlantVillage is comprised of a single leaf, whereas the WheatRust21 images show the entire plant in its natural environment.
c) PlantVillage images have a consistent white paper background, while the background in WheatRust21 images varies based on crop conditions.

The WheatRust21 dataset was divided into three subsets: training set, validation set, and test set, in a ratio of 70:20:10. This split ratio was chosen based on previous research, which found that 70:30 ratios were optimal for building machine learning models (Nguyen et al., 2021; Seidu et al., 2022).

2.2. Data augmentation

Augmentation is a technique used to enhance images in order to prevent neural networks from picking up on unimportant patterns and improving overall performance. The software package Augmentor was used to achieve this goal, ensuring a balanced distribution of classes in the dataset. The process involved rotating the images by 90°, rescaling,

Table 1
A summary of Wheat Rust disease, including its symptoms and the number of relevant images.

Class	Symptoms	Affected area	Collected images	Augmented images
Stripe rust	Small yellow to light orange pustules	Leaf veins, spikes, and leaf sheaths	1536	2500
Stem rust	Large dark orange to red oval-shaped pustules	Lower leaf surface, plant stem, and spikes	1990	2500
Leaf rust	Circular and orange to brown spore pustules	Upper leaf surface and leaf sheath	1330	2500
Healthy	–	–	1700	2500
Total number of images			6556	10,000

Fig. 1. Sample images of a Wheat plant, including healthy and diseased leaves.

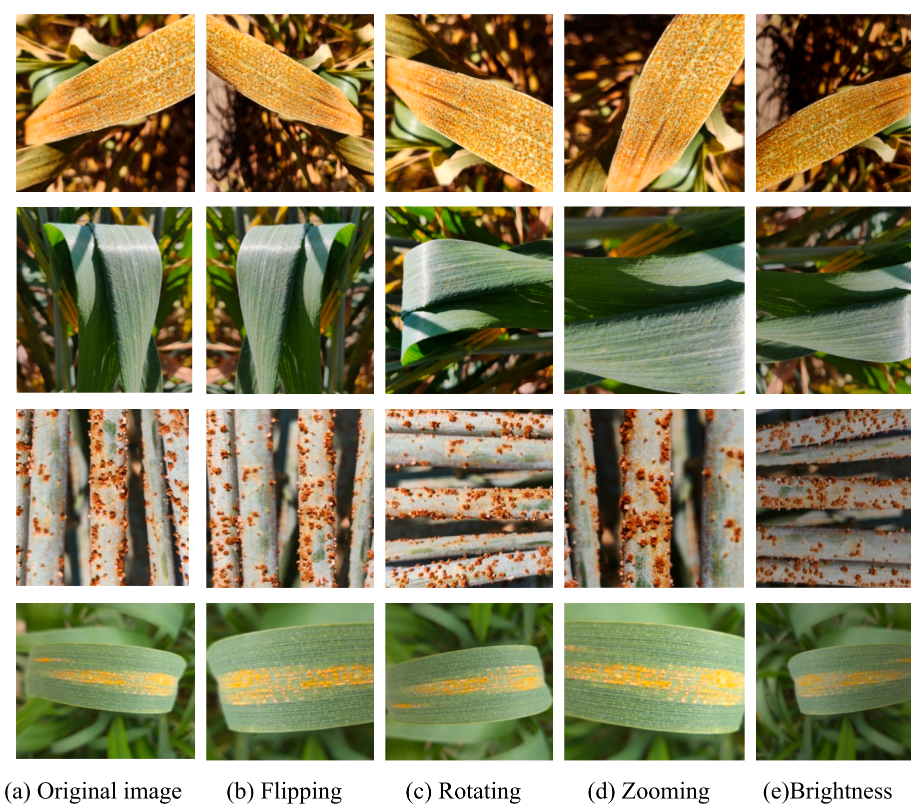


Fig. 2. Samples of augmented dataset images.

distorting, zooming, flipping horizontally and vertically, and adjusting the brightness. The end result was a dataset that evenly distributed 25% of each class. We conducted separate experiments using both the original and augmented datasets.(Fig. 2)

2.3. EfficientNet architecture

Tan and Le introduced a novel baseline network architecture that balances both accuracy and computational efficiency in terms of floating-point operations (Tan, 2018; Tan and Le, 2020). The backbone of the baseline architecture is the Mobile Inverted Bottleneck Convolution (MBConv), which is similar to MobileNetV2 (Sandler et al., 2018) and MnasNet (Tan, 2018). Despite being larger than these previous models, it is designed to be more efficient in terms of FLOPs. To further enhance efficiency, Tan and Le proposed a compound scaling method for their EfficientNet models (B0-B7) that uses coefficients to adjust network depth, width, and resolution uniformly. The method is defined as follows:

$$\text{Depth } (d) = \alpha^{\varphi}$$

$$\text{Width } (w) = \beta^{\varphi}$$

$$\text{Resolution } (r) = \gamma^{\varphi}$$

$$\text{Such that } \alpha.\beta^2.\gamma^2 \approx 2 \text{ and } \alpha \geq 1, \beta \geq 1, \gamma \geq 1 \tag{1}$$

where α, β, γ are the constants obtained by a small grid search. Additionally, the amount of resources that can be used for model scaling is determined by the compound coefficient (φ). The constants α, β, γ determine more resource allocation to network depth, width, and resolution. The FLOPS of a normal convolution operation are proportional to d, w^2, and r^2, meaning that doubling network depth will result in a doubling of FLOPS, but doubling network width or resolution will result in a four-times increase in FLOPS. Since Convolution operations typically account for most of the computation costs in ConvNets, scaling a ConvNet with Eq. (1) will roughly increase total FLOPS by$(\alpha.\ \beta^2.\ \gamma^2)^{\varphi}$. Also, under the constraint of $\alpha.\ \beta^2.\ \gamma^2 \approx 2$, any new φ, the total flops will increase by 2^{φ} approximately. The compound scaling method scales this model in two steps as follows, starting from Baseline EfficientNet-B0 (Tan and Le, 2020):

Step 1: This step assumes that there are twice as many resources available. A grid search is done by fixing $\varphi = 1$, and the best values $\alpha = 1.2$, $\beta = 1.1$, $\gamma = 1.15$ are discovered for EfficientNet B0, under the limitations of $\alpha.\ \beta^2.\ \gamma^2 \approx 2$.

Step 2: By fixing values for the constants α, β, γ and, using Eq. (1) the baseline network is scaled up with various values of φ to create EfficientNet-B1 to B7.

A comparison of the conventional and compound scaling methods is shown in Fig. 3. Fig. 3(a) depicts the primary baseline network, while Figs. 3(b) through (d) illustrate the increase in width, depth, and resolution with conventional scaling. Fig. 3(e) presents the proposed compound scaling method schematically.

The compound scaling in EfficientNet prioritizes the most crucial features in the relevant regions, resulting in a more comprehensive representation of the object details compared to other models that may fail to capture all essential features in images (Tan and Le, 2020). This approach results in a reduction of parameters, making EfficientNet more efficient compared to other models.

Fig. 4, as described by Tan and Le (2020), depicts a graphical representation of the EfficientNet-B0 baseline architecture. The architecture is then uniformly scaled up using fixed scaling coefficients to generate a family of eight models (B1–B7) that have been proven to surpass classical CNN models in terms of accuracy and efficiency (Atila et al., 2021). Each block represents a layer consisting of operators such as Conv 3 × 3, MBConv1, and MBConv6, and the resolution (HxW) and output channels are indicated between the layers.

In order to scale up to the B1-B7 architectures, the blocks of the EfficientNet B0 architecture (as shown in Fig. 4) were expanded with additional sub-blocks. These sub-blocks were added to systematically increase the number of feature maps, layers, and input resolution of the image, resulting in each subsequent variant of EfficientNet. Further information on these sub-blocks can be found in Ahmed and Sabab (2022). The pre-trained EfficientNet models B0-B7 were used for training and fine-tuning on the WheatRust21 dataset to develop a wheat crop disease classification model. However, due to limitations in GPU memory and other hardware constraints, B5-B7 were not feasible to pursue. The performance of B0-B4 is discussed in Section 3.

2.4. Transfer learning-based proposed disease identification model

The central idea behind the experiments in this study is to develop a new network using transfer learning and pre-trained parameters from the ImageNet dataset. As depicted in Fig. 5(a), the layers of EfficientNet were pre-trained by Tan and Le (2020) for the classification of 1000 classes from the ImageNet dataset (Deng et al., 2009). Transfer learning

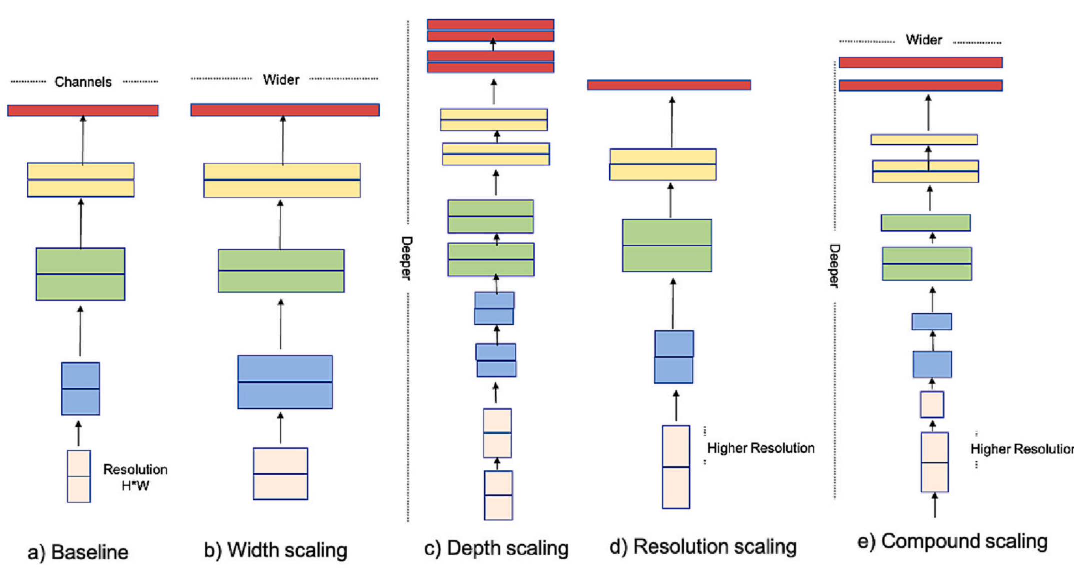


Fig. 3. Comparison of conventional and compound scaling.

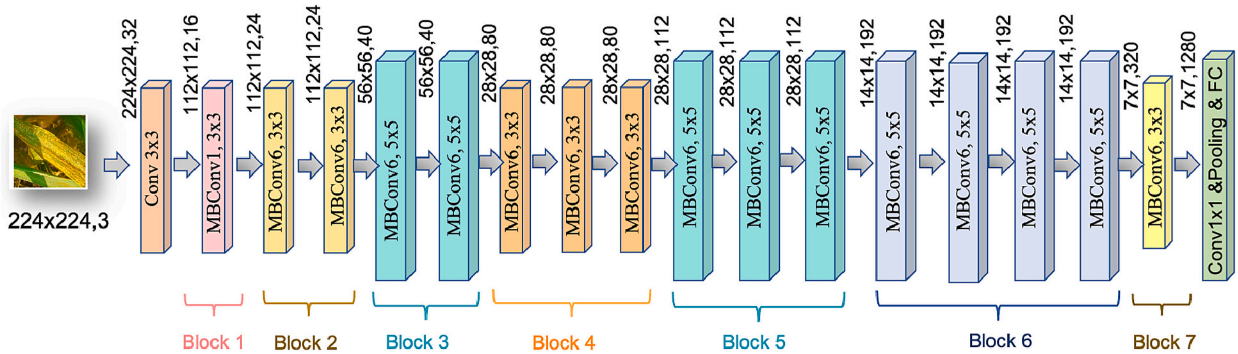


Fig. 4. Baseline architecture of EfficientNetB0.

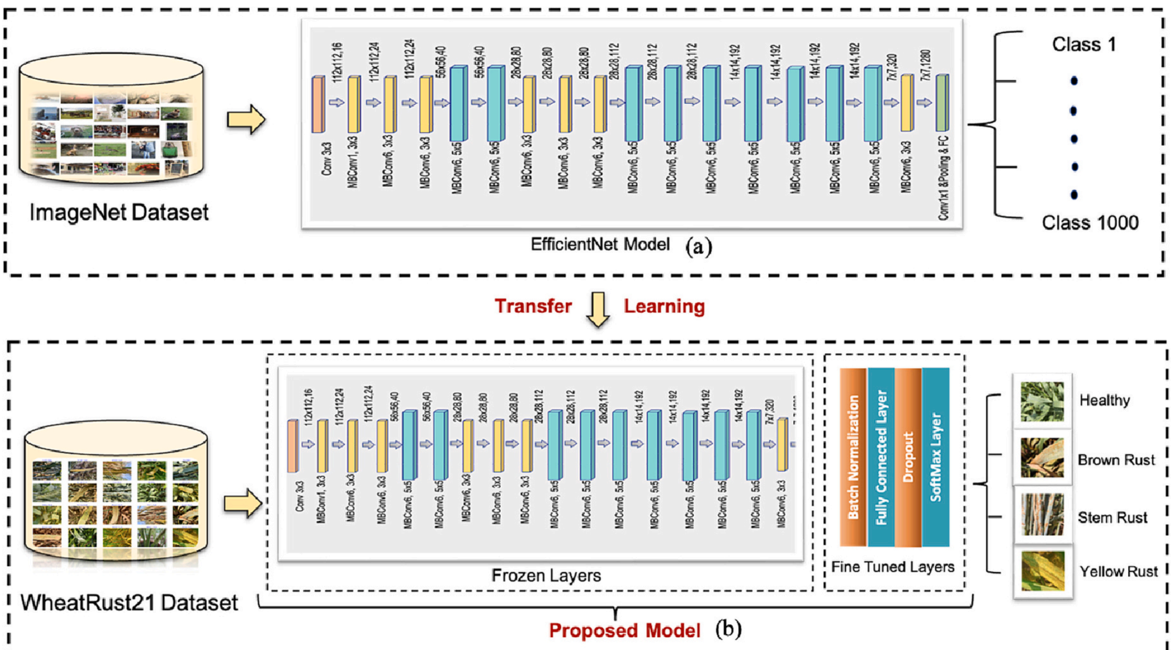


Fig. 5. Transfer learning-based proposed model framework.

allows researchers to utilize these parameters to perform image classification on other related tasks. In this study, the last layer of the pre-trained network was replaced with new layers and fine-tuned for the classification of four classes of diseased leaves on the WheatRust21 dataset (Fig. 5b). The pre-trained weights from Fig. 5a served as default values in Fig. 5b, except for the last layer. In our proposed transfer learning approach, the final layer of each EfficientNet architecture (B0-B7) was replaced with a batch normalization layer, an additional fully connected layer, a dropout, and a SoftMax layer, as illustrated in Fig. 5. The SoftMax layer in the last layer classified the input leaf images into four classes: healthy, yellow, brown, and stem rust.

2.5. State-of-the-art CNN models

During the experimentation process, the proposed model was compared with several classical deep learning models, including VGGNet (Simonyan and Zisserman, 2014), ResNet152 (He et al., 2016), MobileNetV2 (Sandler et al., 2018), DenseNet121 (Huang et al., 2017), and InceptionV3 (Szegedy et al., 2016). The networks were trained to reset the network parameters, and later, the last layer of the trained networks was modified with a new SoftMax layer consisting of four classes from the WheatRust21 dataset (as described in Section 2.1).

2.6. Experimental implementation

The experimental implementation was carried out with GPU support on an NVIDIA DGX server system, using the Ubuntu operating system and an Intel® Xeon® CPU E5–2698 v4 with 32 GB of RAM and an NVIDIA Tesla V100-SXM2 with 528 GB of memory. The models were built using Keras with Tensorflow as the backend.

2.7. Hyperparameters tuning

Hyperparameters are essential components in training a machine learning model and are manually set by the programmer to optimize its performance. The hyperparameters used in this study for the EfficientNet models are presented in Table 2. The Adam optimization algorithm was used in all models with a basic learning rate of 0.001. The batch size was set to 32 in the B0-B5 models, but it was reduced to 8 in the B6-B7 models due to computational limitations. An early stopping technique with patience of 3 was employed to prevent overfitting, and the training epochs for the EfficientNet models ranged from 15 to 25, with a maximum defined epoch of 50. Additionally, a dropout rate of 0.20 was applied to all B0-B7 models, and the momentum was adjusted to 0.9.

Table 2
Set of Hyperparameters identified for training model.

Hyperparameters	Training epochs	Batch sizes	Learning Rate	Loss function	Momentum	Dropout	Optimizer
Values	15–25	32 and 8	0.001	Categorical cross-entropy	0.9	0.2	Adam

2.8. Performance metrics

The Confusion Matrix is a tool used to compare the predicted and actual values of a classification model for each class (Ting, 2017). To perform a comprehensive evaluation of the models, various metrics were used, including accuracy, precision, F1 score, and recall, as presented in Eqs. (2)–(5) (Hossin and Sulaiman, 2015).

$$Accuracy = \frac{True\ Positive(TP) + True\ Negative\ (TN)}{True\ Positive(TP) + True\ Negative(TN) + False\ Positive(FP) + False\ Negative\ (FN)} \tag{2}$$

$$Precision = \frac{True\ Positive(TP)}{True\ Positive(TP) + False\ Positive\ (FP)} \tag{3}$$

$$Recall = \frac{True\ Positive(TP)}{True\ Positive(TP) + False\ Negative\ (FN)} \tag{4}$$

$$F1 = \frac{2{*}True\ Positive(TP)}{2{*}True\ Positive(TP) + False\ Positive(FP) + False\ Negative(FN)} \tag{5}$$

True positives refer to the number of images that were correctly identified in a specific rust category. True negatives represent the total number of images that were correctly classified across all rust categories except for the rust category to which it belongs. False negatives refer to the number of images incorrectly categorized within the relevant rust class. False positives describe the number of images that were incorrectly identified as belonging to another rust category. The F1 score is a summary metric that represents the overall predictive performance of a model.

3. Results

As outlined in Section 2, the performance of the proposed CNN and EfficientNet models was evaluated using a confusion matrix and accuracy curves.

3.1. Performance of state-of-the-art CNN models

The input image size for all models in Table 3 was set to 224 × 224 pixels, with the exception of InceptionV3, which had an input size of 299 × 299 pixels. The batch size and epoch count were set to 32 and 30, respectively, during the experiment. The MobileNetV2 model showed the highest training and testing accuracy on the original dataset, with 96.34% and 95.67%, respectively. The InceptionV3 model had better performance on the augmented datasets, with a training accuracy of 97.89% and a testing accuracy of 97.11% (Table 3).

It was observed that the testing accuracy was significantly lower than the training accuracy, as is commonly seen in machine learning algorithms for both original and augmented datasets. Additionally, the augmented datasets showed higher accuracy than the original, untransformed dataset, which is consistent with findings from other deep learning studies (Atila et al., 2021; Naik et al., 2022).

3.2. Performance of EfficientNet models

The results of each EfficientNet model (B0-B7) on both the original and augmented datasets are shown in Table 4. To improve learning efficiency and speed up the training process, early stopping and learning rate scheduler approaches were utilized. The total training time was considered to be the period until the model loss values started to increase. The training time per epoch was calculated by dividing the total training time by the number of epochs. All hyperparameters were kept the same for both the original and augmented dataset experiments (as shown in Table 2).

Table 4 demonstrates that EfficientNet B4 achieves the highest training and testing accuracy for both the original and augmented datasets. The accuracy of the augmented dataset surpasses that of the training dataset in all instances. The increased time required for the augmented dataset is due to its larger number of images. Among the hardware limitations, EfficientNet B4 on the augmented dataset exhibits the best training and testing accuracy, with results of 99.91% and 99.35%, respectively. It is noted that models B5 to B7 have larger input image sizes compared to models B0 to B4. Due to GPU memory constraints, models B5 to B7 were trained with a batch size of 8, rather than 32 after experimentation with intermediate values through a trial and error process. As a result, models B5 to B7 have lower accuracy and longer training times. In the future, improved GPU memory resources may lead to increased accuracy for models B5 to B7.

Table 5 displays the results of our experiments with various augmentation techniques. It is evident that individual augmentation techniques, such as flipping, zooming, and adjusting brightness, have a

Table 3
Performance of State-of-the-art CNN models.

Model Parameters (Batch size = 32, Epochs = 30)	Original dataset accuracy (%)			Augmented dataset accuracy (%)			Original	Augmented
Model	Train	Val	Test	Train	Val	Test	Training time (s)/epoch	Training time (s)/epoch
VGGNet19	90.36	90.65	90.12	91.28	91.04	90.89	15	33
ResNet152	94.12	92.71	93.91	95.95	95.69	94.25	30	90
MobileNetV2	96.34	96.12	95.67	97.47	97.13	96.91	25	55
MobileNetV3	96.57	96.03	95.12	96.88	96.37	96.04	32	64
DenseNet169	95.59	95.06	94.21	96.74	96.50	95.64	34	81
InceptionV3	96.31	95.77	95.47	97.89	97.45	97.11	41	98

Table 4
Performance comparison of EfficientNet models.

EfficientNet model	Size		Original dataset accuracy (%)			Training Time taken per epoch (*sec*)	Time to predict one image (sec)	Augmented Dataset Accuracy (%)			Training Time taken per epoch (sec)	Time to predict one image (sec)
	Input (pixels)	Batch (#images)	Train	Test	Val			Train	Test	Val		
B0	224	32	97.73	96.55	97.79	32	0.19	97.88	98.10	97.60	48	0.39
B1	240	32	98.84	96.70	97.93	78	0.19	98.82	98.29	98.60	71	0.41
B2	260	32	98.52	97.09	97.47	108	0.20	98.91	98.63	98.60	84	0.42
B3	300	32	99.78	97.10	98.32	122	0.22	98.91	98.85	98.70	129	0.47
B4	380	32	99.83	98.93	98.47	171	0.30	99.91	99.35	99.49	249	0.55
B5	456	8	98.82	97.10	97.23	361	0.36	98.94	98.30	98.09	584	0.60
B6	528	8	98.89	97.70	98.77	443	0.55	98.89	98.70	98.77	840	0.76
B7	600	8	98.91	97.80	97.83	1480	0.61	98.91	98.80	98.83	1674	0.84

Table 5
Performance of EfficientNet B4 model according to the augmentation method.

Augmentation Method	Train Acc (%)	Val Acc (%)	Test Acc (%)	Training time per epoch (s)
Rotation (90° & 270°)	99.87	98.94	99.31	262
Flip (left-right, top-bottom)	99.83	98.90	98.69	238
Zoom	99.81	98.50	98.60	265
Brightness	99.70	98.52	98.55	235
All combined	99.91	99.49	99.35	249

Note: Dataset 10,000 images, Train: Test: Val:70:20:10.

limited impact on the testing accuracy of our dataset. However, rotating the images by 90 and 270° resulted in testing accuracy that approached the overall accuracy of the B4 model, as shown in Table 4. Using a combination of all four augmentation methods resulted in faster training times compared to individual techniques and the highest accuracy for the dataset.

According to a paired *t*-test comparison, our results show that the performance of the augmented dataset is significantly superior to that of the original dataset for all training, testing, and validation datasets (Table 6 and Fig. 6). This is due to the augmentation's introduction of variation among images from different angles, which enhances the model's training. Therefore, we strongly advise augmenting the original dataset to improve the training of the model.

3.3. EfficientNet vs classical CNN models

EfficientNet B4 maintains its top position even when compared to other traditional CNN models (Fig. 6). Our results reveal that VGGNet19 had the lowest testing accuracy of 90.89%, whereas EfficientNet B4 achieved the highest testing accuracy of 99.35%. Additionally, each of the EfficientNet models (B0-B7) performed better than all of the classical CNN models.

3.4. Transfer learning and learning from scratch

In transfer learning, a pre-trained model for one task is utilized as a starting point to train a new model for a similar task (Too et al., 2019). We utilized this concept in our study by reusing the initial few hundred layers of the pre-trained model and adding four fine-tuned layers for disease identification. The hyperparameters of the added layers vary depending on the specific disease or domain. Theoretical research on transfer learning indicates that such models provide higher accuracy in fewer epochs (Section 2.4). To verify these findings empirically, we trained models from scratch on the WheatRust21 dataset using 30 epochs, both for the original and augmented datasets. Our model from scratch consisted of seven convolution layers, five max-pooling layers, and three fully connected layers, with a total of fifteen layers. The testing accuracy of the model from scratch was 89.88% for the original dataset and 95.05% for the augmented dataset.

Building a model from scratch, similar to the EfficientNet models, with several hundred layers is challenging. However, it may be an interesting area for future research to train all of the EfficientNet models B0-B7 from scratch on the WheatRust21 dataset. Our findings, including estimates of accuracy and longer training times, suggest that building a model from scratch for image classification should be avoided in favor of transfer learning-based models, which provide better accuracy, require less training time, and use less intensive hardware and training data.

Therefore, we conclude that training a model from scratch is more challenging and yields lower testing accuracy compared to transfer learning-based models for the same number of epochs. Similar conclusions were also reported by Mohanty et al. (2016) and Lee et al. (2020) in their comparisons of various transfer learning models and models trained from scratch.

3.5. Confusion matrix and other performance measures

The confusion matrices of the EfficientNet B4 model on both the original and augmented datasets are depicted in Fig. 7. The numbers on the diagonal demonstrate the number of correctly classified images, while the off-diagonal numbers indicate misclassification. For instance, in the case of the original dataset (Fig. 7a), two of the 266 brown rust images were misclassified as healthy, and four were misclassified as

Table 6
Comparison of performance of original vs augmented dataset.

Paired samples test									
Pairwise comparison		Paired differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. error mean	95% Confidence interval of the difference				
					Lower	Upper			
Pair 3	Original_Training - Augmented_Training	0.2870	0.4144	0.1465	−0.0595	0.6335	1.958	7	0.091
Pair 1	Original_validation - Augmented_validation	−0.6087	0.4959	0.1753	−1.0233	−0.1941	−3.472	7	0.010
Pair 2	Original_Testing - Augmented_Testing	−0.9312	0.3823	0.1351	−1.2508	−0.6116	−6.889	7	0.000

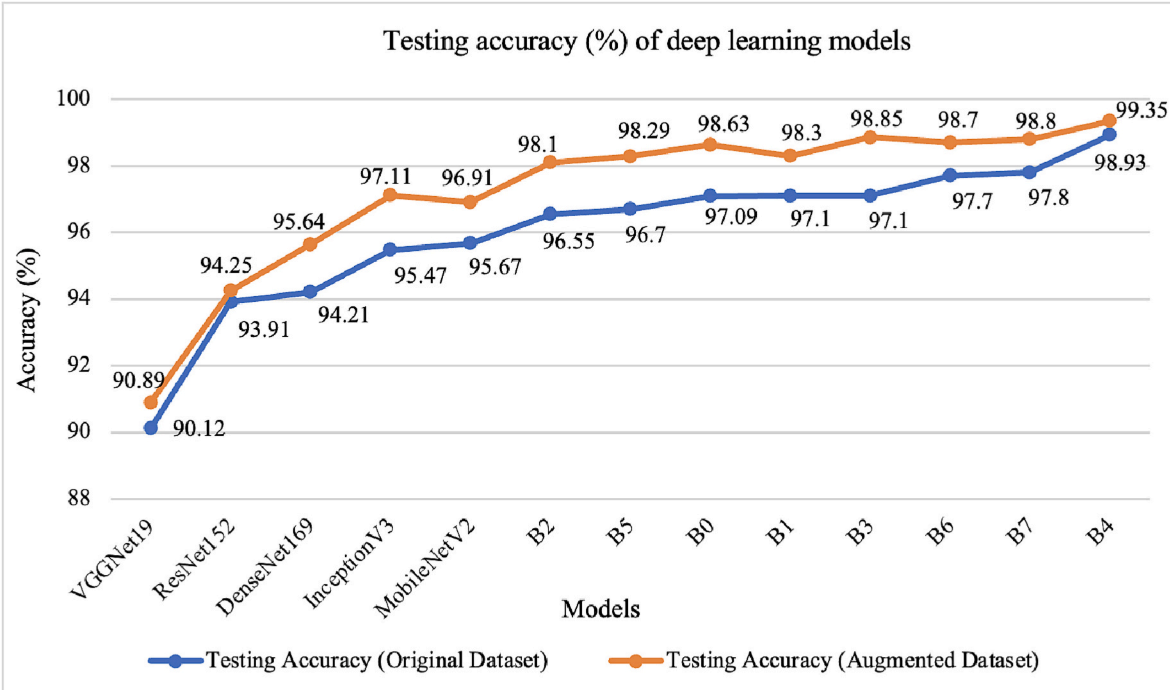
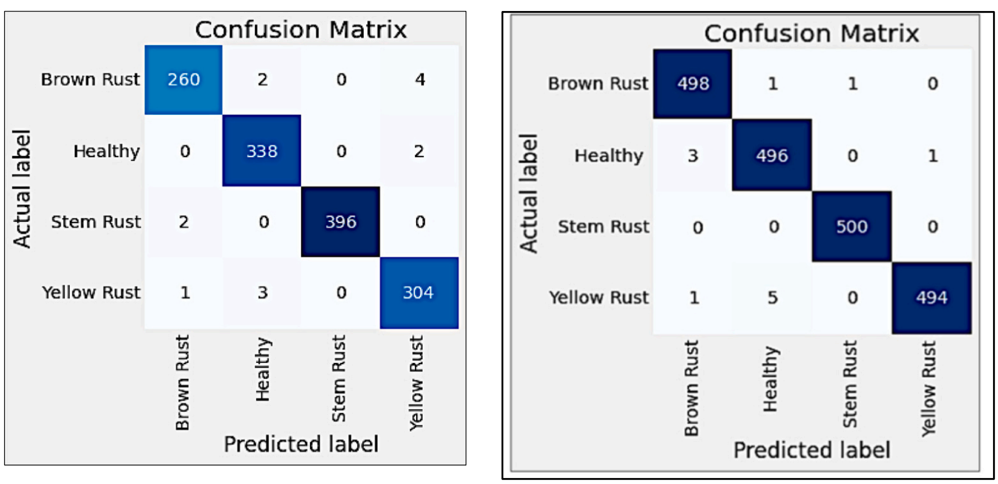


Fig. 6. Comparison of all CNN-based models testing accuracy on Original and augmented dataset.



(a)

(b)

Fig. 7. Confusion Matrix (a) original dataset (b) augmented dataset.

Table 7
Classification report for EfficientNetB4 model on both datasets.

Disease classes	Augmented dataset			Original dataset		
	Precision	Recall	F1	Precision	Recall	F1
Brown rust	0.99	1.00	0.99	0.99	0.98	0.98
Healthy	0.99	0.99	0.99	0.99	0.99	0.99
Stem rust	1.00	1.00	1.00	1.00	0.99	1.00
Yellow rust	1.00	0.99	0.99	0.98	0.99	0.98

yellow rust. The F1 score is a metric that combines Precision and Recall, and a model with an F1 score close to one is considered appropriate.

As shown in Fig. 7(b), in the case of the augmented dataset, only a few images were misclassified out of the 500 images in each category. For example, two brown rust images were misclassified, one as healthy and one as stem rust. All stem rust-infected leaves were correctly identified.

The confusion matrices led to the classification report, which includes the F1, precision, and recall measures for the EfficientNet B4 model, as shown in Table 7. The comparison of these performance measures highlights the following observations: (i) all measures in the augmented dataset are better or comparable to the original dataset, (ii) precision, recall, and F1 measures for stem rust identification using the augmented dataset are 100%, (iii) for brown rust, recall is 100%, while precision and F1 measure 99% each, (iv) in the case of yellow rust, precision is 100%, while recall and F1 are 99% each.

3.6. Accuracy and loss curves

The training-validation accuracy and loss curves for the EfficientNet

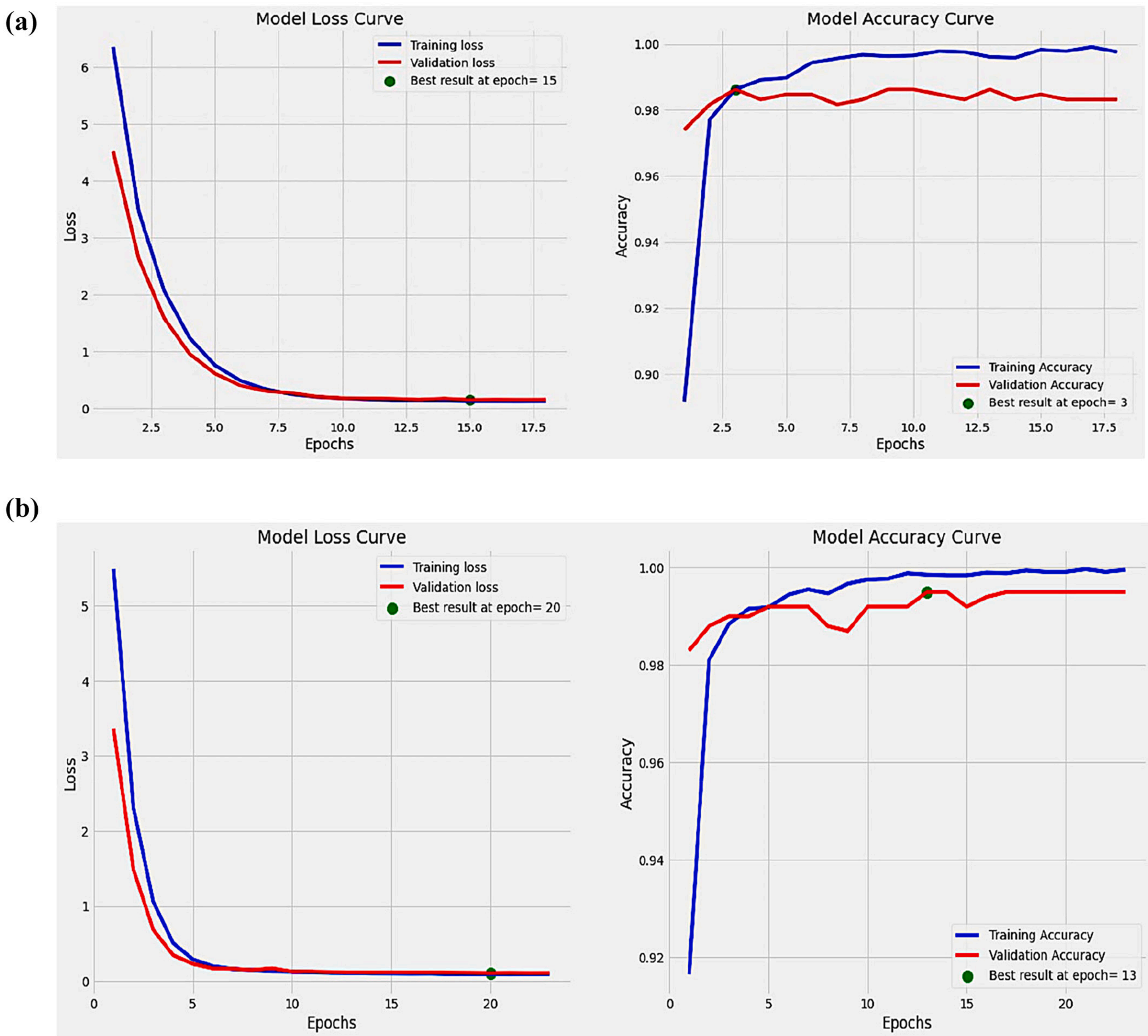


Fig. 8. (a). Training and Validation accuracy curves (Original Dataset).
(b). Training and Validation accuracy curves (Augmented Dataset).

B4 model are shown in Figs. 8a and 8b. According to Hossin and Sulaiman (2015), a model can be considered good and efficient if the training and validation accuracy curves increase steadily without significant fluctuations. The results of the EfficientNet B4 model indicate that it is a suitable option for the automatic identification of wheat diseases. The EfficientNet architecture outperforms other existing architectures due to its compound scaling method and reduced number of parameters (Tan and Le, 2020). The compound scaling approach involves simultaneous scaling of the model's depth, width, and resolution, as opposed to conventional scaling, where only one dimension is scaled at a time.

3.7. Qualitative representation of the incorrect classification of wheat diseases

The unsuccessful cases of disease classification, as depicted in Fig. 7, are presented visually in Fig. 9. The figure provides examples of leaf images from each class that were incorrectly categorized into other classes. The potential reasons for these misclassifications (Fig. 9) are discussed in Table 8.

4. Discussion

The study presents a deep transfer learning model for the image-based identification of major wheat diseases using the EfficientNet B4 architecture. A comprehensive dataset of wheat diseases and healthy plants was required for this task. However, to the best of our knowledge, no publicly available dataset of wheat rust diseases was found. Some previous studies on wheat rust disease identification generated their datasets for specific rusts (as listed in Table 9 by Genaev et al., 2021; Schirrmann et al., 2021; Picon et al., 2019; Mi et al., 2020; Lu et al., 2017). Therefore, a comprehensive dataset of all three rusts of the wheat crop was generated for the experiments in this study. While Genaev et al. (2021) reported 2414 images of wheat diseases and Picon et al. (2019) generated 8178 images of individual leaves, Lu et al. (2017) collected 9230 images of both leaves and plants under field conditions.

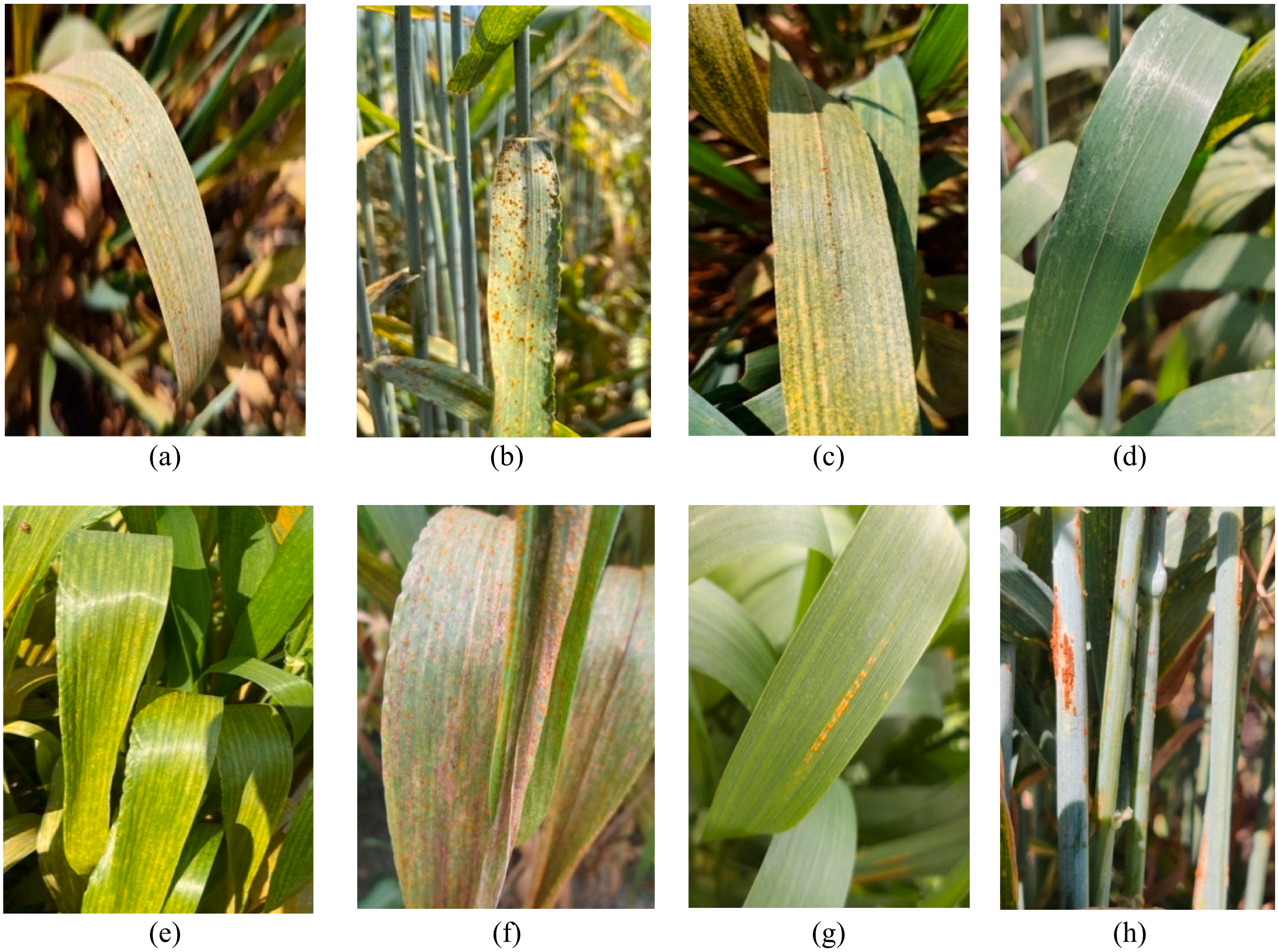


Fig. 9. Sample leaf images in each failed case of disease identification.

Table 8
Qualitative representation of failed cases in disease identification.

Failed case	Actual class	Predicted Class	Possible reason for failure
a)	Brown rust	Healthy	On the leaf, there are only a few brown rust pustules, and 99% of the leaf is healthy.
b)		Stem rust	More stems are visible in the image. Less brown rust pustules can be seen on the leaf.
c)		Yellow rust	The pustules on the leaf image resemble yellow rust pustules.
d)	Healthy	Brown rust	Presence of yellowness in the background leaves
e)		Yellow rust	Leaf chlorophyll content makes the leaf appear more yellow due to brightness, similar to the symptoms of yellow rust.
f)	Yellow rust	Brown rust	Due to the drying of the leaf, later stages of the yellow pustules grew darker over time and resembled the brown rust
g)		Healthy	Minor yellow pustules can be seen, and rest 99% of the leaf area is healthy
h)	Stem rust	Brown rust	The pustule present on the stem looks similar to the brown rust symptom

Unfortunately, our proposed model could not be tested on these datasets from different countries due to the unavailability of data.

Many studies on disease identification have used deep learning models with the PlantVillage dataset, which only includes images of horticultural crops (Atila et al., 2021; Mohanty et al., 2016). These studies have reported accuracy up to 99% on horticultural crops using the PlantVillage dataset, which has over 55,000 images of multiple fruits and 38 classes. However, despite the significance of cereal crops, there are relatively few studies on wheat disease identification using deep learning (Genaev et al., 2021). A summary of previous studies on wheat disease identification is presented in Table 9.

Previous studies have reported accuracy rates ranging from 94 to 97% for the identification of wheat diseases using classical CNN models. Genaev et al. (2021) achieved the best accuracy of 94.2% on their wheat fungi disease dataset using EfficientNet B0. However, our proposed EfficientNet B4 architecture-based model achieved a much higher accuracy rate of 99.35%, surpassing the results reported by other authors (as listed in Table 9). Our approach was evaluated on two datasets (Genaev et al., 2021; Nigam et al., 2021a) and showed an improvement in test accuracy of 1.72% and 3.3%, respectively, compared to previously reported results in the literature (as listed in Table 9-10). Unfortunately, other datasets were not publicly available for evaluation using our proposed methodology.

The limitations of deep learning-based models include the need for powerful computational resources. In this study, we were limited by hardware constraints and GPU memory issues, which prevented us from experimenting with B5-B7 models using a batch size of 32. Despite these limitations, the EfficientNet B4 architecture achieved the highest testing accuracy compared to other classical CNN models.

This model has the potential to be useful for agricultural extension workers and farmers by deploying it on a mobile application. Stakeholders can upload an image of their Wheat plant to the app to receive timely and accurate recommendations. This will also create a large

Table 9
Summary of previous studies for wheat disease identification.

Reference	Wheat Diseases	Dataset (# images)	Model	Country	Accuracy (%)
Mi et al., 2020	Stripe rust	5242	C-DenseNet	China	97.99
Lu et al., 2017	Stripe & leaf rust, leaf blotch, powdery mildew, black chaff, smut	9230	VGG-FCN-VD16 and VGG-FCN-S,	China	97.95 and 95.12
Nigam et al., 2021a	Yellow rust	2000	CNN	India	97.3
Picon et al., 2019	Septoria, tan Spot, and rust	8178	ResNet50	Spain and Germany	97
Schirrmann et al., 2021	Stripe rust	3972	ResNet18	Germany	95
Genaev et al., 2021	Wheat rusts, powdery mildew, and Septoria	2414	EfficientNet B0	Russia	94.2
Bao et al., 2021	Glume blotch, scab	1576	SimpleNet	China	94.1
Proposed Study	Stripe, leaf, and stem rust	10,000	EfficientNet B4	India	99.35

Table 10
Performance evaluation of the proposed model on another available Wheat dataset.

Reference	Classes	Dataset (# augmented images)	Our model performance (Accuracy %)			Previously reported test accuracy (%)
			Train	Val	Test	
Nigam et al., 2021a	Yellow rust, healthy	5000	99.82	98.90	99.02	97.3
Genaev et al., 2021	Yellow, stem and brown rust, healthy	10,000	99.66	97.90	97.50	94.2

database of images that can further improve automated models. The automated system can promote sustainable and eco-friendly agriculture by reducing the excessive use of plant protection chemicals and preventing crop losses.

Future research can focus on identifying the severity level of Wheat diseases. This information can be used to estimate the correct amount of pesticides, which is currently lacking in the literature.

5. Conclusion

The diagnosis and classification of diseases have been crucial contributions of deep learning algorithms in recent years. These algorithms have the capability to automatically extract phenotypic traits of diseased plant leaves, providing a significant improvement over traditional methods. To fully harness the potential of these algorithms and to make them accessible to farmers, it is essential to develop mobile applications that can identify cereal diseases. Currently, a comprehensive dataset of cereal crops is lacking, hindering research in this area.

In this study, we collected 6556 images of Wheat crops infected with rust over a three-year period. We evaluated two deep learning models - EfficientNet and classical CNN-based - for Wheat disease identification and found that the EfficientNet B4-based model was the most accurate, achieving a precision of 99.35%. We also observed that augmenting the original dataset improved model performance.

Moving forward, our aim is to expand the scope of this study by incorporating more classes and severity levels of Wheat diseases. We believe this research will further advance the field of deep learning for disease diagnosis and classification.

Authorship contributions statement

Sapna Nigam: Conceptualization, Data Collection, Data Curation, Methodology Development, Model Development, Writing of Original Draft, Editing. *Rajni Jain*: Supervision, Writing, Review, Editing, Analysis, Formal Analysis, Language Editing, Final Review. *Md. Ashraful Haque*: Visualization, Validation. *Akshay Dheeraj:* Formal Analysis, Validation. *Sudeep Marwaha:* Advisor, Resource Provision. *Alka Arora:* Advisor, Resource Provision. *Vaibhav Kumar Singh:* Field Crop Cultivation.

Declaration of Competing Interest

None.

Data availability

The WheatRust21 dataset will be made available on request to the corresponding author.

Acknowledgement

The authors express their gratitude for the resources and support received from the ICAR-Indian Agricultural Statistical Research Institute (ICAR-IASRI), ICAR-National Institute of Agricultural Economics and Policy Research (ICAR-NIAP), and ICAR-Indian Agricultural Research Institute (ICAR-IARI) in New Delhi, India, that enabled the completion of the research presented in this paper.

References

Abade, A.S., Ferreira, P.A., Vidal, F.D.B., 2020. Plant diseases recognition on images using convolutional neural networks: a systematic review. Computers and Electronics in Agriculture 185, 106125.

Ahmed, T., Sabab, N.H.N., 2022. Classification and understanding of cloud structures via satellite images with EfficientUNet. SN Computer Science 3 (1), 1–11.

Arnal Barbedo, J.G., 2013. Digital image processing techniques for detecting, quantifying, and classifying plant diseases. SpringerPlus 2 (1), 1–12.

Arnal Barbedo, J.G., 2019. Plant disease identification from individual lesions and spots using deep learning. Biosystems Engineering 180, 96–107.

Atila, Ü., Uçar, M., Akyol, K., Uçar, E., 2021. Plant leaf disease classification using EfficientNet deep learning model. Ecological Informatics 61, 101182.

Bao, W., Yang, X., Liang, D., Hu, G., Yang, X., 2021. A lightweight convolutional neural network model for field wheat ear disease identification. *Computers and* Electronics in Agriculture 189, 106367.

Chen, J., Chen, J., Zhang, D., Sun, Y., Nanehkaran, Y.A., 2020. Using deep transfer learning for image-based plant disease identification. Computers and Electronics in Agriculture 173, 105393.

Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L., 2009, June. Imagenet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255.

Ferentinos, K.P., 2018. Deep learning models for plant disease detection and diagnosis. Computers and electronics in agriculture 145, 311–318.

Genaev, M.A., Skolotneva, E.S., Gultyaeva, E.I., Orlova, E.A., Bechtold, N.P., Afonnikov, D.A., 2021. Image-based wheat fungi disease identification by deep learning. Plants 10 (8), 1500.

Haque, M., Marwaha, S., Deb, C.K., Nigam, S., Arora, A., Hooda, K.S., Agrawal, R.C., 2022a. Deep learning-based approach for identification of diseases of maize crop. Scientific reports 12 (1), 1–14.

Haque, M., Marwaha, S., Deb, C.K., Nigam, S., Arora, A., 2022b. Recognition of diseases of maize crop using deep learning models. Neural Computing and Applications 1–15.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Identity mappings in deep residual networks. In: European Conference on Computer Vision. Springer, Cham, pp. 630–645.
Hossin, M., Sulaiman, M.N., 2015. A review of evaluation metrics for data classification evaluations. International Journal of Data Mining & Knowledge Management Process 5 (2), 1.
Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700–4708.
Hughes, D., Salathé, M., 2015. An open-access repository of images on plant health to enable the development of mobile disease diagnostics. arXiv preprint arXiv: 1511.08060 http://arxiv.org/abs/1511.08060.
Kamilaris, A., Prenafeta-Boldú, F.X., 2018. Deep learning in agriculture: a survey. Computers and electronics in agriculture 147, 70–90.
Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems 25.
LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436–444.
Lee, S.H., Goëau, H., Bonnet, P., Joly, A., 2020. New perspectives on plant disease characterization based on deep learning. Computers and Electronics in Agriculture 170, 105220.
Lu, J., Hu, J., Zhao, G., Mei, F., Zhang, C., 2017. An in-field automatic wheat disease diagnosis system. Computers and electronics in agriculture 142, 369–379.
Mi, Z., Zhang, X., Su, J., Han, D., Su, B., 2020. Wheat stripe rust grading by deep learning with attention mechanism and images from mobile devices. Frontiers in Plant Science 11, 558126.
Mohanty, S.P., Hughes, D.P., Salathé, M., 2016. Using deep learning for image-based plant disease detection. Frontiers in plant science 7, 1419.
Naik, B.N., Malmathanraj, R., Palanisamy, P., 2022. Detection and classification of chili leaf disease using a squeeze-and-excitation-based CNN model. Ecological Informatics 69, 101663.
Nguyen, Q.H., Ly, H.B., Ho, L.S., Al-Ansari, N., Le, H.V., Tran, V.Q., Pham, B.T., 2021. Influence of data splitting on performance of machine learning models in prediction of shear strength of soil. Mathematical Problems in Engineering 2021.
Nigam, S., Jain, R., 2020. Plant disease identification using deep learning: a review. Indian Journal of Agricultural Sciences. 90 (2), 249–257.
Nigam, S., Jain, R., Marwaha, S., Arora, A., Singh, V.K., Singh, A.K., Immanuelraj, K., 2021a. Automating yellow rust disease identification in Wheat using artificial intelligence. Indian Journal of Agricultural Sciences. 91 (9), 1391–1395.
Nigam, S., Jain, R., Prakash, S., Marwaha, S., Arora, A., Singh, V.K., Singh, A.K., Prakasha, T.L., 2021b. Wheat disease severity estimation: a deep learning approach. In: International Conference on Internet of Things and Connected Technologies. Springer, Cham, pp. 185–193.
Picon, A., Alvarez-Gila, A., Seitz, M., Ortiz-Barredo, A., Echazarra, J., Johannes, A., 2019. Deep convolutional neural networks for mobile capture device-based crop disease classification in the wild. Computers and Electronics in Agriculture 161, 280–290.
Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C., 2018. Mobilenetv2: Inverted residuals and linear bottlenecks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4510–4520.
Schirrmann, M., Landwehr, N., Giebel, A., Garz, A., Dammer, K.H., 2021. Early detection of stripe rust in winter wheat using deep residual neural networks. Frontiers in Plant Science 12, 469689.
Seidu, J., Ewusi, A., Kuma, J.S.Y., Ziggah, Y.Y., Voigt, H.J., 2022. Impact of data partitioning in groundwater level prediction using artificial neural network for multiple wells. International Journal of River Basin Management 1–26 (just-accepted).
Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. https://doi.org/10.48550/arXiv.1409.1556.
Sutaji, D., Yıldız, O., 2022. LEMOXINET: Lite ensemble MobileNetV2 and Xception models to predict plant disease. Ecological Informatics 101698.
Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016. Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2818–2826.
Tan, M., 2018. MnasNet: towards automating the design of mobile machine learning models. arXiv, 1807.1162v3. doi: 10.48550/arXiv.1807.11626.
Tan, M., Le, Q., 2020. EfficientNet: Rethinking model scaling for convolutional neural networks. In: International Conference on Machine Learning. PMLR, pp. 6105–6114.
Thangaraj, R., Anandamurugan, S., Kaliappan, V.K., 2021. Automated tomato leaf disease classification using transfer learning-based deep convolution neural network. Journal of Plant Diseases and Protection 128 (1), 73–86.
Ting, K.M., 2017. Confusion Matrix. In: Sammut, C., Webb, G.I. (Eds.), Encyclopedia of Machine Learning and Data Mining. Springer, Boston, MA. https://doi.org/10.1007/978-1-4899-7687-1_50.
Too, E.C., Yujian, L., Njuki, S., Yingchun, L., 2019. A comparative study of fine-tuning deep learning models for plant disease identification. Computers and Electronics in Agriculture 161, 272–279.
Zhao, X., Li, K., Li, Y., Ma, J., Zhang, L., 2022. Identification method of vegetable diseases based on transfer learning and attention mechanism. Computers and Electronics in Agriculture 193, 106703.