Journal of Computer Science

Research Article

A Data Mining Driven Framework for Image-Based Crop Stress Analysis Using IPCA, FP-Growth, ConvLSTM, and Sequential Models

A. Josephine and A. Subhashini

Department of Computer Science, PSG College of Arts and Science, Coimbatore- 641014, India

Article history
Received: 24-04-2026
Revised: 07-07-2026
Accepted: 15-07-2026

Corresponding Author:
A. Josephine
Department of Computer Science, PSG College of Arts and Science, Coimbatore- 641014, India
Email: joseantonymca@gmail.com

Abstract: Detecting crop stress before it becomes serious has a major impact on the food supply. It helps in avoiding the loss of crops, increases productivity and helps farmers to practice precision farming. Most technologies today are designed specifically to either detect crop stress or to predict how well the crop will grow. Many of the technologies do not combine the crop's spatial, spectral, and temporal data into a single framework by using an integrated assessment of the overall health of the crop. This paper presents a new framework, a data mining driven framework for image-based crop stress detection and yield prediction. The framework consists of three phases: 1) A new image processing framework utilizing Adaptive Wiener Filtering was developed for improving the quality of images by removing noise and maintaining the important features of the crops. 2): The Independent Principal Component Analysis (IPCA) method for extracting discriminative spectral and textural features from the images of the crops. 3) The association rule mining framework, FP-Growth, were utilized to determine the co-occurrence relationship of the extracted features. A Convolutional Long Short-Term Memory (ConvLSTM) network is then used to identify spatiotemporal patterns in order to classify the crop's stress anomalies. A Hidden Markov Model (HMM) was employed to measure the rate of progress of the crop stress over time, and Multi-Criteria Decision Analysis (MCDA) were used to calculate an overall score of the severity of the crop's stress level, thus enabling the farmer to make informed decisions. The development of this framework was based on an augmented dataset of wheat disease including 5000 images of the five specific classes of crop stress. Through the use of the framework, the results produced a classification accuracy of 98.9%, with an F1-score of 98.7%. The proposed framework outperformed other conventional classification models including CNN, Random Forest, LSTM, and SVM based on performance metrics.

Keywords: Crop Stress, Yield Prediction, IPCA, ConvLSTM, Hmm, MCDA

Introduction

The agricultural industry serves an important part when it comes to guaranteeing world food security and advancing sustainable development. However, global agricultural crop management faces challenges in the form of climate changes, the increase in population, unstable rain pattern, soil degradation, pests, nutrient deficiencies, availability of water, etc., (Burchard-Levine et al., 2025; Park et al., 2024). The various biochemical and biophysical problems had a negative impact on the plant development process resulting in crop quality reduction and yield losses (Mohamed et al., 2024). In light of the findings obtained from modern agricultural studies, one of the most effective means to decrease the agricultural losses and improve agricultural management is timely detection of crop stress (Malounas et al., 2024; Islam et al., 2024). In recent times, the current state of the world agriculture has experienced significant changes as a result of technological advances in remote sensing technology, drones, IoT sensors, multi-spectral imaging, and artificial intelligence (Qiao et al., 2021; Jhajharia et al., 2023). The use of high-quality aerial and satellite images provides the possibility to monitor crops throughout the growing seasons (Gavahi et al., 2021). These findings allow for gathering the required spectral, spatial, and temporal data that help to reflect the changes which take place in plant physiology even when there are no visible signs that show something wrong with crop production (Lee et at al., 2024).

At the same time, artificial intelligence, as well as machine learning technologies, has proven to be highly effective in detecting the symptoms of crop problems, estimating yield accurately (Velázquez-González et al., 2024). Research has been conducted between the years 2021 and 2025 to demonstrate the capability of artificial intelligence for agricultural monitoring. Deep learning models such as CNN, LSTM, ResNet, EfficientNet, YOLO, and attention networks have provided great results in crop disease detection and stress classification (Sharma et al., 2023; Chandraprabha and Dhanraj, 2023). Meanwhile, hybrid learning models have also made it possible to combine remote sensing with environmental data to have a better yield prediction (Gautam and Rani, 2022). Nevertheless, the major problem with these approaches is that they focus either on crop stress classification or crop yield prediction alone (Hemamalini et al., 2025; Kamarudin et al., 2023). Moreover, there are also a number of methods that have relied on one image feature and did not optimize the relations between spectral, textural, spatial, and temporal features (Chandel et al., 2025; Wen and Ma, 2025). Another important limitation of the existing methods is that they do not effectively apply data mining approaches in agriculture. Even though the deep learning models can provide high level of accuracy, they are also regarded as black-box systems (Burchard-Levine et al., 2025). The relation of one's crop parameters such as characteristics of the crops, surrounding environment, and stress indicators is not studied through association mining or pattern recognition. In addition, most of these studies concentrate on the health of the crop only at a certain moment without taking into consideration the evolution of stress conditions (Fevgas et al., 2025; Kapari et al., 2025).

The lack of complete stress assessment makes everything about these systems even less applicable in practice than it would otherwise be. To address this problem, this paper proposes a complete data mining approach to determine crop stress and predict yields based on images obtained during crop planting season. The approach suggested integrates image processing, feature extraction, association rule mining, deep learning, time series analysis, and multi-criteria decision making under one roof. At first, Adaptive Wiener Filtering is used to improve image quality. Secondly, Independent Principal Component Analysis (IPCA) captures useful features by lowering redundancy. Thirdly, FP-Growth algorithm is applied to detect correlation between features in order to find associations that can indicate plant stress. After this stage, the model, based on Convolutional Long Short-Term Memory technology (ConvLSTM), maintains time and space correlation between the data to identify patterns in the way crops grow. In order to boost early crop reliability, early crop yield crop yield association. By exposing hidden correlations between crop attributes, the combination of deep learning and data mining approaches improves both interpretability and forecast accuracy. Additionally, temporal modelling makes it possible to continuously monitor the evolution of stress, which helps farmers and agricultural specialists to optimize resource usage and carry out prompt interventions. The suggested framework outperforms traditional machine learning and deep learning techniques in classification, according to experimental evaluation on the Wheat Disease dataset.

The contributions of this paper are realized here:

- This work employs IPCA for efficient extraction of discriminative spectral and textural features, improving the accuracy of stress characterization in large-scale agricultural fields
- This work applies FP-Growth–based association rule mining to uncover frequent co-occurrence relationships among extracted features, enabling deeper understanding of factors contributing to crop stress
- This work leverages a ConvLSTM-based anomaly detection model to identify unusual or stress-indicative crop patterns over time, combining both spatial and temporal context for superior detection performance
- This work integrates HMM based sequential pattern mining to capture temporal transitions in crop conditions and predict stress progression under dynamic environmental scenarios
- This work incorporates MCDA to compute a holistic stress severity score by fusing multiple variables, offering a structured and interpretable assessment for decision-makers

Literature Review

Joshi et al. (2022) developed RiceBioS, a handheld device with deep learning capabilities that provides real-time stress detection from images of rice on site. RiceBioS utilizes Edge-as-a-Service (EaaS) to determine whether images of rice plants are healthy or stressed, as well as the types of infections present (fungal infection, in this case, rice blast; and bacterial infection, here, bacterial leaf blight, or BLB) by using a "pruned" model and the following techniques: Automated Region of Interest detection, adaptive thresholding, and feature extraction based on a hierarchical masking system.

Hamidon and Ahamed (2022) used the automated deep-learning technique of a single-stage object detection model to automatically find symptoms of tip burn. Multiple images taken in a variety of LED lighting conditions, including red, blue, and white lights, were collected, and a1080322333 images were augmented, allowing for three detectors to be trained: CenterNet, YOLOv4, and YOLOv5.

Mohammadi and Asefpour (2023) created a new method for identifying salinity- and drought-induced stresses in various species of cucumber plants. This method combines information about plant morphology (e.g., photographs of leaf samples), physiological and biochemical traits (e.g., relative water content, chlorophyll concentration, catalase enzyme activity, levels of anthocyanins, phenols, and proline), and microRNA markers (e.g., miRNA-156a, miRNA-166i, miRNA-399g, and miRNA-477b) to produce accurate classifications of plant stress using machine learning models consisting of support vector machine models that were optimized for maximum performance through a combination of genetic algorithms and particle swarm optimization algorithms.

Azimi et al. (2021a) conducted research using deep learning methods to analyze changes in chickpeas' visual representations caused by water stress. They developed a Convolutional Neural Network (CNN)/Long Short-Term Memory (LSTM) pipeline, training it with images taken at three different time periods (early, middle, late) up until five Months post-stress using two varieties of chickpeas (JG-62 and Pusa-372). The authors identified a set of spatiotemporal features that allowed their CNN-LSTM to successfully classify water stressed chickpeas.

Azimi et al. (2021b) published research utilizing a phenotyping methodology based on shoot pictures to quantify plant stress response using deep learning techniques. They reported the effective use of a 23-layered deep neural network, which achieved similar accuracy levels to several advanced models, such as the ResNet18 and NASNet, that had far greater parameter counts.

Wang et al. (2024) published research in measuring four poplar seedlings with differing resistance to drought via five levels of stress, utilizing an autonomous phenotyping platform outfitted with both RGB-D and multispectral cameras to acquire side profile RGB/depth images and top profile multispectral images. Each seedling's color, texture, depth, spectral, and physiological characteristics were documented, and plant variety identification and drought stress diagnosis were performed using LSTM-based multi-label classification models and ResNet18- based LSTM/CBAM-LSTM classification models.

Yang et al. (2025) introduced a deep learning architecture to use RGB leaf pictures to determine the degree of salt stress in soybean seedlings. The model performs dimensionality reduction to eliminate redundancy, combines a CNN with a Convolutional Block Attention Module to extract features, and employs a machine learning classifier to predict stress levels.

Abdel-Salam (2024) used a hybrid model integrating advanced feature selection with an optimized Support Vector Regressor (SVR) is proposed as a framework for crop yield prediction. The design was to create a framework whereby the data is preprocessed, whereby the FMIG-RFE feature selection is incorporated to promote the prediction process by utilizing an Improved Crayfish Optimization Algorithm (ICOA) to optimize SER hyperparameters.

Subramaniam and Marimuthu (2024) proposed an efficient deep learning-based framework for classification and dimensionality reduction of agricultural products in India. This model is a three-step procedure which includes the gathering, preprocessing as well as the dimensionality contraction of the data pertaining to the South Indian crops. In the preprocessing stage, the data obtained are normalized and cleansed and dimensionality reduced by using Squared Exponential Kernel PCA (SEKPCA). Also, a Weight Tuned Deep Convolutional Neural Network (WTDCNN) will be used to forecast the harvests of the most profitable crops.

Nagesh et al. (2024) aimed to develop a successful Machine Learning model for yield prediction based on Precision Agriculture techniques. Various algorithms were tested and a complete automated framework of yield prediction was founded on the most relevant features. The extraction of features was done with the help of the Grey Level Co-occurrence Matrix (GLCM) and classification followed the AdaBoost Decision Trees, Artificial Neural Networks (ANN), and K Nearest Neighbor (KNN) techniques. The dataset consisted of 33 attributes pertaining to yield, rainfall, pesticide application and temperature of major crops all over the world which were retrieved through the Food and Agricultural Organization of the United Nations (FAO) and World Data Bank.

Crop stress detection and crop output prediction have been greatly enhanced by recent developments in artificial intelligence, deep learning, remote sensing, and data mining. In terms of feature representation, temporal modeling, interpretability, and thorough decision support, current methods still have a number of drawbacks. The representative studies, their methods, benefits, and drawbacks are compiled in Table 1.

The comparative study shows that recent developments have made significant advances in crop stress detection and yield estimation using techniques such as machine learning and deep learning. CNN provides good spatial feature extraction; techniques based on LSTM improve the analysis of the temporal sequence. Attention mechanisms and remote sensing technology allow raising accuracy of classification. Not less important are machine learning methods based on optimization. There are some important research issues, despite the advances achieved.

Table 1: Representative studies on crop stress detection and yield prediction: Methodologies, merits, and limitations

Author	Methodology	Merits	Limitations
Joshi et al. (2022)	Edge AI with CNN for rice stress detection	Real-time stress identification with low computational cost	Limited to rice crops and biotic stress; no temporal analysis
Hamidon and Ahamed (2022)	YOLOv4, YOLOv5, CenterNet	Fast object detection with high localization accuracy	Focused only on visible symptoms; no stress progression modelling
Mohammadi and Vakilian (2023)	SVM with GA and PSO optimization	High classification accuracy using physiological and molecular features	Requires laboratory measurements and cannot support large-scale field monitoring
Azimi et al. (2021a)	CNN-LSTM	Captures spatial and temporal crop characteristics	Evaluated only for water stress; limited interpretability
Wang et al. (2024)	ResNet18-LSTM with multisource images	Utilizes RGB, depth, and multispectral information	Does not explore feature association or sequential stress evolution
Yang et al. (2025)	CNN with CBAM	Improved feature extraction using attention mechanisms	Limited to salt stress and single crop species
Abdel-Salam et al. (2024)	Feature Selection with Optimized SVR	Accurate crop yield prediction using optimized regression	Does not incorporate image-based stress detection
Subramaniam and Marimuthu (2024)	SEKPCA with WTDCNN	Efficient dimensionality reduction and crop classification	Ignores temporal crop behaviour and association analysis
Nagesh et al. (2024)	GLCM with AdaBoost	Simple implementation and good prediction performance	Lower robustness under complex agricultural environments

The majority of the presented approaches consider either crop stress detection or crop yield estimation but not both tasks combined. Another problem is that many models focus only on machine learning based on deep learning, while using data mining methodology would allow revealing new relations between the stress degree and the rest of stress-related issues. The models developed have limited explanatory power because they almost never utilize association rule mining or consider severity level of stress as a whole. Moreover, even though some recurrent models utilize temporal parameter, they do not analyze the long-term sequence of stress based on probabilistic learning method. Furthermore, the previous models apply traditional feature extraction techniques, which leads to higher amount of unneeded information and increases computational complexity.

Problem Identification

From the above literature review, it is quite evident that the existing crop monitoring systems suffer from many deficiencies. First, they do not have an integrated framework which can utilize spatial, spectral, temporal, and contextual data for a complete crop health monitoring process. Second, most of the conventional approaches do not include advanced feature extraction, association rule mining, anomaly detection model, sequential pattern mining, and decision-support systems into one comprehensive system. Third, the existing systems hardly have a good interpretability and low capability for the early stress severity assessment, which significantly limits their application within precision agriculture. Another gap is that the temporal evolution of crop stress is rarely modeled, which makes it impossible to predict future stress under the dynamic environmental conditions. Based on the above-mentioned gaps, it is obvious that there is a need for a new framework which would include Adaptive Wiener Filtering, Independent Principal Component Analysis (IPCA), FP-Growth association rule mining, ConvLSTM-based anomaly detection, Hidden Markov Models (HMM), and Multi-Criteria Decision Analysis (MCDA).

Novel Contributions

Even though much research has been done on using ML/DL techniques for detecting crop stress and predicting crop yield, most of the methods that currently exist focus mainly on either crop image classification or yield/disease prediction as standalone tasks. The proposed framework is a new, integrated, data mining-based architecture for detecting crop stress and predicting crop yield that allows for feature extraction, pattern discovery, temporal anomaly detection, sequential modeling of different types of stressors, and multi-criteria decision analysis to be performed in one combined functional process. The uniqueness of this framework comes from the way IPCA, FP-Growth, ConvLSTM, HMM, and MCDA are seamlessly integrated, allowing for a thorough analysis of spatial, spectral, and temporal characteristics of crops. The proposed framework has taken the approach of combining explainable, data mining techniques with deep learning in order to enhance the interpretation capability of the data and provide a more complete understanding of crop stress detection and yield estimation. By integrating MCDA into the proposed framework, a structured method for assessing the level of stress by using aggregated data from multiple stress indicators has been introduced, instead of depending solely on the classification probabilities of stress and/or yield models for that assessment. Since both crop stress detection and yield estimation can now be done together in the proposed framework, it provides farmers with sufficient time to evaluate stress progression at an early stage and use this information to make timely agricultural decisions.

Proposed Methods

Crop stress detection and accurate yield prediction are indispensable constituents of precision agriculture that allow for the early detection of plant health issues and reliable production outcome forecasts. Modern techniques analyze spatial, spectral, and temporal data to help in continuous monitoring of crop conditions and in making effective decisions to improve agricultural productivity. However, farmers are mainly faced with several challenges, such as subtle or late visible symptoms of stress, variable weather conditions, noisy remote-sensing data, and limitations of traditional manual monitoring. These problems often culminate in delayed interventions, yield estimates inaccuracies, and inefficient resource management that create difficulties regarding sustaining stable and high crop productivity. This work proposes ways to overcome some of the challenges discussed by integrating multisource image data using advanced machine learning, deep learning, and data-mining techniques for early detection of stress indicators and modeling complex temporal patterns affecting yield. These intelligent methods can be applied to enhance the accuracy and provide real-time insights that could support proactive decisions, thus enhance crop health management and ensure sustainable yield prediction. The overall proposed architecture is depicted in Fig. 1.

The framework's processing pipeline contains sequential modules, each with the output of the previous module as an input to the next. The input acquired crop images go through a preprocessing stage to remove noise, improve contrast, normalize pixel intensity, and standardize dimensions.

Next, the preprocessed images are processed in the IPCA module (to extract compact spectral and textural features). The extracted features are sent to the FP-Growth algorithm to mine for frequently occurring feature associations related to crop's stress.

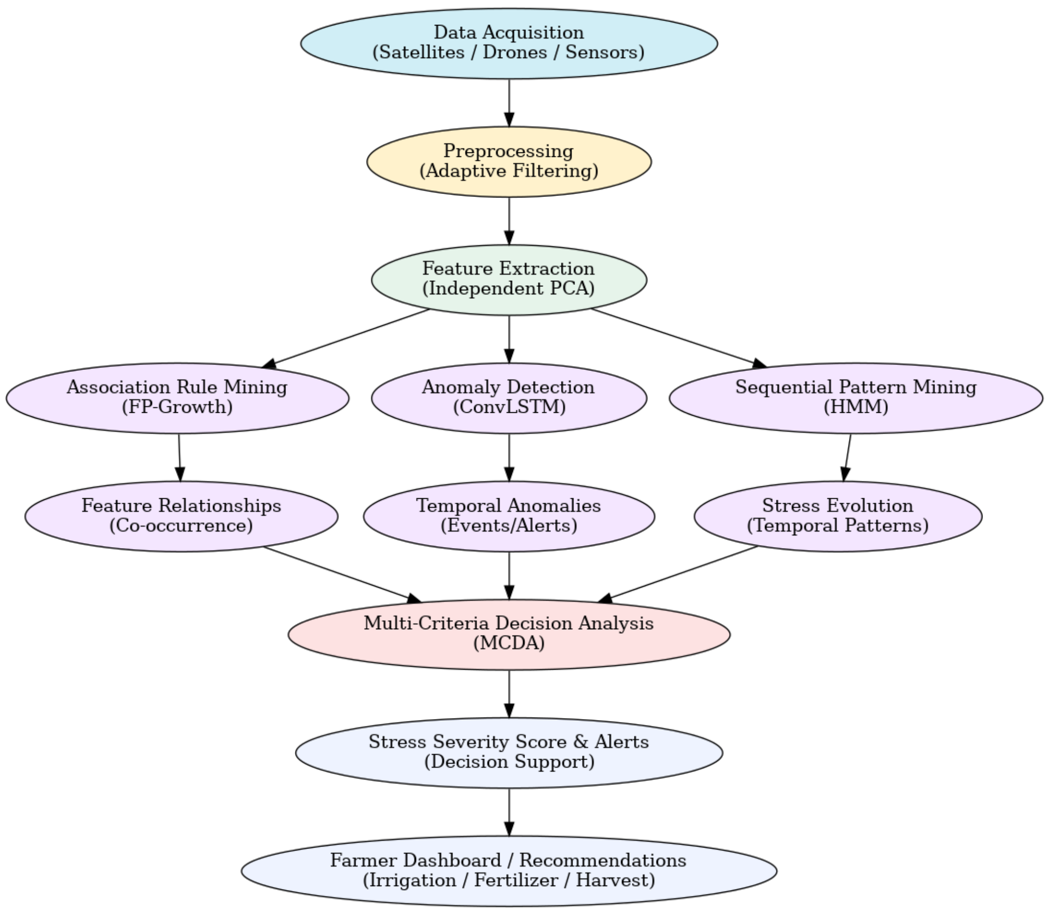

Fig. 1: Overall Proposed Architecture

The mined associations are then processed with the ConvLSTM model to output detected anomalies and passed on to the HMM to link stress progression/changes through time. Lastly, the temporal predictions and extracted stress indicators will be used with the MCDA model to compute the stress severity score for accurate crop stress assessment and yield prediction.

Pre-Processing

Preprocessing the raw captured crop images using (field sensors, drones, and/or satellites) to enhance their overall quality, as well as provide more uniformity for use in the analytic applications will take place. The initial step in conducting this procedure is the resizing of all images to a pixel size of 224×224 for both standardization of spatial resolution and reducing computationally intensive requirements for training the machine learning models. This resizing allows all of the raw images from the crop field to be equalized in their dimensions while still preserving key details about the types of stress that crops undergo. Subsequent to resizing, an Adaptive Wiener Filter will remove noise that results from both atmospheric disturbances and sensor limitations along with illumination variances within the raw images while retaining important details about edges and textures. After this step, a contrast-enhancement operation will be completed to allow the visual identification of healthy and stressed areas of crops in the images through increased contrast of the images themselves, as well as enhanced visibility of the various details that indicate where there are lesions, discoloration, or other types of texture changes. To provide robustness of the model to inconsistencies that may result from variations in lighting, an RGB Normalization will be performed using $X_{norm} = X/255$. This operation will also provide a remedy for any potential variances due to brightness in crop images as well. In addition to the above operations, data augmentation procedures (rotation, flipping, scaling, cropping, and brightness changes) will be completed on the crop images to create additional unique images for inclusion within the overall dataset and to reduce the likelihood that any one crop image will be the primary cause of overfitting to one of the training models. Once this preprocessing has occurred for all raw crop images, they will then be available for feature extraction using an IPCA-based method to allow for the extraction of discriminating features from the data. The filtered output at pixel $(i, j)$is computed using:

$$\hat{X}(i,j) = \mu_L + \frac{\sigma_L^2 - \sigma_n^2}{\sigma_L^2}(I(i,j) - \mu_L) \quad (1)$$

Where, $I(i,j)$is the observed noisy pixel,$\mu_L$and σ_L^2represent the local mean and local variance in the neighborhood window,σ_n^2 is the estimated noise variance, and $\hat{X}(i,j)$is the noise-reduced pixel value.This adaptive formulation ensures that regions with high variance (edges or textured crop areas) are minimally smoothed, while homogeneous regions receive stronger noise reduction.

As a result, the pre-processed images exhibit improved clarity and enhanced feature integrity, providing a robust foundation for subsequent feature extraction, pattern mining, and stress detection within the proposed agricultural analysis pipeline.

Feature Extraction

Independent Principal Component Analysis is used as a robust feature extraction technique in the proposed framework to extract discriminative spectral and textual features out of high-resolution agricultural images. In contrast to the classical PCA, which only maximizes the variance in the description of the complex data, IPCA expands the latter idea and adds the concept of the identification of statistically independent components that are capable of describing the underlying structure of the complex agricultural data with more regularity. The images which are obtained with the help of remote sensers by satellites, drones and field sensors are always associated with redundant information, correlated spectral bands, and mixed textural patterns that may suggest poor health of the classification and stress detection analysis. In response to the above difficulties, on the one hand, IPCA will focus on breaking down the data of the input image into a list of independent components, each of which represents a distinct spectral signature or changes in texture that may have existed in the crop fields. This type of decomposition enables one to isolate meaningful patterns and noise and irrelevant information. The main elements are the orthogonal linear combinations of features that are originally present, and PCA is able to do so, by retrieving the eigenvectors and eigenvalues of the covariance matrix of the data. The eigenvectors are the major components and the nonpositive eigenvalues are the extent of the amount of varying explained by every principal component. After having computed the principal components, PCA projects the data onto the principal components, and achieves a potentially significant dimensional reduction of data. This projection has the largest possible amount of variability in the original data. Kernel PCA addresses the linear constraints by nonlinear kernel functions: It transforms data to a high dimensional space, which enables the extraction of nonlinear patterns that are complex. Improvement of this kind increases the power of the feature extraction process especially in identifying weak variation in agricultural images.

Covariance Matrix Calculation

Given a dataset X with n observations and p features, calculate the covariance matrix C is expressed as per Eq. (2):

$$C = \frac{1}{n-1}(X - \bar{X})^T(X - \bar{X}) \quad (2)$$

Eigenvalue Decomposition

The core operation of Kernel PCA is represented by the eigenvalue equation $\widetilde{K}\alpha = \lambda\alpha$, where $\widetilde{K}$is the centered kernel matrix, $\alpha$is the eigenvector, and λ is the corresponding eigenvalue. This equation allows Kernel PCA to identify principal components in a high-dimensional feature space without explicitly computing the nonlinear mapping. By using a kernel function such as the Gaussian RBF, the algorithm captures complex, nonlinear relationships present in agricultural images. Solving this eigenvalue problem reveals the directions in the transformed space that contain the most significant variation. These nonlinear principal components provide more discriminative spectral and textural features compared to traditional PCA, enabling improved detection of subtle crop stress patterns and more accurate yield prediction.

Sort Eigenvalues

Sort the eigenvalues in descending order and rearrange the corresponding eigenvectors accordingly.

Select Principal Components

Choose the first k eigenvectors corresponding to the k largest eigenvalues to form the transformation matrix M.

Project Data

Project the original data X onto the new subspace defined by M to obtain the transformed dataset Y. In general, PCA diminishes the dimensionality of the dataset while maintaining the maximum variance. It finds the directions in the data where it varies the most-largely known as principal components-and then projects data onto these principal components to get all the vital information captured within this lower-dimensional space. The resultant independent components become compact and non-redundant feature representations that considerably enhance the power of subsequent data mining tasks such as association rule discovery, anomaly detection, and stress prediction. Besides this, IPCA enhances computational efficiency due to the reduction of dimensionality and preserves the important features useful for accurate crop health assessment.

Association Rule Mining

Mining association rules is a basic method in data mining and aims at finding appropriate rules that define relationships, correlations, and co-occurrences among large numbers of variables. The association rule mining technique is adopted within the proposed system for discovering meaningful associations among the spectral and textural features extracted from images representing the crops. The associations discovered help identify better insights into the inherent patterns causing stress conditions on the respective crops. The FP-Growth algorithm efficiently determines frequent itemsets with less generation and scan operations, as compared to Apriori, and creates a Frequent Pattern Tree. The algorithm contains:

- Database Scan: Calculate the frequency of every item while eliminating those with frequency less than minimum support
- Build FP-Tree: Develop a tree structure with a root node labeled as “null” and arrange the items based on frequency and then add them to the tree with common prefixes and connect same items
- Mine FP-Tree: Beginning with low-root nodes, conditional pattern bases are created and conditional FP-Trees are made for frequent pattern mining
- FP-Growth algorithm needs two passes on the database and does not have to generate candidates, making it more efficient and scalable, particularly for large databases. Yet, mining and traversing FP-Tree may be intricate and memory-consuming. Various criteria such as support, Rank Index, disassociation, antecedence, and confidence values are employed for ordering sets, assisting in understanding set frequency, relevance, independence, position, and association strength

Anomaly Detection

Anomaly detection on temporal agricultural images needs a model that can address spatial and temporal characteristics concurrently. The Convolutional Long Short Term Memory network, a method that effectively handles spatiotemporal sequences, will be used for this research. The ConvLSTM network is an extension of the regular LSTM network and uses convolution operations instead of fully connected operations for transitioning from input to state and from state to state. It becomes possible for ConvLSTM networks to maintain spatial locality while learning temporal relationships within a sequence of images. CNN represents a family of deep learning models that have shown an aptitude for learning and identifying meaningful features within raw data, including images. A CNN structure primarily consists of Convolutional Layers, Pooling Layers, Activation Functions, Fully Connected Layers, and an Output Layer. Convolution Layers apply filters on images, pooling Layers reduce feature spatial information, Activation functions introduce non-linearities within images, and Fully Connected Layers generate predictions. During learning sessions, CNNs optimize parameters via functions like Backpropagation and Gradient Descent methods. By so doing, CNNs have the possibility and ability to improve and optimize identifying and classifying objects within images.

A type of Recurrent Neural Network named LSTM (RNN) was developed with the intention of resolving the issue associated with the identification of long-term dependencies within input sequences. Due to its memory blocks.

- Input Layer: It captures serial input data
- Hidden Layer: It holds the LSTM cells for processing and storing information
- Output Layer: It generates the resultant output based on the processed data

The memory cells replace the basic cells that comprise regular RNNs. The basic cells in LSTMs include three gates: the forget gate, input gate, and output gate:

- Input Gate: It controls the amount of new information that enters a memory cell
- Forget Gate: It determines what needs to be removed from the memory cell
- Output Gate: It varies the output based on the previous state and current value of the input

The activation of each LSTM unit at time $l_t$is calculated using Eq. (3):

$$l_t = \sigma(wm_{i,l} \cdot x_t + wm_{h,l} \cdot l_{t-1} + bi) \quad (3)$$

The ConvLSTM model looks at sequences of crop images that were already prepared for use and learns how crops grow in a normal way. It will use a variety of different attributes to figure out how crops are changing during time. Any "anomalies" or changes from what the model previously learned to be the normal growth of crops can then be flagged as issues. Because ConvLSTM has the ability to deal with long-term time series data, it will help identify very small changes in the stress level of crops before the stress level is visible. Therefore, using the proposed method will provide a greater level of reliability and accuracy for monitoring stress on crops using large area monitoring systems.

Sequential Pattern Mining

Under this framework we will use Sequential Pattern Mining and Hidden Markov Models (HMMs) in order to model how crop stress evolves over time. HMMs are based on the principle of using probabilities to describe a sequence of events when we do not have access to the exact states of the system that generated the sequence. The hidden states represent different levels of crop stress or "stages," while the observed features are derived from analysis of images taken with various spectral and textural indicators. During the training phase, we build a dataset by taking time-series data from images taken consecutively, and this dataset is then used to train the HMM by estimating three key components of the model:

(i) The transition probabilities of moving from one hidden stress state to another
(ii) The emission probabilities which link hidden states to the patterns of features observed in each image
(iii) The initial state distribution. After training, we can use the viterbi algorithm, or other similar algorithms, to make inferences about the most likely sequence of hidden stress states at any given time

A core component of the Hidden Markov Model is the state transition probability, defined as:

$$a_{ij} = P(S_{t+1} = j \mid S_t = i) \quad (4)$$

where:$S_t$is the hidden stress state at time t,$i$and $j$represent possible stress levels (e.g., healthy, mild stress, severe stress),a_{ij}denotes the probability of transitioning from state $i$to state $j$in the next time step.

Stress Severity Scoring

The second stage in the proposed framework is to evaluate the amount of stress placed on each crop using the structured MCDA approach. The MCDA process provides an organized way to evaluate the number of variables affecting crops at one time and assigns a single severity score for crop stress via the aggregation of multiple, heterogeneous variables. These variables typically include the following: Spectral Index, Textural Index, Anomaly Probability, Time Transition, and Environmental Conditions that have been determined from the prior stages of evaluation. The above referenced indicators will initially need to be normalized prior to the aggregation process so that they will all be on an equal scale. Weighting for each variable will be determined based on how important the variable is to crop stress as determined by expert opinion or by statistical analysis of the prior evaluation. The individual criteria are then combined using a weighted aggregation function, allowing the model to take into consideration both direct and indirect factors that contribute to plant stress. The entire level of stress in a particular area or crop segment is accurately represented by the composite score that is produced.

The overall stress severity score Scan be calculated as:

$$S = \sum_{i=1}^{n} w_i \ x_i \quad (5)$$

Where, n= number of stress-related criteria, x_i= normalized value of the i^{th}criterion (e.g., spectral index, anomaly probability, texture feature), w_i= weight assigned to the i^{th}criterion, such that

$$\sum_{i=1}^{n} w_i = 1 \quad (6)$$

Each criterion contributing to crop stress is first normalized and weighted according to its importance. The weighted sum then produces a single composite score S, representing the overall severity of stress. Higher values of Sindicate more severe crop stress. This scoring mechanism enhances decision-making by offering a transparent, multi-dimensional assessment that facilitates early intervention, resource prioritization, and adaptive management in precision agriculture environments.

Results and Discussion

The experiments were implemented in Python and the results demonstrate the effectiveness of the proposed framework in detecting crop stress and predicting yield with high reliability. Using Python 3.10 and deep learning and machine learning libraries like TensorFlow, Keras, and Scikit-learn, the suggested framework was put into practice. A machine with an Intel Core i7 CPU, 16 GB RAM, and a 12 GB NVIDIA RTX 3060 GPU was used for the experiments. With a batch size of 32, an Adam optimizer, and a learning rate of 0.001, the ConvLSTM model was trained for 100 epochs. To get the best results, hyperparameters were experimentally adjusted.

For IPCA, 50 principal components that accounted for the majority of variance in the data were retained to extract features. For FP-Growth, the levels of minimum support and minimum confidence were established at 0.30 and 0.80, respectively, in order to provide a trustworthy method of generating association rules. The ConvLSTM network consisted of three layers of ConvLSTM with 64, 128, and 256 filters, and all three layers used a Kernel of size 3 × 3. Each layer was enhanced by batch normalization and dropout. The HMM (hidden markov model) included three hidden states for healthy condition, moderate stress condition, and severe stress condition (each of these being defined as a hidden state). The weights of the criteria with MCDA were determined using expert defined weights based on how much weight was assigned to the spectral, temporal, and anomaly-based indicators of crop stress evaluation.

Dataset Collection

For this research study, the data utilized came from the Mendeley Data repository. Field images of wheat crops dataset (Rahman and Rokonozzaman, 2025) that were annotated as high-resolution RGB, were collected under real agricultural conditions in Bangladesh. This data was chosen due to providing unambiguous definitions of the different categories of crop stress they contain and the various types of disease symptoms found in the data were taken under a variety of lighting and environmental conditions. Therefore, the annotated images in this dataset would serve well for conducting image-based crop stress analysis. In its original form, the dataset consists of 1,603 images spread evenly over five categories of crop stress: Black Point, Fusarium Foot Rot, Healthy Leaf, Leaf Blight, and Wheat Blast. To rectify class imbalance issues and improve the robustness of a model predicting crop stress based on the images in the dataset, the following augmentation methods were implemented: Rotation, flipping, scaling, shifting, cropping, and colour/intensity variations. Therefore, the number of images in each category increased to 1,000 images per category, for an overall total of 5,000 images in the final dataset. The preparation of the dataset included only the highest quality images, as defined by: Distinguishable crop disease symptoms; good image illumination; and correct annotations of the images. Images were then removed from the dataset based on the existence of one or more of the following: Blurry images; duplicate images; occluded images; or incomplete images. The removal of bad images from the dataset served to reduce overall noise and ambiguity in the collected image dataset. The final dataset was divided into three subsets (70%, 10%, 20%) using stratified random sampling techniques to ensure equal representation of classes within each of the three subsets, including: Training data (3,500 images); Validation data (500 images); and Testing data (1,000 images). Utilizing stratified random sampling techniques ensured that learning was at least balanced and that reliable evaluation of the proposed crop stress detection framework could be achieved.

Fig. 2 shows pictures of the five wheat diseases, which are Black Point, Fusarium Foot Rot, Healthy Leaf, Leaf Blight, and Wheat Blast. Such pictures represent the differences in disease symptoms, lesion morphology, color, and texture. The differences used for detecting distinguishing features needed in classification. Fig. 3 shows the usefulness of the method used for preprocessing images. The Adaptive Wiener Filter allows obtaining better results because of illumination correction, contrast enhancement, noise reduction, and the through color normalization. The improvement of image quality allows classifying plants through feature extraction and other purposes. The good quality images enable feature extraction and classification of crops. For example, Black Point has high contrast ratio in healthy tissue and areas affected.

Healthy leaf images have more uniform color distribution and crisper venation patterns, which reduces the risk of false positives during categorization. Preprocessing improves disease borders, discoloration patterns, and necrotic areas for Leaf Blight and Wheat Blast, making it possible to differentiate subtle infections from stress states that are visually identical. As a result, the extracted spectral and textural features become more discriminative, enhancing the performance of the ensuing FP-Growth, ConvLSTM, HMM, and MCDA modules and allowing IPCA to provide higher-quality feature representations. This enhancement is reflected in the high classification accuracy of 98.9%, suggesting that adequate preprocessing plays a vital role in strengthening the resilience and reliability of the proposed crop stress analysis framework.

DOI: 10.3844/jcssp.2026.2503.2518

Fig. 2: Sample Images of Wheat Crop Stress Classes

Fig. 3: Original vs. Preprocessed Image Comparison for Each Class

Performance Evaluation Metrics

To comprehensively evaluate the performance of the proposed crop stress analysis framework, several classification and regression evaluation metrics were employed. Accuracy measures the proportion of correctly classified samples and is computed as:

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

Where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively.

Precision evaluates the correctness of positive predictions:

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

While Recall measures the ability of the model to correctly identify positive samples:

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

The F1-score, which balances Precision and Recall, is calculated as:

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

For yield prediction, regression performance is evaluated using Root Mean Square Error (RMSE):

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}$$

Mean Absolute Error (MAE):

$$MAE = \frac{1}{N}\sum_{i=1}^{N} | y_i - \hat{y}_i |$$

And the Coefficient of Determination (R^2):

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$

Where $y_i$is the actual yield, $\hat{y}_i$is the predicted yield, and $\bar{y}$is the mean of the actual values. In addition, the Area Under the Receiver Operating Characteristic Curve (AUC) is used to assess the model's ability to distinguish between crop stress classes, with higher AUC values indicating better classification performance.

Fig. 4 depicts the training process of the model and its accuracy and loss convergence. The loss values for both training and validation datasets are reducing uniformly, and it clearly shows that an efficient learning process with valid pattern recognition is being achieved.

DOI: 10.3844/jcssp.2026.2503.2518

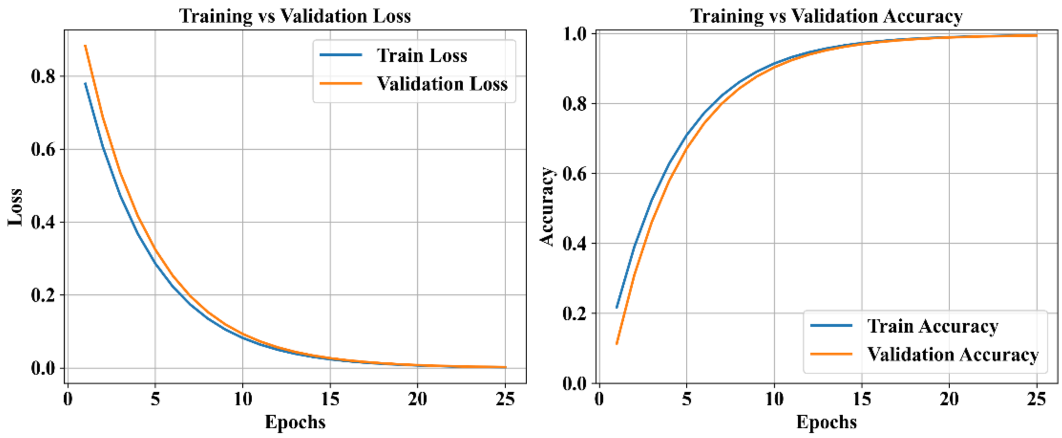

Fig. 4: Training vs. Validation Loss and Accuracy Curve

The fact that the gaps between these two graphs are reducing signifies efficient generalization with negligible overfitting. Additionally, it can be seen that the accuracy values for both the training and validation datasets are rising with every epoch, indicating that efficient discriminatory feature extractions are being achieved for successful classification of wheat stress types. The accuracy graphs are almost converging, signifying efficient learning.

Figure 5 shows a confusion matrix representing classification accuracy on five classes of wheat types: Healthy, Leaf Blight, Wheat Blast, Fusarium Foot Rot, and Black Point. The values on the diagonals identify perfectly classified instances for every class, reflecting superior classification abilities. Little misclassifications exist, signifying efficient addressing of inter-class similarities. Figure 6 highlights the accuracy of the designed yearly prediction model for grain production. The plot identifies instances' actual production against predicted production. Tight grouping with the regression equation on the diagnostic plot establishes an extremely correct prediction rate. RMSE (0.27), MAE (0.22), and R^2 (0.975) confirms that it effectively identifies production rates and establishes highly correct predictions with negligible error margins.

Fig. 7 highlights the accuracy and F1-score value of the devised model and traditional approaches like CNN, RF, LSTM, and SVM. From Fig. 7, it can be seen that the devised model represents maximum accuracy and F1-score value and thus outperforms all other approaches. Fig. 8 depicts an ablation study that analyzes the efficacy of the devised components, namely IPCA and FP-Growth, and their combination with a traditional CNN model. All three modules increase accuracy and F1-score value. However, it can be seen that the devised model with all three modules represents maximum accuracy and an F1-score value of 98.8%.

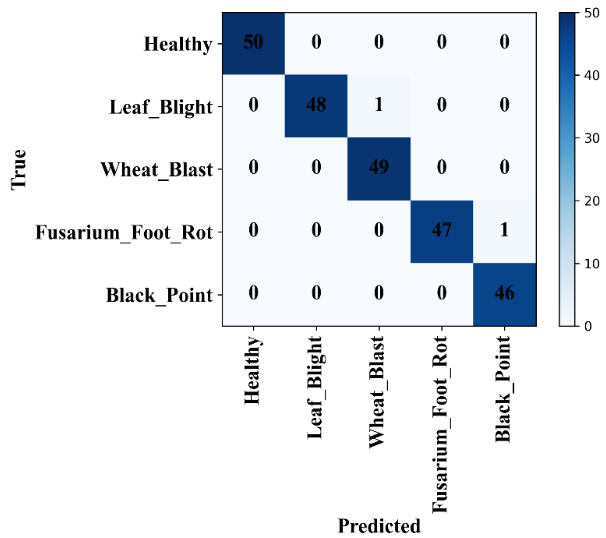

Fig. 5: Confusion Matrix of Crop Stress Classification

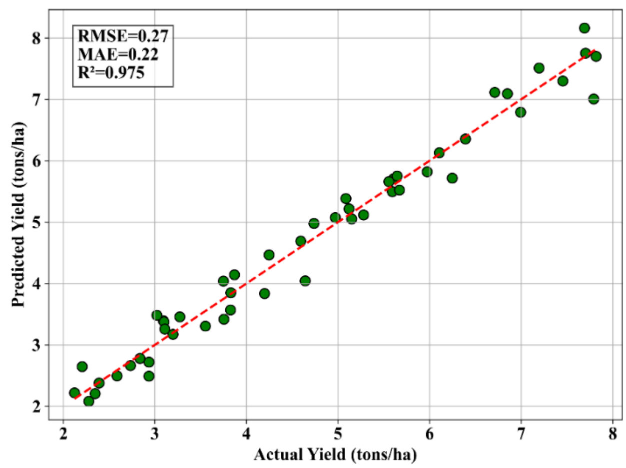

Fig. 6: Actual vs. Predicted Yield Scatter Plot

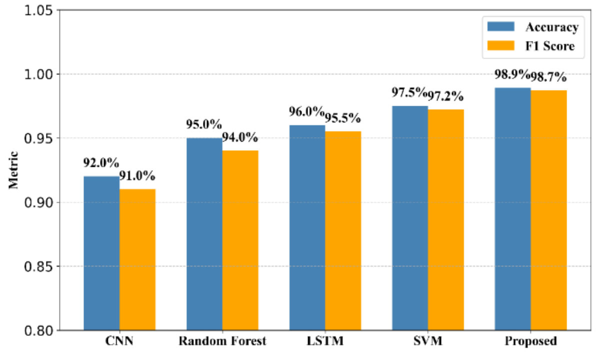

Fig. 7: Classification Performance Comparison with Existing Models

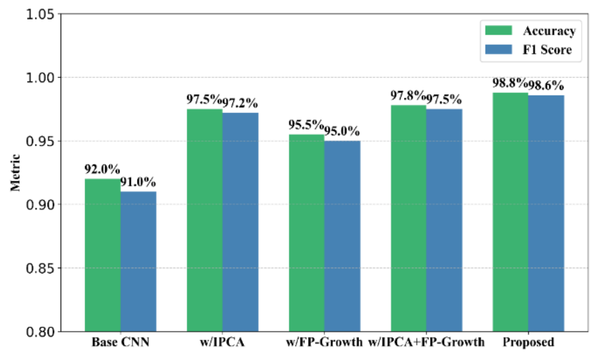

Fig. 8: Ablation Study: Impact of Different Components

The performance comparison in Table 2 shows the accuracy and F1-score obtained using various machine learning algorithms for crop stress classification. The traditional methods like CNN, RF, LSTM, and SVM result in superior performance with an accuracy ranging from 92.0 to 97.5%. Out of these models, it can be seen that SVM performs better with an accuracy value of 97.5%, and the F1-score value of 97.2%. But it is clear that the proposed method performs significantly better compared to previous methods with an accuracy value of 98.9%, and an F1-score value of 98.7%, which clearly illustrates the superiority and accuracy of the proposed method compared to traditional ones.

Enhanced effectiveness of suggested system is due to combination of number of techniques from data mining and deep learning. Implementation of Adaptive Wiener Filter helps to filter noise from crop images without neglecting essential objects in the crops. Employing IPCA makes it possible to filter independent and informative spectral-textural features from the input data and to get rid of duplicates because assurance of sufficient discriminative ability of the obtained data is very important for the process. Application of FP-Growth algorithm combines stress-related features together and allows to obtain the additional information to make distinction between two or more classes. Use of ConvLSTM makes it possible to analyse the initial data in terms of both space and time. Hidden Markov model enables tracking stress evolution for some period of time and this provides more accuracy in prediction.

Simulation of the above-mentioned models increases the overall performance of the new methodology in comparison with the conventional CNNs and support vector machines.

As observed in Fig. 9, each curve visualises the Receiver Operating Characteristic (ROC) of each class of crop stress. Each ROC curve is positioned such that the higher the curve, the closer the curve is towards the top left corner, thus demonstrating a high sensitivity and specificity.

The respective Area Under the Curve (AUC) of each of the classes support excellent ability to optimally separate the classes i.e., Black Point (0.98), Fusarium Foot Rot (0.97), Leaf Blight (0.95), Wheat Blast (0.96) and Healthy (0.99). Overall, the results outlined above show that the models demonstrated excellent multi-class discrimination power.

The precision-recall behavior across classes is shown in Fig. 10, which is especially helpful for datasets that are unbalanced. The model correctly identifies stress classes with few false positives and false negatives, as shown by the high precision and recall curves. Healthy (0.97), Leaf Blight (0.90), Wheat Blast (0.94), Fusarium Foot Rot (0.95), and Black Point (0.96) AUC-PR results support performance stability under different class distributions.

The results from the ablation study shown in Table 2 prove that all components have a positive impact on the accuracy. The use of IPCA achieves a representation of the features that is free from redundant spectral components, enhancing the classifications by 2.8%. FP-Growth enhances the discriminative features through the identification of frequently occurring relationships among the features.

Table 1: Performance Comparison of Existing and Proposed Models

Model	Accuracy (%)	F1-Score (%)
CNN	92.0	91.0
Random Forest	95.0	94.0
LSTM	96.0	95.5
SVM	97.5	97.2
Proposed Model	98.9	98.7

Table 2: Ablation Study

Method	Accuracy	F1-score
CNN	92.0	91.0
CNN + IPCA	94.8	94.2
CNN + IPCA + FP-Growth	96.5	96.0
CNN + IPCA + ConvLSTM	97.6	97.2
CNN + IPCA + ConvLSTM + HMM	98.4	98.1
Full Proposed	98.9	98.7

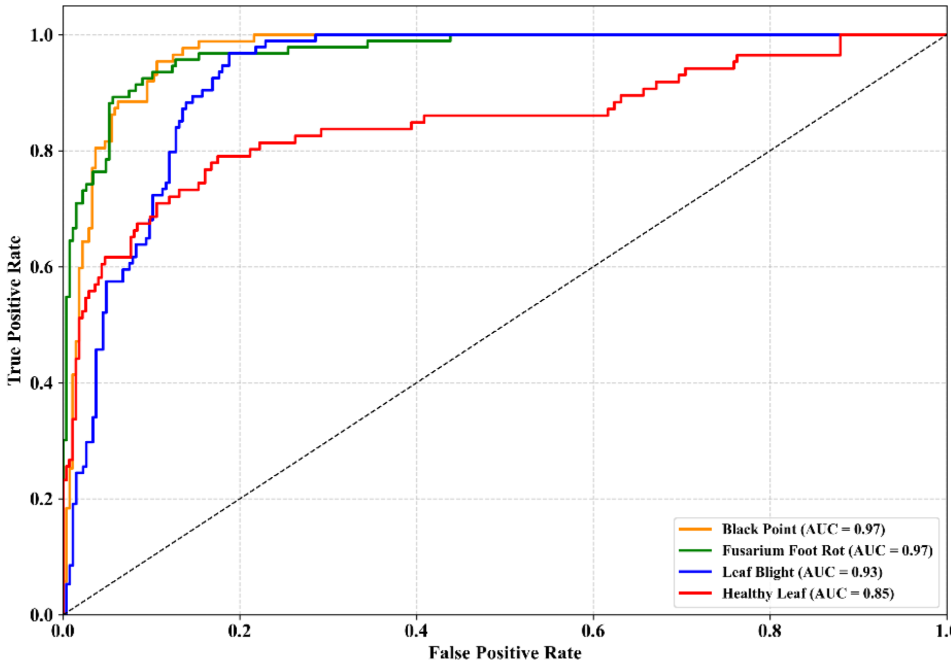

Fig. 9: ROC Curves for Different Crop Stress Classes

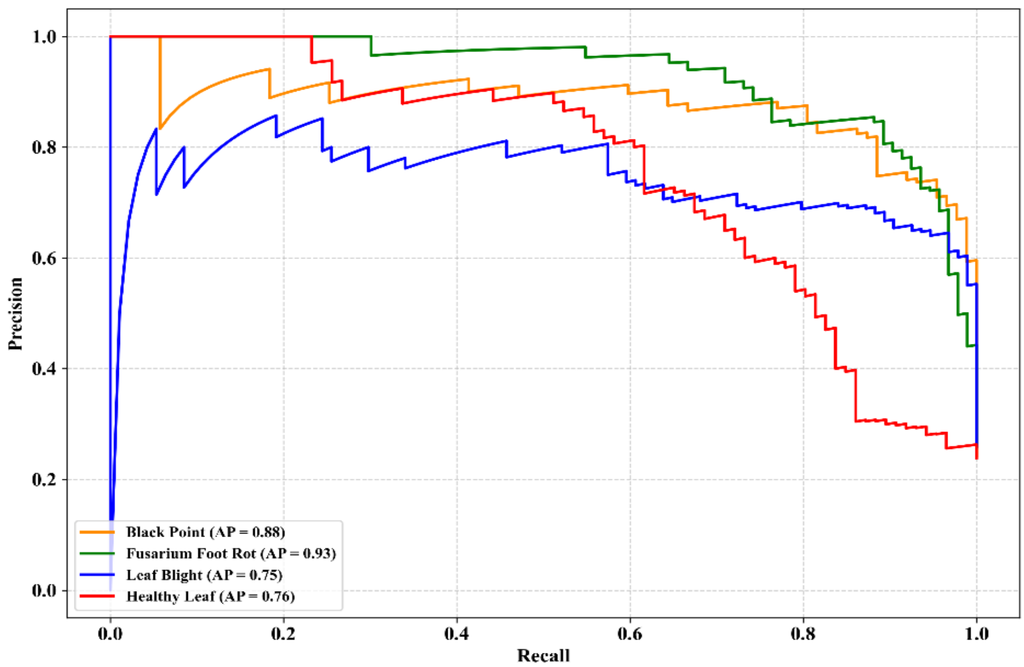

Fig. 10: Precision Recall Curves for Each Crop Stress Class

Use of ConvLSTM enhances the detection of temporal anomaly detection, while HMM detects evolutions of stress over time. Deleting the HMM reduces the overall accuracy of detection from 98.9% to 98.4%, showing that the sequential approach significantly impacts the accuracy of prediction.

The proposed design takes about 0.54 seconds to analyze one high-resolution image; thus, it is acceptable for almost real-time crop analysis. The majority of the time is taken by ConvLSTM inference while the additional time needed by IPCA, FP-Growth and MCDA is insignificant.

Conclusion

Research has established a framework for paradigm of image-driven crop stress and yield prognosis, which combines the use of data mining methods and deep learning techniques in the analysis of crops. The application of the framework has shown that the proposed model provides 98.9% of correctness in classification together with 98.7% of the F1 score in the given conditions. The experimental results show that the proposed framework yields better results that any of the previously known data mining approaches to the problem, including CNN, Random Forest, LSTM, and SVM models. Though the framework shows its great potential, it also has the drawbacks as this study has been conducted using one wheat disease dataset that contains only images collected in RGB format. The use of the one crop dataset greatly limits the application of the proposed system to other crop kinds, regions, and conditions. The application of the proposed system has also been limited due to its offline experimental application, and it is still not used for the real-time agriculture. The directions of future research will involve the drawing on multispectral, hyperspectral, thermal, and other UAV-based images. The integration of Transformer-based deep learning models, Explainable Artificial Intelligence (XAI) approaches, Federated Learning, and edge-enabled IoT systems will further boost model interpretability, scalability, privacy, and real-time decision support. Additionally, it facilitates the collection of practical crop-accurate practical crop systems.

Limitations and Scalability

Although the framework proposed is effective at detecting crop stress and yield forecasting, there are a number of computational factors that need to be considered in order to enable large-scale implementation. For example, the ConvLSTM model will require a lot of computation power and GPU memory capabilities to carry out its operation on a long sequence of temporally-ordered images. This fact will likely increase training time due to the large amount of agricultural image data present today. Additionally, the FP-Growth will create large FP-trees with increases in the number of features and transactions extracted, causing memory usage to increase considerably. Even though utilizing IPCA to reduce dimensionality of features and thus improve computational efficiency, the entire framework is computationally intensive as it integrates several analytical modules into one framework. Therefore, at the currently available computational resources, edge devices (especially those with less available computational resources) would not be capable of deploying the model in real-time. Therefore, model-compression, lightweight architectures, or hardware acceleration techniques will likely be the only means through which models could be deployed from edge devices in real-time. However, the framework can be effectively deployed to cloud-based or high performance computing environments to provide extensive agricultural monitoring of large geographic areas.

Practical Applications and Future Work

The new framework will help farmers use modern technology to improve farming efficiency by using drone imagery, satellite monitoring, and on-site sensors to continuously check how crops are growing across huge farms. Farmers often apply irrigation, fertilizers, or pesticides to large portions of their fields and it can be difficult for them to see which crops need help because of all the land they have to manage. The Stress Severity Scores that this system creates will help farmers and other agricultural professionals identify which areas of their farms have poor crop conditions, so that they can prioritize which areas to inspect in the field and only apply products where needed. This process will help them reduce the cost of applying these inputs and make better use of their resources. By using cloud-based applications, this could allow for continuous monitoring across multiple farms, and also enable the use of lightweight deep learning models and edge computing in the future to make real-time decisions directly at the farm. Coming up, we will be using multiple crop datasets from different environments to validate our framework, and improve the efficiency of our framework for real-world use in smart agriculture settings.

Acknowledgment

There is no acknowledgement in the work.

Funding Information

On Behalf of all authors, the corresponding author states that they did not receive any funds for this project.

Author’s Contributions

A. Josephine: Performed the analysis the overall concept, write and edited.

A. Subhashini: Participated in the methodology, conceptualization, data collection and write the study.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and no ethical issues involved.

Conflicts of Interest

The authors declare that we have no conflict of interest.

Competing Interests

The authors declare that we have no competing interest.

Data Availability Statement

All the data is collected from the simulation reports of the software and tools used by the authors. Authors are working on implementing the same using real world data with appropriate permissions.

List of Abbreviations

Abbreviation	Full Form
AI	Artificial Intelligence
RGB	Red Green Blue
IPCA	Independent Principal Component Analysis
PCA	Principal Component Analysis
FP-Growth	Frequent Pattern Growth
ConvLSTM	Convolutional Long Short-Term Memory
HMM	Hidden Markov Model
MCDA	Multi-Criteria Decision Analysis
CNN	Convolutional Neural Network
LSTM	Long Short-Term Memory
SVM	Support Vector Machine
RF	Random Forest
IoT	Internet of Things
UAV	Unmanned Aerial Vehicle
RNN	Recurrent Neural Network
AUC	Area Under Curve
ROC	Receiver Operating Characteristic
PR	Precision-Recall
RMSE	Root Mean Square Error
MAE	Mean Absolute Error

References

Abdel-Salam, M., Kumar, N., & Mahajan, S. (2024). A proposed framework for crop yield prediction using hybrid feature selection approach and optimized machine learning. *Neural Computing and Applications*, *36*(33), 20723–20750. https://doi.org/10.1007/s00521-024-10226-x

Azimi, S., Kaur, T., & Gandhi, T. K. (2021a). A deep learning approach to measure stress level in plants due to Nitrogen deficiency. *Measurement*, *173*, 108650. https://doi.org/10.1016/j.measurement.2020.108650

Azimi, S., Wadhawan, R., & Gandhi, T. K. (2021b). Intelligent Monitoring of Stress Induced by Water Deficiency in Plants Using Deep Learning. *IEEE Transactions on Instrumentation and Measurement*, *70*, 1–13. https://doi.org/10.1109/tim.2021.3111994

Burchard-Levine, V., Borra-Serrano, I., Peña, J. M., Kustas, W. P., Guerra, J. G., Dorado, J., Mesías-Ruiz, G., Herrezuelo, M., Mary, B., McKee, L. M., de Castro, A. I., Sanchez-Élez, S., & Nieto, H. (2025). Evaluating the precise grapevine water stress detection using unmanned aerial vehicles and evapotranspiration-based metrics. *Irrigation Science*, *43*(1), 65–85. https://doi.org/10.1007/s00271-024-00931-9

Chandel, N. S., Jat, D., Chakraborty, S. K., Upadhyay, A., Subeesh, A., Chouhan, P., Manjhi, M., & Dubey, K. (2025). Deep learning assisted real-time nitrogen stress detection for variable rate fertilizer applicator in wheat crop. *Computers and Electronics in Agriculture*, *237*, 110545. https://doi.org/10.1016/j.compag.2025.110545

Chandraprabha, M., & Dhanraj, R. K. (2023). Ensemble Deep LearniAlgorithm for Forecasting of Rice Crop Yield based on Soil Nutrition Levels. *ICST Transactions on Scalable Information Systems*, *10*(4), e7. https://doi.org/10.4108/eetsis.v10i3.2610

Fevgas, G., Lagkas, T., Papadopoulos, P., Sarigiannidis, P., & Argyriou, V. (2025). Integrating thermal infrared and RGB imaging for early detection of water stress in lettuces with comparative analysis of IoT sensors. *Smart Agricultural Technology*, *11*, 100881. https://doi.org/10.1016/j.atech.2025.100881

Gautam, V., & Rani, J. (2022). Mango Leaf Stress Identification Using Deep Neural Network. *Intelligent Automation & Soft Computing*, *34*(2), 849–864. https://doi.org/10.32604/iasc.2022.025113

Gavahi, K., Abbaszadeh, P., & Moradkhani, H. (2021). DeepYield: A combined convolutional neural network with long short-term memory for crop yield forecasting. *Expert Systems with Applications*, *184*, 115511. https://doi.org/10.1016/j.eswa.2021.115511

Hamidon, M. H., & Ahamed, T. (2022). Detection of Tip-Burn Stress on Lettuce Grown in an Indoor Environment Using Deep Learning Algorithms. *Sensors*, *22*(19), 7251. https://doi.org/10.3390/s22197251

Hemamalini, P., Chandraprakash, M. K., Laxman, R. H., Rathinakumari, C., Senthil Kumaran, G., & Suneetha, K. (2025). Thermal canopy segmentation in tomato plants: A novel approach with integration of YOLOv8-C and FastSAM. *Smart Agricultural Technology*, *10*, 100806. https://doi.org/10.1016/j.atech.2025.100806

Islam, S., Samsuzzaman, Reza, M. N., Lee, K.-H., Ahmed, S., Cho, Y. J., Noh, D. H., & Chung, S.-O. (2024). Image Processing and Support Vector Machine (SVM) for Classifying Environmental Stress Symptoms of Pepper Seedlings Grown in a Plant Factory. *Agronomy*, *14*(9), 2043. https://doi.org/10.3390/agronomy14092043

Jhajharia, K., Mathur, P., Jain, S., & Nijhawan, S. (2023). Crop Yield Prediction Using Machine Learning and Deep Learning Techniques. *Procedia Computer Science*, *218*, 406–417. https://doi.org/10.1016/j.procs.2023.01.023

Joshi, P., Das, D., Udutalapally, V., Pradhan, M. K., & Misra, S. (2022). RiceBioS: Identification of Biotic Stress in Rice Crops Using Edge-as-a-Service. *IEEE Sensors Journal*, *22*(5), 4616–4624. https://doi.org/10.1109/jsen.2022.3143950

Kamarudin, M. H., Ismail, Z. H., Saidi, N. B., & Hanada, K. (2023). An augmented attention-based lightweight CNN model for plant water stress detection. *Applied Intelligence*, *53*(18), 20828–20843. https://doi.org/10.1007/s10489-023-04583-8

Kapari, M., Sibanda, M., Magidi, J., Mabhaudhi, T., Mpandeli, S., & Nhamo, L. (2025). Assessment of the Maize Crop Water Stress Index (CWSI) Using Drone-Acquired Data Across Different Phenological Stages. *Drones*, *9*(3), 192. https://doi.org/10.3390/drones9030192

Lee, K.-H., Samsuzzaman, Reza, M. N., Islam, S., Ahmed, S., Cho, Y. J., Noh, D. H., & Chung, S.-O. (2024). Evaluation of Machine Learning Models for Stress Symptom Classification of Cucumber Seedlings Grown in a Controlled Environment. *Agronomy*, *15*(1), 90. https://doi.org/10.3390/agronomy15010090

Malounas, I., Paliouras, G., Nikolopoulos, D., Liakopoulos, G., Bresta, P., Londra, P., Katsileros, A., & Fountas, S. (2024). Early detection of broccoli drought acclimation/stress in agricultural environments utilizing proximal hyperspectral imaging and AutoML. *Smart Agricultural Technology*, *8*, 100463. https://doi.org/10.1016/j.atech.2024.100463

Mohamed, H. I., Khan, A., & Basit, A. (2024). CRISPR-Cas9 System Mediated Genome Editing Technology: An Ultimate Tool to Enhance Abiotic Stress in Crop Plants. *Journal of Soil Science and Plant Nutrition*, *24*(2), 1799–1822. https://doi.org/10.1007/s42729-024-01778-x

Mohammadi, P., & Asefpour, K. V. (2023). Machine learning provides specific detection of salt and drought stresses in cucumber based on miRNA characteristics. *Plant Methods*, *19*(1), 123. https://doi.org/10.1186/s13007-023-01095-x

Nagesh, O. S., Budaraju, R. R., Kulkarni, S. S., Vinay, M., Ajibade, S.-S. M., Chopra, M., Jawarneh, M., & Kaliyaperumal, K. (2024). Boosting enabled efficient machine learning technique for accurate prediction of crop yield towards precision agriculture. *Discover Sustainability*, *5*(1). https://doi.org/10.1007/s43621-024-00254-x

Park, B., Wi, S., Chung, H., & Lee, H. (2024). Chlorophyll Fluorescence Imaging for Environmental Stress Diagnosis in Crops. *Sensors*, *24*(5), 1442. https://doi.org/10.3390/s24051442

Qiao, M., He, X., Cheng, X., Li, P., Luo, H., Zhang, L., & Tian, Z. (2021). Crop yield prediction from multi-spectral, multi-temporal remotely sensed imagery using recurrent 3D convolutional neural networks. *International Journal of Applied Earth Observation and Geoinformation*, *102*, 102436. https://doi.org/10.1016/j.jag.2021.102436

Rahman, M. I. R., & Rokonozzaman, A. (2025). Disease Dataset of Wheat: Original, Augmented, and Balanced for Deep Learning. *Mendeley Data*. https://doi.org/10.17632/5gc7hwydwg.1

Sharma, P., Dadheech, P., Aneja, N., & Aneja, S. (2023). Predicting Agriculture Yields Based on Machine Learning Using Regression and Deep Learning. *IEEE Access*, *11*, 111255–111264. https://doi.org/10.1109/access.2023.3321861

Subramaniam, L. K., & Marimuthu, R. (2024). Crop yield prediction using effective deep learning and dimensionality reduction approaches for Indian regional crops. *E-Prime - Advances in Electrical Engineering, Electronics and Energy*, *8*, 100611. https://doi.org/10.1016/j.prime.2024.100611

Velázquez-González, J. R., Pérez-Patricio, M., de Jesús Osuna-Coutiño, J. A., Camas-Anzueto, J. L., Aguilar-González, A., Morales-Navarro, N. A., & Hernández-Gutiérrez, C. A. (2024). Supervisory Configuration of Deep Learning Networks for Plant Stress Detection and Synthetic Dataset Generation. *IEEE Access*, *12*, 186255–186276. https://doi.org/10.1109/access.2024.3509814

Wang, L., Zhang, H., Bian, L., Zhou, L., Wang, S., & Ge, Y. (2024). Poplar seedling varieties and drought stress classification based on multi-source, time-series data and deep learning. *Industrial Crops and Products*, *218*, 118905. https://doi.org/10.1016/j.indcrop.2024.118905

Wen, G., & Ma, B.-L. (2025). Decoding canola and oat crop health and productivity under drought and heat stress using bioelectrical signals and machine learning. *Artificial Intelligence in Agriculture*, *15*(4), 696–706. https://doi.org/10.1016/j.aiia.2025.04.006

Yang, H., Fei, L., Wu, G., Deng, L., Han, Z., Shi, H., & Li, S. (2025). A novel deep learning framework for identifying soybean salt stress levels using RGB leaf images. *Industrial Crops and Products*, *228*, 120874. https://doi.org/10.1016/j.indcrop.2025.120874