

Leaf disease detection and classification with a CNN-LSTM hybrid model

Yibing Pan

Manus International School of Engineering, Fuzhou University, No. 2 Wulongjiang North Avenue, Fuzhou University City, Fujian Province

731413525@qq.com

Abstract. Smart agriculture refers to the use of modern information technology, Internet of Things technology, and artificial intelligence to achieve accurate management, efficient operation and sustainable development of the entire process of agricultural production. Smart agriculture mainly includes the application of data collection and analysis, intelligent agricultural machinery, precise fertilization, disease and pest detection and agricultural products traceability. Leaf disease detection and classification are considered as challenging yet important tasks in smart agriculture. Deep learning methods have been proven effective for these image-based recognition tasks. In this study, two advanced deep learning methods, namely, Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) are combined together to achieve a further improvement. Numerical results demonstrate that the proposed method outperforms both CNN and LSTM variants.

Keywords: Leaf Disease Detection, Deep Learning, CNN, LSTM

1. Introduction

With global agriculture facing increasing challenges such as population growth, climate change, and diminishing land resources, achieving efficient, sustainable, and environmentally friendly agricultural production has become paramount. In response to these challenges, researchers are exploring a range of advanced techniques and methods, particularly in the realms of plant health, crop disease detection, and agricultural resource management. From deep learning's convolutional neural networks, drone remote sensing technology to the Internet of Things (IoT) and edge intelligence, these technologies are continuously revolutionizing agricultural practices, making them more precise, resource-saving, and high-yielding. Furthermore, hyperspectral imaging technology, smartphone detection methods, and various machine and deep learning models have demonstrated tremendous potential for enhancing agricultural productivity, minimizing losses from diseases and pests, and achieving sustainable development goals. Hence, the integrated application of modern agricultural technology holds profound significance not only in bolstering global food security but also in conserving the environment and promoting sustainable agricultural advancement.

Numerous studies have focused on diverse techniques and methods in the field of agriculture, particularly for plant and crop health detection. Bhagat and Kumar [1] reviewed classification methods based on leaf diseases, emphasizing the importance of features, such as shape, color, and texture, and outlined the current state of various technologies. To this end, Dudi and Rajesh [2] proposed a plant leaf

classification model based on hybrid optimization algorithms and enhanced segmentation, demonstrating high accuracy. Kanda et al. [3] and Albattah et al. [4] employed deep learning approaches, such as Conditional Generative Adversarial Networks, convolutional neural networks, and frameworks based on CenterNet and DenseNet-77, for efficient plant and plant disease categorization.

In specific applications, Gashaw Ayalew and his team [5] explored the Gabor wavelet technique for wild blueberry field scenes. Huang et al. [6] employed CNN models to classify a tomato pest image dataset, complemented by traditional machine learning methods for optimization. By integrating it with traditional agriculture, Liu et al. [7] examined the applications of edge intelligence and drone remote sensing in precision agriculture, emphasizing the value of edge intelligence. Nagaraju et al. [8] proposed two novel learning algorithms for image preprocessing and segmentation to address dataset and model overfitting issues. Moreover, several deep learning models, such as the EfficientNetB0 and DenseNet121 combination by HASSAN AMIN et al. [9] and the MCPE method and EnsembleNet combination by Bo Li et al. [10], have provided high-accuracy solutions for plant disease categorization.

More broadly, Prabira Kumar Sethy et al. [11] discussed the applications of hyperspectral imaging technology in agriculture, underscoring its significance in agricultural product information. Qazi et al. [12] investigated the integration of the IoT and artificial intelligence in smart agriculture and projected future development trends. Various specific technologies, such as MobOca_Net by Chen [13], RGB and thermal imaging techniques by Hespeler et al. [14], convolutional neural networks combined with support vector machines by Turkoglu [15], and smartphone detection methods by Butera [16], have been applied to various agricultural challenges.

Although there have been many relevant studies, we found that the current mainstream model is a CNN-based deep-learning model. However, a hybrid model of a CNN and LSTM has not yet been developed. The proposed CNN-LSTM hybrid model in this study can be effectively used to detect and classify leaf diseases. The evaluation results based on a publicly available dataset show that the proposed CNN-LSTM hybrid model outperforms advanced CNN models, including MobileNe and DenseNet.

The remainder of this paper is organized as follows. Related work is discussed in Section 2. The proposed CNN-LSTM hybrid model is introduced in Section 3. Numerical experiments are presented in Section 4. Finally, conclusions are presented in Section 5.

2. Related work

In this section, we provide a literature summary related to image recognition in the domain of smart agriculture. The aim of this chapter is to review and compare some of the key relevant publications to assist readers in quickly grasping the essence of this research area. Given space constraints, we limited our scope to include only articles published within the last three years, discussing the most recent advancements in this research field.

2.1. Relevant Surveys and Reviews

This chapter presents a comprehensive overview of key reviews relevant to the domain of smart agriculture. Given the extensive literature available on the subject, revisiting these seminal reviews is essential for providing readers with a structured understanding of the broader research landscape. The reviews focused not only on image recognition in smart agriculture but also extended to interdisciplinary areas, including the Internet of Things (IoT), edge computing, and remote sensing.

Bhagat and Kumar [1] conducted a systematic review of literature pertaining to leaf disease detection and classification. They begin by delineating various types of leaf diseases, such as Biotic and Abiotic, and substantiate their categorizations with relevant images. Furthermore, 179 studies, including noise removal, image conversion, and segmentation, were analyzed, providing insights into the preprocessing techniques employed. Techniques, such as SURF, SIFT, PCA, and LDA, have been discussed in the realm of feature extraction and selection. Their findings indicate a discernible trend towards the utilization of statistical learning models, notably SVM, and deep learning approaches primarily based on CNNs. Their review concludes by identifying gaps in current research and suggesting avenues for future exploration, emphasizing the potential of online detection and classification systems.

Liu et al. [7] offered insights into deep learning applications for smart agriculture, particularly focusing on the integration of Unmanned Aerial Vehicles (UAVs) and Edge Intelligence. They adopted a systems-oriented approach, classifying and discussing various available drones and sensors. Their analysis revealed the advantages and limitations of each type. With respect to UAV remote sensing image processing, prevalent methodologies, specifically CNNs and RNNs, have been underscored. In addition, they delved into the potential of edge computing, providing an overview of the common hardware utilized in this domain.

Sethy et al. [11] focused their review on the application of remote sensing images in precision agriculture with a specific emphasis on hyperspectral imagery. They highlighted the capability of hyperspectral images to capture both the external physical and internal chemical properties of crops rapidly. Their analysis identified the predominant role of CNN-based methodologies in processing such images.

Qazi et al. [12] presented a holistic review of the confluence of the IoT and artificial technologies in smart agriculture. They provide a structured overview of IoT communication, control, and sensing technologies, juxtaposing them with contemporary technological solutions. In parallel, they discuss AI applications in smart agriculture and pinpoint areas such as plant disease detection. Their review underscores certain limitations of current methodologies and advocates for future research trajectories, particularly the integration of AI image recognition with edge-computing systems.

2.2. *Relevant Studies*

This section discusses extant research efforts pertinent to this topic. Given the central theme of this paper, deep learning methodologies the studies encompassed here predominantly pertain to image classification models and approaches rooted in deep learning.

2.2.1. *Artificial Neural Networks (ANNs)*

ANNs epitomize the foundational structure of neural networks and primarily leverage feed-forward connections coupled with backpropagation. With their straightforward architecture and ubiquitous applicability, when combined with conventional feature extraction techniques, ANNs can efficiently facilitate agricultural image recognition with minimal computational overheads. However, a limitation of ANNs is their performance, which tends to be subpar with that of RNNs and CNNs.

Ayalew et al. [5] proposed an agricultural scene classification scheme based on Gabor wavelet features that achieved an approximate classification accuracy of 90% using image data. Subsequently, Ayalew et al. introduced a feature selection strategy based on Linear Discriminant Analysis (LDA), which not only ameliorated the model's classification capabilities but also curtailed computational expenses.

2.2.2. *Recurrent Neural Networks (RNNs)*

RNNs, originally conceptualized for processing sequential data, encompass enhanced architectures such as LSTM and Gated Recurrent Units (GRU). These networks have also been tailored for image data processing and have demonstrated promising outcomes.

Dudi and Rajesh [2] unveiled a novel algorithm called Crow-Electric Fish Optimization (C-EFO) tailored for efficacious feature extraction. By harnessing the extracted features, they introduced an augmented RNN model dedicated to plant-leaf classification. When juxtaposed with traditional machine learning methodologies, their avant-garde approach manifested an accuracy enhancement ranging from 2% to 5%.

2.2.3. *Convolutional Neural Networks*

CNNs have become the mainstream deep learning technique for image processing, as confirmed by the aforementioned overview, and have proven to be particularly effective in agricultural image processing. A myriad of studies have focused on CNN-based agricultural image processing. This section provides an overview of the most representative studies in this domain.

Model structures and their enhancements are often a focus of research. Kanda et al. [3] introduced a plant leaf classification approach based on data generation and CNNs, where a Conditional Generative Adversarial Network (CGAN) was employed for generating new image samples, CNNs for feature extraction, and Logistic Regression for classification. This approach achieves an average accuracy of 96%. Albattah et al. [4] proposed an improved DenseNet model for plant disease detection and classification, by introducing an improved CenterNet for deep keypoint extraction.

Furthermore, there is an abundance of research centered on data augmentation and sample generation. Hu and Fang [18] addressed the problem of tea leaf disease detection and proposed a solution using multi-convolutional neural networks. To address the small-sample problem inherent in this issue, Hu and Fang suggested a SinGAN model based on GANs to generate new training samples. Nagaraju et al. [8] addressed the challenges of acquiring agricultural image data because of their high cost and scarcity. They introduced two learning algorithms to expand and enhance the dataset. The experimental results revealed that the data augmentation techniques further improved the detection performance of CNN models.

Regarding specific agricultural challenges and their corresponding solutions, Al-Gaashani et al. [17] focused on the classification of tomato leaf disease. They presented a solution that combines transfer learning and feature concatenation, which is suitable for deployment on mobile devices. Specifically, they utilized pre-trained MobileNetV2 and NASNetMobile for feature extraction, followed by kernel principal component analysis, culminating in classification using conventional learning models. The evaluation results show an impressive accuracy of up to 97%. Butera et al. [16] employed various object detection models on leaf images from different sources, and assessed their performance based on the detection accuracy and computational resources required. They found that, among the prevalent detection models, FasterRCNN based on MobileNetV3 exhibited superior performance, making it suitable for smart agriculture scenarios. Turkoglu et al. [15] introduced a CNN ensemble model called PlantDiseaseNet for plant disease and pest detection. The model incorporated six different CNNs for feature extraction, with each feature set serving as the input for the SVM models, resulting in a classification accuracy of 97.56%. Hespeler et al. [14] explored non-destructive thermal images and employed object detection models such as Mask-RCNN and YOLOv3 for robotic inspection and harvesting of chili peppers. They highlighted the superior efficiency of the YOLOv3 algorithm over Mask-RCNN, boasting a computational speed of over ten times faster. This study offers novel insights for deep learning applications in smart agriculture. Building on the MobileNet model, Chen et al. [13] designed an improved model, MobOca_Net, for potato leaf disease recognition. In MobOca_Net, an attention mechanism was integrated with the pre-trained MobileNet model, achieving a reasonable accuracy rate of 97.73%. Li et al. [10] presented a CNN-based method called the multi-class plant ensemble net (MCPE), which employed data augmentation strategies to enhance the dataset and utilized an EnsembleNet consisting of four CNNs for plant species classification. By introducing a novel activation function and training individual disease classifiers for each plant, they were able to identify disease types and severities across over 40,000 images of 10 plant species with 61 distinct diseases, achieving a recognition rate of 97.5%, surpassing existing technologies. Amin et al. [9] investigated the corn leaf disease classification problem and assessed a hybrid deep-learning model based on EfficientNetB0 and DenseNet121. Additional data augmentation techniques are introduced to diversify the training images. Their proposed approach achieved an accuracy of 98.56%. Huang et al. [11] focused on the application of deep learning in the growth stages of chili seeds, rather than on conventional disease detection. They utilized a model based on DenseNet201 and employed transfer learning, achieving a remarkable growth stage recognition accuracy of 95.5%. For soybean cultivation, Sankaran et al. [12] leveraged a Faster R-CNN-based method for weed detection and classification, enabling a more efficient herbicide application process. Their deep learning-based method proved capable of distinguishing between various weeds and soybean plants.

Although deep learning has made significant progress in agricultural image processing, numerous challenges remain [19]. For example, constraints on computational resources, and time restrictions must be considered in real-time processing applications [20]. Moreover, a wealth of research has underscored

the importance of data augmentation and multimodel fusion techniques in enhancing model performance, paving the way for future research [21].

2.2.4. Summary

Based on our analysis of the existing literature, we observe that the predominant solutions are centered on CNN-based models, whereas research on models based on RNNs remains relatively sparse. By integrating CNN with LSTM models, we can further harness the strengths of different models, achieving superior performance compared to using CNN models alone.

3. Model Description

The proposed CNN-LSTM hybrid model combines convolutional layers for feature extraction and an LSTM layer for sequence processing, efficiently addressing image classification tasks.

3.1. Convolutional Neural Network

CNNs have been naturally crafted to specifically cater to the processing of two-dimensional data structures, most notably images. Within the CNN framework, the convolutional layer plays a pivotal role, operating as a latent layer. Within this convolutional stratum, each cluster of neurons, colloquially referred to as filters, executes a convolution operation on the image data. It is imperative to note that although each neuron within a filter connects to a unique region of the image, they uniformly share identical weight parameters.

Given the potentially expansive resolution of input images, an operation known as max pooling is frequently incorporated into CNN architectures. The objective of max pooling is to downscale the spatial dimensions of the original image. This was achieved by applying a maximum filter to the non-overlapping segments of the preliminary image. The intricacies of both the convolution and max-pooling operations can be elucidated through schematic representations. This specialized network design demonstrates the efficacy of feature extraction from images, thereby exhibiting superior performance in tasks related to image recognition and categorization.

3.2. Long Short-Term Memory

A RNN is designed to handle sequential dependencies by adding previous outputs as inputs while maintaining hidden states. This makes RNNs adept at managing sequential data such as natural language or time series. To address the vanishing gradient problem, LSTM units were introduced. LSTMs are variants of RNNs designed specifically to mitigate the vanishing gradient issue. The LSTM architecture consists of a series of cells, each containing three essential gating mechanisms: a Forget Gate, Input Gate, Output Gate.

Forget Gate: Determines which information should be discarded or retained from the cell state. It is defined by the equation

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f)$$

Input Gate: Decides which new information will be stored in the cell state. This gate can be described by the following set of equations.

$$\begin{aligned} i_t &= \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \\ \tilde{c}_t &= \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \\ c_t &= f_t \times c_{t-1} + i_t \times \tilde{c}_t \end{aligned}$$

where denotes a matrix of new candidate values that can be added to the cell state.

Output Gate: Determines the information output from the cell state. It can be described by:

$$\begin{aligned} o_t &= \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \\ h_t &= o_t \times \tanh(c_t) \end{aligned}$$

LSTM units are designed to remember information over long periods, making them particularly suitable for tasks that require the model to learn from long sequences of data. The gating mechanisms

ensure that the network learns which information to store, which to update, and which to discard, thereby mitigating the vanishing gradient problem faced by traditional RNNs.

3.3. CNN-LSTM Hybrid Model

The CNN-LSTM hybrid model structure begins with an input layer whose dimensions are defined by the specific application at hand. For instance, this could entail the height, width, and color channels of an image or the number of frames and dimensions of each frame in a video sequence. In the context of image processing, the input layer is responsible for conveying the raw pixel values of the image data.

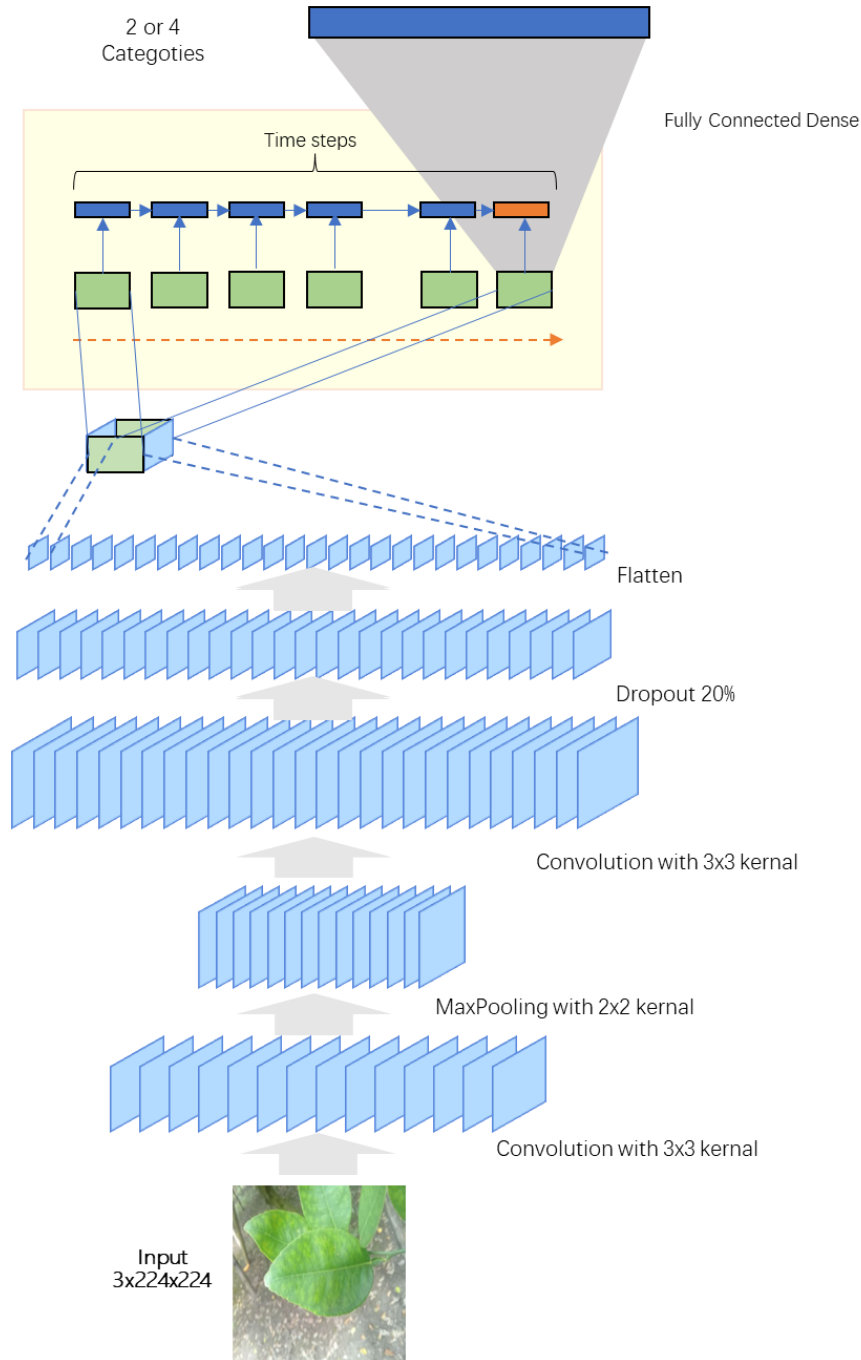


Figure 1. The CNN-LSTM hybrid model structure.

Following the input layer, the first principal component of the model was a series of convolutional layers. Within the specified structure, there were two convolutional layers, each utilizing 32 filters with a kernel size of 3×3 . These filters slide across the input image to learn features ranging from the edges and corners to more complex patterns. Each convolutional layer is typically succeeded by a nonlinear activation function, such as ReLU, to introduce nonlinearity, enabling the network to learn more intricate patterns.

Subsequent to the convolutional layers is a max-pooling layer utilizing a 2×2 window to reduce the spatial dimensions of the feature maps. This max-pooling operation aids in decreasing the number of parameters, prevents overfitting, and makes the feature detection invariant to minor positional changes.

After the second convolutional layer, the model incorporated a dropout layer set at a rate of 20%. The Dropout layer randomly zeroes a portion of the network connections during training to reduce overfitting and to enhance the generalization capability of the model. Following the Dropout layer, a flattened layer converts the two-dimensional feature maps into a one-dimensional feature vector, which is suitable for processing by the LSTM layer.

The transformed one-dimensional feature vector is then fed into a LSTM layer containing 128 neurons. The LSTM layer is capable of processing sequential data and is adept at remembering and utilizing temporal information, which is crucial for tasks involving a time dimension.

Finally, the model culminates in a fully connected output layer. Depending on the application requirements, the output layer can be configured with either two or four neurons. For instance, in a binary classification task, the output layer may consist of only two neurons corresponding to the probabilities of the two classes. In a quadruple classification scenario, there are four neurons, each representing the probability of one class. The output layer typically employs a softmax activation function to provide a probability distribution, allowing the network output to be interpreted directly as the likelihood of each category.

In summary, the convolutional layers were tasked with extracting useful spatial features from the input data, while the LSTM layer was designed to handle the dynamics of these features over time. This architecture is particularly suited to tasks that encompass both spatial and temporal features, such as video processing and sequential image recognition.

4. Experiments and Discussion

4.1. Dataset Description

The dataset referred to in this paper is the LeLePhid dataset, which is a collection of 665 photographs that capture the top and back of lemon tree leaves. This dataset is particularly focused on the detection of leaf damage caused by aphids, a significant pest that causes damage to crops worldwide by leaving visible spots on the leaves. Such a characteristic feature allows for the potential application of the dataset not only for lemon trees, but also for other crops affected by aphids.

The LeLePhid dataset was balanced in terms of its composition, containing an equal division between healthy leaves and those infested by aphids, with 335 images for each category. The data collection process was meticulous and carried out manually in citrus crops under varying weather conditions conducive to crop growth, that is, warm and rainy periods. Notably, the capture of data did not occur in a controlled environment but rather under natural conditions, including cloudy, sunny, rainy, and windy weather, introducing a variety of uncontrolled variables into the dataset.

The photographs within the dataset show high color similarity and are characterized by the presence of noise and variation. These variations are manifested in terms of the position, chromaticity, structure, and size of lemon leaves. This complexity ensures that the dataset presents a realistic challenge for the CNN-LSTM hybrid model, with the aim of robustly detecting and classifying the presence and severity of aphid infestation despite noise and variations in the dataset.



(a) infested image sample



(b) healthy image sample

Figure 2. Data examples.

4.2. Settings

a) Programming Language & Hardware Configuration

The computational experiments in this study were executed using Python as the programming language of choice, leveraging robust libraries such as TensorFlow and Keras for the implementation of deep learning models. The hardware configuration on which the models were trained and tested consisted of a desktop running Windows 10 equipped with an Intel Core i5-9600K CPU operating at 3.70 GHz. The system also included a GPU, specifically an NVIDIA GeForce RTX 2070, which provided the necessary computational power for handling intensive tasks associated with deep learning. In addition, the machine was provisioned with 32 GB of RAM to ensure efficient data-handling and processing capabilities.

To achieve a thorough training process, each model underwent a training period spanning 30 epochs, with a batch size of 8. This setup was chosen to balance the computational load and granularity of the model updates, thereby facilitating a conducive learning environment for the neural network to learn accurately from the dataset.

b) Evaluation metrics

In the methodology section of the research paper, we describe the evaluation metrics used to assess the performance of the proposed model. The primary metrics include accuracy, F1 score, and confusion matrix, which together offer a comprehensive view of the model's predictive capabilities.

The confusion matrix is a fundamental tool in machine learning classification, providing insights into the true positive (TP), false positive (FP), false negative (FN), and true negative (TN) predictions made by the classifier. This matrix forms the basis for computing the various performance metrics.

Accuracy is the most intuitive performance measure, and is simply the ratio of correctly predicted observations to the total number of observations. This is formulated as follows:

$$accuracy = \frac{TP+TN}{TP+FN+FP+TN}$$

The F1 score is the harmonic mean of the precision and recall, providing a balance between the two in cases where one may be more important than the other. It is defined as:

$$F1 = 2 \times \frac{precision \times recall}{precision + recall}$$

where precision and recall are given by:

$$precision = \frac{TP}{TP + FP}$$

$$recall = \frac{TP}{TP + FN}$$

The choice to use these particular metrics is justified by the nature of the dataset and the binary classification task at hand, allowing us to effectively measure both the accuracy of the model and its

ability to manage the trade-off between precision and recall, which is particularly important in medical and agricultural diagnostic applications, where the costs of false negatives and false positives can be significant.

c) Baselines – MobileNet & DenseNet

MobileNet stands out against traditional deep learning models because of its unique convolution approach. Instead of using standard convolutions, it employs depth-wise separable convolutions that separate the filtering and combination steps. This results in a model with fewer parameters and computational steps, enabling faster operation with less impact on the performance. Its architecture is tailored for environments where computational resources are limited, such as mobile devices or embedded systems, allowing for the deployment of complex models that previously could not run effectively.

DenseNet, on the other hand, excels in feature utilization and conservation. By connecting every layer to every other layer, DenseNet ensures that the features extracted at every level are used to their fullest extent. Moreover, because each layer has access to all preceding layers' feature-maps, it leads to improved feature propagation and helps in reducing the vanishing gradient problem. This makes DenseNet particularly effective in tasks that benefit from deep feature exploration, like detailed image classification, without the excessive computational burden usually associated with deeper networks.

4.3. Results

The proposed CNN-LSTM hybrid model is compared to standalone CNN models, specifically MobileNet and DenseNet. Performance metrics such as the number of parameters, accuracy, and F1 scores are discussed, providing a quantitative basis for model comparison.

The complexity of a model is often indicated by its number of parameters. Table 1 illustrates the parameter count for each model. The CNN-LSTM hybrid model has significantly fewer parameters than DenseNet, suggesting a more efficient model that could be advantageous in resource-constrained environments.

Table 1. Model parameters.

Model	Parameters
MobileNet	18,315,074
DenseNet	42,407,234
CNN-LSTM Hybrid	8,872,692

As shown in Table 2, the proposed CNN-LSTM hybrid model achieved the highest accuracy and F1 scores for leaf disease detection. Similarly, the proposed CNN-LSTM hybrid model achieved the highest accuracy and F1 scores for leaf disease classification, as shown in Figure 3.

Table 2. The accuracy and F1 scores for leaf disease detection.

Model	Accuracy	F1 Score
MobileNet	0.8241	0.8240
DenseNet	0.8948	0.8946
CNN-LSTM Hybrid	0.9714	0.9713

Table 3. The accuracy and F1 scores for leaf disease classification.

Model	Accuracy	F1 Score
MobileNet	0.8195	0.8137
DenseNet	0.8932	0.8921
CNN-LSTM Hybrid	0.9549	0.9547

The confusion matrices for leaf disease detection and classification of the proposed CNN-LSTM hybrid model are further shown in Figure 3.

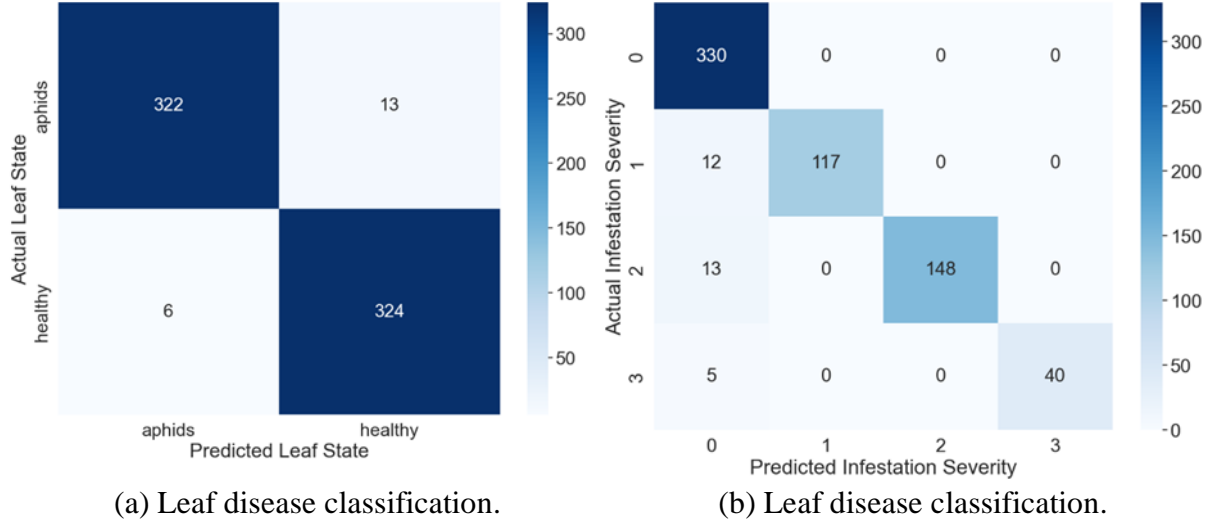


Figure 3. Confusion matrices of the proposed CNN-LSTM hybrid model.

In evaluating the impact of model parameters on the performance of the CNN-LSTM hybrid model, we consider the number of layers and neurons within the network. Here, we present a brief discussion on how varying these parameters may influence the model's effectiveness in detecting and classifying leaf diseases.

(a) CNN Layer Number

The number of convolutional layers in a CNN affects the model's ability to capture hierarchical features, as shown in Figure 4.

- 1 Layer: May only capture basic features and might be insufficient for complex image classification tasks.
- 2 Layers: Can extract more sophisticated features by building on the initial layer's outputs.
- 3 Layers: Allows for an even deeper hierarchy of features, potentially improving classification accuracy for more complex patterns.

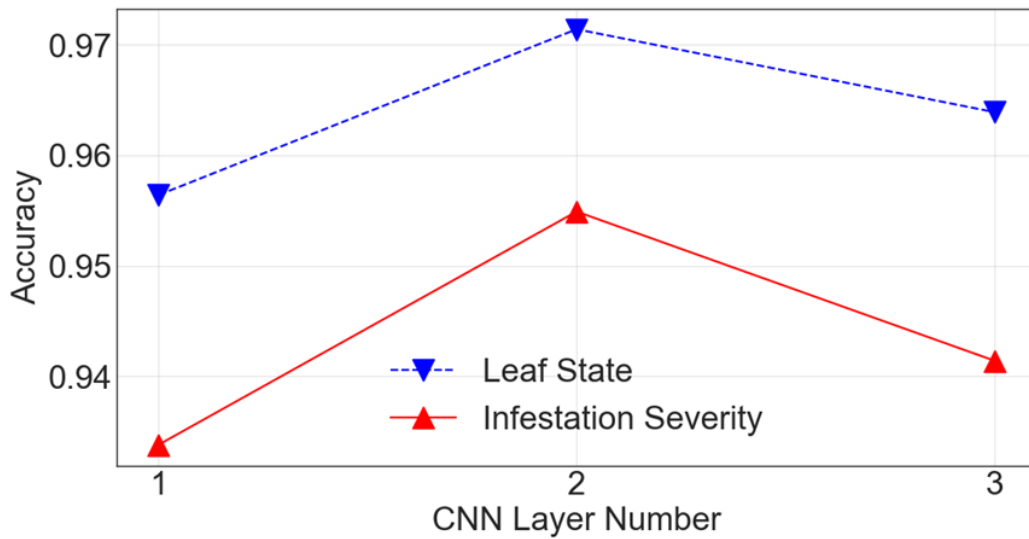


Figure 4. The influence of the CNN layer number.

(b) LSTM Layer Number

The number of LSTM layers impacts the model's temporal feature extraction capabilities, as shown in Figure 5.

- 1 Layer: Suitable for capturing short-term dependencies in sequence data.
- 2 Layers: Improves the model's ability to understand longer-term dependencies, which may be necessary for more complex sequences.
- 3 Layers: Enables the model to learn even longer and more abstract temporal patterns, although it may increase the risk of overfitting and require more data.

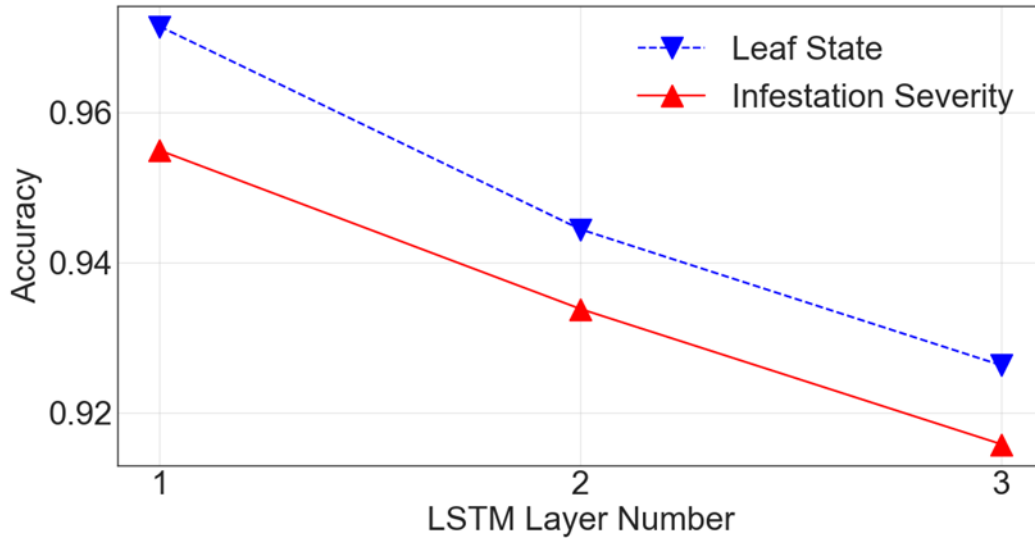


Figure 5. The influence of the LSTM layer number.

(c) CNN Filter Number

The number of filters in each convolutional layer determines the breadth of features the model can detect, as shown in Figure 6.

- 16 Filters: May be sufficient for simple or less varied images.
- 32 Filters: Provides a broader set of features that can be useful for more varied data.
- 64 Filters: Captures a wide variety of features, which is beneficial for complex images but increases computational complexity.

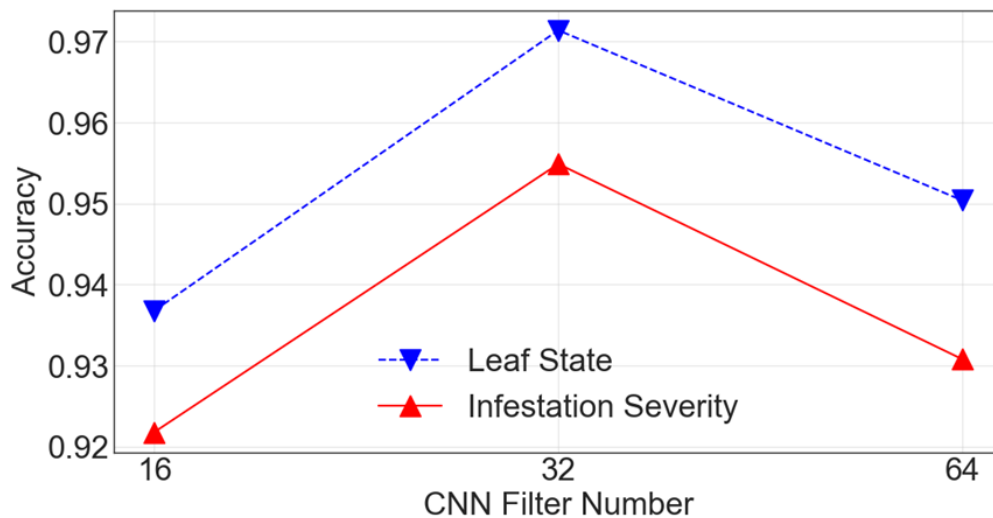


Figure 6. The influence of the CNN filter number.

(d) LSTM Neuron Number

The number of neurons in LSTM layers affects the model's capacity to process and remember information, as shown in Figure 7.

- 64 Neurons: May limit the model's memory capability for complex patterns.
- 128 Neurons: A balanced choice that offers improved memory while keeping computational demands reasonable.
- 256 Neurons: Maximizes the model's memory capability, allowing it to handle very complex sequences at the cost of increased computation.

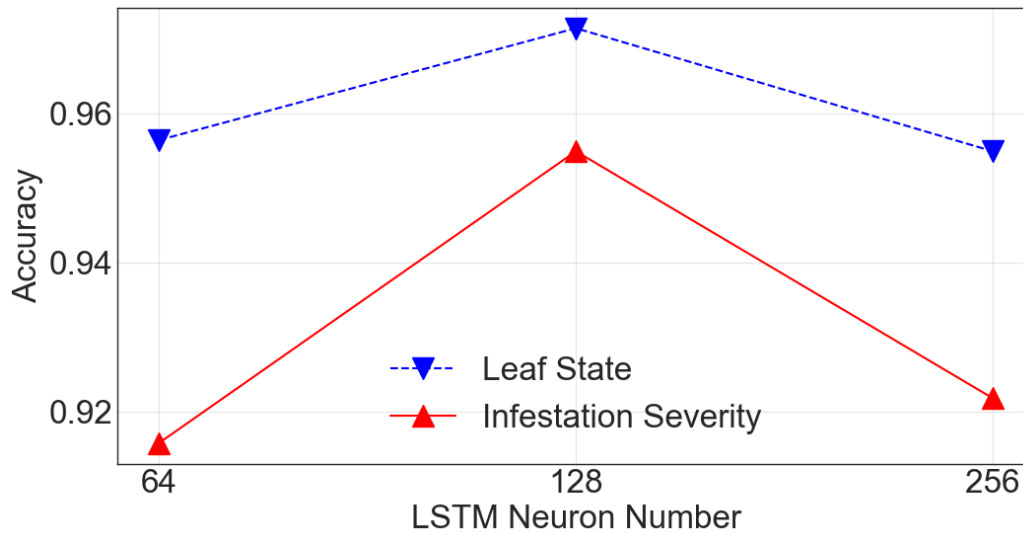


Figure 7. The influence of the LSTM neuron number.

The CNN-LSTM hybrid model's performance suggests that the integration of temporal sequence processing via LSTM layers with spatial feature extraction via CNN layers results in a more powerful model for image classification tasks. The results demonstrate the potential of this hybrid approach in accurately identifying and classifying leaf diseases, which could significantly impact precision agriculture practices.

Careful attention should be paid to the model's limitations, including potential biases in the dataset and the generalizability of the model to different crops or disease types. Future work may explore the application of this model to a broader range of agricultural challenges and the integration of additional data sources for enhanced model robustness.

5. Conclusion

This research has presented a comprehensive study on the application of CNN and LSTM networks for the detection and classification of leaf diseases. The developed CNN-LSTM hybrid model demonstrates a significant improvement in accuracy and F1 scores over traditional CNN models such as MobileNet and DenseNet, thereby highlighting the effectiveness of combining spatial and temporal feature extraction techniques for image classification tasks.

The evaluation of model parameters, such as the number of CNN and LSTM layers, number of filters, and number of LSTM neurons, indicates that a balanced approach to model complexity can yield high accuracy while maintaining computational efficiency. The hybrid model, with its fewer parameters compared to the more complex DenseNet, affords a more practical solution in resource-constrained environments, potentially facilitating real-time analysis and decision-making in precision agriculture.

The confusion matrices derived from the model's performance offer valuable insights into its predictive capabilities, ensuring that the model not only achieves high overall accuracy but also

maintains a low rate of false positives and negatives, which is crucial in the agricultural context where misdiagnoses can lead to significant economic losses.

Future work may focus on further optimization of the model architecture, expanding the dataset to include a more diverse range of leaf diseases, and exploring the model's generalizability to other crops [22, 23]. Additionally, integrating the model into a real-world agricultural decision support system would be a logical next step towards operationalizing the benefits of this research.

References

- [1] Bhagat, M., & Kumar, D. (2022). A comprehensive survey on leaf disease identification & classification. *Multimedia Tools and Applications*, 81(23), 33897-33925.
- [2] Dudi, B., & Rajesh, V. (2023). A computer aided plant leaf classification based on optimal feature selection and enhanced recurrent neural network. *Journal of Experimental & Theoretical Artificial Intelligence*, 35(7), 1001-1035.
- [3] Kanda, P. S., Xia, K., & Sanusi, O. H. (2021). A deep learning-based recognition technique for plant leaf classification. *IEEE Access*, 9, 162590-162613.
- [4] Albattah, W., Nawaz, M., Javed, A., Masood, M., & Albahli, S. (2022). A novel deep learning method for detection and classification of plant diseases. *Complex & Intelligent Systems*, 1-18.
- [5] Ayalew, G., Zaman, Q. U., Schumann, A. W., Percival, D. C., & Chang, Y. K. (2021). An investigation into the potential of Gabor wavelet features for scene classification in wild blueberry fields. *Artificial Intelligence in Agriculture*, 5, 72-81.
- [6] Huang, M. L., Chuang, T. C., & Liao, Y. C. (2022). Application of transfer learning and image augmentation technology for tomato pest identification. *Sustainable Computing: Informatics and Systems*, 33, 100646.
- [7] Liu, J., Xiang, J., Jin, Y., Liu, R., Yan, J., & Wang, L. (2021). Boost precision agriculture with unmanned aerial vehicle remote sensing and edge intelligence: A survey. *Remote Sensing*, 13(21), 4387.
- [8] Nagaraju, M., Chawla, P., Upadhyay, S., & Tiwari, R. (2022). Convolution network model based leaf disease detection using augmentation techniques. *Expert Systems*, 39(4), e12885.
- [9] Amin, H., Darwish, A., Hassanien, A. E., & Soliman, M. (2022). End-to-end deep learning model for corn leaf disease classification. *IEEE Access*, 10, 31103-31115.
- [10] Li, B., Tang, J., Zhang, Y., & Xie, X. (2022). Ensemble of the deep convolutional network for multiclass of plant disease classification using leaf images. *International Journal of Pattern Recognition and Artificial Intelligence*, 36(04), 2250016.
- [11] Sethy, P. K., Pandey, C., Sahu, Y. K., & Behera, S. K. (2022). Hyperspectral imagery applications for precision agriculture-a systemic survey. *Multimedia Tools and Applications*, 1-34.
- [12] Qazi, S., Khawaja, B. A., & Farooq, Q. U. (2022). IoT-equipped and AI-enabled next generation smart agriculture: A critical review, current challenges and future trends. *IEEE Access*, 10, 21219-21235.
- [13] Chen, W., Chen, J., Zeb, A., Yang, S., & Zhang, D. (2022). Mobile convolution neural network for the recognition of potato leaf disease images. *Multimedia Tools and Applications*, 81(15), 20797-20816.
- [14] Hespeler, S. C., Nemati, H., & Dehghan-Niri, E. (2021). Non-destructive thermal imaging for object detection via advanced deep learning for robotic inspection and harvesting of chili peppers. *Artificial Intelligence in Agriculture*, 5, 102-117.
- [15] Turkoglu, M., Yanikoğlu, B., & Hanbay, D. (2022). PlantDiseaseNet: Convolutional neural network ensemble for plant disease and pest detection. *Signal, Image and Video Processing*, 16(2), 301-309.
- [16] Butera, L., Ferrante, A., Jermini, M., Prevostini, M., & Alippi, C. (2021). Precise agriculture: Effective deep learning strategies to detect pest insects. *IEEE/CAA Journal of Automatica Sinica*, 9(2), 246-258.

- [17] Al-gaashani, M. S., Shang, F., Muthanna, M. S., Khayyat, M., & Abd El-Latif, A. A. (2022). Tomato leaf disease classification by exploiting transfer learning and feature concatenation. *IET Image Processing*, 16(3), 913-925.
- [18] Hu, G., & Fang, M. (2022). Using a multi-convolutional neural network to automatically identify small-sample tea leaf diseases. *Sustainable Computing: Informatics and Systems*, 35, 100696.
- [19] Jiang, W., Han, H., Zhang, Y., & Mu, J. (2023). Federated split learning for sequential data in satellite-terrestrial integrated networks. *Information Fusion*, 102141.
- [20] Jiang, W., Han, H., He, M., & Gu, W. (2023). ML-based Pre-deployment SDN Performance Prediction with Neural Network Boosting Regression. *Expert Systems with Applications*, 2023.
- [21] Liu, J., Jiang, W., Han, H., He, M., & Gu, W. (2023, September). Satellite Internet of Things for Smart Agriculture Applications: A Case Study of Computer Vision. In *2023 20th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)* (pp. 66-71). IEEE.
- [22] Zheng, Y., & Jiang, W. (2022). Evaluation of vision transformers for traffic sign classification. *Wireless Communications and Mobile Computing*, 2022.
- [23] Jiang, W., & Luo, J. (2022). An evaluation of machine learning and deep learning models for drought prediction using weather data. *Journal of Intelligent & Fuzzy Systems*, 43(3), 3611-3626.