

TABLE IV
COMPARISON OF DIFFERENT TYPES OF MODELS [42]

Model name	Input parameters	Performance of the models	Model complexity	Scalability with respect to time and geographical location	Recommended applications	References
Linear Regression Model	-One variable such as vegetation index or reflectance values -One response variable such as chlorophyll measurement	Low-high (depends upon the input variable)	Low	Low	-To choose the range of predictor variables and to identify the input which performs best -To remove noise in the training data	[289–291]
Advanced Linear Regression Model	-Multiple inputs -Multiple response variable/output variable	Medium-high	Medium	Low	-To include spectral and temporal information -Hyperparameter tuning can result in optimal performance	[292–296]
Machine Learning based Model	-Multiple input/predictor variables -Multiple response variables -number of trees/loss function (depends upon the selected model)	Medium-high	Medium	Low	-To include spectral and temporal information -Hyperparameter tuning can result in optimal performance	[74, 161, 297, 298]
Deep Learning based Model	-Multiple input variables -Based on the input of the size of the model is determined -Size refers to the number of layers and weights -Pre-trained models can be used (if available)	Medium-high	High	High	-To employ a large number of training data/samples to prevent overfitting/underfitting issues -Optimize the model to obtain an optimal performance	[299–301]

spectral features [263], [264]. RF is an ensemble-based learning method that combines multiple decision trees to produce robust and accurate classification or regression results. RF can be employed for plant disease detection to classify hyperspectral data into healthy or diseased classes [265], [266]. kNN is a nonparametric-based algorithm for classification and regression tasks. For plant disease detection, it compares the plant sample with its neighboring features to identify whether the sample is infected or not [266], [267]. SAM is an algorithm that classifies hyperspectral imagery based on the angle between the spectra. Based on the spectral angle difference, pixels are classified into categories such as healthy or diseased plants [268]. Maximum likelihood classification is a probabilistic classifier that assigns a pixel to the class that has the maximum likelihood. It can be used for plant disease detection to classify hyperspectral data into categories like healthy or diseased [269]. On the other hand, PLSR is a statistical technique that can model the relationship between hyperspectral data and plant disease severity [270]. Dirichlet aggregation regression uses Dirichlet distributions for modeling compositional data, and it can also be used to model the relationship between hyperspectral data and plant disease severity [271]. Logistic regression is used for cases with binary classification, such as disease or no disease detection [272], whereas multiple linear regression can provide multiple classes classification [273].

The machine-learning-based classification models can be broadly categorized into supervised and unsupervised learning techniques. Supervised learning techniques require labeled training images/samples to detect the pixel of the region of interest (ROI). A label is associated with each ROI to map the class categories. The labeling of the images is visually conducted and sent to the classification model for training. This aids in improving the detection accuracy of the models but at the

expense of high labor costs. On the other hand, unsupervised learning techniques require no prior knowledge and learning through clustering and distribution rules of the spectral features in the images. An example of unsupervised learning would be the K-means clustering algorithm. Rice sheath blight diseases were detected using the K-means clustering algorithm in [274]. A study by Yuan et al. [275] proposed an unsupervised learning technique to identify anthracnose diseases in tea leaves. The study could achieve around 96% of accuracies for disease detection [274].

Besides machine learning models, deep learning, a subset of machine learning algorithms, is now applied for hyperspectral image analysis. Over the past decades, deep learning algorithms have improved image classification performance, object detection, and so on. Recently, remote sensing image processing algorithms have also been developed to process complex images such as multispectral and hyperspectral images. Some popular algorithms are CNN [276], deep belief networks [277], stacked autoencoder [278], and so on. Table IV compares different models used in image processing techniques for hyperspectral imaging.

B. Applications of Hyperspectral Imaging in Plant Disease Detection

Once the image acquisition is completed and the images are processed through objective-based selected models, the next step is to analyze the models. The term objective-based selected models refer to the analytical models chosen based on the specific goals or objectives such as disease detection and identification, classification, severity, and genetic resistance for the diseases [166], [292], [293]. The analyses are carried out in different levels and arrangements. For instance, to observe