

Research Paper

Assessing fusarium oxysporum disease severity in cotton using unmanned aerial system images and a hybrid domain adaptation deep learning time series model

Alwaseela Abdalla^a, Terry A. Wheeler^b, Jane Dever^c, Zhe Lin^d, Joel Arce^e, Wenxuan Guo^{a,c,*}

^a Department of Plant and Soil Science, Texas Tech University, Lubbock, TX, 79409, United States

^b Department of Plant Pathology and Microbiology, Texas A&M AgriLife Research, Lubbock, TX, 79403, United States

^c Department of Soil and Crop Sciences, Texas A&M AgriLife Research, Lubbock, TX, 79403, United States

^d Environmental Model and Data Optima, Laurel, MD, 20707, United States

^e Department of Entomology, Texas A&M AgriLife Extension, El Paso, TX, 79905, United States

ARTICLE INFO

Keywords:

FOV4

Spatiotemporal information

Domain adaptation

Bidirectional long-short term memory

Deep features

ABSTRACT

Currently, deep learning has achieved remarkable success in estimating plant disease from unmanned aerial system (UAS) images. However, two critical challenges remain unexplored: spatiotemporal variations in disease symptoms and the domain shift between source and target datasets. To overcome these challenges, this paper proposes an approach that incorporates temporal aspects of disease progression using time series analysis. Spatiotemporal information is integrated by combining convolutional neural networks and bidirectional long-short term memory (CNN-BiLSTM) to classify the disease into five severity levels. Various feature extraction methods, including both handcrafted and CNN-based feature extractors, are evaluated. Furthermore, to tackle the problem of domain shift, a feature-level domain adaptation method is proposed. This method aims to learn transferable feature representations that remain consistent despite variations between source and target datasets. This approach enhances the spatiotemporal transferability of the CNN-BiLSTM model, enabling the effective utilisation of historical datasets. The study demonstrates that the CNN-BiLSTM model outperforms traditional time-independent machine-learning methods that rely on handcrafted features. Specifically, the Resnet101-BiLSTM model achieves the highest overall classification accuracy of 89.7% among all tested models evaluated on a one-year dataset. Moreover, it shows superior generalisation with 72.7% accuracy for cross-spatiotemporal disease severity classification using domain adaptation, as demonstrated through two-year experiments. By reducing the domain shift of source and target datasets and harnessing time series high-resolution images obtained throughout the crop growing season, this hybrid approach has substantial potential to advance the assessment of crop disease severity in field conditions.

1. Introduction

Fusarium oxysporum f. sp. *vasinfectum* race 4 (FOV4) is a fungal pathogen that causes an early-season disease called fusarium wilt where the fungus invades the root vascular system of plants, leading to wilting, necrosis, and plant death (Zhang, Abdelraheem, et al., 2022). It is a soil-borne disease where the fungus can survive in the soil for many years, making it difficult to control (Zhang, Zhu, et al., 2022). In many cotton-producing countries, Fusarium wilt (multiple fungal races) causes significant yield and economic losses for farmers (Blasingame & Patel, 2013, pp. 1242–1245). However, race 4, which is only found in

California, New Mexico, and Texas in the U.S.A., is very aggressive and can cause significant stand loss virtually every year in infested fields. Other races of FOV found in the U.S.A., cause damage more sporadically, and usually to a much lesser degree than FOV4. Other FOV races also typically cause damage later in the season (45–90 days after planting) than FOV4 that will kill emerging and young seedlings. Accurate and timely assessment of fusarium wilt disease caused by FOV4 can help farmers make informed decisions about preventative measures and mitigation strategies and can also help plant breeders identify and select cultivars with resistance or tolerance to the disease (Ulloa et al., 2023; Zhu et al., 2022). Traditionally, cotton FOV4 disease severity has been

* Corresponding author. Department of Plant and Soil Science, Texas Tech University, Lubbock, TX, 79409, United States.

E-mail address: wenxuan.guo@ttu.edu (W. Guo).

<https://doi.org/10.1016/j.biosystemseng.2023.12.014>

Received 12 July 2023; Received in revised form 20 November 2023; Accepted 25 December 2023

Available online 1 January 2024

1537-5110/© 2023 IAgRE. Published by Elsevier Ltd. All rights reserved.

assessed by visual inspection of plants for symptoms of wilting and root necrosis (Zhu et al., 2021), which is time-consuming, labour-intensive, and subjective, leading to potential inconsistencies in disease severity assessment. Therefore, it is essential to develop more efficient and reliable methods of assessing FOV4 disease severity.

The rapid advancement of computer vision and machine-learning (ML) has opened new possibilities for monitoring the health status of plants and detecting and assessing plant disease severity. Nevertheless, applying these techniques in outdoor field environments presents significant challenges that demand advanced feature extraction techniques and sophisticated ML classifiers. One approach to using ML for plant disease severity assessment is to extract handcrafted (HC) features from images of plants and use these features as input to a classifier. For instance, Aqel et al. (2021) proposed an approach that uses a gray-level co-occurrence matrix (GLCM) to extract texture features from images, and these features were passed to an extreme learning machine for plant disease classification. Similarly, Harakannanavar et al. (2022) applied wavelet transform, principal component analysis, and GLCM to extract features from tomato leaves to detect the disease using the support vector machine (SVM) and *k*-nearest neighbour (*k*-NN). Kaur (2021) presented a ML approach for the classification of plant diseases. The approach involved utilizing various feature extraction algorithms, such as local binary pattern, GLCM, shift-invariant feature transform, and Gabor, to extract features from input images. Several ML classifiers, namely, SVM, *k*-NN, artificial neural network, and random forest (RF), were trained to perform the task of categorizing plant diseases.

However, FOV4 disease symptoms primarily manifest in the roots, making infection not immediately visible in the plant canopy. This can result in wilting and necrosis in the above-ground tissues during the later stages, particularly with susceptible Pima cotton. In such cases, relying on HC features to detect these symptoms may not be effective for disease assessment. This is because HC features may need expert knowledge to identify and extract all the relevant information from an input image. Alternatively, convolutional neural networks (CNNs) can automatically learn features from raw pixel data through a process known as convolution. CNNs can learn features in a hierarchical manner, enabling them to capture more complex and abstract representations of the image. This makes CNNs flexible and effective for various image-based agricultural tasks, including disease detection, plant image segmentation (Abdalla et al., 2019a, 2019b), Abdalla, Cen, Wan, et al., 2019nd nutrient status diagnosis (Abdalla et al., 2021).

Several studies have reported satisfactory results in identifying different types of plant diseases using CNNs (Bao et al., 2022; Borhani et al., 2022; Krizhevsky, Sutskever, & Hinton, 2012; Nawaz et al., 2022; Qian et al., 2022). These studies have demonstrated that CNNs can be trained to recognise specific patterns of damage, such as leaf spots, wilting, or discoloration, which can help identify the type of disease with high accuracy. In contrast, accurate assessment of disease severity can be challenging because it involves quantifying the degree of damage caused by the disease on the plant, which requires a more detailed analysis of the overall condition of plants, including the extent and severity of symptoms, disease progression, and the impact on plant growth (Shi et al., 2023). This can make it difficult to accurately identify and quantify the severity of a disease using CNNs. An automatic method for assessing the severity of plant diseases using CNNs was first proposed by Wang et al. (2017). They employed various CNN models to classify images of apple black rot into four levels of severity and obtained an overall accuracy of 90.4%, suggesting that CNNs have great potential for fully automating the classification of plant disease severity. In a recent study by Patil et al. (2023), a CNN-based approach was used to estimate the severity of three rice diseases (brown spot, blast, and bacterial blight). The authors manually annotated a dataset of images of diseased leaves with disease zones and trained a CNN to predict the severity scale of a given image. Similarly, Divyanth et al. (2022) developed a deep learning model to categorise images of pea roots into three groups based on the severity of *Aphanomyces* root rot disease. The disease severity

was rated using a scale ranging from 0.0 to 5.0, which was determined by visually examining root discoloration and the softness of the hypocotyl. In another study, Verma et al. (2022) proposed a method for accurately detecting the severity of late blight disease in tomato crops using deep learning techniques. The method combined squeeze and excitation networks with capsule networks to improve feature computation and classification accuracy. The proposed approach achieved high accuracy measures and is robust with noisy datasets.

Current CNN-based disease severity estimation methods rely on ground-based images, such as PlantVillage datasets (Hughes & Salathé, 2015), or images acquired under controlled environments (Esgario et al., 2020), which may not capture spatiotemporal changes that occur in crop diseases. To overcome this limitation, unmanned aerial systems (UASs) have been found to provide a more comprehensive view of crops, enabling deep learning models to capture spatial and temporal variations (Bouguettaya et al., 2023). By combining UASs with deep learning, automated disease severity assessment can be achieved, reducing the need for human intervention and enabling quicker and more accurate decisions. However, models trained on controlled images may have reduced accuracy when used to predict images collected under natural environments (Thapa et al., 2020; Zeng et al., 2018, pp. 1–5).

FOV4 root disease manifests differently depending on the growth stage of the plant canopy, with unique symptom sets exhibited by different developmental stages. For example, younger plants may display stunted growth, yellowing leaves, and wilting, while older plants may exhibit chlorosis, browning of vascular tissue, and defoliation. As such, the severity and presentation of FOV4 root disease is highly influenced by the plant growth stage (Zhu et al., 2023). Although CNNs have demonstrated excellent performance in detecting plant disease symptoms from images, they may not be suitable for capturing temporal dependencies in disease progression over time. On the other hand, bidirectional long short-term memory (BiLSTM) networks have demonstrated great potential for analyzing dynamic systems in various applications, such as fault diagnosis in water management systems (Liu et al., 2019), medical imaging processing (Hwang et al., 2022) and mood recognition (Son, 2017). The BiLSTM has the ability to remember information over long periods of time, making it well-suited for modeling time-series data. In a study conducted by Crisóstomo de Castro Filho et al. (2020), the application of time-series analysis in agricultural imaging was investigated. The study utilised both long short-term memory (LSTM) and BiLSTM models for crop classification, and the results revealed that the BiLSTM model was the most effective. To date, BiLSTM has not yet been applied to assess plant disease severity levels. The hypothesis of this study is that BiLSTM can effectively capture temporal changes in disease severity over time.

Despite the success of BiLSTM models, their performance may be hindered when they are trained solely on the source domain, as their ability to generalise to the target domain, characterised by dissimilar data distributions, can be compromised. This phenomenon, known as domain shift, arises due to a range of factors, including variations in field conditions, crop phenology, and agricultural management practices. Such diversities across domains make it challenging to develop deep learning models capable of effectively generalising across diverse real-world scenarios. Consequently, the manifestation and severity of diseases may vary significantly in different agricultural management contexts.

The main objective of this study is to develop an effective deep learning time series model to predict the severity of cotton FOV4 disease accurately and efficiently from UAS images. The main contributions of this article are as follows. Firstly, to investigate the effectiveness of different CNN-based feature extractors in comparison to traditional handcrafted (HC) features for describing the visual symptoms associated with FOV4 disease in UAS images. Secondly, to monitor the progression of the disease in cotton plants, the BiLSTM model was employed. This model is designed to capture temporal dependencies within UAS images by processing data in both forward and backward directions, facilitating

a more comprehensive understanding of the disease development over time. Lastly, to address the challenge of domain shift and enhance the generalisation capability of the proposed model, a domain adaptation technique is introduced to improve the transferability of feature representations across diverse datasets. The proposed approach shows promise in providing more accurate and consistent disease severity estimates, reducing manual inspections, and enabling efficient crop management practices.

2. Materials and methods

2.1. Experimental site and image acquisition

An experiment was carried out during the summer seasons of 2020 and 2021 at a commercial field located 10 km northwest of Fabens, Texas (31°31'42.03" N and 106°11'42.76" W, Fig. 1). This region has an arid to semi-arid climate with hot summers, warm and dry winters, and low precipitation. Soil classification information was obtained from the NRCS Soil Survey Geographic Database (SSURGO). The soils in this field are dominated by the Glendale silty clay (Fine-silty, mixed, superactive, calcareous, thermic Typic Torrifluvents), approximately 90% of the area on the north side of the field. On the south side, the soil is Harkey silty clay loam (Coarse-silty, mixed, superactive, calcareous, thermic Typic Torrifluvents). The surface texture of this soil is very fine sandy loam.

The field was known to be infested with FOV4 and had been in test plots since 2018. The field was planted on May 4th in both 2020 and 2021. The experiment was implemented using a randomised complete block design with three replications. Plots were one row wide (1 m centres), 4.88 m in length, with a 1.21 m alley between blocks. Throughout the experiment, surface irrigation was employed, starting from three weeks prior to the planting date, and then from mid-June onwards at approximately 3 weeks intervals until September.

A DJI Phantom Pro 4 (DJI Technology Co., Ltd., Shenzhen, Guangdong, China) UAS was employed to acquire high-resolution images. The camera on this UAS has a focal length of 35 mm and an aperture range between f/2.8 and f/11. The camera captured RGB images at a resolution of 5472 × 3078 pixels at a height of 25 m with 80% front and side overlaps. In 2020, images were acquired on June 10, July 1, July 29, and August 31, 2020. In 2021, images were acquired on June 9, July 1, August 2, and September 28, 2021 (Table 1). An automated flight plan

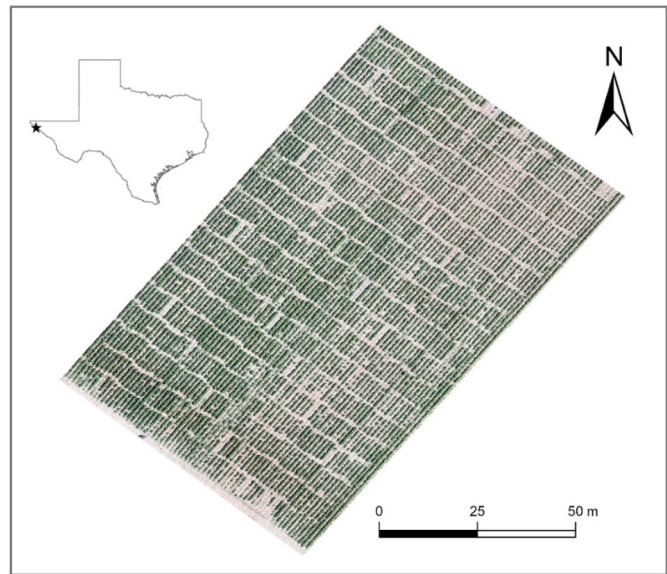


Fig. 1. Location of the experimental field near Fabens, Texas (indicated by star in the state map of Texas), and a sample UAS image of the field captured on July 29, 2020.

Table 1
UAS image datasets obtained from a field near Fabens, Texas, in 2020 and 2021.

Year	Sampling date	No. of plots	Total
2020	06/10	741	2964
	07/01	741	
	07/29	741	
	08/31	741	
2021	06/09	240	720
	08/02	240	
	09/28	240	

was created using a pre-prepared shapefile made in Pix4Dcapture, and the flight duration was approximately 15 min. Ground control panels were placed at 12 locations within the field for georeferencing images during each flight. The raw UAS images were stitched using Pix4Dmapper (Version 4.3.3, Pix4D, Lausanne, Switzerland). A polygon shapefile was created and overlaid on each stitched image, with each polygon representing a plot of a genotype entry. This shapefile was then used to clip the images into individual plot images, such that each plot was saved as a file in the jpg format. The total number of sample images corresponding to the number of plots and the dates of image acquisition for the two-year experiment are summarised in Table 1.

2.2. FOV4 disease severity rating and image labeling

The process of image labeling was based on the severity of plant disease at the plot level. Each plot was assigned a label ranging from 1 to 5, reflecting the level of disease severity in the plants. To assess the severity of FOV4 disease in cotton plants, a qualitative evaluation of the root systems from the plots was performed after the cotton had been harvested on December 14–15, 2020, and December 8–10, 2021. The damage to the vascular system of roots was cumulative and could be observed as early as late June. However, end-of-season ratings provide a maximum estimate of the damage that developed throughout the season. The taproots were excavated and subsequently sectioned down the centre to expose the vascular system. The root ratings followed a scale ranging from 0 to 5, with 0 indicating no discoloration of the root centre, 1 indicating limited discoloration that did not run the length of the taproot, 2 representing discoloration that is relatively narrow in width but runs the length of the root system core, and 3 to 4, which is similar to a 2 but with progressively wider discoloration across the vascular system, and 5 indicating a dead taproot. To determine the disease severity of each lot, five root samples were rated and the average score was obtained from the samples. Class 5 had few samples and was combined with class 4 to train the deep learning model, which requires a large dataset for each class. Additionally, the UAS images of the plots were labeled and categorised according to the disease severity scale assigned to each plot.

2.3. Data pre-processing

Several pre-processing techniques were employed to prepare the images before feeding them to the final prediction model. Image segmentation, which separates the plant canopy from the soil background, was a critical step in this process. To achieve this, an encoder model based on VGG16 was utilised, as described in the work of Abdalla, Cen, Wan, et al. (2019). Fig. 2 depicts examples of segmented images at each imaging time and the corresponding FOV4 disease severity. In addition, the 2020 and 2021 datasets presented a class imbalance, with the mild class having a substantially higher number of instances than the other classes. This imbalance posed a challenge in generating effective learning patterns in the less frequent classes and could lead to poor performance when predicting samples from those classes. To address this issue, median frequency balancing was employed, as proposed by Eigen and Fergus (2015, pp. 2650–2658). This technique involves adjusting the weights in the cross-entropy loss function to account for

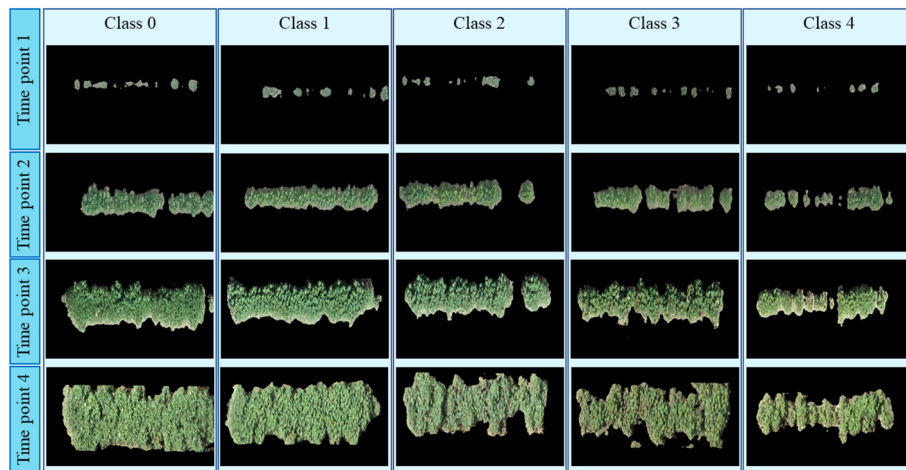


Fig. 2. Sequences of images that depict the growth stages of the cotton canopy at different times. Each image represents the severity of FOV4 disease at the respective time point.

class imbalances. The weight assigned to each image for a given class c is determined by the formula:

$$\alpha_c = \frac{\text{median frequency}}{\text{frequency}(c)}$$

Here, *frequency* (c) is the frequency of class c , calculated as the number of images of class c divided by the total number of images. The term *median frequency* represents the median of these frequencies across all classes. By employing this technique, the loss function is weighted in a manner that gives more emphasis to classes that are under-represented in the dataset. This is achieved by increasing the loss contribution of less

frequent classes during training.

2.4. CNN-BiLSTM for disease severity classification

The proposed hybrid CNN-BiLSTM approach for assessing disease severity involved a multi-step process. Firstly, CNNs were employed to extract features from the pre-processed images. These features were then fed into a BiLSTM network that was responsible for analyzing the sequence of features over time and predicting the severity of the disease. Fig. 3 illustrates how this approach combined the strengths of CNNs and the sequential BiLSTM model to achieve an accurate assessment of

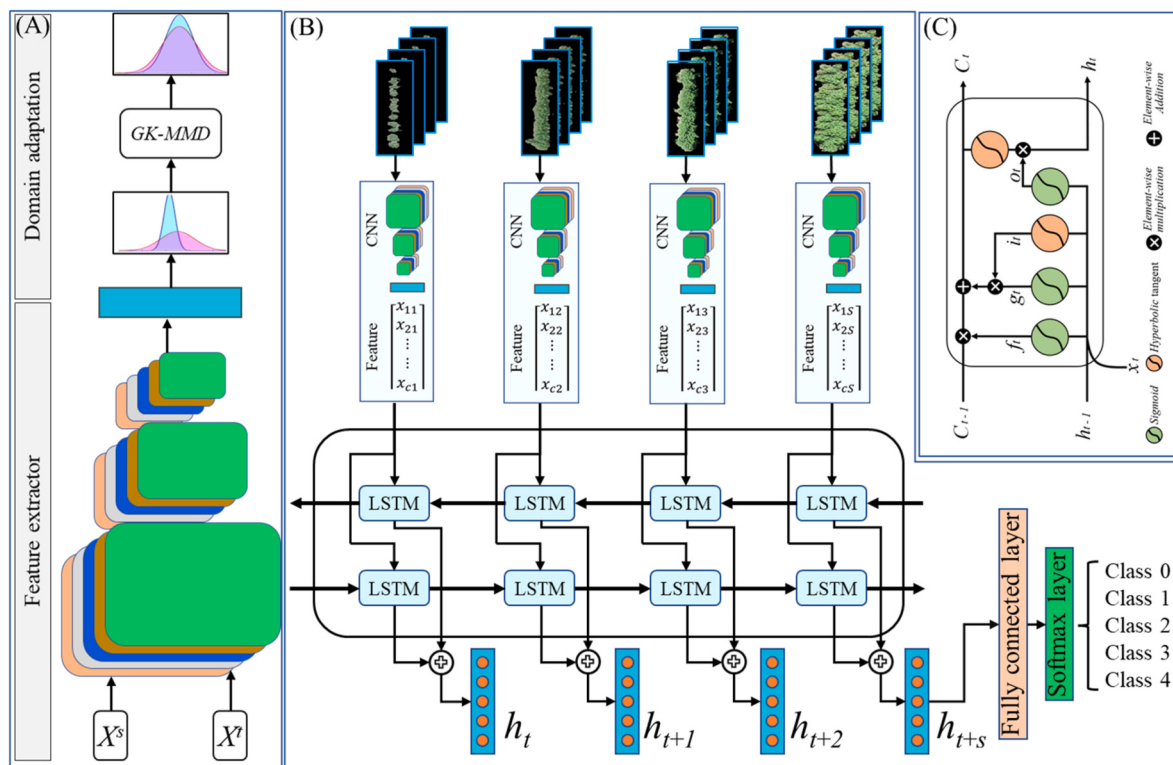


Fig. 3. A schematic of a domain adaptation model for time-series data using deep learning. (A) The domain adaptation process using a Gaussian kernel-based maximum mean discrepancy (GK-MMD) to align the feature distributions of the source domain (X^S) and target domain (X^T). (B) outlines a convolutional neural network (CNN) feature extraction (x_{cs}) followed by a bidirectional long short-term memory (BiLSTM) network for processing sequential data. (C) illustrates the internal structure of a single LSTM unit, including the input gate (i_t), forget gate (f_t), output gate (o_t), the previous cell state (C_{t-1}), the current cell state (C_t), the previous hidden state (h_{t-1}), and the current hidden state (h_t).

disease severity based on the distribution and progression of symptoms over time. The following provides a more detailed explanation of each step.

2.4.1. Feature extraction using CNNs

CNNs are powerful tools for extracting informative and efficient features from images through fine-tuning of pretrained models. To determine the most appropriate CNN architecture to extract features from pre-processed images, a comparative analysis of the performance of five popular pretrained models, namely Alexnet ((Krizhevsky et al., 2012), Inceptionv3 (Szegedy et al., 2016, pp. 2818–2826), VGG (Simonyan & Zisserman, 2015), Resnet18, and ResNet101 (He et al., 2016, pp. 770–778) was conducted. These networks were modified by replacing their final classification layers with five neurons to classify UAS images into five FOV4 disease severity classes. A transfer learning approach was employed that fine-tuned the weights of pretrained models, using a stochastic gradient descent optimiser with a learning rate of 0.0005, a momentum of 0.09, and a mini-batch size of 40 images to optimise the model parameters. The training was stopped after a maximum of 10 epochs or when there was no improvement in validation loss over four consecutive epochs. These features were extracted from the last fully connected layers of the CNN model and utilised as input for the BiLSTM network in a sequential manner, as described in section 2.4.2. It is important to note that, in this stage, the CNN architectures were only utilised to extract features and did not incorporate the temporal aspect of the image datasets or the potential temporal correlations, which were incorporated in the subsequent step using a BiLSTM model.

2.4.2. Feature-level domain adaptation

Domain adaptation plays a critical role in training models on a source domain and deploying them on a target domain, where the underlying data distributions may exhibit dissimilarities. Let's consider a source domain feature denoted as $X_s = [x_{s1}, x_{s2}, \dots, x_{sn}]$, consisting of n samples, and a target domain feature denoted as $X_t = [x_{t1}, x_{t2}, \dots, x_{tm}]$, consisting of m samples. These features are characterised by probability distributions P and Q , respectively. Notably, $P \neq Q$ due to environmental conditions and variations in crop phenology. Consequently, a model trained on the dataset X_s exhibits reduced performance when directly applied to the dataset from X_t . To quantitatively measure the dissimilarity between the distributions P and Q , Gaussian kernel-based maximum mean discrepancy (GK-MMD) is employed within a reproducing kernel Hilbert space (RKHS). The fundamental principle behind GK-MMD is rooted in the notion that if P and Q are identical, then the expected value of any function computed from samples drawn from these distributions should be equal. Given samples generated independently from P and Q , the GK-MMD can be calculated using the following Eq. (1)

$$GK - MMD(P, Q) = \left\| \frac{1}{n(n-1)} \sum_{i \neq j} k(X_{si}, X_{sj}) + \frac{1}{m(m-1)} \sum_{i \neq j} k(X_{ti}, X_{tj}) - \frac{2}{mn} \sum_{i,j} k(X_{si}, X_{tj}) \right\|_H \quad (1)$$

The first term in the equation represents the mean kernel value between samples from P ; the second term represents the mean kernel value between samples from Q , and the third represents the mean kernel value between samples from P and Q . The kernel k is the positive definite kernel function that defines the RKHS associated with the distributions. The choice of the kernel function depends on the problem at hand and the characteristics of the data. This study employed a Gaussian kernel as one of the most widely used kernel functions. It is defined as:

$$K(x, y) = \exp\left(-\frac{\|X_s - X_t\|^2}{2\sigma^2}\right) \quad (2)$$

Where $\|X_s - X_t\|$ represents the Euclidean distance, and σ is the bandwidth parameter controlling the width of the kernel.

By utilizing the GK-MMD, the dissimilarity between the source and target domains can be precisely quantified, enabling effective domain adaptation to address the performance degradation observed when directly applying models trained on X_s to X_t .

To mitigate the domain shift and reduce the dissimilarity between the source and target domains, a feature-level domain adaptation approach called correlation alignment (CORAL) is proposed. The primary objective of CORAL is to enhance the transferability of feature representations across diverse datasets by aligning the second-order statistics or covariances of the source and target domains, denoted as C_s and C_t , respectively. Given the source domain features X_s and target domain features X_t , CORAL aims to transform the source domain features (X_s) to reduce the domain shift and align them with the target domain (X_t), using Eqs. (3)–(5).

$$X_s^{coral} = (X_s - \mu_s) \times D_s D_t + \mu_t \quad (3)$$

$$D_s = (C_s + \lambda I)^{-\frac{1}{2}} \quad (4)$$

$$D_t = (C_t + \lambda I)^{\frac{1}{2}} \quad (5)$$

Where X_s^{coral} represents the CORAL-transformed source domain samples, μ_s and μ_t denote the mean vectors of X_s and X_t , respectively, D_s is the square root of the inverse of the source domain covariance matrix plus a regularization term (in this study $\lambda = 1e-2$) added to avoid singularities, and $\sqrt{C_t}$ represents the square root of the target domain covariance matrix, and I represents the identity matrix which has the same size as C_s . By applying this transformation, the aim is to align the second-order statistics of the source and target domains, enabling the learning of more transferable feature representations that effectively bridge the gap between different datasets. The effectiveness of CORAL is demonstrated through experimental evaluations, i.e., CNN-BiLSTM trained on one dataset (source domain) and tested on another dataset (target domain) before and after applying CORAL, illustrating its potential for mitigating the effects of domain shift.

2.4.3. BiLSTM model

To classify sequential images of the cotton plots based on disease severity and to account for temporal aspects, the BiLSTM network was utilised. The BiLSTM network is a type of recurrent neural network (RNN) that incorporates two LSTM units. One of the LSTM units processes the input sequence in the forward direction, while the other

processes it in the backward direction, followed by concatenating their outputs at each time step. This design allows BiLSTM to update the cell state using both past and future time steps. The LSTM architecture was originally proposed by Hochreiter and Schmidhuber (1997) to address the problems of vanishing and exploding gradients encountered in traditional RNN. The LSTM overcomes these issues by introducing a gating mechanism and memory cells in the hidden layers of the network, which enhances its memory and storage capabilities. As shown in Fig. 3C, each LSTM unit has a cell state (c) and a hidden state (output state, h), which are controlled by four gates: forget gate (f), candidate

gate (g), input gate (i), and output gate (o). These gates concatenate the input of the current time step (x_t) with the hidden state of the previous time step (h_{t-1}). The forget gate determines which information to discard from the cell state, while the input and candidate gates decide what new information should be stored in the cell state. The output gate determines what information should be passed to the next time step.

Extracted features can be mathematically represented by a vector D composed of a cell array containing T time steps. Each time step within this vector is represented by a matrix X_T with dimensions of (N, M) , where N represents the total number of instances and M represents the number of features for a given instance. The values within each element of this vector denote the specific feature value at a particular time step for a given instance. To train the BiLSTM model on these features, an end-to-end backpropagation strategy is employed, using cross-entropy loss as the loss function. After training, the final hidden state, h_{t-s} , encodes the most important information for sequential data. The h_{t-s} is utilised as a representative vector, which is transformed into a vector of the same length as the number of classes using a fully connected layer. A softmax layer is then attached to the end of the BiLSTM, with the number of neurons in the softmax layer set to match the number of classes. This ensures that the output of the model is compatible with the classification task.

To comprehensively assess the efficacy of the proposed framework, three separate evaluations were conducted. The first evaluation involved randomly dividing the features extracted from the 2020 dataset, which consists of 2,964 images, using a 70:30 ratio for training and testing, respectively. This means that 2,075 images were used for training and 889 images for testing. The second evaluation aimed to demonstrate the generalisation ability of the model to new data by training it on the features extracted from the complete 2020 dataset (2,964 images) and testing it on the features derived from the entire 2021 dataset, comprising 720 images. The third evaluation involved comparing the performance of this framework with other ML methods. These evaluations were conducted to demonstrate the effectiveness and superiority of the proposed approach. The frameworks were implemented in MATLAB R2023a (MathWorks, Inc., Natick, MA, USA) on an NVIDIA Quadro M5000 GPU with 2048 CUDA Cores and memory of 8 GB GDDR5.

2.5. Hand-crafted features

HC features refer to manually selected visual attributes of the plant canopy, which are used to estimate the severity of the disease. Typically, these features are identified by domain experts with specialised knowledge of plant diseases. In this study, three HC features were extracted from segmented images, including vegetation fraction, color, and texture. The vegetation fraction was calculated by determining the ratio of canopy pixels to the total number of image pixels, expressed as a percentage. The average of the red, green, and blue color intensities from the segmented images were computed. Additionally, four textural features were extracted using the GLCM, namely energy, homogeneity, contrast, and entropy. To create high-dimensional HC features, the textural, vegetation fraction, and color features were combined. These HC features were then compared with deep learning-based features to assess their efficacy in classification models.

2.6. Performance evaluation of the proposed method

A study was conducted to evaluate the impact of image segmentation as a pre-processing step in FOV4 cotton disease assessment. Additionally, the study compared the performance of several pre-trained CNN architectures (Alexnet, Inceptionv3, VGG, Resnet) to determine the most suitable deep CNN for feature extraction. To assess the contribution of deep features in evaluating FOV4 disease severity in cotton plants, they were compared with well-known high-dimensional HC features. The effectiveness of the feature extractors was evaluated visually by

projecting high-dimensional features into three dimensions using t -distributed stochastic neighbour embedding (t -SNE) (Van der Maaten & Hinton, 2008). Additionally, the proposed time-series model, BiLSTM, was compared with a well-known standard SVM to highlight its capability in understanding the temporal relationships in the datasets. The best model configuration was then combined with domain adaptation CORAL and utilised to test the generalisation capability and reliability of the proposed framework. For this purpose, the classification model was trained on the entire dataset of 2020 and evaluated on the dataset of 2021.

To evaluate the performance of the competing methods, evaluation metrics including accuracy (Ac), precision (Pr), recall (Re), and F-score (Fs) were used. The prediction time (PT) of deep learning models was also compared since they were intended for real-time application. The PT of hybrid models, such as CNN-BiLSTM and CNN-SVM, was measured for both classification and feature extraction processes. For single models, such as CNN, and HC feature-based BiLSTM or SVM, only the PT for classification was measured since no feature extraction process was incorporated, and the time required for calculating HC features was negligible. The training time required for all network implementations was not considered in this study, as it can be conducted offline and may not need to be frequently repeated.

3. Results

3.1. Impact of image segmentation on the prediction accuracy

Image segmentation was found to be a crucial pre-processing step in accurately estimating plant disease severity, as it noticeably impacted the performance of different learning models (Table 2). It should be noted that the effect of segmentation varied across the models used, but in general, it was found to enhance the performance of the tested models when compared using different performance metrics. For instance, CNNs, commonly used for image classification, were particularly effective for plant disease detection when segmentation was applied. Other models, such as BiLSTM and SVM, also benefited from the use of segmentation. This is because segmentation can improve the extraction of relevant features from the image, leading to more robust and accurate predictions. By eliminating the irrelevant background information and focusing on the plants, segmentation not only enhances the feature extraction process but also improves the accuracy of disease severity predictions. For the subsequent discussions, focus will only be on the model that incorporated image segmentation as a pre-processing step.

3.2. Training and validation processes

The training and validation loss and accuracy of six models are shown in Fig. 4. The results showed that the training accuracy of the models fluctuated to a certain extent. Models A, B, C, D, and E achieved high accuracy on both the training and validation sets. Among the models, model E (Resnet101-BiLSTM) achieved the highest accuracy on the validation set, and it was relatively stable compared to the other models. The stability in the training accuracy and losses was observed to have an impact on the final predicted outputs of the model (Table 2). However, it was observed that model E exhibited overfitting behavior, as the validation loss started to increase after 150 iterations while the training loss kept decreasing. This indicated that the model that used HC features did not generalise well to unseen data and provided lower training and validation accuracy compared to BiLSTM models that relied on deep learning-based features. This phenomenon could be explained by the lowest level of discriminability of HC features between each class (Fig. 4F).

3.3. Visualizing features extracted from CNN using t -SNE

An evaluation was conducted to assess the discriminability of

Table 2

Performance comparison of various classification models for assessing FOV4 cotton disease severity. The best performances are highlighted in bold. The metrics include Ac for accuracy, Pr for precision, Re for recall, Fs for the F-score, and PT for prediction time.

Feature extractor	Classifier	Original images				Segmented images				PT (sec/image)
		Ac (%)	Pr (%)	Re (%)	Fs (%)	Ac (%)	Pr (%)	Re (%)	Fs (%)	
Alexnet	BiLSTM	58.7	65.7	65.3	65.5	73.1	76.5	75.7	76.1	0.10
Inceptionv3		61.9	61.7	71.8	66.4	74.8	82.0	71.2	76.2	0.14
Mobilenetv2		69.2	70.6	72.1	71.4	77.1	91.2	81.1	85.8	0.13
Resnet18		77.3	81.8	82.6	82.2	86.7	93.1	93.4	93.3	0.11
Resnet101		81.8	86.0	86.1	86.0	89.7	87.5	87.9	87.7	0.13
HC		39.6	52.6	50.0	51.3	41.5	65.2	71.0	68.0	0.15
Alexnet	SVM	44.7	55.9	84.9	67.4	50.6	56.4	79.2	65.9	0.10
Inceptionv3		43.6	54.9	77.4	64.2	50.7	57.2	78.5	66.1	0.16
Mobilenetv2		45.4	57.0	82.9	67.6	56.5	63.5	83.8	72.2	0.14
Resnet18		53.2	65.2	85.6	74.0	65.6	72.1	85.4	78.2	0.10
Resnet101		60.9	66.3	86.9	75.2	74.5	74.5	89.4	81.2	0.14
HC		24.9	44.3	77.7	56.5	27.9	47.2	45.4	46.3	0.18
Alexnet		27.0	32.2	27.2	29.5	30.3	26.2	45.8	33.4	0.02
Inceptionv3		35.1	50.0	50.1	50.1	35.8	33.3	31.5	32.4	0.08
Mobilenetv2		23.7	28.3	18.5	22.4	31.4	31.1	30.2	30.7	0.04
Resnet18		25.7	16.0	29.2	20.6	30.1	51.3	39.6	44.7	0.04
Resnet101		27.1	17.9	33.3	23.3	33.7	20.3	50.3	28.9	0.11

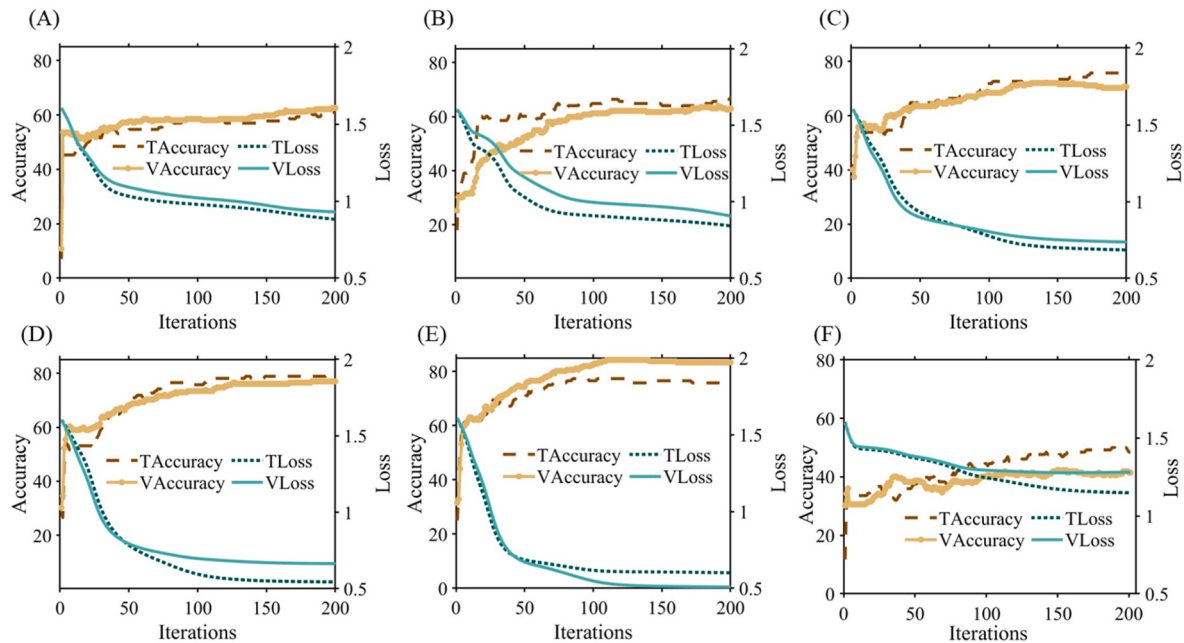


Fig. 4. Training progress of five different models ((A) Alexnet-BiLSTM, (B) Inceptionv3-BiLSTM, (C) VGG16-BiLSTM, (D) Resnet18-BiLSTM, (E) ResNet101-BiLSTM, and (F) HC-BiLSTM) on the 2020 dataset, as measured by training accuracy (TAccuracy), validation accuracy (VAccuracy) training loss (TLoss), and validation loss (VLoss) metrics over the course of training.

features extracted from different layers of several CNN architectures, namely AlexNet, Inceptionv3, VGG16, VGG19, ResNet18, and ResNet101. The analysis focused on the early max pooling layer, final convolutional layer, and softmax layer of each architecture. The most effective layer for distinguishing between different classes was identified by utilizing a *t*-SNE-based visualisation technique in a three-dimensional feature space. The results of the study revealed that the features extracted from the final convolutional layer of ResNet101 exhibited the highest level of discriminability (Fig. 5F, last column). This finding suggests that the deeper architecture of ResNet101 captured more discriminative information, enabling improved differentiation between classes. Consequently, selecting the appropriate layer for feature extraction was crucial in deep learning models, with the final convolutional layer of ResNet101 showing particular promise in achieving high discriminability. It was also observed that AlexNet, one of the early deep

learning architectures, demonstrated competitive performance in the analysis. This can be attributed to its successful application in various image classification tasks and its ability to extract meaningful features at different layers. However, compared to ResNet101, AlexNet has a relatively shallower structure, potentially limiting its capacity to model complex patterns and intricate details. Inceptionv3, known for its inception modules and auxiliary classifiers, also performed well in the experiments. The capability of architecture to capture multi-scale features through its inception modules facilitates effective representation learning. Although Inceptionv3 did not surpass ResNet101 in this study, its strong performance suggests its suitability for tasks where multi-scale information is crucial. The VGG architectures (VGG16 and VGG19) have a simple and uniform structure comprising repeated convolutional layers. While VGG16 and VGG19 did not exhibit the same level of discriminability as ResNet101 in this study, they have been extensively

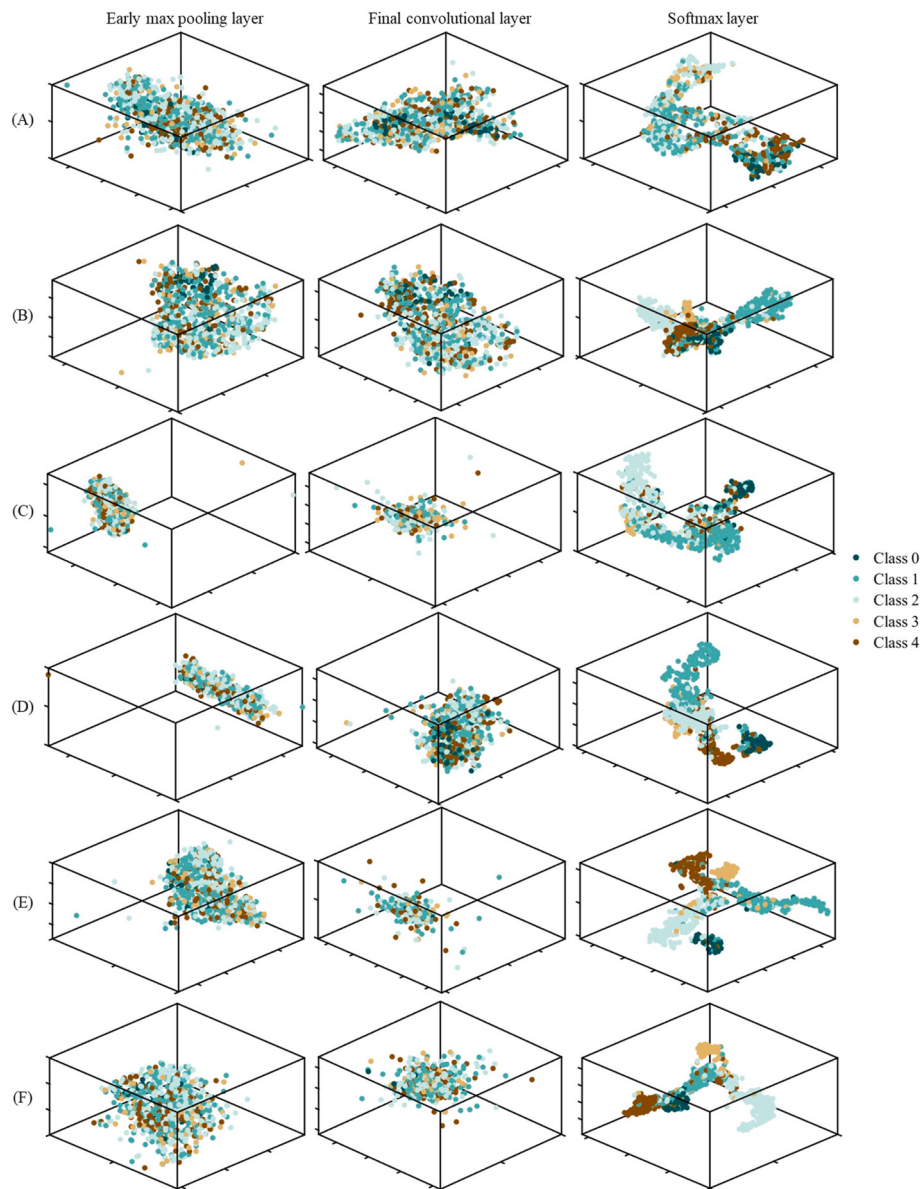


Fig. 5. *t*-SNE maps of six different feature extraction methods: (A) Alexnet, (B) Inceptionv3, (C) VGG16, (D) VGG19, (E) Resnet18, and (F) Resnet101. The features were extracted from the early max pooling layer, final convolutional layer, and softmax layer of each architecture.

utilised in various image classification tasks due to their simplicity and interpretability. The VGG architectures are often preferred when computational resources are limited or when interpretability is a key consideration. ResNet18, despite being a shallower version of ResNet, demonstrated notable performance in this analysis. Although it may not capture intricate details as effectively as ResNet101, ResNet18 can be a suitable choice for tasks where a balance between model complexity and computational efficiency is desired. Generally, the findings emphasised the critical role of selecting the appropriate CNN architecture for feature extraction and discriminability.

To gain a better understanding of the role of deep features in assessing the FOV4 cotton disease, a comparative experiment and quantitative analysis was conducted based on four performance metrics: Ac, Pr, Re, and Fs. Despite efforts to incorporate many HC descriptors, the deep feature-based models outperformed the classification models that relied on these descriptors (as shown in Table 2). Specifically, when the HC features were replaced with deep features in the same classification model (i.e., Resnet101-BiLSTM), substantial improvements were observed in Ac, Pr, Re, and Fs, by 116.3%, 34.3%, 23.8%, and 29.0%,

respectively. These findings demonstrated the effectiveness of deep features for image classification, which was also supported by the *t*-SNE visualisation analysis as shown in Fig. 5.

3.4. Traditional machine-learning versus BiLSTM

This study also investigated the importance of considering temporal dependence in the analysis of UAS image datasets for FOV4 cotton disease severity assessment. To this end, the performance of the BiLSTM model was compared with a static ML model, specifically an SVM, for assessing the severity of FOV4 cotton disease. The results indicated that the BiLSTM model outperformed the SVM in all classification metrics, regardless of the type of features used as input. The Resnet101-BiLSTM model achieved the highest Ac of 89.7%, Pr of 87.5%, Re of 87.9%, and Fs of 87.7% among all models tested. In contrast, the HC-SVM model exhibited the lowest Ac of 27.9%, Pr of 47.2%, Re of 45.4%, and Fs of 46.3%. The reported results were based on an average performance of over 10 different experimental dataset executions, with the training and testing sets randomly selected each time. The findings of this study

suggest that the proposed BiLSTM model is highly effective for assessing the severity of FOV4 cotton disease and holds significant potential as a valuable tool for disease assessment.

Furthermore, the impact of using a pure CNN-based classifier was evaluated and compared to an approach with BiLSTM to capture temporal information in the image datasets. It was found that the use of a pure CNN-based classifier (i.e., Resnet101) without a sequence learning module resulted in a lower classification Ac of 33.7% compared to the approach with BiLSTM. The study highlights the significance of considering temporal dependence in the analysis of image datasets for disease severity assessment, particularly in cases where plant disease evolves over time. The findings suggested that the proposed BiLSTM model was a highly effective tool for assessing the severity of FOV4 cotton disease, which has significant potential for disease assessment.

3.5. Domain adaptation and generalisation capabilities of the proposed model

Fig. 6 presents a quantitative analysis of the distributions before and after CORAL transformation. The results of applying the CORAL technique to the source and target domains are promising, as evidenced by the observed reduction in the dissimilarity between their distributions. Prior to applying CORAL, the normal distributions of the two domains exhibited noticeable disparities, suggesting a pronounced domain shift (Fig. 6A). However, after the CORAL transformation, the distributions of the source and target domains became more aligned, suggesting a successful reduction in the domain shift (Fig. 6B). GK-MMD values were calculated as a measure of dissimilarity between the distributions. A comparison of the GK-MMD values before and after applying CORAL (0.221 and 0.075, respectively) indicated a substantial reduction in the domain shift after applying CORAL.

To further investigate the impact of reduced distribution differences between domains on the generalisation capabilities of the proposed model, experiments were conducted to evaluate the performance of the ResNet101-BiLSTM models trained on the source domain (2020 dataset) and tested on the target domain (2021 dataset), both with and without CORAL. As shown in Fig. 7, the results revealed that the ResNet101-BiLSTM model with CORAL exhibited superior generalisation performance compared to the model without CORAL. The model with CORAL achieved an Ac of 72.7%, Pr of 75.2%, Re of 77.8%, and Fs of 76.5%. These metrics indicated that the model with CORAL was able to accurately classify the severity levels of cotton diseases in the target domain, even though it was trained on data from a different year. This finding has significant implications for real-world applications, where deploying a model trained on one dataset to classify instances in a different dataset is a common scenario.

To further assess the implications of the reduced distribution

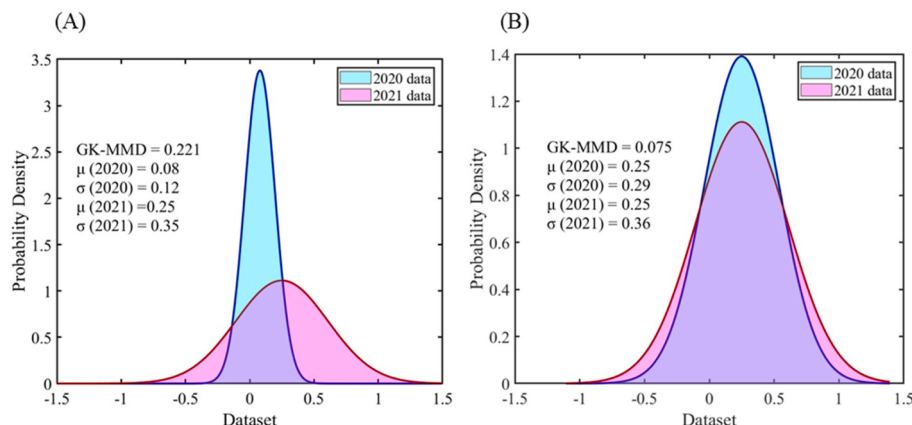


Fig. 6. Comparison of normal distributions of the source and target domains before (A) and after (B) applying the correlation alignment (CORAL) technique.

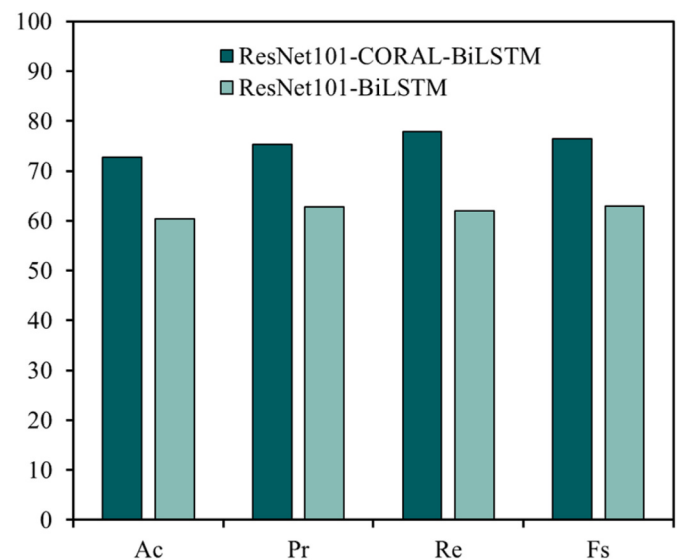


Fig. 7. The cross-domains performance results of a model that was trained on the 2020 dataset and tested on the 2021 dataset. The metrics include Ac standing for accuracy, Pr for precision, Re for recall, and Fs for the F-score.

difference between the two domains on the generalisation capabilities of the proposed model, a receiver operating characteristic (ROC) curve analysis was performed that compared the performance of ResNet101-BiLSTM models with (Fig. 8A) and without (Fig. 8B) domain adaptation using CORAL. The results, as demonstrated by the ROC curve analysis, showed that the ResNet101-BiLSTM model with CORAL achieved a superior generalisation performance. This is supported by the higher area under the ROC curve obtained for the model with CORAL compared to the model without CORAL, indicating improved classification performance across the severity levels of cotton diseases in the target domain. Nonetheless, a relatively lower AUC was observed for Class 2; this may be attributed to inherent complexities associated with distinguishing this class, imbalanced class distribution, and limited training samples.

4. Discussion

Assessing the severity of FOV4 cotton disease from UAS images is a complex task that requires advanced techniques capable of capturing temporal and spatial variations in disease symptoms. Temporal variations arise due to disease progression over time, which can take several days to fully manifest, and capturing these variations is crucial for accurate disease severity assessment. Spatial variations occur due to the

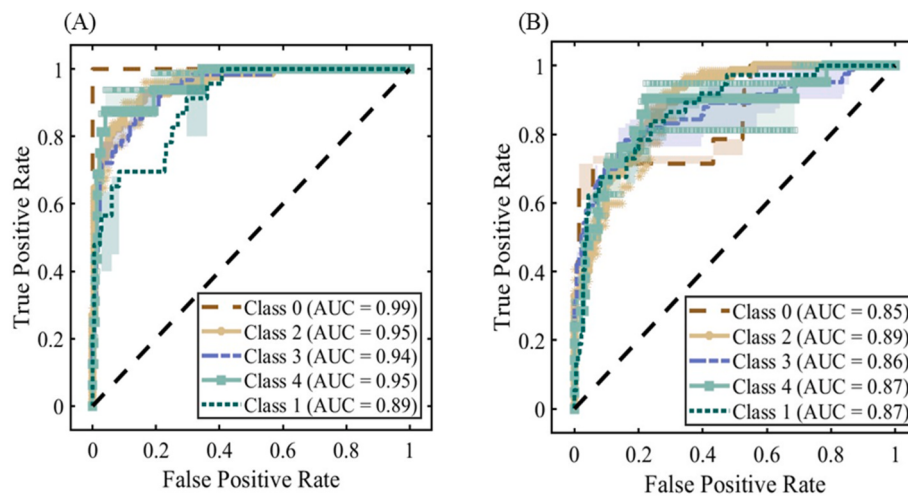


Fig. 8. Receiver operating characteristic (ROC) curve analysis comparing the performance of ResNet101-BiLSTM models with (A) and without (B) domain adaptation using CORAL.

disease affecting different parts of the cotton plant, making it challenging to assess disease severity accurately. Consequently, a sophisticated technique that combined CNNs, BiLSTM, and UAS images was developed in this study to overcome these challenges. It was shown that the combination of these techniques can lead to a more accurate disease severity assessment than using any single technique alone. CNNs are good at extracting features from images, while BiLSTM networks can capture temporal dependencies in the data. Using both techniques together can help capture both spatial and temporal information from the images, leading to better accuracy. To the best of the authors' knowledge, this is the first research to leverage the strength of both CNN and BiLSTM to exploit both spatial and temporal information from the UAS images to estimate the FOV4 disease severity.

Current state-of-the-art methods employed for disease detection and classification can be categorised into two categories: classical CNN frameworks and CNN-based feature extraction coupled with traditional ML models. The first category includes the use of pretrained CNN models like VGG, DenseNet, InceptionV3, and various versions of ResNet, as introduced in studies by Haque et al. (2022) and Tang et al. (2023). The second category focuses on using CNN only for feature extraction, followed by the application of traditional ML models, as demonstrated in research by Saeed et al. (2021). This study presented a comparative analysis of these methods. In Table 2 of the study, the proposed approach, which integrated various feature extractors, was first compared with BiLSTM. The second part of the table examines the most prevalent CNN-based feature extraction methods combined with traditional ML models (e.g., SVM), while the final section compares classical CNN frameworks.

To support the claim of advancements being made by CNNs to identify complex symptom patterns of FOV4 disease, this study compared the effectiveness of CNN-based deep features with traditional HC feature extraction methods. The findings revealed that CNN-based deep features outperformed HC features in terms of capturing intricate patterns in plant images. As a result, the CNN-based approach was more effective at accurately assessing disease severity. This finding has significant implications, as it suggests that CNNs offer a more efficient and accurate approach to assessing plant disease severity compared to traditional HC feature extraction methods. The reason for this is that CNNs can automatically learn discriminative features directly from raw input data. This eliminates the need for manual feature engineering and makes the method more robust to variations in illumination conditions and image scale. By leveraging CNN-based deep features, human error can be reduced and classification accuracy improved, which is crucial for the efficient detection and monitoring of plant diseases in field

conditions. This study identified ResNet101 as the most effective CNN-based feature extraction method among those tested. This model demonstrated the highest degree of discriminability and was more effective than traditional HC features in detecting FOV4 disease severity. This finding is consistent with previous research that has demonstrated the effectiveness of ResNet101 in various computer vision tasks, including object recognition and segmentation. To further confirm the claim of improvements being made by using CNNs, it is necessary to compare various deep learning techniques. However, it may not be appropriate to associate the previously reported methods with the current study as they have primarily been utilised to predict diseases from images captured under controlled conditions or using ground-based platforms (Chen et al., 2020; Haque et al., 2022; Saeed et al., 2021), which may not accurately capture subtle changes in disease symptoms. By leveraging UAS images in conjunction with CNN and BiLSTM models, this study can offer a more comprehensive understanding of crops than conventional ground-based methods. UAS images, which have high-resolution capabilities, can capture nuanced changes in crops that may go unnoticed by the human eye or proximal sensing techniques. Maes and Steppe (2019) also reported that high-resolution time-series UAS imagery, when combined with deep learning, can automate disease severity assessment, reducing the need for human intervention and enabling quicker and more accurate decisions.

While CNNs have demonstrated proficiency in extracting features from images, they lack the ability to capture the temporal dynamics of disease progression over time. This observation motivated us to incorporate BiLSTM model that considers multiple growth stages to address this limitation. The results of the comparison suggest that the BiLSTM model is a more effective approach than the SVM model for FOV4 disease severity detection, particularly when using deep learning-based features. The BiLSTM model consistently outperformed the SVM model in all classification metrics, regardless of the type of features used. This is because the BiLSTM captures the temporal dependencies within the UAS images, enabling the model to track the progression of the disease over time. This is critical as FOV4 disease severity can vary over time, and an accurate assessment of disease severity requires monitoring the temporal changes in the disease symptoms. In contrast, the SVM model is a static ML model that does not consider temporal dependencies. Although the challenges of multitemporal data analysis have been documented for land-cover observation and dynamic changes detection (Belward & Skoien, 2015), the use of BiLSTM and deep learning-based features for multitemporal analysis to estimate disease severity in plants has not been reported in the literature.

The accurate segmentation of plants from their background is crucial

for precise disease severity assessment. In this study, an advanced deep learning method was employed to segment cotton plants from their background. This approach notably improved the accuracy of various learning models in estimating plant disease severity. The findings indicated that image segmentation facilitated the extraction of relevant features from plant images while eliminating irrelevant background information, leading to more focused inputs for deep learning models. Consequently, CNN-BiLSTM models can better capture the spatial and temporal features of the disease, resulting in more robust and accurate predictions. The CNN-BiLSTM models with image segmentation demonstrated superior performance compared to the CNN-BiLSTM model without image segmentation.

The implications of these research findings have noteworthy practical implications for the assessment of cotton disease severity in real-world applications. The capability of the model to analyze datasets with inter-annual variations highlights its potential for use in other domains where data may vary over time. By reducing the distribution discrepancy between the source and target domains, the Resnet101-BiLSTM model becomes more robust and capable of accurately classifying cotton disease severity levels in unseen data, demonstrating superior performance in cross-dataset evaluation and suggesting that domain adaptation using CORAL can effectively mitigate the adverse effects of domain shift and enhance the usability and effectiveness of trained models in practical applications. The ability of the hybrid approach (i.e. Resnet101-CORAL-BiLSTM) to perform well on new and unseen data is crucial for deploying vision-guided autonomous field robots with variable-rate fungicides. However, future research could explore additional domain adaptation techniques and investigate their effectiveness in addressing the challenges posed by domain shift in agricultural applications. It would be beneficial to evaluate the performance of the proposed model on larger and more diverse datasets to ascertain its scalability and effectiveness across different geographical regions and cotton disease populations.

It is interesting to note that Class 2 was the most challenging to distinguish. This suggests that further improvements may be necessary in distinguishing between this class and other classes. Additionally, environmental factors, such as weather conditions, can result in misclassification. Therefore, it is crucial to consider these factors in developing and training ML models for crop disease severity assessment. The findings of this study emphasise the need for continued research and development in improving the accuracy and reliability of ML models for crop disease severity assessment. Additionally, further research is needed to explore the capabilities of the model and its limitations and to optimise its performance for other relevant applications.

5. Conclusions

In this study, a hybrid approach combining the strengths of CNNs and BiLSTM architectures was leveraged to assess the severity of cotton FOV4 disease using UAS images. The results showed that the hybrid approach outperformed single architecture or HC feature-based methods, indicating the significance of deep learning-based feature extraction techniques for complex disease severity evaluation in outdoor field conditions. The study also highlighted the effectiveness of the BiLSTM model in capturing temporal dependencies among UAS image datasets compared to static ML models like SVM, resulting in improved disease severity assessment accuracy. The Resnet101-BiLSTM model achieved the highest overall accuracy of 89.7%, demonstrating the efficacy of the hybrid CNN-BiLSTM approach in accurately estimating the severity of cotton FOV4 disease from UAS images. Furthermore, the application of the CORAL approach successfully aligned the distributions and reduced the domain shift, as evidenced by quantitative analysis using GK-MMD values. This alignment of distributions resulted in superior performance in classifying cotton diseases in the target domain, even when trained and tested on a different dataset. The overall accuracy of the model with CORAL was 72.7%, highlighting its high

generalisation capability and practical significance in real-world applications with diverse datasets. In summary, this study demonstrates that hybrid CNN-BiLSTM techniques combined with domain adaptation are effective for evaluating plant disease severity. Our findings provide a strong foundation for future research in this field, leading to the development of more effective and efficient disease monitoring and management.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was supported by Texas A&M AgriLife Research Strategic Research Initiative 'Fov4 in Texas Cotton' FY'20 FY'21.

References

- Abdalla, A., Cen, H., El-manawy, A., & He, Y. (2019). Infield oilseed rape images segmentation via improved unsupervised learning models combined with supreme color features. *Computers and Electronics in Agriculture*, 162, 1057–1068.
- Abdalla, A., Cen, H., Wan, L., Mehmood, K., & He, Y. (2021). Nutrient status diagnosis of infield oilseed rape via deep learning-enabled dynamic model. *IEEE Transactions on Industrial Informatics*, 17(6), 4379–4389.
- Abdalla, A., Cen, H., Wan, L., Rashid, R., Weng, H., Zhou, W., & He, Y. (2019). Fine-tuning convolutional neural network with transfer learning for semantic segmentation of ground-level oilseed rape images in a field with high weed pressure. *Computers and Electronics in Agriculture*, 167, Article 105091.
- Aqel, D., AlZu'bi, S., Mughaid, A., & Jararweh, Y. (2021). *Extreme learning machine for plant diseases classification: A sustainable approach for smart agriculture* (Vol. 25). Cluster Computing.
- Bao, W., Fan, T., Hu, G., Liang, D., & Li, H. (2022). Detection and identification of tea leaf diseases based on AX-RetinaNet. *Scientific Reports*, 12(1), 2183.
- Belward, A. S., & Skoien, J. O. (2015). Who launched what, when and why: trends in global land-cover observation capacity from civilian earth observation satellites. *ISPRS Journal of Photogrammetry and Remote Sensing*, 103, 115–128.
- Blasingame, D., & Patel, M. (2013). *Cotton disease loss estimate committee report*.
- Borhani, Y., Khoramdel, J., & Najafi, E. (2022). A deep learning based approach for automated plant disease classification using vision transformer. *Scientific Reports*, 12(1), Article 11554.
- Bouguettaya, A., Zarzour, H., Kechida, A., & Taberkit, A. M. (2023). A survey on deep learning-based identification of plant and crop diseases from UAV-based aerial images. *Cluster Computing*, 26(2), 1297–1317.
- 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, Article 105393.
- Crisóstomo de Castro Filho, H., Abílio de Carvalho Júnior, O., Ferreira de Carvalho, O. L., Pozzobon de Bem, P., dos Santos de Moura, R., Olino de Albuquerque, A., Rosa Silva, C., Guimarães Ferreira, P. H., Fontes Guimarães, R., & Trancoso Gomes, R. A. (2020). Rice crop detection using LSTM, Bi-LSTM, and machine learning models from sentinel-1 time series. *Remote Sensing*, 12(16), 2655.
- Divyant, L. G., Marzougui, A., González-Bernal, M. J., McGee, R. J., Rubiales, D., & Sankaran, S. (2022). Evaluation of effective class-balancing techniques for CNN-based assessment of Aphanomyces root rot resistance in pea (*Pisum sativum* L.). *Sensors*, 22(19), 7237.
- Eigen, D., & Fergus, R. (2015). *Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture*.
- Esgario, J. G., Krohling, R. A., & Ventura, J. A. (2020). Deep learning for classification and severity estimation of coffee leaf biotic stress. *Computers and Electronics in Agriculture*, 169, Article 105162.
- Haque, M. A., Marwaha, S., Deb, C. K., Nigam, S., Arora, A., Hooda, K. S., Soujanya, P. L., Aggarwal, S. K., Lall, B., Kumar, M., Islam, S., Panwar, M., Kumar, P., & Agrawal, R. C. (2022). Deep learning-based approach for identification of diseases of maize crop. *Scientific Reports*, 12(1), 6334.
- Harakannanavar, S. S., Rudagi, J. M., Puranikmath, V. I., Siddiqua, A., & Pramodhini, R. (2022). Plant leaf disease detection using computer vision and machine learning algorithms. *Global Transitions Proceedings*, 3(1), 305–310.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). *Deep residual learning for image recognition*.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.
- 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.
- Hwang, I.-C., Choi, D., Choi, Y.-J., Ju, L., Kim, M., Hong, J.-E., Lee, H.-J., Yoon, Y. E., Park, J.-B., Lee, S.-P., Kim, H.-K., Kim, Y.-J., & Cho, G.-Y. (2022). Differential diagnosis of common etiologies of left ventricular hypertrophy using a hybrid CNN-LSTM model. *Scientific Reports*, 12(1), Article 20998.

- Kaur, N. (2021). Plant leaf disease detection using ensemble classification and feature extraction. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 12 (11), 2339–2352.
- Krizhevsky, A., Sutskever, I., & Hinton, E. G. (2012). *ImageNet Classification with Deep Convolutional Neural Networks*, 25.
- Liu, J., Li, Q., Yang, H., Han, Y., Jiang, S., & Chen, W. (2019). Sequence fault diagnosis for PEMFC water management subsystem using deep learning with t-SNE. *IEEE Access*, 7, 92009–92019.
- Maes, W. H., & Steppe, K. (2019). Perspectives for remote sensing with unmanned aerial vehicles in precision agriculture. *Trends in Plant Science*, 24(2), 152–164.
- Nawaz, M., Nazir, T., Javed, A., Masood, M., Rashid, J., Kim, J., & Hussain, A. (2022). A robust deep learning approach for tomato plant leaf disease localization and classification. *Scientific Reports*, 12(1), Article 18568.
- Patil, R. R., Kumar, S., Chiwhane, S., Rani, R., & Pippal, S. K. (2023). An artificial-intelligence-based novel rice grade model for severity estimation of rice diseases. *Agriculture*, 13(1), 47.
- Qian, X., Zhang, C., Chen, L., & Li, K. (2022). Deep learning-based identification of maize leaf diseases is improved by an attention mechanism: Self-attention. *Frontiers in Plant Science*, 13, Article 864486. <https://doi.org/10.3389/fpls.2022.864486>
- Saeed, F., Khan, M. A., Sharif, M., Mittal, M., Goyal, L. M., & Roy, S. (2021). Deep neural network features fusion and selection based on PLS regression with an application for crops diseases classification. *Applied Soft Computing*, 103, Article 107164.
- Shi, T., Liu, Y., Zheng, X., Hu, K., Huang, H., Liu, H., & Huang, H. (2023). Recent advances in plant disease severity assessment using convolutional neural networks. *Scientific Reports*, 13(1), 2336.
- Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In *The 3rd international conference on learning representations (ICLR2015)*. <https://arxiv.org/abs/1409.1556>.
- Son, H. H. (2017). Toward a proposed framework for mood recognition using LSTM Recurrent Neuron Network. *Procedia Computer Science*, 109, 1028–1034.
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). *Rethinking the inception architecture for computer vision*.
- Tang, Z., Wang, M., Schirrmann, M., Dammer, K. H., Li, X., Brueggeman, R., Sankaran, S., Carter, A. H., Pumphrey, M. O., Hu, Y., Chen, X., & Zhang, Z. (2023). Affordable high throughput field detection of wheat stripe rust using deep learning with semi-automated image labeling. *Computers and Electronics in Agriculture*, 207, Article 107709.
- Thapa, R., Zhang, K., Snavely, N., Belongie, S., & Khan, A. (2020). The Plant Pathology Challenge 2020 data set to classify foliar disease of apples. *Applications in Plant Sciences*, 8(9), Article e11390.
- Ulloa, M., Huttmacher, R., Zhang, J., Schramm, T., Roberts, P. A., Ellis, M. L., Dever, J., Wheeler, T. A., Witt, T., Sanogo, S., Hague, S., Keely, M. P., Arce, J., Angeles, J., Hake, K., & Payton, P. (2023). Registration of 17 upland cotton germplasm lines with improved resistance to Fusarium wilt race 4 and good fiber quality. *Journal of Plant Registrations*, 17(1), 152–163.
- Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(11).
- Verma, S., Chug, A., Singh, R. P., Singh, A. P., & Singh, D. (2022). SE-CapsNet: Automated evaluation of plant disease severity based on feature extraction through Squeeze and Excitation (SE) networks and Capsule networks. *Kuwait Journal of Science*, 49(1), 1–31.
- Wang, G., Sun, Y., & Wang, J. (2017). Automatic image-based plant disease severity estimation using deep learning. *Computational Intelligence and Neuroscience*, 2017.
- Zeng, W., Li, M., Zhang, J., Chen, L., Fang, S., & Wang, J. (2018). *High-order residual convolutional neural network for robust crop disease recognition*.
- Zhang, J., Abdelraheem, A., Ma, J., Zhu, Y., Dever, J., Wheeler, T. A., Hake, K., Wedegaertner, T., & Yu, J. (2022). Mapping of dynamic QTLs for resistance to Fusarium wilt (Fusarium oxysporum f. sp. vasinfectum) race 4 in a backcross inbred line population of Upland cotton. *Molecular Genetics and Genomics*, 297(2), 319–332.
- Zhang, J., Zhu, Y., Elkins-Arce, H. D., Wheeler, T., Dever, J. K., Whitelock, D., Wedegaertner, T., Hake, K., & Bissonnette, K. (2022). Efficiency of selection for resistance to Fusarium wilt race 4 in cotton when conducted in the field versus greenhouse. *Euphytica*, 218(11), 165.
- Zhu, Y., Abdelraheem, A., Lujan, P., Idowu, J., Sullivan, P., Nichols, R., Wedegaertner, T., & Zhang, J. (2021). Detection and characterization of Fusarium wilt (Fusarium oxysporum f. sp. vasinfectum) race 4 causing Fusarium wilt of cotton seedlings in New Mexico. *Plant Disease*, 105(11), 3353–3367.
- Zhu, Y., Elkins-Arce, H., Wheeler, T. A., Dever, J., Whitelock, D., Hake, K., Wedegaertner, T., & Zhang, J. (2023). Effect of growth stage, cultivar, and root wounding on disease development in cotton caused by Fusarium wilt race 4 (Fusarium oxysporum f. sp. vasinfectum). *Crop Science*, 63(1), 101–114.
- Zhu, Y., Thyssen, G. N., Abdelraheem, A., Teng, Z., Fang, D. D., Jenkins, J. N., McCarty, J. C., Wedegaertner, T., Hake, K., & Zhang, J. (2022). A GWAS identified a major QTL for resistance to Fusarium wilt (Fusarium oxysporum f. sp. vasinfectum) race 4 in a MAGIC population of Upland cotton and a meta-analysis of QTLs for Fusarium wilt resistance. *Theoretical and Applied Genetics*, 135(7), 2297–2312.