ABSTRACT    Huanglongbing (HLB) is one of the most threatening diseases for citrus production and it has caused significant economic damage worldwide. Hence, computer-vision systems that are based on convolutional neural networks (CNNs) can detect HLB accurately. Moreover, the detection system should be able to discriminate between HLB and other citrus abnormalities to ensure that any treatments are effective. Besides, the causal pathogen of HLB is usually detected and diagnosed by the quantitative real-time polymerase chain reaction (qPCR) test, which is costly. Consequently, it is difficult to collect large datasets to train CNN-based systems. In this case, transfer learning from pre-trained CNNs is a solution for building an HLB-detection system using small-sized datasets. This paper evaluates two kinds of CNN architectures: series network (represented by AlexNet, VGG16, and VGG19 models) and directed acyclic graph (DAG) network (represented by ResNet18, GoogLeNet, and Inception-V3 models). These pre-trained CNNs are fine-tuned to distinguish HLB, healthy cases, and 10 kinds of abnormalities of the Citrus sinensis species, which is commonly known as sweet orange. The dataset includes 953 color images, where the leaf samples were collected from orange groves in north Mexico. The 10-fold cross-validation results show that all the CNNs present a 95% or higher HLB sensitivity. However, the number of trainable parameters impacts HLB detection more than the network’s depth. Specifically, VGG19, with 19 layers and 144 M parameters, reached a perfect sensitivity for all cross-validation experiments; whereas Inception-V3, with 48 layers and 24 M parameters, reached 95% sensitivity to HLB detection. This outcome happens because a higher number of parameters compensates for the limited number of HLB cases, so VGG19 can successfully transfer the learned characteristics to new cases. This study gives guidance when choosing an adequate CNN to efficiently detect HLB and other orange abnormalities. Besides, a detection scheme is proposed to be further implemented in a portable system to detect HLB in situ, potentially helping to reduce economic losses for small growers from low-income regions.
of citrus plants and it has been detected in South and North America, Asia, and South Africa. HLB’s visible symptoms include blotchy mottle leaves, stunted growth, yellow shoot, reduced fruit size, corky veins, root decline, and dieback [3].

The suspicious causal pathogen of this disease is a bacterium of the genus Candidatus. Three bacterial species are known to cause HLB: Ca. L. asiaticus, Ca. L. africanus, and Ca. L. americanus. Transmission of the bacteria occurs through an insect vector. Both Ca. L. asiaticus and Ca. L. americanus are transmitted by the Asian citrus psyllid Diaphorina citri, whereas Ca. L. africanus is transmitted by the African citrus psyllid Trioza erytreae [4].

Ca. L. asiaticus is the most pathogenic and widely distributed species [5]. HLB caused by this bacterium has been identified in several states of the USA (e.g., Florida, Texas, California, and Arizona), but also in South and Central America, Mexico, and the Caribbean [3]. The quantitative real-time polymerase chain reaction (qPCR) diagnostic test has been routinely used to detect and diagnose Ca. L. asiaticus bacterium [3].

Currently, there is no cure for HLB. Therefore, managing this disease requires prevention and control strategies that include physical, chemical, and biological methods. Meanwhile, HLB’s spread should be reduced by cultivating pathogen-free seedlings, removing diseased trees, and eradicating the Asian citrus psyllid [5].

The early detection of HLB is challenging because of its long asymptomatic stage. Consequently, efforts have been made to detect traces of the causal pathogen before HLB symptoms appear using methods such as gas chromatography, mass spectrometry, epifluorescence, and confocal microscopy [6], [7]. However, these approaches require specialized laboratories and professionals to handle the equipment, whose costs are often unfeasible to be afforded by small citrus producers in low-income regions.

Expert diagnosis is a popular way to identify HLB symptoms in situ. However, human diagnostic errors can be higher than 30% [8]. Hence, to reduce economic losses, citrus growers are currently looking for technologies that are capable of detecting HLB automatically and accurately, in a short time, and at a low cost, including image analysis and machine learning (ML) methods [9].

Conventional detection systems use image analysis techniques that incorporate prior domain knowledge into handcrafted features that are tailored by an engineer to describe HLB’s visual symptoms using shape, texture, color, and information about leaves and fruits [9]. These conventional systems help to differentiate between HLB-infected and healthy vegetal samples. Nevertheless, HLB coexists with other citrus abnormalities (i.e., diseases, nutritional deficiencies, and pests). This makes HLB detection more complex because treatments focus on ministering a specific abnormality [10]. Therefore, creating hand-crafted discriminant features to distinguish between the visual subtleties of citrus abnormalities is challenging. In addition, these features inherently depend on the designer’s expertise and skills.

Several automated plant disease detection systems have been developed with the rise of deep learning techniques for image classification [11]. In general, convolutional neural networks (CNNs) have shown superior generalization capacity when compared to conventional methods because CNNs are able to automatically learn high-level features from raw images [12], [13]. However, as reviewed in the next section, the use of CNNs still needs to be explored to detect HLB and other citrus-specific nutritional deficiencies, diseases, and pest symptoms. Additionally, the CNN models are usually unshared with the research community, limiting their reproducibility and improvement.

CNN architectures usually have many layers and convolution kernels, which dramatically increase the number of trainable parameters (i.e., convolution coefficients). This characteristic makes CNNs prone to overfitting when training from scratch using a small-sized dataset. In this regard, because qPCR is a costly diagnostic test to confirm HLB, collecting a large number of labeled images is challenging. Thus, this limitation should be considered when developing a CNN-based HLB-detection system. Transfer learning is a feasible alternative to overcome this issue, where a CNN model is trained on a large dataset and its learned parameters are then transferred to a smaller dataset through a fine-tuning procedure [14].

Given that HLB detection is critical to avoiding economic losses, mainly for small producers from low-income regions, this study aims to experimentally determine a CNN architecture using transfer learning that accurately detects 12 orange abnormalities, including HLB, which is one of the most threatening diseases for orange groves. We also propose an automatic detection system involving three main modules: 1) image acquisition of leaves with a portable studio, 2) automatic image segmentation, and 3) abnormality detection with a CNN model.

According to our literature review, detecting 12 orange abnormalities has not been performed because current CNN-based HLB-detection approaches detect up to four classes. Besides, the proposed detection system can potentially be implemented in a portable system to detect HLB and other abnormalities in situ.

The significance of this work is to potentially reduce the economic losses of small citrus producers from low-income regions by introducing a CNN-based technological alternative to detect HLB, and other pathologies and deficiencies that affect orange trees.

II. LITERATURE REVIEW

Table 1 shows an overview of current detection systems based on computer vision for classifying HLB and other abnormalities. In most works, conventional detection systems are proposed, although more CNN-based systems have recently emerged. The general pipeline of these methods is as follows.

First, the input image is drawn from an image acquisition system, which is often based on RGB cameras to obtain color images of leaves and fruits. Specialized acquisition systems
TABLE 1. HLB detection systems based on computer vision, where \( n \) is the number of images in the dataset, and \( c \) is the number of classes. They are sorted in descending order regarding the number of classified abnormalities. The symbols ‘♣’ and ‘•’ indicate that the dataset includes leaf and fruit samples, respectively.

| Imaging system                              | \( n \) | \( c \) | System type | Features                       | Classifier         | Accuracy | Ref  |
|--------------------------------------------|--------|--------|-------------|--------------------------------|--------------------|----------|------|
| Color camera                               | 190 (♣) | 6      | Conventional| Texture                      | Random forest      | 81%      | [15] |
| Color camera                               | 580 (♣, •) | 5     | Conventional| Color, texture, and shape     | SVM                | 90%      | [16] |
| Color camera                               | 199 (♣) | 4      | Conventional| Color and texture             | Bagged tree        | 99%      | [17] |
| Color and hyperspectral camera             | 1325 (♣) | 4     | CNN-based   | Learned-fusion features       | Auxiliary classifier| 97%      | [18] |
| Polarizing lens filter                      | 60 (♣)  | 4      | Conventional| Intensity statistics          | SVM                | 97%      | [19] |
| Fluorescence and multispectral camera       | 1489 (♣) | 4     | CNN-based   | Learned features              | Softmax layer      | 96%      | [20] |
| Color camera                               | 477 (♣) | 4      | CNN-based   | Learned features              | Softmax layer      | 94%      | [21] |
| Color camera                               | 1200 (♣) | 4     | CNN-based   | Learned features              | Softmax layer      | 92%      | [22] |
| Hyperspectral camera                        | 2592 (♣) | 3     | Conventional| Texture                      | SVM                | 92%      | [23] |
| Polarization camera                         | 220 (♣) | 2      | Conventional| Texture and intensity         | Random forest      | 96%      | [24] |
| Fluorescence spectroscopy                   | 200 (♣) | 2      | Conventional| Texture                      | SVM                | 95%      | [25] |
| Hyperspectral camera                        | 180 (♣) | 2      | Conventional| Spectrum features             | SVM                | 93%      | [26] |
| Fluorescence spectroscopy                   | 200 (♣) | 2      | Conventional| Texture                      | ANN                | 92%      | [27] |
| Color camera                               | 898 (♣) | 2      | Conventional| Color and texture             | SVM                | 92%      | [28] |
| Color camera                               | 800 (♣) | 2      | Conventional| Color and texture             | SVM                | 92%      | [29] |
| Color camera                               | 106 (♣) | 2      | CNN-based   | Learned features              | Softmax layer      | 90%      | [30] |

(e.g., fluorescence spectroscopy and hyperspectral cameras) can also be used to obtain images that enhance the reflectance properties of vegetal samples.

Next, in a conventional detection system, image classification is performed in two stages. The first stage extracts from the input image a set of hand-crafted features (e.g., color and texture) to create a feature vector that is classified in the second stage by an ML method (e.g., a support vector machine (SVM) or artificial neural network (ANN)) that determines the type of abnormality.

In contrast, in a CNN-based detection system, image classification is performed in a single process in which learned high-level features are extracted from the input image through several convolutional layers, and an output softmax layer gives the posterior probabilities to distinct classes of abnormalities.

Generally, conventional detection systems consider samples with HLB, healthy, other diseases, and nutritional deficiencies. However, it is common to train the detection system with two superclasses: HLB-positive and HLB-negative [24], [28]. In multiclass classification, two-stage methods have been proposed that first detect if a sample is normal or abnormal. In the case of abnormality, the classification between different diseases is performed in the second stage [17], [19]. Other approaches have performed multiclass classification in a single stage to differentiate between several classes of abnormalities and healthy cases, where the number of classes ranges from three to six [15], [16], [23].

On the other hand, CNN-based systems usually detect four classes, including HLB, healthy, mineral deficiencies (e.g., magnesium, nitrogen, and zinc) [18], [20], [22], and other diseases (e.g., canker and greasy spot) [21]. Moreover, only one CNN-based approach distinguishes between HLB-positive and HLB-negative classes in orange crops [30].

It can be seen in Table 1 that conventional and CNN-based systems were evaluated using particular sets of images with a different number of samples and classes, and therefore a paired comparison between methods would be unfair. Consequently, in this study, we directly compare CNN-based models and a conventional method based on hand-crafted color and texture features using the same image dataset. Moreover, we increased the classification to 12 classes of orange leaves, which is the largest number of classes to be considered so far.

It is worth mentioning that CNNs have been widely used to detect a wide range of diseases in many crops. Recent works use the public PlanVillage dataset (up to 58 classes) to train CNN-based disease detection systems [31], [32], [33], [34]. Although these works present relevant advances in detecting diseases in a wide variety of plants using CNNs, the PlanVillage dataset only includes HLB images—it includes no other orange diseases, nutritional deficiencies, and pests whose symptoms could overlap with HLB. Consequently, it is crucial to distinguish between HLB and other kinds of abnormalities to apply an adequate treatment. Thus, because CNN models have demonstrated a remarkable classification performance, it is convenient to develop CNN-based methods by considering other citrus diseases, pest symptoms, and nutritional deficiencies that are endemic to the citrus-producing region.

III. PROPOSED APPROACH

A. DETECTION SYSTEM SCHEME

Figure 1 shows a block diagram of the proposed HLB detection scheme, which includes three modules: 1) image acquisition, which is a portable studio with controlled illumination that is used to acquire leaf images in situ; 2) segmentation, which is a region of interest that is automatically obtained from the inner part of the leaf; and 3) CNN model, in which leaf classification is performed to discriminate between 12 classes of abnormalities, including HLB. The technical details of the modules are given in the following sections.

B. IMAGE ACQUISITION AND LEAF DATASET

The leaf samples that are used in this study were collected from orange trees of *Citrus sinensis* (L.) Osbeck species. The sample collection required the support of experts from the
State Plant Health Committee of the states of Tamaulipas and San Luis Potosi in Mexico to identify abnormalities in orange groves and then transport the samples to the laboratory for the qPCR test to confirm HLB. Because the sample collection is subject to the experts’ availability, the vegetal samples were collected in three batches: August 2016, March 2017, and March 2019. The sample collection date does not affect the results of this study because the symptoms are preserved, regardless of the time of year.

The protocol to obtain the image dataset consists of the following steps:

1) Cut leaf samples in full development from four branches of the orange tree.
2) Take color images of samples with an RGB camera using an image acquisition system with controlled lighting and dark background.
3) Place the samples on absorbent paper towels and put them in sealed plastic bags with their corresponding identification number.
4) Transport the samples in ice chests to the Molecular Detection Laboratory (Empacadora Santa Engracia, Tamaulipas, Mexico).
5) Perform the diagnosis of HLB (caused by Ca. L. asiaticus) using qPCR analysis by following the protocol of the National Phytosanitary Reference Center of the General Directorate of Plant Health (CNRF, Mexico) [35].

In the second step of the collection protocol, the samples were photographed in situ using a portable studio. Every leaf sample is placed on the dark bottom of the box, which is internally illuminated by white LEDs. The box is closed with a lid with a hole for the camera lens, as shown in Figure 2. This image acquisition system has two purposes: 1) to control the illumination conditions and 2) to contrast the leaf’s colors from the dark background. Hence, these elements help the segmentation method to automatically define a global intensity threshold, as detailed in Section III-C. The pictures were shot with a Samsung Galaxy phone (Samsung Electronics, Suwon, South Korea), model SM-J730 Pro, with 13 megapixels, without zoom and flash, and saved in JPEG format.

The generated dataset that is used in this study comprises 953 leaf samples and 12 classes, including distinct diseases, nutritional deficiencies, and pest symptoms (as summarized in Table 2). The observed class imbalance reflects the frequency of the abnormalities that are detected in the orange groves. Additionally, Figure 3 shows representative samples of each abnormality for didactic purposes to illustrate the expected differences in texture and color patterns that are present on the leaves.

### C. REGION OF INTEREST SEGMENTATION

An image segmentation method was devised to automatically extract the maximum inscribed circular region of interest (ROI) inside the leaf. Because a CNN has a fixed-size input layer (as summarized in Table 3), the ROI’s width and height are scaled to fit the specific dimensions of the input layer. Therefore, a circular ROI allows us to preserve the leaf patterns’ aspect ratio when uniform scaling is performed to fit the CNN’s input.
The proposed ROI segmentation is based on the Otsu method to efficiently separate the leaf from its background and then detect the maximum inscribed circle within the leaf region. The Otsu method determines the optimum gray-level threshold that maximizes class variance between the foreground and background pixels. Hence, the image histogram should present a bimodal distribution to find an adequate threshold between modes [36]. In this regard, we evaluate different color spaces to determine the best chromatic component that produces a well-contrasted image such that the Otsu method provides a good leaf segmentation. The following color spaces are evaluated: RGB, HSV, YCbCr, LAB, LUV, and YIQ. Every component of these color spaces is considered separately to obtain an intensity image that is used as an input to the Otsu method. Finally, the Intersection over Union (IoU) metric (also known as the Jaccard index) compares the segmented image versus its reference, which is determined by the manual outlining of the leaf region.

The experiments on the 953 leaf samples of the dataset revealed that the B channel of the LAB color space is the best option, with an IoU value of 0.991 ± 0.006. We refer to the reader to Appendix to consult the complete results for all the evaluated chromatic channels. Based on these results, the proposed automatic ROI segmentation method follows the pipeline shown in Figure 4, which comprises the following steps:

1) Get an image of the leaf sample using the portable studio, as shown in Figure 2.
2) Convert the input image from RGB to LAB color space to decorrelate the luminance component from the chrominance components.
3) Separate the chromatic B component and reduce the noise with a Gaussian filter with \( \sigma = 1 \) and a \( 5 \times 5 \) pixel kernel size.
4) Apply the Otsu method to obtain the binary mask of the leaf.
5) Calculate the distance map of the binary mask [37].
6) Obtain the maximum distance value \( d_{\text{max}} \) from the distance map.
7) Extract the \((x, y)\) coordinates in the distance map with values \( d_{\text{max}} \).
8) Average the \((x, y)\) coordinates to get the centroid \((\bar{x}, \bar{y})\) of the circular ROI.
9) Calculate the Euclidean distance from the point \((\bar{x}, \bar{y})\) to all pixels in the image.
10) Obtain the points whose Euclidean distance is less or equal to \( d_{\text{max}} \) to define the circular ROI.
11) Mask and crop the input RGB image to obtain the ROI that a CNN will classify.

**D. CNN ARCHITECTURES AND TRANSFER LEARNING**

A CNN consists of multiple layers that perform convolution and pooling operations to extract feature maps, which are similar to neurons in a biological brain’s primary visual cortex [38]. As the CNN’s depth increases (i.e., the number of layers augments), the feature maps’ dimension gradually decreases to activate more subtle features. Finally, a fully connected layer with the softmax function performs the classification task [12]. Thus, CNN automatically learns high-level features to discriminate between different objects.

Currently, CNN architectures can be divided into two main types [39]:

- Series network: this architecture has a strictly sequential path, where single layers are arranged one after the other from input to output, as shown in Figure 5(a).
- Directed acyclic graph (DAG) network: this architecture allows path bifurcations, where layers have inputs from multiple layers and outputs to multiple layers, as shown in Figure 5(b).
TABLE 3. Characteristics of the pre-trained CNN models that are used in this study. Symbols † and ‡ indicate that the CNN is a series or DAG network, respectively.

| CNN model   | Input size | Number of layers | Trainable parameters | Ref. |
|-------------|------------|------------------|----------------------|------|
| AlexNet†    | 227 × 227 × 3 | 8                | 60 M                 | [40] |
| VGG16‡      | 224 × 224 × 3 | 16               | 138 M                | [44] |
| VGG19†      | 224 × 224 × 3 | 19               | 144 M                | [44] |
| ResNet18‡   | 224 × 224 × 3 | 18               | 11 M                 | [45] |
| GoogLeNet‡  | 224 × 224 × 3 | 22               | 7 M                  | [46] |
| Inception-V3‡ | 299 × 299 × 3 | 48               | 24 M                 | [47] |

Training very deep series networks is computationally infeasible because the number of parameters grows rapidly as the depth increases. Thus, series networks have less depth and more parameters than DAG network architectures. In contrast, DAG networks have fewer parameters and are deeper than the series networks, which allows the efficient training of very deep CNN models.

When the size of the training set is limited, designing a CNN from scratch has the risk of overfitting. This happens because the CNN contains many trainable parameters and in small datasets there is a higher chance of finding a solution that fits the training set but does not generalize well in new images. In this scenario, a CNN trained with millions of images can be reused to perform a new classification task, which is called transfer learning. It has been observed that the first layers learn general features, similar to Gabor filters, regardless of the training dataset. Therefore, in transfer learning, the convolutional layers of a pre-trained CNN are kept to extract features, whereas the last fully connected layer is replaced with a new one matching the number of classes in the new classification problem. Next, a fine-tuning procedure is performed to update the network’s parameters [14].

The characteristics of the six pre-trained CNNs that are used in this study are summarized in Table 3. These CNNs were initially trained to classify 1000 classes of objects in the ImageNet dataset [40]. To reuse these deep convolutional networks in our classification problem, the last fully connected layer and the softmax layer are replaced to match the 12 classes of orange leaves that are summarized in Table 2. The stochastic gradient descent algorithm then performs the fine-tuning of CNN parameters during 100 training epochs with a learning rate of $1 \times 10^{-4}$ and a momentum factor of 0.9. The mini-batch size is set to 16. At the input layer, the z-score standardization adjusts the distribution of pixel values in input images, which facilitates the activation and gradient descent progress [41].

Online data augmentation was used to reduce overfitting by applying random geometric transformations to the training set, including scaling, reflections, rotations, and translations. This procedure assumes that more information can be extracted from the original dataset through augmentation [42], [43].

IV. CONVENTIONAL HLB DETECTION SYSTEM

In a conventional HLB detection system, features are defined by an engineer, considering that they are discriminant for distinguishing different classes. These hand-crafted features are then calculated from an input image to form a feature vector that feeds a classifier [48].

The abnormality detection system based on hand-crafted features shown in Figure 6 is implemented for comparative purposes. This approach extracts 216 texture and color features using four methods that are usually employed to build HLB detection systems:

1) Ranklet co-occurrence matrix (RCM): Intensity invariant GLCM-based (gray-level co-occurrence matrix) features have been used to detect HLB in orange leaves. First, the input RGB image is converted to a gray-scale image, from which the ranklet transform is calculated. Four resolutions (2, 4, 8, and 16) and three orientations (horizontal, vertical, and diagonal) are considered to obtain 12 ranklet images. Next, a fine-tuning procedure is performed to update the network’s parameters [14].

The characteristics of the six pre-trained CNNs that are used in this study are summarized in Table 3. These CNNs were initially trained to classify 1000 classes of objects in the ImageNet dataset [40]. To reuse these deep convolutional networks in our classification problem, the last fully connected layer and the softmax layer are replaced to match the 12 classes of orange leaves that are summarized in Table 2. The stochastic gradient descent algorithm then performs the fine-tuning of CNN parameters during 100 training epochs with a learning rate of $1 \times 10^{-4}$ and a momentum factor of 0.9. The mini-batch size is set to 16. At the input layer, the z-score standardization adjusts the distribution of pixel values in input images, which facilitates the activation and gradient descent progress [41].

Online data augmentation was used to reduce overfitting by applying random geometric transformations to the training set, including scaling, reflections, rotations, and translations. This procedure assumes that more information can be extracted from the original dataset through augmentation [42], [43].

2) Auto-mutual information (AMI): Similar to the RCM method, 12 ranklet images are obtained from the original gray-scale image. Next, the normalized mutual information measures the similarity between successive displacements of a single ranklet image, which is performed in the image’s horizontal and vertical directions. Hence, a total of 24 AMI-based features are calculated [49].

3) Local binary patterns variance (LBPV): Traditional local binary patterns (LBP) have been explored to detect HLB in citrus leaves [28]. However, in this study, we use LBPV, which is an improved version of the traditional LBP method that reduces the feature vector’s dimensionality and incorporates local contrast information. Radius sizes of one (eight neighbors) and two (12 neighbors) are considered to obtain 24 LBPV-based features [50].

4) Histogram statistics: The original RGB image is converted to the HSI color space. Next, eight histogram statistics (e.g., mean, standard deviation, entropy, etc.)
are calculated from each HSI image channel. Hence, a total of 24 color-based features are calculated [28].

A Multilayer Perceptron (MLP) network with two hidden layers is trained to classify the 12 classes of orange leaves, as shown in Table 2. The input layer has 216 nodes that distribute the input feature vector to the first hidden layer. The optimal number of hidden nodes is determined by minimizing the validation error under a grid search scheme, where the search ranges are [5, 165] and [5, 125] for the first and second hidden layers, respectively, and where the maximum number of nodes corresponds to the 75% of nodes in the previous layer. The output layer is a softmax function with 12 nodes (one node per each class of orange leaves). The MLP network is trained with the backpropagation algorithm with a learning rate of 0.01, a momentum factor of 0.9, and a maximum number of training epochs of 1000. Moreover, before the training procedure, the texture features are rescaled to the range $[-1, 1]$ by the softmax normalization to reduce the influence of extreme feature values [51].

V. CLASSIFICATION ASSESSMENT

For a classification problem with $c$ classes, the corresponding confusion matrix $C$ is a square matrix of size $c$-by-$c$ whose $ij$th entry $C_{ij}$ is the number of elements of the actual class $i$ that have been assigned to class $j$ by the classifier.

Accuracy is probably the most frequently used measure to evaluate the overall effectiveness of a classifier, which is expressed by [52]

$$\text{ACC} = \frac{\text{tr}(C)}{n}, \quad (1)$$

where $\text{tr}(\cdot)$ is the trace operator and $n$ is the total number of test observations. The accuracy should tend toward unity to indicate an adequate success rate.

Because the leaf image dataset presents class imbalance, the accuracy tends to be optimistic due to the high hit rate of the majority class. Thus, to deal with imbalanced classes, the Matthews correlation coefficient (MCC) is used [53]:

$$\text{MCC} = \frac{n \cdot \text{tr}(C) - \sum_{kl} C_{kl} C_{lk}}{\sqrt{n^2 - \sum_{kl} C_{kl} (C^T)_{lk} n^2 - \sum_{kl} (C^T)_{kl} C_{lk}}}, \quad (2)$$

where $C_{kl}$ is the $k$th row of $C$, $C_{lk}$ is the $l$th column of $C$, and $C^T$ is the transpose of $C$. This index should tend toward unity to indicate an adequate classification performance.

HLB detection capability can also be evaluated by dividing the image dataset into positive (HLB-infected) and negative (HLB-negative) classes. The latter includes all of the orange leaf classes except the HLB cases. In this case, sensitivity (SEN) and specificity (SPE) are calculated as [52]

$$\text{SEN} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (3)$$

and

$$\text{SPE} = \frac{\text{TN}}{\text{TN} + \text{FP}}, \quad (4)$$

where TP is a true positive, TN is a true negative, FN is a false negative, and FP is a false positive. SEN and SPE indices measure the classifier’s effectiveness in identifying positive and negative classes. They should tend toward unity to indicate an adequate classification performance.

The 10-fold cross-validation method creates disjoint training and test sets [54]. To determine statistical differences between methods, McNemar’s test ($\alpha = 0.05$) is used to check the disagreements between any two classification methods, where the null hypothesis is that the predictive performance of two models is equal [55]. Finally, the Holm-Bonferroni method performs the correction for multiple comparisons (i.e., post-hoc analysis) [56].

The computing platform employed an Intel i9 processor with eight cores at 3.60 GHz, 64 GB of RAM, and a graphics card NVIDIA GeForce RTX 2070. All of the programs were developed in MATLAB R2020a (The Mathworks, Boston, Massachusetts, USA).

VI. RESULTS

Figure 7 shows the classification performance results of the evaluated CNNs, including the conventional method based on hand-crafted features and an MLP classifier. It is observed that the VGG19 network obtained the best performance for the classification of 12 classes of orange leaves with $\text{ACC}=0.991$ and $\text{MCC}=0.990$. In contrast, the MLP-based method obtained the lowest classification performance with $\text{ACC}=0.940$ and $\text{MCC}=0.935$. It is also notable that all of the CNNs seem to have a similar classification performance. To verify if there are significant differences in accuracy between the methods, McNemar’s test, with Holm-Bonferroni correction, revealed that AlexNet ($p = 0.0073$) and ResNet18 ($p = 0.0349$) performed statistically significantly differently from VGG19. Besides, the conventional method presented significant differences with all of the CNNs ($p < 0.001$). However, the other pairwise comparisons between CNNs did not present significant differences ($p > 0.05$). Table 4 summarizes the $p$-values of all pairwise comparisons between the classification methods.
GoogLeNet has modules called “inception” with stacked 1 × 1 convolutions, which reduces the number of trainable parameters.

In general, all the evaluated CNNs presented an acceptable performance in classifying 12 classes of orange leaves, with accuracy values greater than 97%. However, the number of network parameters has a noticeable impact on HLB detection. The VGG19 network, with 144 million parameters, reached a perfect sensitivity: all HLB cases for all cross-validation experiments were classified correctly. The VGG16 network (with 138 million parameters) was the second-best method for detecting HLB because it reached an average sensitivity of 97.5%. In contrast, Inception-V3, with 48 layers and 24 million parameters, obtained an average sensitivity of 95%, which is the same level as AlexNet achieved with eight layers and 60 million parameters.

The number of trainable parameters has a more direct relationship with HLB detection than the network’s depth. This behavior could happen because a higher number of parameters compensates for the limited number of HLB cases, so VGG19 can successfully transfer the learned characteristics to new cases. Therefore, increasing the number of HLB cases could improve the networks’ classification performance with fewer trainable parameters.

Concerning the conventional method with hand-crafted features and an MLP classifier, it is notable that the accuracy depends on the quality of texture and color features extracted from the images. Determining these features is human-dependent, and therefore this procedure involves subjectiveness. This study obtained 216 texture and color features from four description methods that were previously used for HLB detection. This strategy improved HLB detection results concerning our previous work [15], in which 144 RCM-based features were used. Consequently, the sensitivity of HLB detection increased from 83% to 92.5%. This result indicates that combining different feature description methods improves HLB detection. However, achieving competitive results between hand-crafted feature-based and CNN-based approaches is challenging. Nevertheless, an advantage of conventional classification systems is that they need fewer computational resources than CNNs, which require GPU-based platforms to calculate thousands of convolutions. Therefore, further research should evaluate and even create other hand-crafted features that could improve the classification performance at a lower computational cost.

Notably, hand-crafted features with the MLP classifier achieved classification results around those of the other literature methods shown in Table 1. However, as expected,
the experiments revealed that CNN-based methods achieved higher classification performance. Incorporating various types of nutritional deficiencies, pest symptoms, and diseases of orange leaves in the classification system could potentially decrease human errors due to the confusion of HLB characteristics with some other citrus abnormalities that can be treated with remedies. In addition, the proposed classification system can detect diseases that are typical of the citrus-producing region. In this study, we used leaf samples from the North of Mexico; hence, the generated CNN models involve a limited set of instances of the universe of orange abnormalities. Fortunately, it is feasible to improve the network models with new samples and classes through transfer learning. All of the CNN models generated in this study are available on request from the authors to fulfill this purpose. Besides, our image dataset is publicly available in [57].

VIII. CONCLUSION

This paper presented a comparative study of six different pretrained CNNs to classify 12 different abnormalities of orange leaves of the *Citrus sinensis* species, including HLB, nutritional deficiencies, and pest symptoms. In addition, a conventional method based on hand-crafted features and MLP was evaluated. From the experimental results, it is concluded that the best network is VGG19, which obtained an overall accuracy of 99% in detecting 12 classes of orange leaves. Moreover, VGG19 reached a sensitivity of 100% in detecting HLB-positive cases. In general, it was observed that the classification quality was better when the number of trainable parameters was higher. Conversely, the conventional method based on texture and color features presented the lowest classification performance, which demonstrates the difficulty of extracting subtle features with high discriminant power among classes.

This study gives guidance for choosing an adequate CNN to efficiently detecting HLB. We also provide an alternative solution for small citrus producers to detect in situ HLB and other orange abnormalities. Our future work considers implementing the proposed CNN-based detection system in an embedded system-on-module such as NVIDIA's Jetson family for the in-field detection of orange tree abnormalities.
We also plan to include more leaf samples and other kinds of citrus abnormalities. Furthermore, the proposed method can be feasibly extended to other citrus crops affected by HLB, such as lemon, lime, and grapefruit.

APPENDIX
LEAF SEGMENTATION RESULTS
Figure 10 shows the Intersection over Union (IoU) results obtained with the Otsu method when the components of six color spaces are binarized independently (i.e., RGB, HSV, LAB, YCbCr, LUV, and YIQ). The boxplots concentrate the IoU results on the 953 images of the dataset. The best result is obtained by the B component of the LAB color space.

ACKNOWLEDGMENT
The authors would like to thank the Universidade Autónoma de Tamaulipas and Cinvestav Unidad Tamaulipas for their economic support.

REFERENCES
[1] M. Khan and F.-U.-D. Razi, “Citrus greening disease (huanglongbing): a perilous threat to global citrus industry,” J. Horticulture, vol. 5, no. 3, p. 1000, 2018.
[2] S. H. Futch, Identification of Mites, Insects, Diseases, Nutritional Syndromes and Disorders on Citrus. Gainesville, FL, USA: Univ. Florida, IFAS Extension, 2011.
[3] S. Kogenaru, Q. Yan, N. Riera, C. Roper, X. Deng, T. Ebert, M. Rogers, M. Irey, G. Pietersen, C. Rush, and N. Wang, “Repertoire of novel sequence signatures for the detection of Candidatus Liberibacter asiaticus by quantitative real-time PCR,” BMC Microbiol., vol. 14, no. 39, pp. 1–11, Feb. 2014.
[4] N. Killiny, “Made for each other: Vector-pathogen interfaces in the huanglongbing pathosystem,” Phytopathology, vol. 112, no. 1, pp. 26–43, Jan. 2022.
[5] X. Li, H. Kuang, C. Zhou, X. Meng, and W. Chen, “Controlling citrus huanglongbing: Green sustainable development route is the future,” Frontiers Plant Sci., vol. 12, Nov. 2021, Art. no. 760481.
[6] A. A. Aksenov, A. Pasamontes, D. J. Peirano, W. Zhao, A. M. Dandekar, O. Fiehn, R. Ehsani, and C. E. Davis, “Detection of huanglongbing disease using differential mobility spectrometry,” Anal. Chem., vol. 86, no. 5, pp. 2481–2488, Feb. 2014.
[7] S. S. Pandey and N. Wang, “Targeted early detection of citrus huanglongbing causal agent ‘Candidatus Liberibacter asiaticus’ before symptom expression,” Phytopathology, vol. 109, no. 6, pp. 952–959, Jun. 2019.
[8] F. M. V. Pereira, D. M. B. P. Miliori, E. R. Pereira-Filho, A. L. Venâncio, M. D. S. T. Russo, M. C. D. B. Cardinini, P. K. Martins, and J. Freitas-Astúa, “Laser-induced fluorescence imaging method to monitor citrus greening disease,” Comput. Electron. Agric., vol. 79, no. 1, pp. 90–93, Oct. 2011.
[9] Z. Iqbal, M. A. Khan, M. Sharif, J. H. Shah, M. H. U. Rehman, and K. Javed, “An automated detection and classification of citrus plant diseases using image processing techniques: A review,” Comput. Electron. Agric., vol. 153, pp. 12–32, Oct. 2018.
[10] J. G. A. Barbedo, “A review on the main challenges in automatic plant disease identification based on visible range images,” Biosyst. Eng., vol. 144, pp. 52–60, Apr. 2016.
[11] A. Kamilaris and F. X. Prenafeta-Boldu, “Deep learning in agriculture: A survey,” Comput. Electron. Agric., vol. 147, pp. 70–90, Aug. 2018.
[12] Y. LeCun, Y. Bengio, and G. E. Hinton, “Deep learning,” Nature, vol. 521, pp. 436–444, Dec. 2015.
[13] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.
[14] M. Hussain, J. J. Bird, and D. R. Faria, “A study on CNN transfer learning for image classification,” in Advances in Computational Intelligence Systems, vol. 840, A. Lotfi, H. Bouchachia, A. Gegov, C. Langensiepen, and M. McGinnity, Eds. Cham, Switzerland: Springer, 2019, pp. 191–202.
[15] W. Gómez-Flores, J. J. Garza-Saldaña, and S. E. Varela-Fuentes, “Detection of huanglongbing disease based on intensity-invariant texture analysis of images in the visible spectrum,” Comput. Electron. Agric., vol. 162, pp. 825–835, Jul. 2019.
[16] M. Sharif, M. A. Khan, Z. Iqbal, M. F. Azam, M. I. U. Lali, and M. Y. Javed, “Detection and classification of citrus diseases in agriculture based on optimized weighted segmentation and feature selection,” Comput. Electron. Agric., vol. 150, pp. 220–234, Jul. 2018.
[17] H. Ali, M. I. Lali, M. Z. Nawaz, M. Sharif, and B. A. Saleem, “Symptom based automated detection of citrus diseases using color histogram and textural descriptors,” Comput. Electron. Agric., vol. 138, pp. 92–104, Jun. 2017.
[18] D. Yang, F. Wang, Y. Hu, Y. Lan, and X. Deng, “Citrus huanglongbing detection based on multi-modal feature fusion learning,” Frontiers Plant Sci., vol. 12, Dec. 2021, Art. no. 809506.
[19] A. Pourreza, W. S. Lee, R. Ehsani, J. K. Schueller, and E. Raveh, “An optimum method for real-time in-field detection of huanglongbing disease using a vision sensor,” Comput. Electron. Agric., vol. 110, pp. 221–232, Jan. 2015.
[20] C. He, X. Li, Y. Liu, B. Yang, Z. Wu, S. Tan, D. Ye, and H. Weng, “Combining multicolor fluorescence imaging with multispectral reflectance imaging for rapid citrus huanglongbing detection based on lightweight convolutional neural network using a handheld device,” Comput. Electron. Agric., vol. 194, Mar. 2022, Art. no. 106808.
[21] S. F. Syed-Ab-Rahman, M. H. Hesamian, and M. Prasad, “Citrus disease detection and classification using end-to-end anchor-based deep learning model,” Int. J. Speech Technol., vol. 52, no. 1, pp. 927–938, Jan. 2022.
[22] J. Berger, C. A. Preussler, and J. P. Agostini, “Identificación de síntomas de huanglongbing en hojas de cítricos mediante técnicas de deep learning,” in Proc. 10th Congreso de AgroInformática, Buenos Aires, Argentina, 2018, pp. 90–103.
[23] H. Weng, J. Ly, H. Cen, M. He, Y. Zeng, S. Hua, H. Li, Y. Meng, H. Fang, and Y. He, “Hyperspectral reflectance imaging combined with carbohydrate metabolism analysis for diagnosis of citrus huanglongbing in different seasons and cultivars,” Sens. Actuators B, Chem., vol. 275, pp. 50–60, Dec. 2018.
[24] Q. Xu, J.-R. Cai, W. Zhang, J.-W. Bai, Z.-Q. Li, B. Tan, and L. Sun, “Detection of citrus huanglongbing (HLB) based on the HLB-induced leaf starch accumulation using a home-made computer vision system,” Biosystems Eng., vol. 218, pp. 163–174, Jun. 2022.
[25] C. B. Wetterich, R. F. de Oliveira Neves, J. Belasque, and L. G. Marcassa, “Detection of citrus canker and huanglongbing using fluorescence imaging spectroscopy and support vector machine technique,” Appl. Opt., vol. 55, no. 2, pp. 400–407, Jan. 2016.
[26] K. Wang, D. Guo, Y. Zhang, L. Deng, R. Xie, Q. LV, S. Yi, Y. Zheng, Y. Ma, and S. He, “Detection of huanglongbing (citrus greening) based on hyperspectral image analysis and PCR,” Frontiers Agric. Sci. Eng., vol. 6, no. 2, pp. 172–180, Jun. 2019.
[27] C. B. Wetterich, R. F. de Oliveira Neves, J. Belasque, R. Ehsani, and L. G. Marcassa, “Detection of huanglongbing in Florida using fluorescence imaging spectroscopy and machine-learning methods,” Appl. Opt., vol. 56, no. 1, pp. 15–23, Jan. 2017.

FIGURE 10. IoU results for the components of six color spaces. At the top is the mean IoU value of 953 images in the dataset.
W. Gómez-Flores et al.: HLB Detection Method for Orange Trees Based on Deep Neural Networks and Transfer Learning

[28] X. Deng, Y. Lan, T. Hong, and J. Chen, “Citrus greening detection using visible spectrum imaging and C-SVC,” *Comput. Electron. Agricult.*, vol. 130, pp. 177–183, Nov. 2016.

[29] D. Xiaoling, Y. Lan, X. Xiaojing, M. Huilan, L. Jiakai, and H. Tiansheng, “Detection of citrus huanglongbing based on image feature extraction and two-stage BPNN modeling,” *Int. J. Agric. Biol. Eng.*, vol. 9, no. 6, pp. 20–26, Nov. 2016.

[30] C. T. Sonin, S. Fellah, and M. R. Abid, “Citrus greening infection detection (CiGiD) by computer vision and deep learning,” in *Proc. 3rd Int. Conf. Inf. Syst. Data Mining (ICSDM)*, New York, NY, USA, 2019, pp. 21–26.

[31] M. Brahim, M. Arsenovic, S. Laraba, S. Sladojevic, K. Boughalfa, and A. Moussauini, “Deep learning for plant diseases: Detection and saliency map visualisation,” in *Human and Machine Learning*, J. Zhou and F. Chen, Eds., Cham, Switzerland: Springer, 2018, pp. 93–117.

[32] K. P. Ferentinos, “Deep learning models for plant disease detection and diagnosis,” *Comput. Electron. Agricult.*, vol. 145, pp. 311–318, Feb. 2018.

[33] G. Geetharamani and A. Pandian, “Identification of plant leaf diseases using a nine-layer deep convolutional neural network,” *Comput. Electr. Eng.*, vol. 76, pp. 323–338, Jun. 2019.

[34] S. H. Lee, H. Goeau, P. Bonnet, and A. Joly, “New perspectives on plant disease characterization based on deep learning,” *Comput. Electron. Agricult.*, vol. 170, Mar. 2020, Art. no. 105220.

[35] W. Li, J. S. Hartung, and L. Levy, “Quantitative real-time PCR for detection and identification of candidatus liberibacter species associated with citrus huanglongbing,” *J. Microbiol. Methods*, vol. 66, no. 1, pp. 104–115, Jul. 2006.

[36] N. Otsu, “A threshold selection method from gray-level histograms,” IEEE Trans. Syst., Man, Cybern., Syst., vol. SMC-9, no. 1, pp. 62–66, Feb. 1979.

[37] R. Kimmel, N. Kiryati, and A. M. Bruckstein, “Sub-pixel distance maps and weighted distance transforms,” *J. Math. Imag. Vis.*, vol. 6, nos. 2–3, pp. 223–233, Jun. 1996.

[38] G. W. Lindsay, “Convolutional neural networks as a model of the visual system: Past, present, and future,” *J. Cognit. Neurosci.*, vol. 33, no. 10, pp. 2017–2031, Sep. 2021.

[39] M. L. Richter, J. Schoning, A. Wiedenroth, and U. Krumm, “Should you go deeper? Optimizing convolutional neural network architectures without training,” in *Proc. 20th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2021, pp. 964–971.

[40] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.

[41] Y. Zhong, F. Fei, Y. Liu, B. Zhao, H. Jiao, and L. Zhang, “SatCNN: Satellite image dataset classification using agile convolutional neural networks,” *Remote Sens. Lett.*, vol. 8, no. 2, pp. 136–145, Oct. 2017.

[42] J. A. Pandian, G. Geetharamani, and B. Annette, “Data augmentation on plant leaf disease image dataset using image manipulation and deep learning techniques,” in *Proc. IEEE 9th IACC*, Tiruchirappalli, India, Dec. 2019, pp. 199–204.

[43] C. Shorten and T. M. Khoshgoftaar, “A survey on image data augmentation and two-stage BPNN modeling,” in *Proc. 3rd Int. Conf. Inf. Syst. Data Mining (ICSDM)*, Tiruchirappalli, India, Dec. 2019, pp. 20–26, Nov. 2016.

[44] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[45] D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 2818–2826.

[46] N. O’Mahony, S. Campbell, A. Carvalho, S. Harapanahalli, G. V. Hernandez, L. Krpalkova, D. Riordan, and J. Walsh, “Deep learning vs. traditional computer vision,” in *Advances in Computer Vision*, K. Arai and S. Kapoor, Eds., Cham, Switzerland: Springer, 2020, pp. 128–144.

[47] W. Gómez-Flores, A. Rodríguez-Cristerna, and W. C. de Albuquerque Pereira, “Texture analysis based on auto-mutual information for classifying breast lesions with ultrasound,” *Ultrasound Med. Biol.*, vol. 45, no. 8, pp. 2213–2225, Aug. 2019.

[48] Z. Guo, L. Zhang, and D. Zhang, “Rotation invariant texture classification using LBP variance (LBPV) with global matching,” *Pattern Recognit.*, vol. 43, no. 3, pp. 706–719, Mar. 2010.

[49] K. L. Priddy and P. E. Keller, *Artificial Neural Networks: An Introduction*. Bellingham, WA, USA: SPIE, 2005.

[50] M. Sokolova and G. Lapalme, “A systematic analysis of performance measures for classification tasks,” *Inf. Process. Manage.*, vol. 45, no. 4, pp. 427–437, Jul. 2009.

[51] J. Gorodkin, “Comparing two K-category assignments by a K-category correlation coefficient,” *Comput. Biol. Chem.*, vol. 28, no. 5, pp. 367–374, Dec. 2004.

[52] Z.-H. Zhou, *Ensemble Methods: Foundations and Algorithms*. Boca Raton, FL, USA: CRC Press, 2012.

[53] T. G. Dietterich, “Approximate statistical tests for comparing supervised classification learning algorithms,” *Neural Comput.*, vol. 10, no. 7, pp. 1895–1923, Oct. 1998.

[54] S. Holm, “A simple sequentially rejective multiple test procedure,” *Scandin. J. Statist.*, vol. 6, no. 2, pp. 65–70, Jan. 1979.

[55] W. Gómez-Flores, J. J. Garza-Saldana, and S. E. Varela-Fuentes, *Citrus UAT Dataset: An Image Dataset of Orange Leaves for Disease Detection*. Accessed: Nov. 5, 2022. [Online]. Available: https://www.tamps.cinvestav.mx/~wgomez/uatdataset.html

WILFRIDO GÓMEZ-FLORES (Member, IEEE) received the B.S. degree in electronics and communications engineering from the Instituto Tecnológico de Ciudad Madero, in 1994, the M.Sc. and D.Sc. degrees in electrical engineering from the Centro de Investigación y de Estudios Avanzados del IPN (Cinvestav), in 2006 and 2009, respectively. Since 2010, he has been a Full-Time Professor with the Cinvestav, Unidad Tamaulipas. His main research interests include digital image analysis and pattern recognition.

JUAN JOSE GARZA-SALDANA received the B.S. degree in computer systems engineering from the Instituto Tecnológico de Ciudad Madero, in 1994, the M.Sc. degree in computer sciences from the Centro de Investigación y de Estudios Avanzados del IPN (Cinvestav), Unidad Tamaulipas, in 2009, and the Ph.D. degree in gestión y transferencia del conocimiento from Universidad Autónoma de Tamaulipas, in 2018. Since 2014, he has been a Full-Time Professor and a Researcher with the Faculty of Engineering and Sciences, Universidad Autónoma de Tamaulipas. His main research interests include digital image analysis, pattern recognition, computer systems security, mobile application development, and web application development.

SÓSTENES EDMUNDO VARELA-FUENTES received the B.S. degree in agronomy engineering from the Facultad de Engineering and Sciences, Autonomous University of Tamaulipas, and the master’s and Ph.D. degrees in agricultural sciences in agricultural parasitology from the ITESM Monterrey Unit. He is a Research Professor with the Faculty of Engineering and Sciences, Autonomous University of Tamaulipas. His research interests include agricultural parasitology and entomology, integrated pest, and pesticide management.

**IEEE Access**

VOLUME 10, 2022

116696