Tomato Disease Recognition: Advancing Accuracy Through Xception and Bilinear Pooling Fusion

Hoang-Tu Vo, Nhon Nguyen Thien, Kheo Chau Mui
Software Engineering Department, FPT University, Cantho City, Vietnam

Abstract—Accurate detection and classification of tomato diseases are essential for effective disease management and maintaining agricultural productivity. This paper presents a novel approach to tomato disease recognition that combines Xception, a pre-trained convolutional neural network (CNN), with bilinear pooling to advance accuracy. The proposed model consists of two parallel Xception-based CNNs that independently process input tomato images. Bilinear pooling is applied to combine the feature maps generated by the two CNNs, capturing intricate interactions between different image regions. This fusion of Xception and bilinear pooling results in a comprehensive representation of tomato diseases, leading to improved recognition performance. Extensive experiments were conducted on a diverse dataset of annotated tomato disease images to evaluate the effectiveness of the suggested approach. The model achieved a remarkable test accuracy of 98.7%, surpassing conventional CNN approaches. This high accuracy demonstrates the efficacy of the integrated Xception and bilinear pooling model in accurately identifying and classifying tomato diseases. The implications of this research are significant for automated tomato disease recognition systems, enabling timely and precise disease diagnosis. The model’s exceptional accuracy empowers farmers and agricultural practitioners to implement targeted disease management strategies, minimizing crop losses and optimizing yields.

Keywords—Tomato disease recognition; Xception; Bilinear pooling; convolutional neural networks; disease management

I. INTRODUCTION

Tomato plants are a vital crop worldwide, serving as a staple in numerous cuisines and contributing to global food security. Besides, regular consumption of tomatoes can contribute to improved health and a decreased susceptibility to various ailments, including cancer, osteoporosis, and cardiovascular disease. Those who incorporate tomatoes into their diet on a consistent basis are found to have a lower likelihood of developing cancer, encompassing lung, prostate, stomach, cervical, breast, oral, colorectal, esophageal, pancreatic, and several other forms of cancer [1]. However, tomato harvests are at risk of a variety of diseases that can cause significant yield losses and quality deterioration. Early and reliable identification of such diseases is necessary for implementing timely disease management strategies and minimizing agricultural losses. Tomato diseases are a significant concern for tomato growers worldwide, including Vietnam [2], [3], [4].

Techniques derived from Machine Learning (ML) and Deep Learning (DL) have been extensively utilized in image recognition across diverse domains, including agriculture [5], [6], medicine [7], [8], self-driving cars [9], [10], etc. Numerous studies have explored the application of these technologies to achieve accurate image recognition in these respective fields.

The purpose of this paper is to address the crucial need for accurate classification of tomato diseases in order to effectively manage them and maintain agricultural productivity. The study presents a novel approach to tomato disease recognition by combining Xception, a pre-trained convolutional neural network (CNN), with bilinear pooling to enhance accuracy. Extensive experiments were conducted on a diverse dataset of annotated tomato disease images to evaluate the effectiveness of the suggested approach. The model obtained an impressive test accuracy of 98.7%.

The paper is structured as follows: Section II offers an extensive literature review, presenting pertinent background information. Section III details the methodology utilized for Tomato Disease Recognition, including the Data Set, Data Preparation, and Model Evaluation Metrics. Section IV describes the experimental system and final results. Lastly, Section V concludes the study by summarizing the findings and providing concluding remarks. Section VI gives future directions of research.

II. RELATED WORKS

In recent years, improvements in DL techniques, particularly convolutional neural networks (CNNs) and transfer learning models (TL), have shown promising results in automating disease recognition tasks in agriculture. Zahid Ullah et al. [11] presents a hybrid deep learning approach, EffiMobNet, combining EfficientNetB3 and MobileNet models with techniques to handle overfitting, achieving a 99.92% accuracy in accurately detecting tomato leaf diseases. This study [12] utilizes pre-trained CNNs, specifically Inception V3 and Inception ResNet V2, to classify healthy and unhealthy tomato leaf images, achieving high accuracy (99.22%) and low loss (0.03%) with dropout rates of 50% and 15%, respectively. Sachin Kumar et al. [13] utilized a dataset of 6,594 tomato leaves, including six disease classes and one healthy class, from Plant Village, and achieved a significant accuracy of 96.35% using the ResNet-50 model. The authors in [14] suggest a method for classifying tomato leaf diseases employing transfer learning and feature concatenation by leveraging pre-trained kernel from MobileNetV2 and NASNetMobile models. They extract features from these models, concatenate them, and then reduce the dimensionality using kernel principal component analysis. The effectiveness of the concatenated features is confirmed through experimental results, with multinomial logistic regression achieving the best performance among the evaluated traditional machine learning classifiers, achieving an average accuracy of 97%.

The authors in the article [15] focus on developing a combined model for identifying tomato diseases utilizing image recognition across diverse domains, including agriculture [5], [6], medicine [7], [8], self-driving cars [9], [10], etc. Numerous studies have explored the application of these technologies to achieve accurate image recognition in these respective fields.
data. Seven architectures, including VGG16, ResNet50, and various EfficientNet models, are evaluated for performance using transfer learning. The best-performing models are then combined using a weighted average ensemble, resulting in a suggested model with an accuracy of 98.1%. Nagamani H S et al. in the study [16] explores the identification of diseases affecting tomato leaves using ML techniques, including Fuzzy-SVM, CNN, and Region-based Convolutional Neural Network (R-CNN). Various image processing and feature extraction methods are employed, and R-CNN achieves the greatest accuracy of 96.735% in classifying different disease types. In this study [17], a DL model combining CNN and SVM is deployed to recognize and categorize tomato leaf images into 8 classes, including 7 prevalent diseases and a healthy class. The model is trained on a dataset of 8,000 photos and achieves an accuracy of 92.6% by utilizing CNN for feature extraction and SVM for classification.

Sanjeela Sagar et al. in this paper [18] presents an experimental study comparing traditional ML algorithms (RF, SVM, NB) with a deep learning CNN algorithm in order to classify tomato leaf disease. The results show that the CNN approach outperforms traditional methods, achieving over 95% accuracy in detection and classification. The authors in the article in this study [19] focus on using CNN methods, specifically the VGG model, for the detection of Multi-Crops Leaf Disease (MCLD). The trained model successfully classifies disease-affected leaves with high accuracy, achieving 98.40% accuracy for grapes and 95.71% accuracy for tomatoes. In this paper [20], the authors conduct a thorough evaluation of deep-learning approaches utilizing pre-trained CNN models and the PyTorch framework for classifying instances of diseases affecting tomato plants. Several models, including EfficientNet-B0, ResNext-50-32x4d, and MobileNet-V2, were tested, and ResNet-50-32x4d achieved the highest accuracy of 90.14%.

The paper in [21] presents an approach for classifying seven different types of tomato illnesses using DL models trained on a dataset of 10,448 images. The trained models demonstrated high accuracy, with the best testing precision reaching 95.71%. Sakkavarthi, Gnanavel, et al. in the article [22] propose a deep-learning-based agricultural disease detection technique, employing a CNN approach in order to detect and classify diseases. The model, consisting of a pair of convolutional and pooling layers, exceeded the performance of the pre-trained InceptionV3, ResNet 152, and VGG19 models, achieving 98% training accuracy and 88.17% testing accuracy. In this paper [23], Singh, Rahul, et al. utilizes transfer learning with the EfficientNetB3 model for leaf classification, using a dataset of 11 different leaf types collected from an internet database. With a batch size of 32, the model is trained for 15 iterations and evaluated using the Adam optimizer, achieving an accuracy of 0.94. Sultana, Irene, et al. [24] present a dataset of 14,529 tomato leaf images containing ten different infections. InceptionV3 and ResNet-50 serve as employed learning algorithms, leveraging transfer learning techniques to train a classifier. The proposed deep learning model achieves promising results with an 85.52% accuracy rate for InceptionV3 and 95.41% for ResNet-50. The study in [25] aims for the purpose of identifying the presence of early blight infestation affecting tomato plants using a CNN approach. Various image processing techniques are applied to refine the dataset, and the CNN model is trained and evaluated using different performance metrics, achieving a high accuracy of 98.10% with specific hyperparameters. The article [26] presents a CNN model that combines elements of different approaches for classifying diseases in tomato leaf images, utilizing well-known CNN architectures and feature transfer techniques. The suggested approach obtains high accuracy rates of 98.3% and 96.3% for both the dataset specifically designed for detecting tomato leaf diseases and the dataset collected in Taiwan, respectively.

The purpose of this study is to introduce an innovative method for tomato disease recognition by combining Xception, a pre-trained CNN, with bilinear pooling to enhance accuracy. The proposed model incorporates two parallel Xception-based CNNs that independently process tomato images and utilizes bilinear pooling to capture complex interactions between image regions. This fusion of Xception and bilinear pooling yields a comprehensive representation of tomato diseases, resulting in improved recognition performance.

III. METHODOLOGY

A. Data Collection and Preparation

In this research, a dataset comprising 32,535 images was used, obtained from the PlantVillage dataset [27] and Kaggle. The Tomato Disease dataset consists of 10 diseases and 1 healthy class, including Late blight (Class1), healthy (Class2), Early blight (Class3), Septoria leaf spot (Class4), Tomato Yellow Leaf Curl Virus (Class5), Bacterial spot (Class6), Target Spot (Class7), Tomato mosaic virus (Class8), Leaf Mold (Class9), Spider mites Two-spotted spider mite (Class10), and Powdery Mildew (Class11).

The examples of tomato disease pictures from the dataset are displayed in Fig. 1, and Fig. 2 illustrates the distribution of the dataset. Before training and evaluating the model, the images are preprocessed by resizing them to 224x224 and applying an image preprocessing function. The dataset is then split into 25,851 photos for the training dataset, 4,010 photos for the validation dataset, and 2,674 photos for the test dataset.

B. Proposed Model

This paper introduces a novel approach for accurate classification of tomato diseases using a model that combines Xception [28], a pre-trained CNN, with bilinear pooling. The proposed model consists of two parallel Xception-based CNNs that independently process tomato images, with bilinear pooling capturing intricate interactions between different regions of the images. The model aims to accurately classify images into 11 different classes.

The proposed model is based on the Xception architecture. It takes an input image of size 224x224x3. The model consists of two parallel Xception layers that process the input image independently, resulting in two sets of feature maps with dimensions 7x7x2048. These feature maps are then combined by the so-called bilinear pooling layer (by taking their outer product). An average pooling layer is applied to reduce the spatial dimensions to 1x1 while preserving the depth of 2048. The output is flattened to a 1D vector of size 2048.

Next, a batch normalization layer is used, followed by a 256-unit dense layer. A dropout layer is introduced to prevent
Fig. 1. Sample tomato disease in tomato disease dataset.

Fig. 2. A dataset distribution.
overfitting, and another dense layer with 11 units (matching the number of classes) is added as the final output layer. Fig. 3 shows the suggested model.

In the suggested model, a learning rate (LR) scheduling strategy [29] is integrated into the training phase. LR scheduling is a dynamic technique that adapts the LR, determining the size of the steps taken in the gradient descent optimization algorithm, during the model's training process. The objective of this approach is to improve the model's convergence by gradually decreasing the LR as training progresses.

In this specific model, the LR is reduced by a factor of 0.5 after two epochs, allowing for adjustments if training accuracy does not show improvement. To mitigate the risk of overfitting, the architecture of the model utilizes regularization techniques, which aid in avoiding overfitting and improving the overall performance of the model. These measures are crucial for real-world applications where optimal performance is of utmost importance.

The proposed model exhibits a substantial parameter count, with a total of 42,258,523 parameters. Among these, 42,145,371 parameters are trainable, meaning they are optimized and adjusted during the training process to enhance the model's performance. Additionally, there are 113,152 non-trainable parameters, which consist of fixed or pre-defined values that remain unchanged during training.

The suggested model and its underlying architecture can be visualized in Fig. 4, while Fig. 5 provides a comprehensive representation of important details. This includes information about the model's layers, the output shape, the trainable parameter count, and the overall number of trainable parameters that are examined for each layer in the model.

C. Performance Evaluation Measures

The performance of the proposed integrated Xception and bilinear pooling model was comprehensively evaluated using a range of essential metrics. Precision, which quantifies the ratio of correctly predicted positive instances to the total...
predicted positive instances, reflected the model’s ability to minimize false positives in classifying tomato diseases. Recall, capturing the ratio of correctly predicted positive instances to the actual total positive instances, demonstrated the model’s proficiency in identifying all relevant disease cases. The F1-score, a harmonic mean of precision and recall, provided a balanced assessment of the model’s precision-recall trade-off. Accuracy, a fundamental measure in classification tasks, gauges the proportion of correctly predicted instances out of the total instances in the dataset. In the context of tomato disease recognition, accuracy indicates the model’s overall correctness in identifying and classifying different disease types from input tomato images.

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}
\]

\[
\text{Precision} = \frac{TP}{TP + FP} \tag{2}
\]

\[
\text{Recall} = \frac{TP}{TP + FN} \tag{3}
\]

\[
F_1 - \text{Score} = \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}
\]

In which, TP represents True Positive, TN signifies True Negative, FP represents False Positive, and FN stands for False Negative.

IV. RESULTS

A. Environmental Settings

The experimental results were obtained by conducting the experiments on the Kaggle platform. The system used for the experiments had 13GB of RAM and a GPU Tesla P100-PCIE with 16GB of memory. The training of the model spanned across 30 epochs, and a batch size of 32 was used during the training process.

B. Evaluation Overall

The confusion matrix, showcasing the results of the proposed model, is presented in Fig. 6 and Fig. 7. Additionally, Fig. 8 and Fig. 9 illustrate the performance metrics, such as loss and accuracy, that were evaluated during both the model’s training and validation stages. The model achieves its highest accuracy at the 17th epoch and exhibits the lowest loss at the 28th epoch.

The classification report Table I provides a detailed analysis of the evaluation metrics for each tomato disease class. It includes various metrics like precision, recall, and F1-score, which assess the model’s accuracy in identifying specific diseases.

Table II provides a comprehensive analysis of the suggested model in comparison to other state-of-the-art methods tackling similar problems. The results indicate that the proposed model
TABLE I. CLASSIFICATION REPORT

<table>
<thead>
<tr>
<th>Class</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-Score</th>
<th>Support</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacterial spot</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
<td>293</td>
</tr>
<tr>
<td>Early_blight</td>
<td>0.97</td>
<td>0.95</td>
<td>0.96</td>
<td>257</td>
</tr>
<tr>
<td>Late_blight</td>
<td>0.99</td>
<td>0.98</td>
<td>0.99</td>
<td>296</td>
</tr>
<tr>
<td>Leaf_Mold</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
<td>296</td>
</tr>
<tr>
<td>Septoria_leaf_spot</td>
<td>0.97</td>
<td>0.99</td>
<td>0.98</td>
<td>298</td>
</tr>
<tr>
<td>Spider_mite_Two-spotted_mite</td>
<td>1.00</td>
<td>0.99</td>
<td>1.00</td>
<td>174</td>
</tr>
<tr>
<td>Target_Spot</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
<td>183</td>
</tr>
<tr>
<td>Tomato_Yellow_Leaf_Curl_Virus</td>
<td>1.00</td>
<td>0.99</td>
<td>1.00</td>
<td>199</td>
</tr>
<tr>
<td>Tomato_mosaic_virus</td>
<td>0.99</td>
<td>0.98</td>
<td>0.98</td>
<td>234</td>
</tr>
<tr>
<td>healthy</td>
<td>0.98</td>
<td>1.00</td>
<td>0.99</td>
<td>322</td>
</tr>
<tr>
<td>powdery_mildew</td>
<td>1.00</td>
<td>0.99</td>
<td>1.00</td>
<td>101</td>
</tr>
</tbody>
</table>

TABLE II. A COMPARISON OF OUR PROPOSED MODEL WITH CONTEMPORARY APPROACHES ON SIMILAR PROBLEMS

<table>
<thead>
<tr>
<th>The study</th>
<th>Dataset</th>
<th>Method of Use</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PlantVillage dataset</td>
<td>CNN model</td>
<td>95.71%</td>
</tr>
<tr>
<td>[21]</td>
<td>PlantVillage dataset</td>
<td>MobileNetV2 and NASNetMobile</td>
<td>97%</td>
</tr>
<tr>
<td>[14]</td>
<td>PlantVillage dataset</td>
<td>ResNet-50</td>
<td>96.35%</td>
</tr>
<tr>
<td>[13]</td>
<td>PlantVillage dataset</td>
<td>Wavelet-like Auto-Encoder (WAE)</td>
<td>98.1%</td>
</tr>
<tr>
<td>[19]</td>
<td>PlantVillage dataset</td>
<td>VGG16</td>
<td>95.71%</td>
</tr>
<tr>
<td>[18]</td>
<td>PlantVillage dataset</td>
<td>Inception v3</td>
<td>95%</td>
</tr>
<tr>
<td>[22]</td>
<td>PlantVillage dataset</td>
<td>CNN model</td>
<td>88.17%</td>
</tr>
<tr>
<td>This study</td>
<td>PlantVillage dataset</td>
<td>Xception and Bilinear Pooling</td>
<td>99.7%</td>
</tr>
</tbody>
</table>

achieved superior performance surpassing all other techniques mentioned in the table.

The recall values reflect the ability of the model to correctly identify positive instances for each class. The range of recall values varies from 0.95 to 1.00, indicating a high level of accuracy in capturing true positive instances.

The F1-scores, which represent the precision-recall harmonic mean, provide an overall measure of the model’s performance for each class. The F1-scores range from 0.96 to 1.00, indicating a strong balance between precision and recall.

This comparison data demonstrates that the suggested model performs well across multiple classes, with consistently high precision, recall, and F1-scores. These metrics indicate the model’s effectiveness in accurately identifying and classifying the different classes in the dataset.

This integrated model empowers farmers and agricultural practitioners with timely and precise disease diagnosis, enabling them to implement targeted disease management strategies and optimize yields. The successful integration of these techniques showcases the potential of advanced deep learning methods in automated tomato disease recognition, contributing to the advancement of agricultural systems.

VI. FUTURE WORKS

Building upon the innovative approach presented in this paper, future research directions in the field of agricultural disease recognition can explore several promising avenues. Firstly, investigating the generalizability of the proposed Xception and bilinear pooling model to other crops and diseases holds great potential. Additionally, refining the model’s interpretability and explainability could enhance its usability by providing insights into the features and regions contributing to disease classification. Exploring techniques like attention mechanisms or saliency maps could shed light on the decision-making process of the model, enabling users to trust and fine-tune its predictions.
REFERENCES


